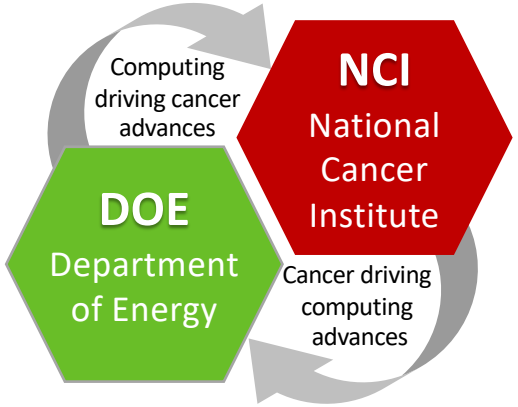
DOE-NCI Pilot 2:
RAS Biology on Membranes

DOE-NCI partnership to advance exascale development through cancer research



Advanced Scientific Computing Advisory Committee (ASCAC)

Washington, DC
September 17, 2018

Dwight Nissley, FNLCR

Fred Streitz, LLNL

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. This work was performed under the auspices of the U.S. Department of Energy by Argonne National Laboratory under Contract DE-AC02-06-CH11357, Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, Los Alamos National Laboratory under Contract DE-AC5206NA25396, and Oak Ridge National Laboratory under Contract DE-AC05-00OR22725. LLNL-PRES-754451.

Oncogenic KRAS is responsible for many human cancers

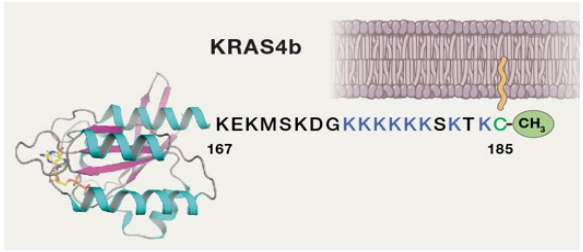
93% of all pancreatic

42% of all colorectal

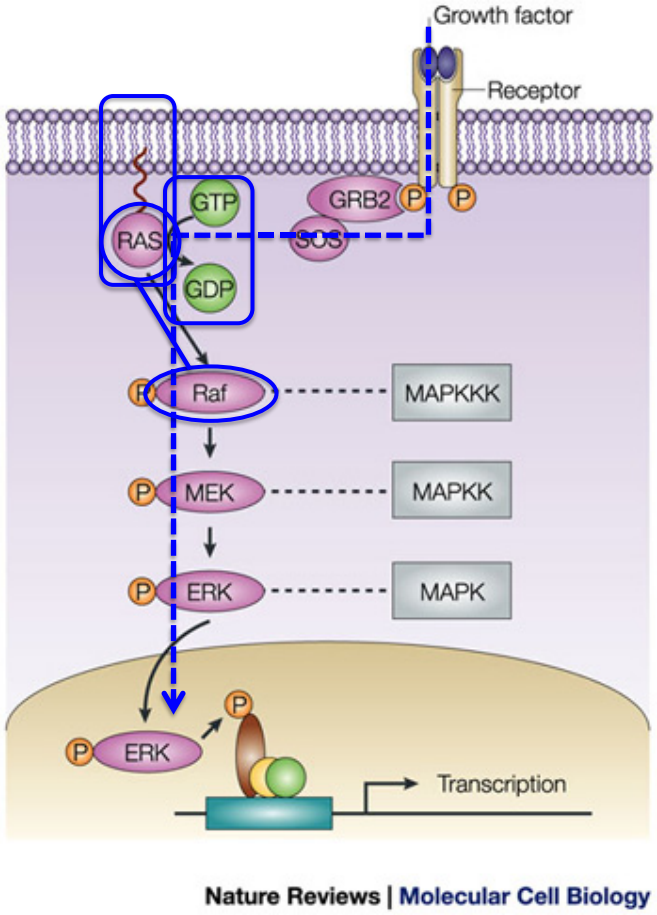
33% of all lung cancers

1 million deaths/year world-wide

No effective inhibitors



Simanshu,Cell 170, 2017



Nature Reviews | Molecular Cell Biology

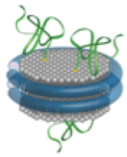
Pathway transmits signals

RAS is a switch
oncogenic RAS is “on”

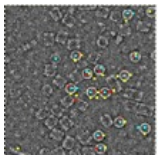
RAS localizes to the plasma membrane

RAS binds effectors (RAF) to activate growth

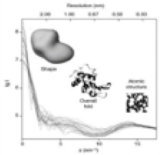
Cancer Moonshot Pilot 2: RAS biology on membranes



Experiments on nanodisc



CryoEM imaging

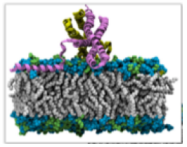


X-ray/neutron scattering

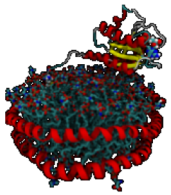
Multi-modal experimental data, image reconstruction, analytics

Protein structure databases

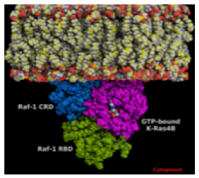
Predictive simulation and analysis of Ras



