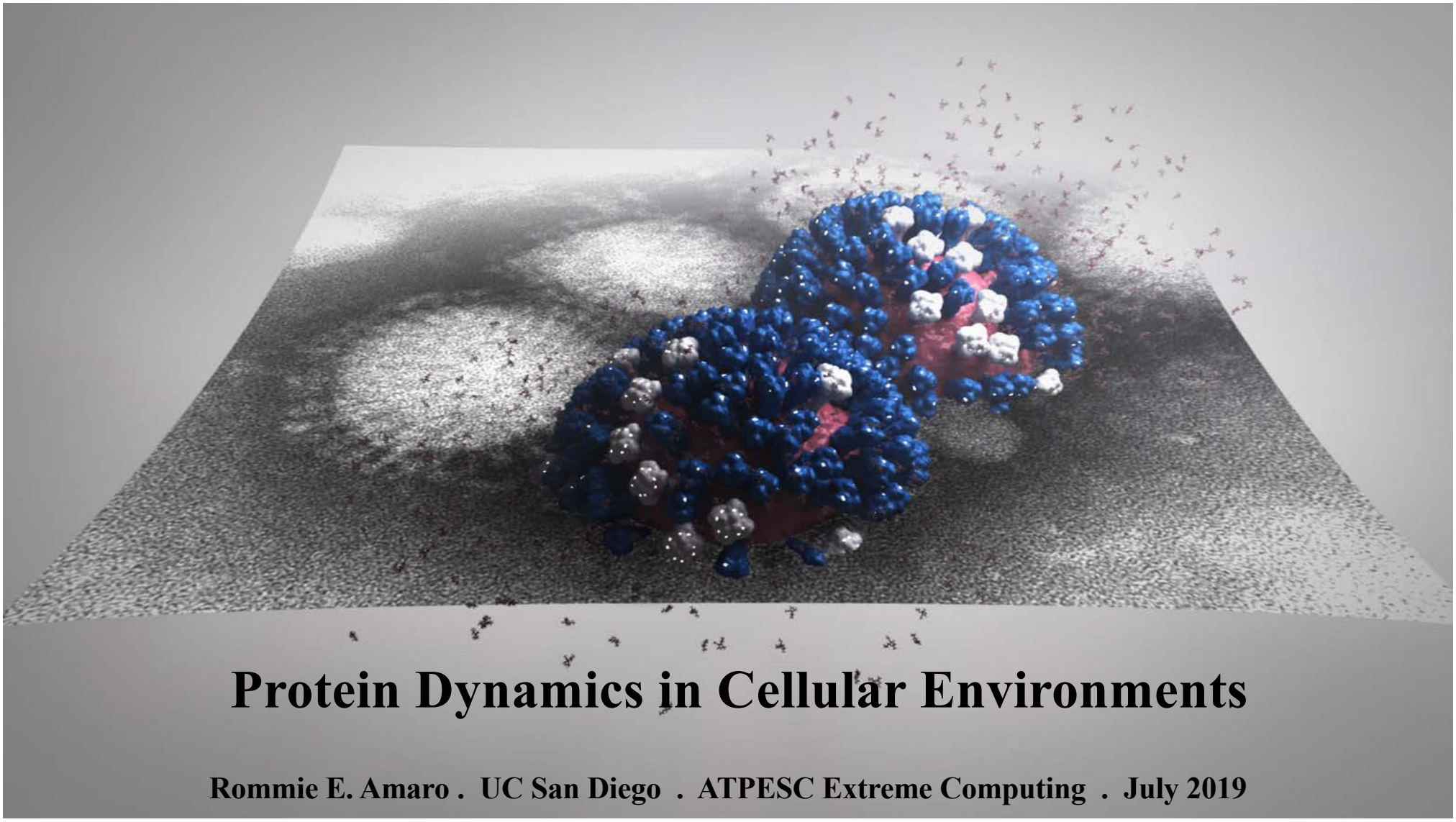


Protein Dynamics in Cellular Environments

Rommie E. Amaro . UC San Diego . ATPESC Extreme Computing . July 2019

Convergence of HPC, data science, & data enabling transformative advances at the intersection of observational and simulation sciences

360,000 cores + GPU acceleration

National Strategic Computing Initiative (NSCI)

Exascale

MolSSI

Anton

Compute Power

HP 735
12 CPUs

1993
protein
10k atoms
100s ps

SGI Origin
128 CPUs

1997
ion channel
100k atoms
1 ns

LeMieux
3k CPUs

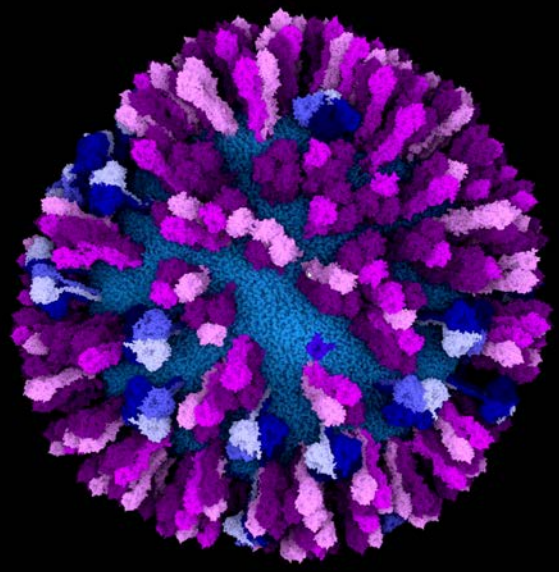
2002
ATPase
500k atoms
10s ns

Ranger
60k CPUs

2007
ribosome
2 mil atoms
100s ns

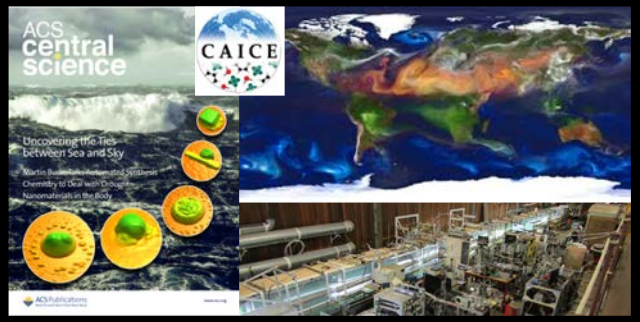
2015
Enveloped virus
160 mil+ atoms
1-100 μs

time

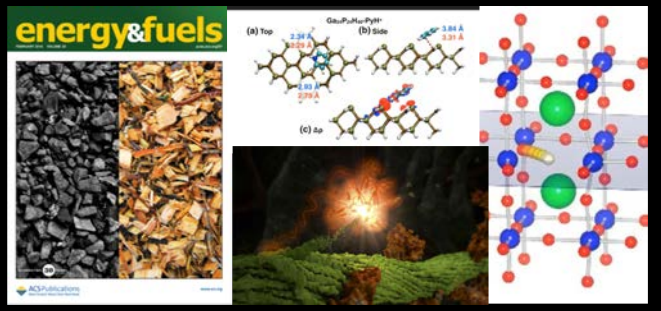


100 nm, 160 million atoms, μs

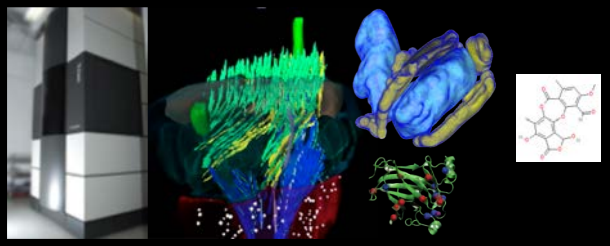
Research breakthroughs will occur at the interface of observational and simulation science



Environmental & atmospheric chemistry
(Climate Change)



Energy & Materials
(Sustainable Energy)



Biology & medicine
(New Therapeutics)

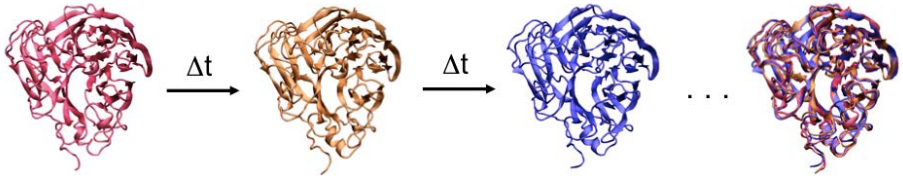

Influenza
Cancer
Chlamydia
Trypanosomiasis

Molecular Dynamics Simulations as a Computational Microscope

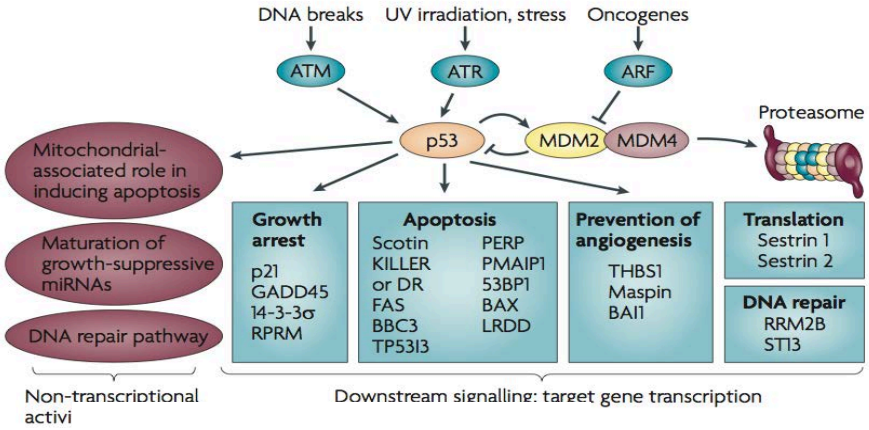
Chemistry, Physics, Math, Software Supercomputers & GPUs

$$U(\vec{R}) = \sum_{bonds} k_{bond}(r - r_o)^2 + \sum_{angles} k_\theta(\theta - \theta_o)^2 + \sum_{dihedrals} k_{dihed}[1 + \cos(n\phi + \delta)] + \sum_{\substack{nonbonded \\ paris}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{\substack{nonbonded \\ pairs}} \frac{q_i q_j}{\varepsilon r_{ij}}$$

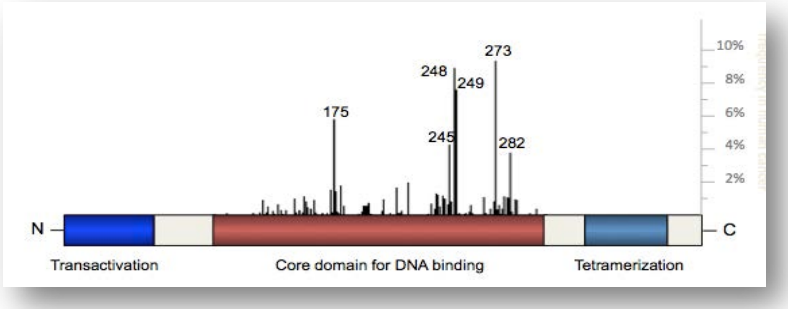
$$\vec{F}_i = ma = m_i \frac{d^2\vec{r}_i}{dt^2} = -\vec{\nabla}U(\vec{R})$$