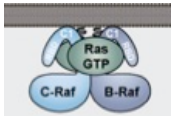
Phase Field model of lipid membrane



Granular Ras membrane interaction simulations



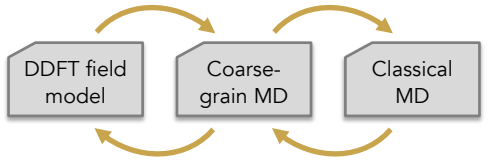
Atomic resolution Ras-RAF interaction



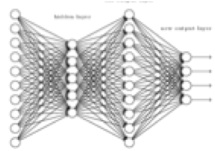
Ras Activation

Adaptive time stepping

Adaptive spatial resolution



High-fidelity subgrid modeling



Unsupervised deep feature learning

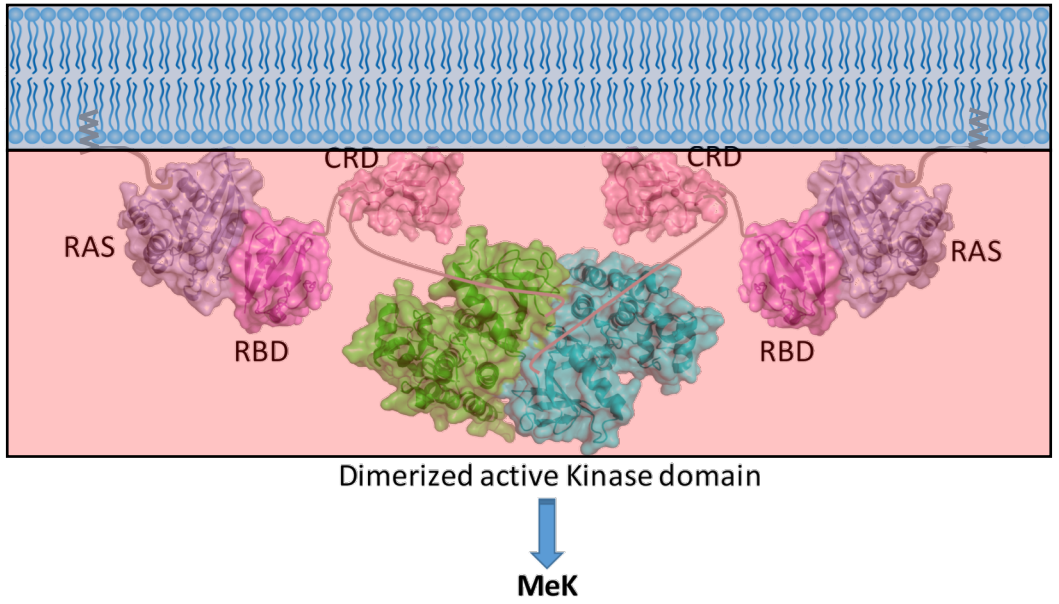
Mechanistic network models

Uncertainty quantification

Essential strategy: utilize appropriate scale methodology for each component

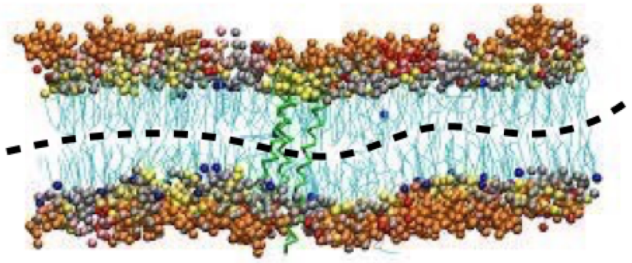
Model membrane with RAS at micron (continuum) scale

Model protein behavior at molecular scale



Multiscale Model of Lipid Bilayer

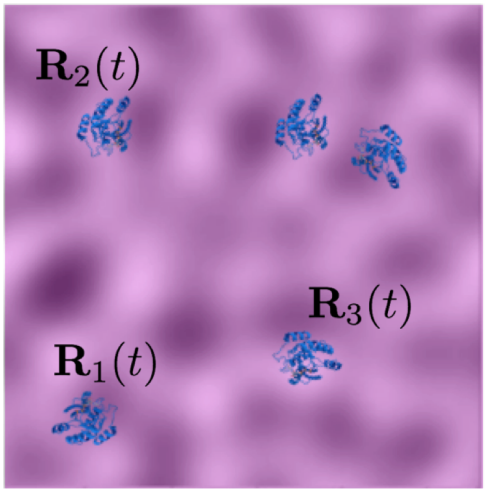
$$\mathbf{n}_2 = (n_{2,1}, n_{2,2}, \ldots, n_{2,N})$$



$$z = h(x, y, t)$$

$$\mathbf{n}_1 = (n_{1,1}, n_{1,2}, \ldots, n_{1,N})$$

$n_{i,j}$ – lipid number densities
h – membrane deformation
$\mathbf{R}_k$ – protein coordinate
s – protein "states"



$$\mathbf{R}_k = \mathbf{R}_k(t; \mathbf{s})$$

Multiscale Model of Lipid Bilayer

To bridge the particle and continuum scales, the relevant degrees of freedom can be described through the framework of a free energy functional.

$$\mathscr{F}\left[\{n_i(\mathbf{r},t)\}\right] = \int_{\mathbb{R}^2} \left(f_{\mathrm{mm}}(\{n_i\}) + \sum_{i=1}^{P} \left[\sum_{j=1}^{N} u_{\mathrm{pm}}^{(i)}(\mathbf{r}-\mathbf{R}_i) n_j(\mathbf{r}) + \frac{1}{2} \sum_{i'=1}^{P} u_{\mathrm{pp}}^{(i)}(\mathbf{r}-\mathbf{R}_i)\delta(\mathbf{r}-\mathbf{R}_{i'}) \right] \right) d\mathbf{r}$$

$$f_{\mathrm{mm}}(\{n_i\}) = \sum_{i=1}^{N} \left(T n_i(\mathbf{r},t) \log\left(\Lambda^2 n_i(\mathbf{r},t)\right) + \frac{1}{2} T \sum_{i'=1}^{N} \int_{\mathbb{R}^2} \Delta n_i(\mathbf{r},t) c_{i,i'}(\mathbf{r}-\mathbf{r}') \Delta n_{i'}(\mathbf{r}',t)\, d\mathbf{r}' + \dots \right)$$

Protein-membrane interaction

Membrane-membrane interaction

Protein-protein interaction

Multiscale Model of Lipid Bilayer

Evolution equations can be obtained from the free energy for both the lipid densities (Dynamic Density Functional Theory) and the proteins (Langevin).