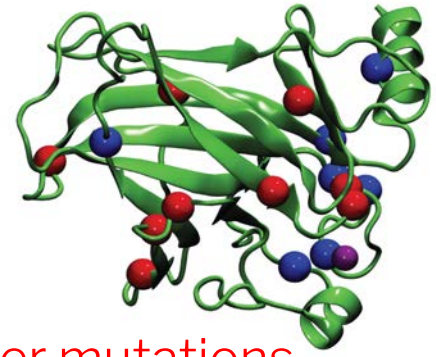
p53: Guardian of the genome



Frequency of p53 mutations in cancer



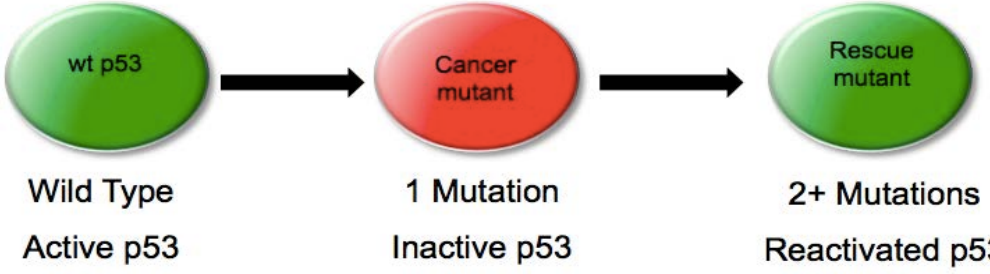
>600,000 new cancer patients annually in the US with p53 point mutations



Cancer mutations

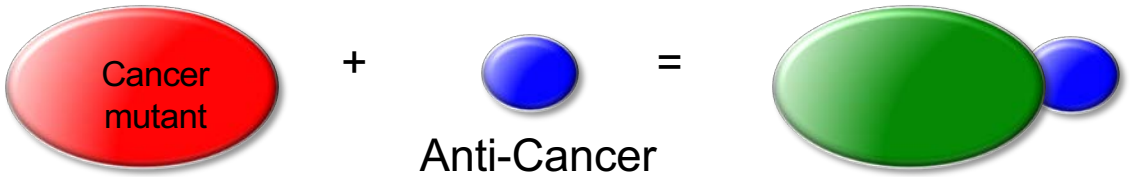
Cancer rescue mutations

Susceptible to oncogenic mutations that inactivate by lowering its stability

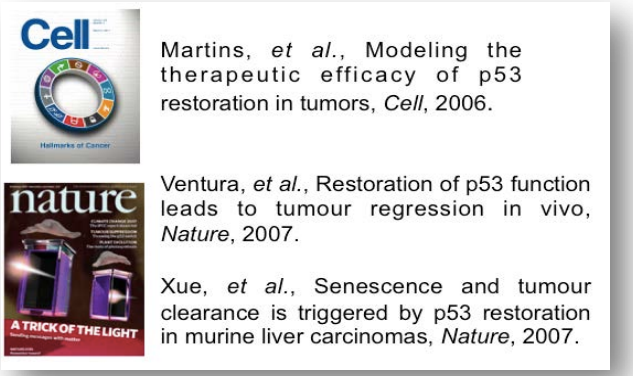


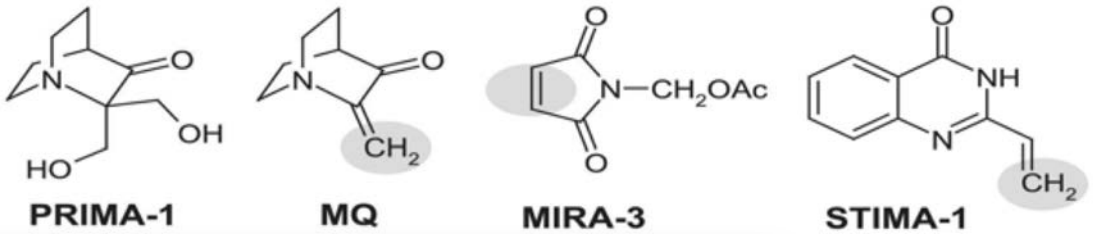
Brown *et al.* (2009) *Nature Reviews. Cancer, 9(12), 862–873*

Dream of cancer biologists: small-molecule p53 reactivation

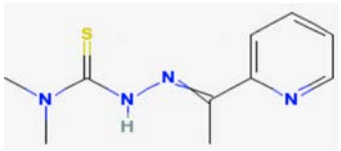


Inactive p53 | Anti-Cancer Drug | Reactivated p53

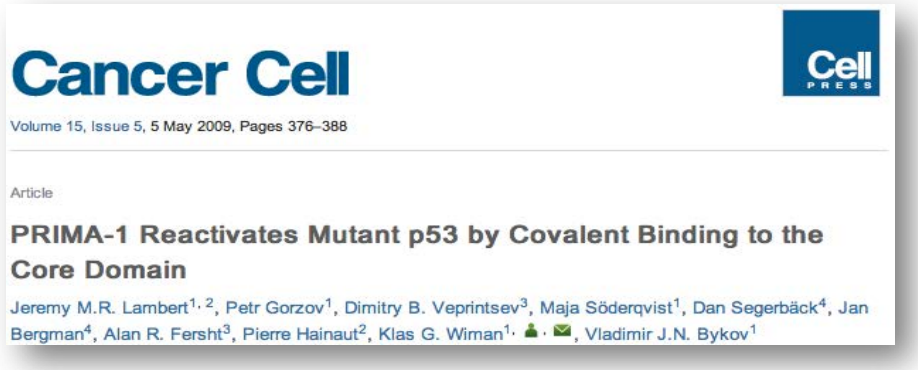




PRIMA-1 **MQ** **MIRA-3** **STIMA-1**

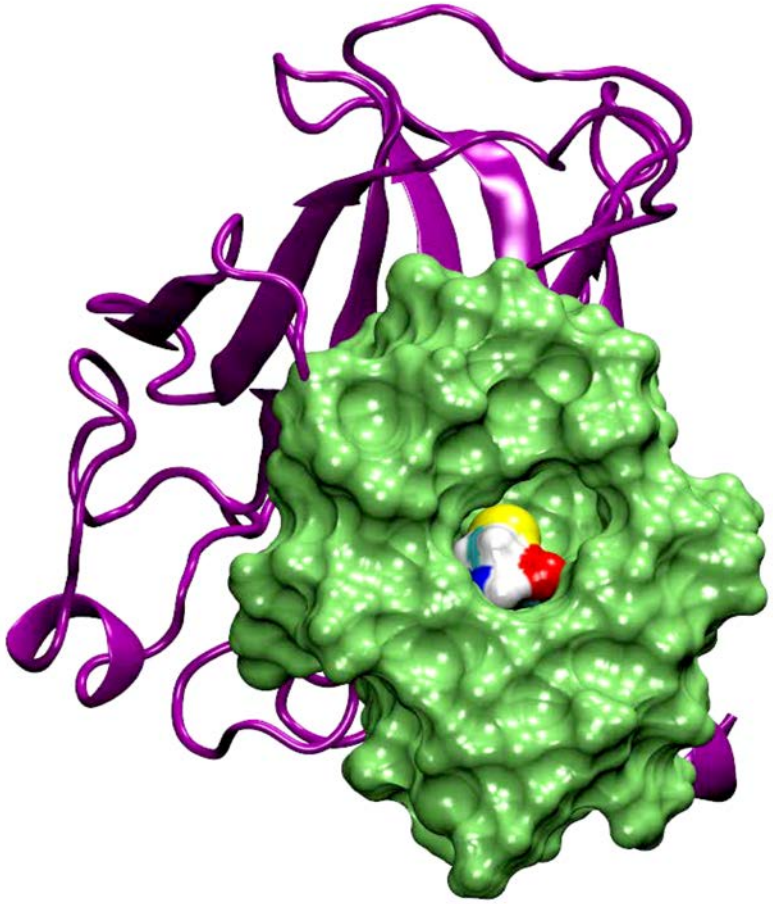


NSC319725



Identified covalent attachment of products, but could not discern which of 10 cysteine residues

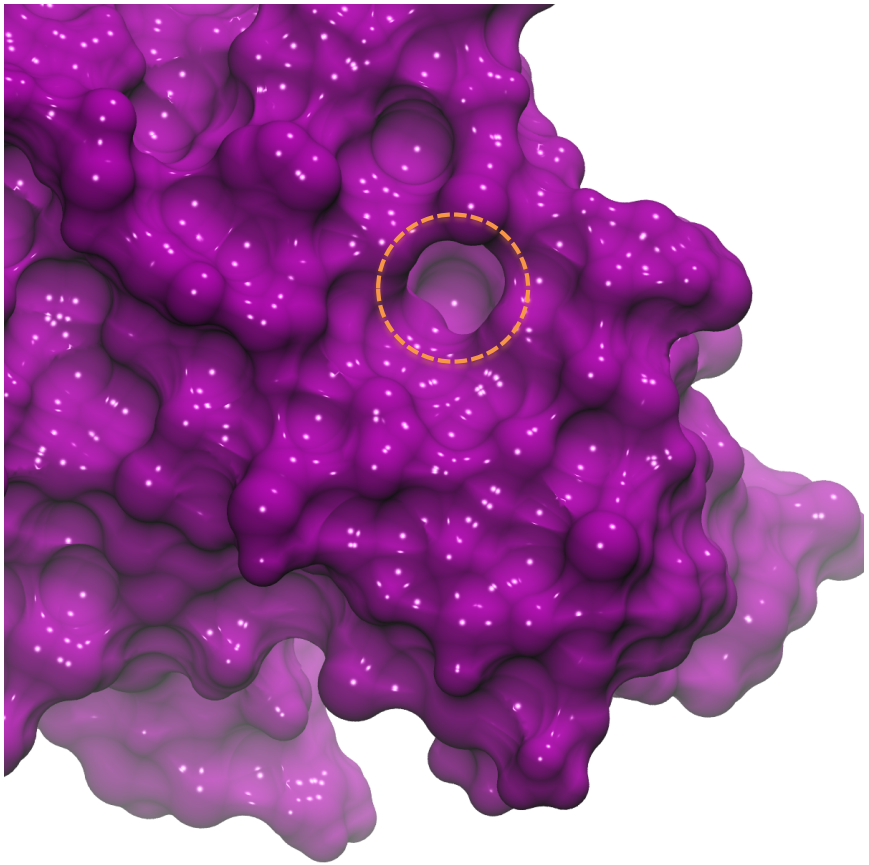
Simulations Reveal Target Flexibility



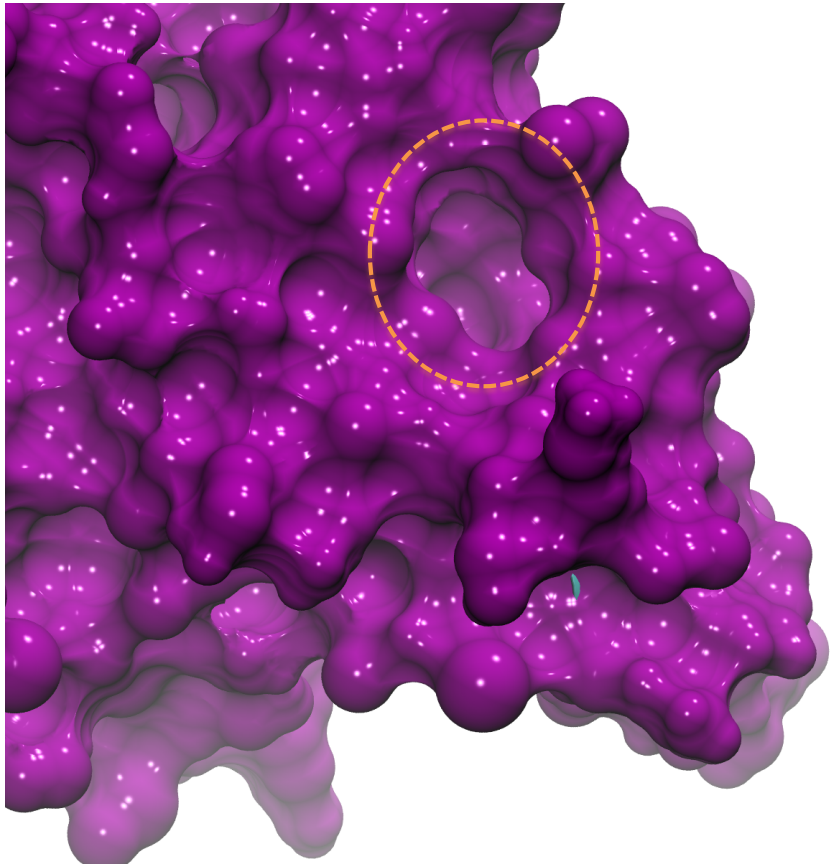
5% exposed,
matches NMR

Wassman, Baronio, Demir, et al. Nature Comm., (2013)