DDFT:
$$\frac{\partial n_i}{\partial t} = \nabla \cdot \left(\beta D_i n_i \nabla \left(\frac{\delta \mathscr{F}}{\delta n_i} \right) \right) + \xi_i$$

Langevin:
$$M_k \frac{d^2 \mathbf{R}_k}{dt^2} = -\nabla \mathscr{F} - \gamma_k \frac{d\mathbf{R}_k}{dt} + \boldsymbol{\eta}$$

- Of course, all parameters must be calculated from the MD simulations.

Multiscale Model of Lipid Bilayer

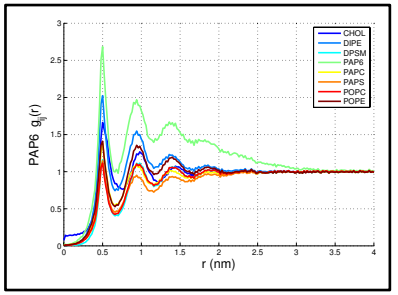
$$\mathscr{F}\left[\{n_i(\mathbf{r},t)\}\right]$$

- Dynamic density functional theory
- Langevin dynamics

$$\frac{\partial n_i}{\partial t} = \nabla \cdot \left(\beta D_i n_i \nabla \left(\frac{\delta \mathscr{F}}{\delta n_i}\right)\right) + \xi_i$$

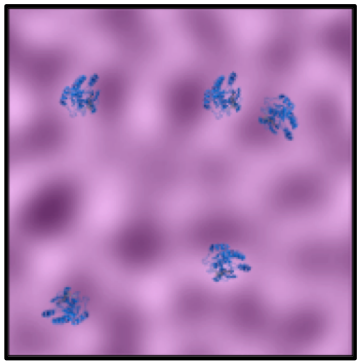
$$M_k \frac{d^2\mathbf{R}_k}{dt^2} = -\nabla \mathscr{F} - \gamma_k \frac{d\mathbf{R}_k}{dt} + \boldsymbol{\eta}$$

- Transport coefficients

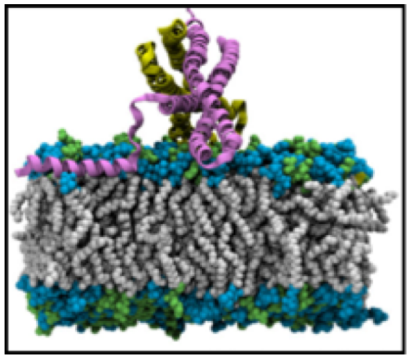


- Statistical correlations
- Potentials of mean force

continuum scale

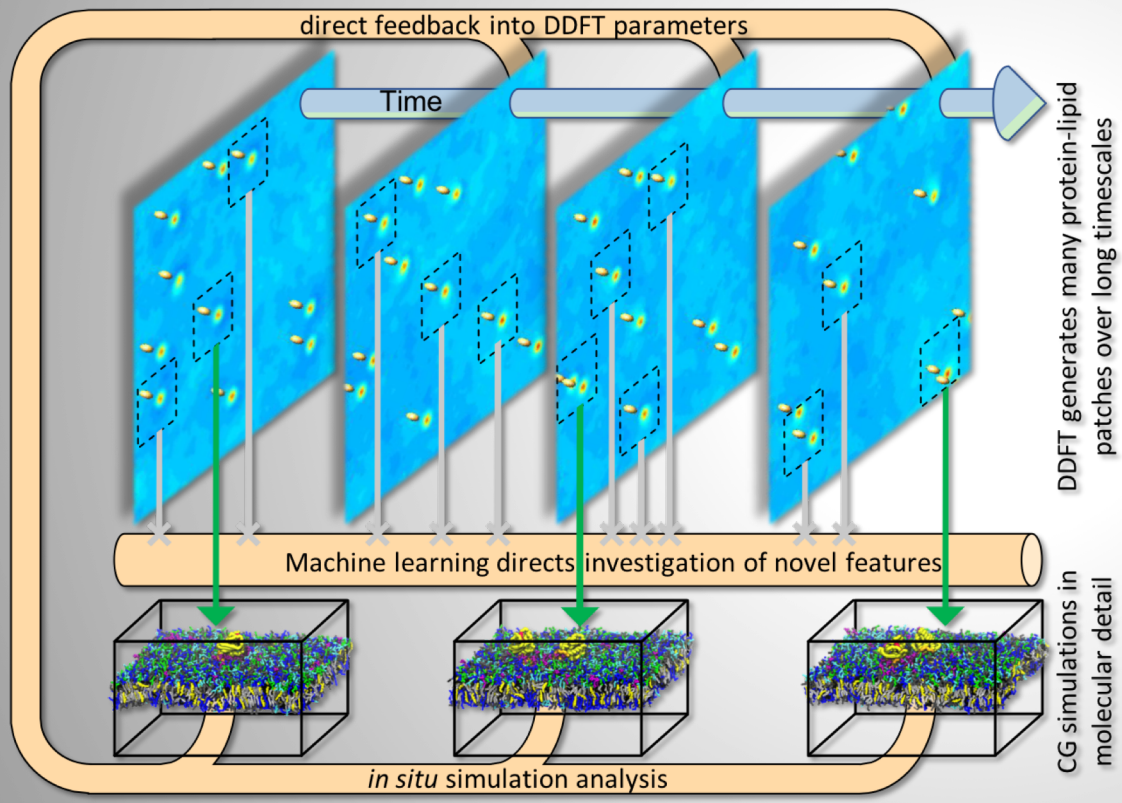


- Bulk parameters
- Boundary conditions
- Significant events detected by Machine Learning