New Site Opens

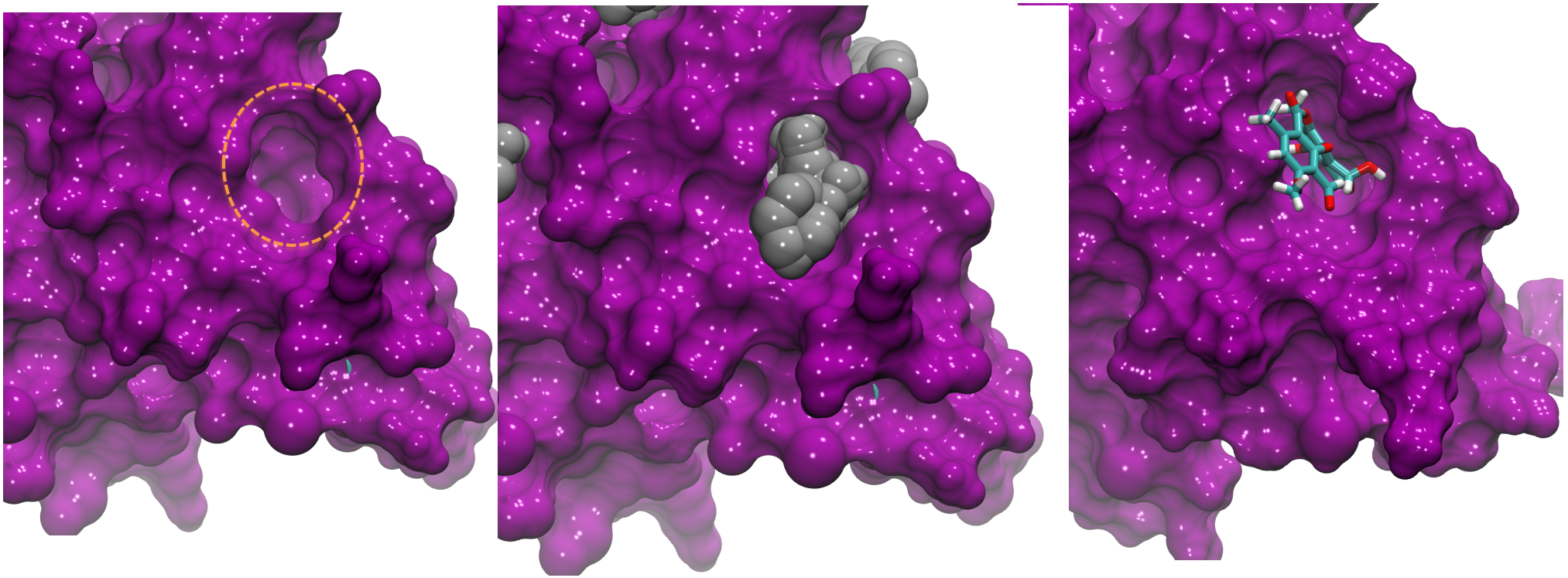


> 95 X-ray structures



“Open” MD structure

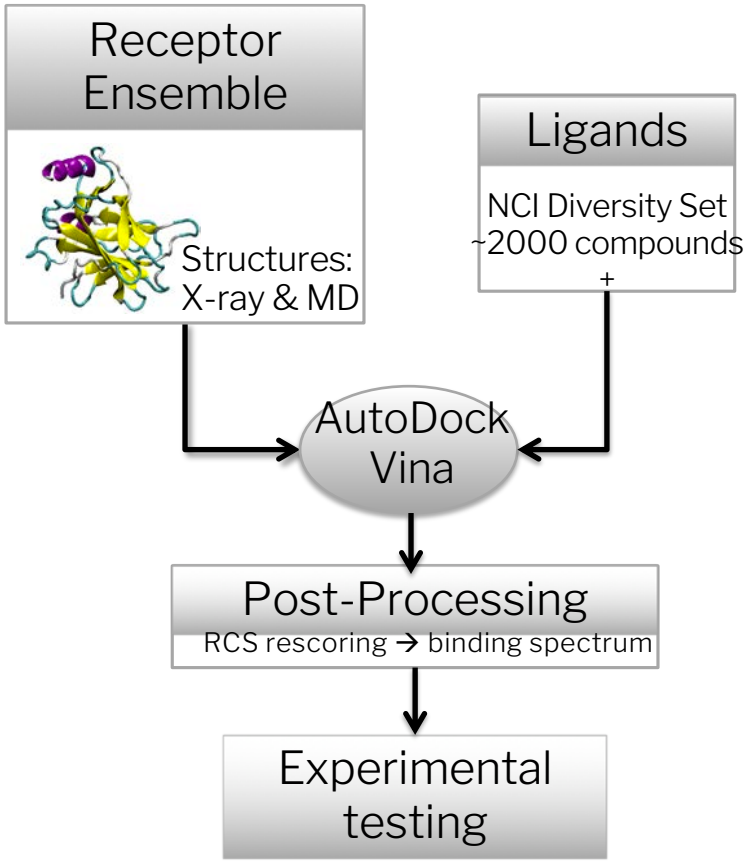
New Site is Druggable

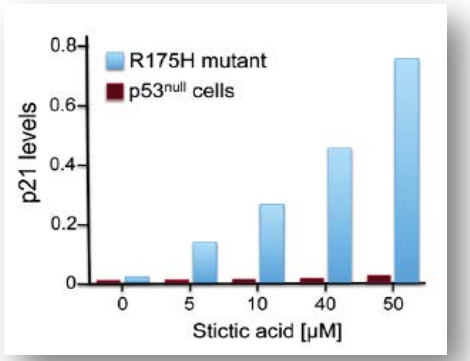


Wassman, Baronio, Demir, et al. Nature Comm., (2013)

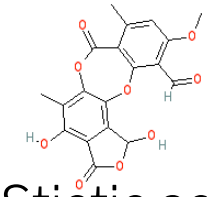
Vajda et al., Computational Solvent Mapping: http://ftmap.bu.edu/

Discovery of novel reactivation compound & rationalization of clinical trial compound

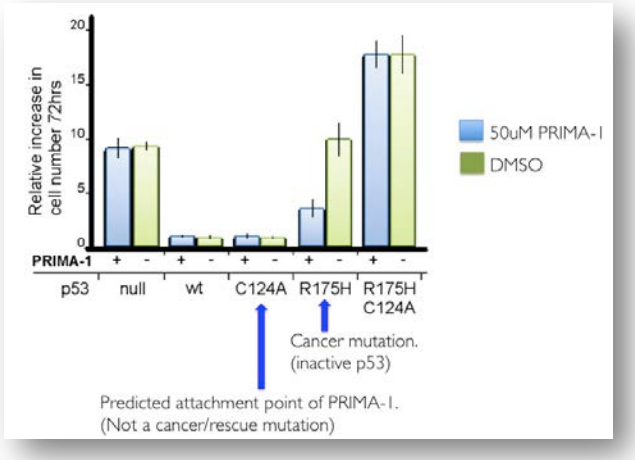




Dose-dependent rescue in mammalian cancer cells

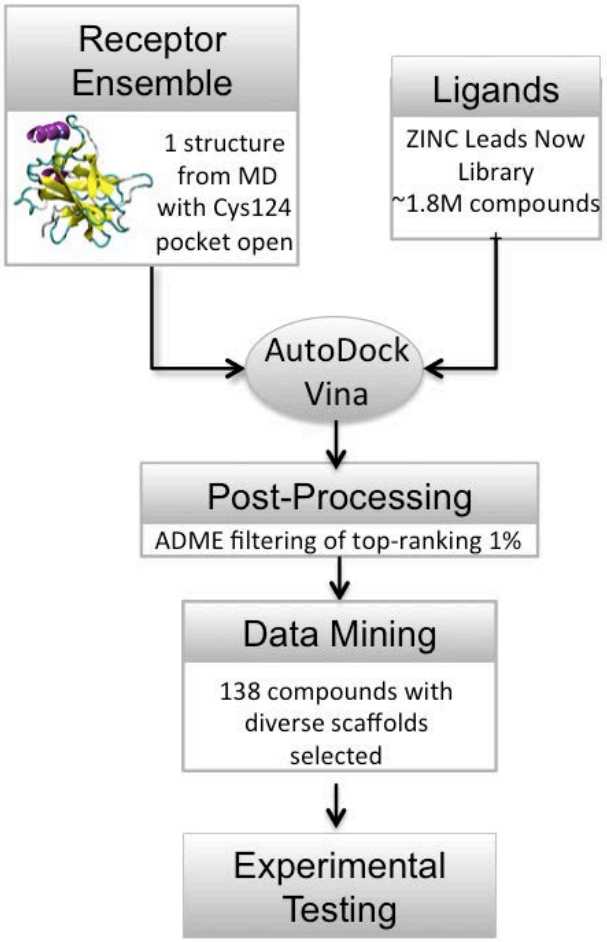


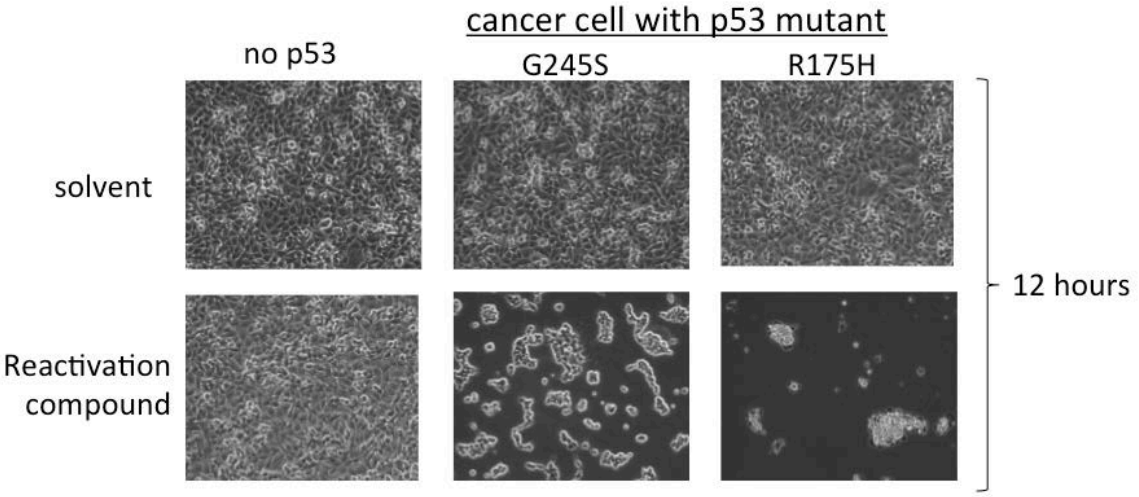
Stictic acid



Wassman, Baronio, Demir, et al. Nature Comm., (2013)