particle scale

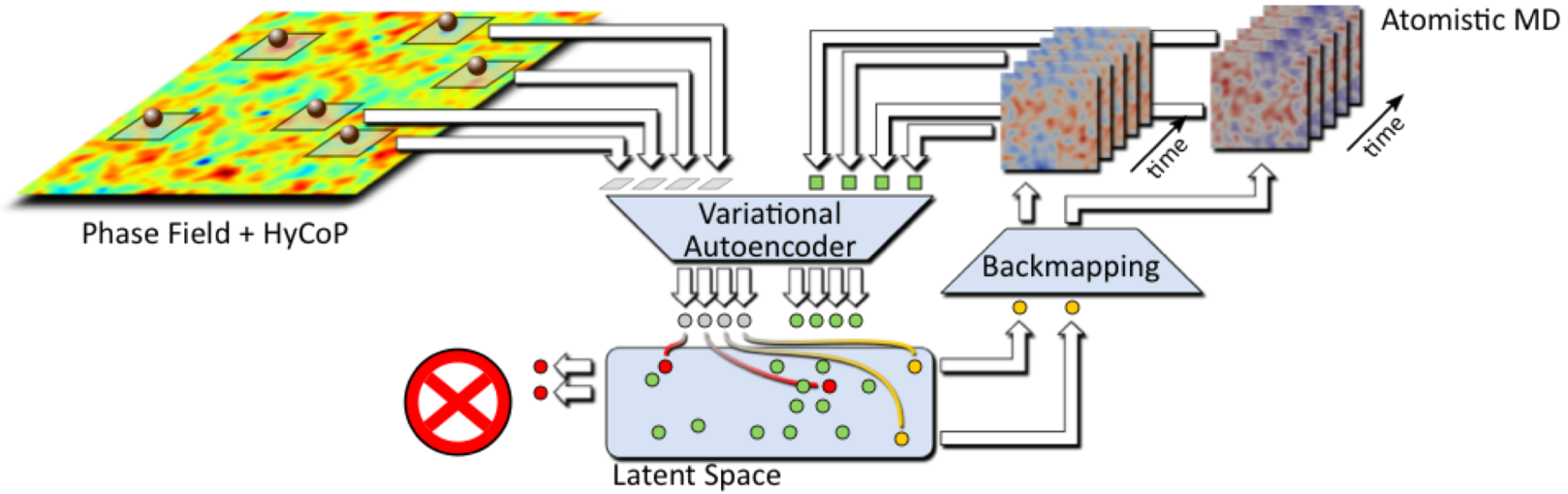
Exploit machine learning to guide simulation investigation



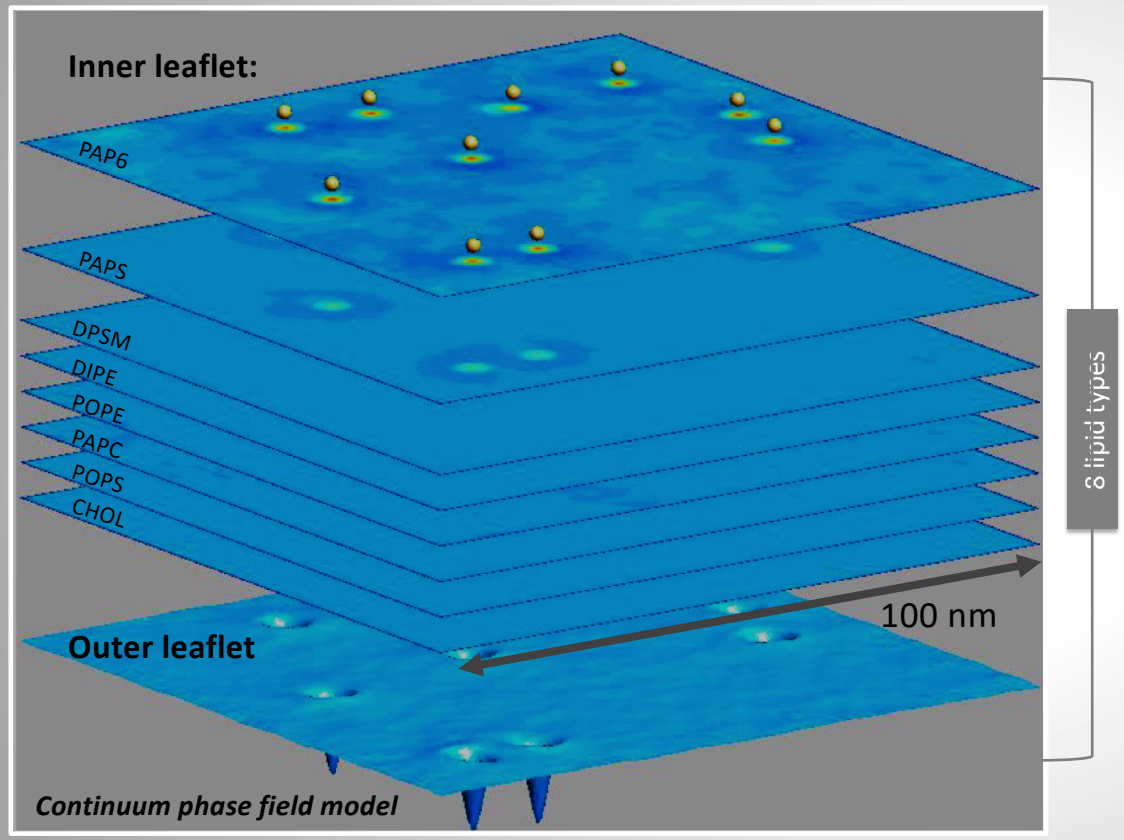
- Learn reduced order representation of high dimensional parameter space
- Define “similarity” as Euclidian distance in reduced dimensions
- Identify areas that are dissimilar in continuum simulation
- Initiate molecular dynamics simulations to explore maximally dissimilar conditions