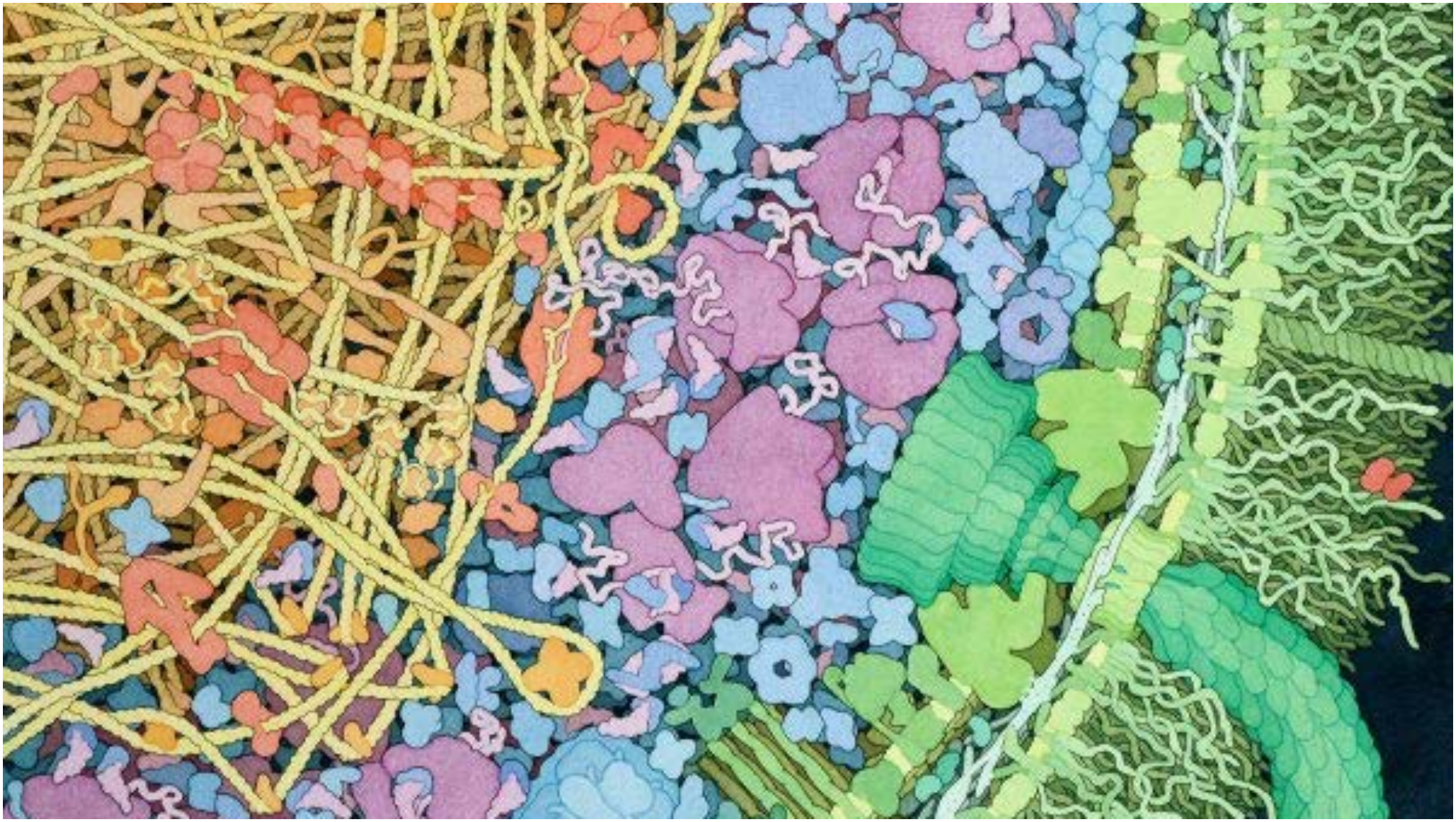
Our computational approach discovers more novel p53 reactivation compounds in 6 months than all the research efforts of the previous 20 years combined