Steering Multi-Scale Simulations By Adaptively Sampling Data Driven Latent Spaces

- Train latent space representing space of relevant lipid configurations
- Dynamically sample configuration space to understand RAS-membrane interactions at macro time-scales with micro precision



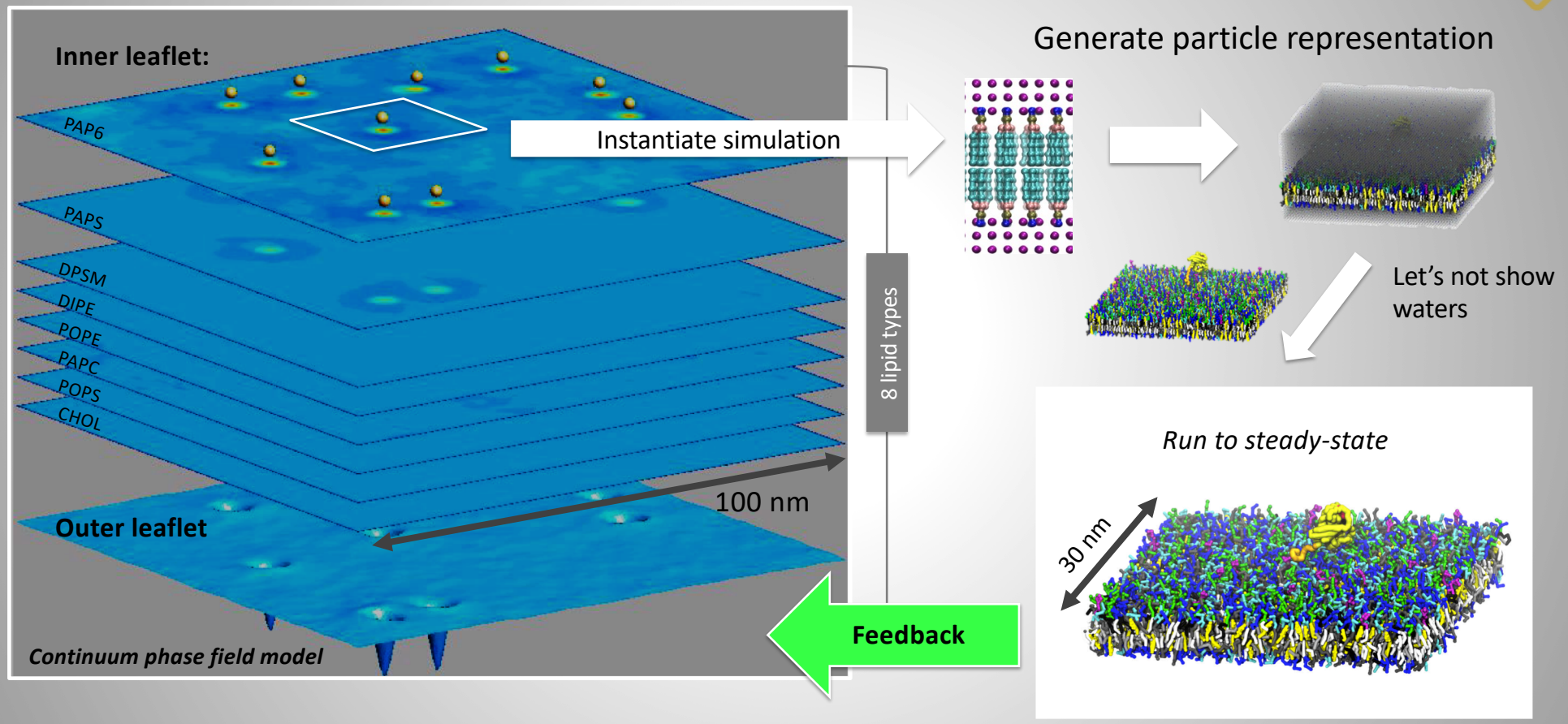
Demonstrated multi-scale lipid/protein modeling capability



- Incorporate particle degrees of freedom into continuum (phase field) model
- Use AI techniques to identify "most interesting" region in continuum simulation
- Initiate fine-scale simulation using continuum environment
- Rigorously self-consistent interaction energies

10 k-Ras proteins in 100 nm X 100 nm membrane

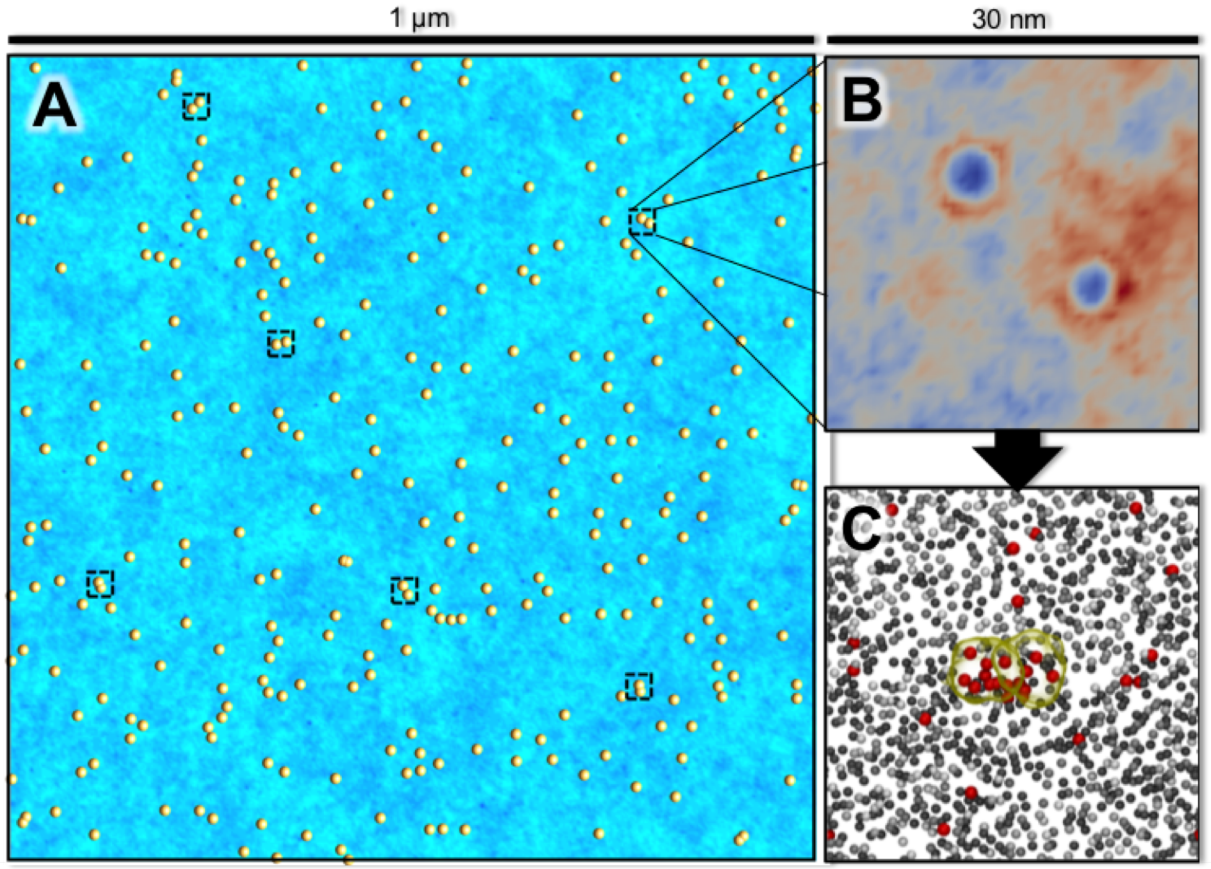
Demonstrated multi-scale lipid/protein modeling capability



Initial Science Runs on Sierra Supercomputer

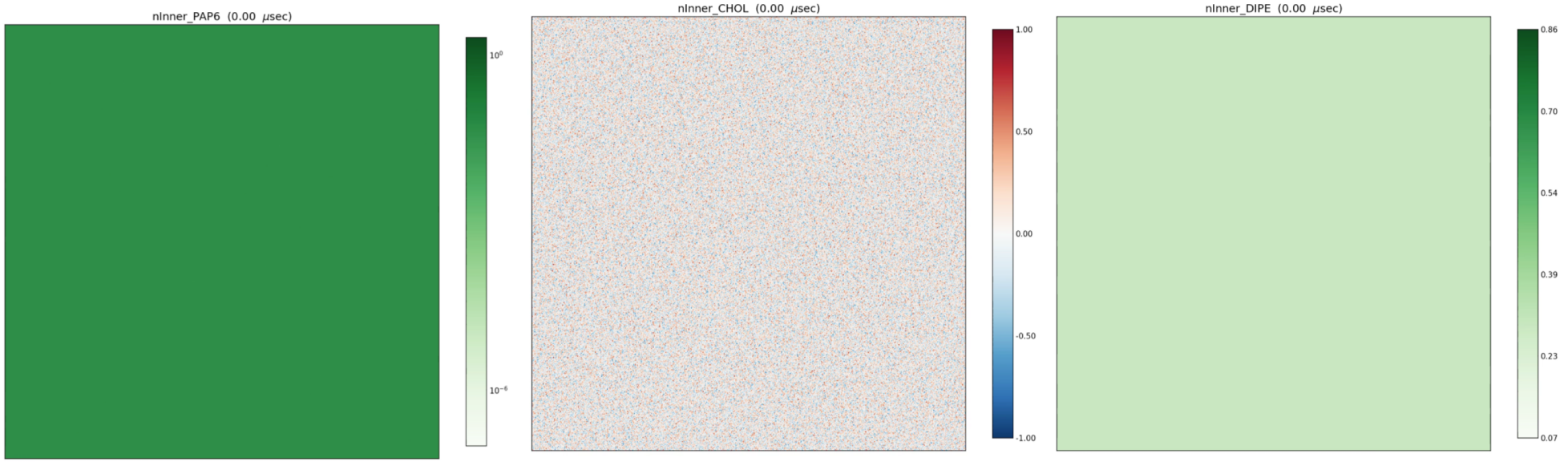
First-of-a-kind simulations will explore:

- Dependence of RAS mobility and dynamics as a function of membrane environment
- Aggregation of RAS in context of realistic membrane
- Effect of RAS concentration on local membrane composition



Preliminary Results

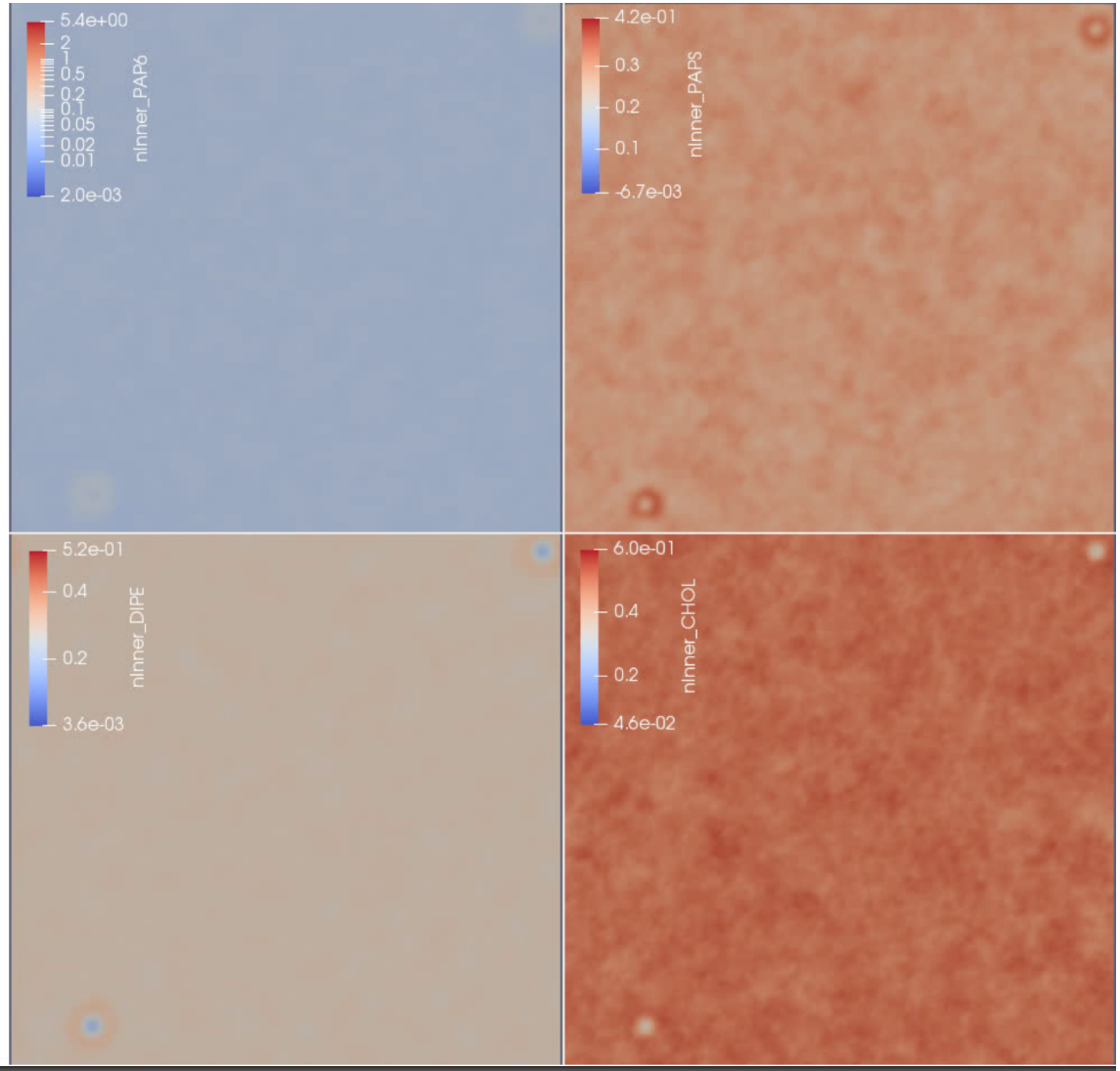
Witnessing formation of large-scale fluctuations in lipid structure through long-time scale simulations



PAP6 Concentration (log scale) **Chol Concentration** **DIPE Concentration**

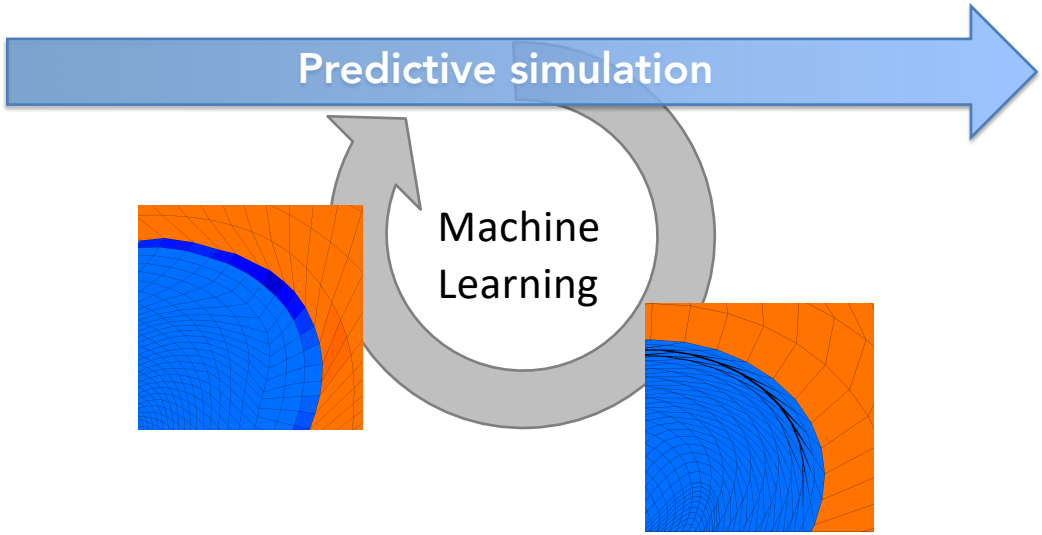
Preliminary Results

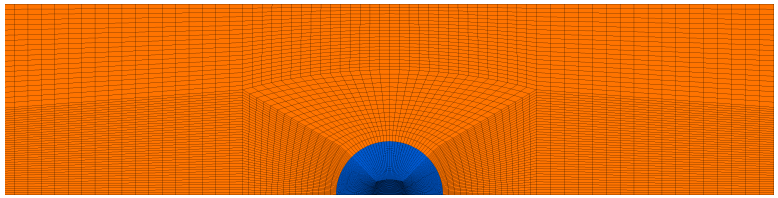
- 50nm X 50nm high-res study w/ 2 RAS proteins (40 μs)
- Investigate phenomena witnessed originally in μm X μm scale simulation
- Aggregation/repulsion of charged lipids (PAP6, PAPS, DIPE, CHOL) following "collision" of RAS
- Unusual stability of formation is unexpected – currently under investigation
- Results demonstrate importance of time and length-scale for simulation



Two ways we envision using machine learning techniques along with predictive simulation

Machine learning "on the inside"

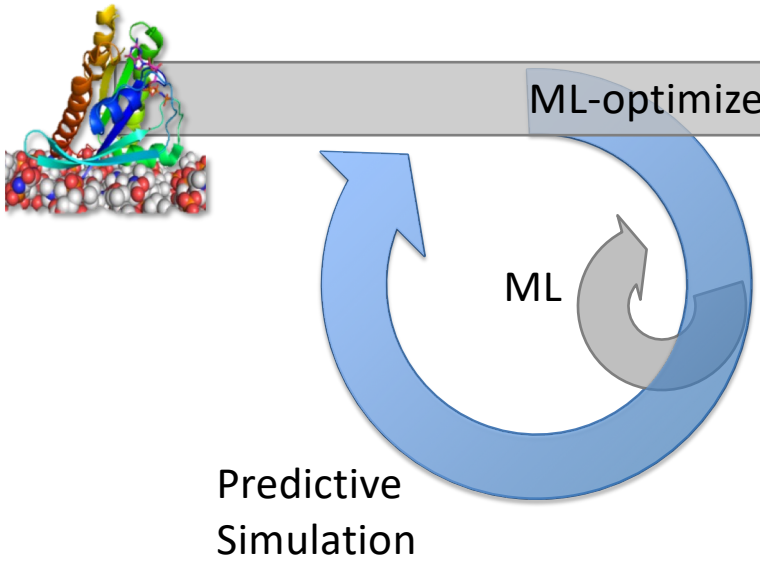




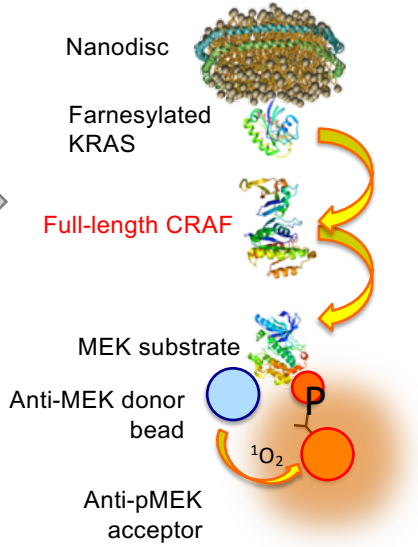
Machine learning to identify bottlenecks and optimize running simulations

Two ways we envision using machine learning techniques along with predictive simulation

Machine learning "on the outside"



Optimize solutions with significant reduction in compute requirements



Integrated workflows develop insight faster

NCI-DOE Pilots: Multi-institution/multi-disciplinary teams

FNLCR / NCI: Debanjan, Goswami, Gulcin Gulten, Rebika Shrestha, **Andrew Stephen**, Tommy Turbyville, Que Van

Oak Ridge National Lab: Debsindhu Bhowmik, **Arvind Ramanathan**

Los Alamos National Lab: Boian Alexandrov, **Angel Garcia**, Nick Hengartner, Jeevapani Hettige, Christoph Jungans, Cesar Lopez, Chris Neale, Sandrasegaram Gnanakaran, Tim Travers, Art Voter

Lawrence Livermore National Lab: Ryan Berg, Harsh Bhatia, Timo Bremer, Tim Carpenter, Gautham Dharuman, Francesco Di Natale, **Jim Glosli**, Helgi Ingolfsson, Piyush Karande, **Felice Lightstone**, Tomas Oppelstrup, Liam Stanton, Shiv Sundram, Michael Surh, Brian Van Essen, Xiaohua Zhang