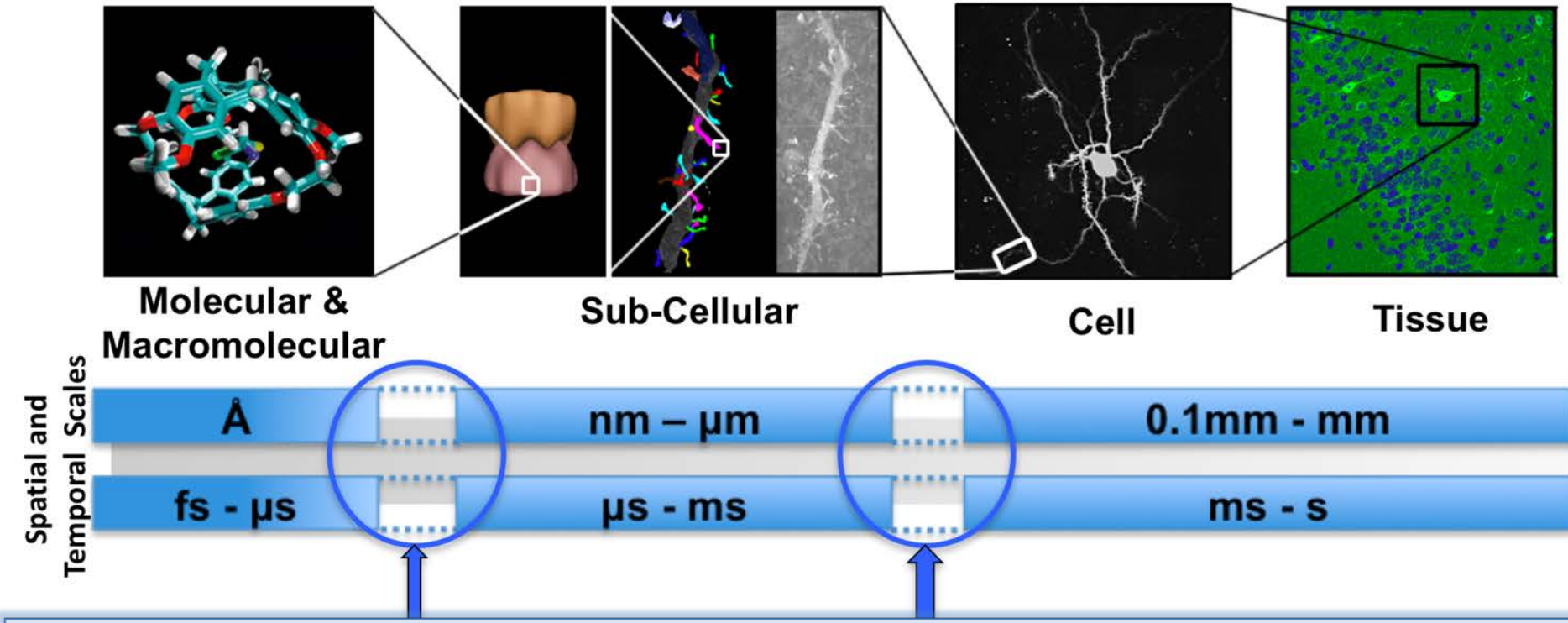


15/138 compounds tested in mammalian cancer cell lines rescue p53 activity and kill cancer cell

Actavalon



Computational biophysics bridges gaps across scales



Need to go into and across key "capability gaps";
Computational methods to give unseen views
into the inner workings of cells at the molecular level

Volume 2, Article No. 0148, 2018

PERSPECTIVES

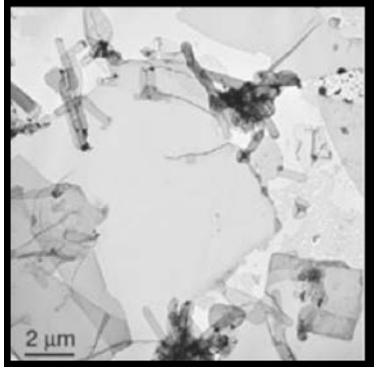
Multiscale methods in drug design bridge chemical and biological complexity in the search for cures

Rommie E. Amaro and Adrian J. Mulholland

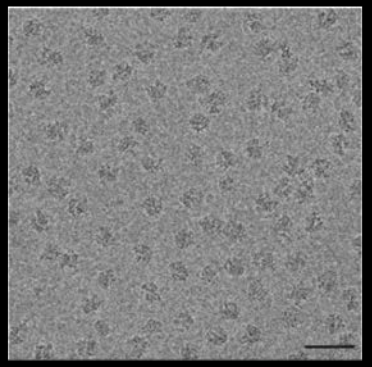
Abstract | Drug action is inherently multiscale: it connects molecular interactions to emergent properties at cellular and larger scales. Simulation techniques at each of these different scales are already central to drug design and development, but methods capable of connecting across these scales will extend our understanding of complex mechanisms and our ability to predict biological effects. Improved algorithms, ever-more-powerful computing architectures and the accelerating growth of rich data sets are driving advances in multiscale modelling methods capable of bridging chemical and biological complexity from the atom to the cell.

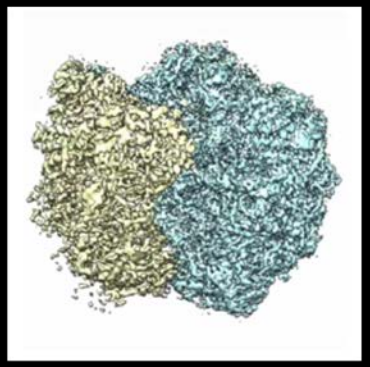
3D structural data to build visible virtual cells

Electron crystallography
2-D crystals of membrane proteins in their native environment

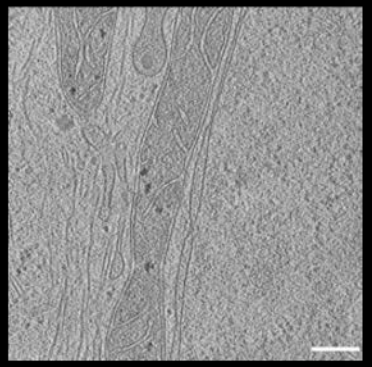


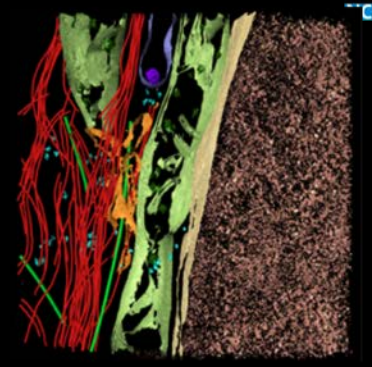
Single-particle analysis
Purified molecules in solution ~0.2-10 MDa



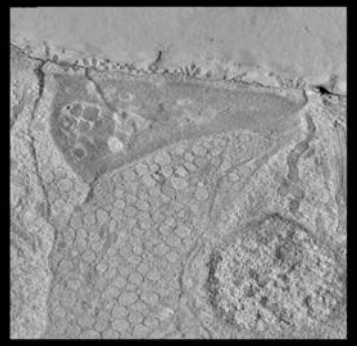


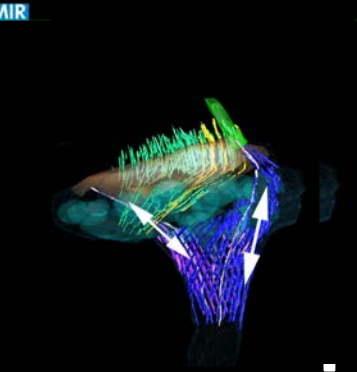
Electron tomography
Pleomorphic samples, e.g., cells and organelles



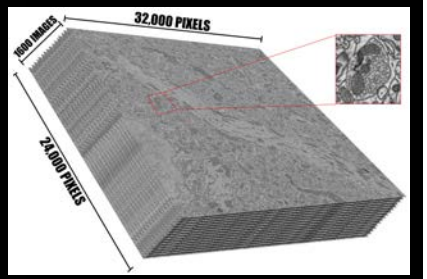


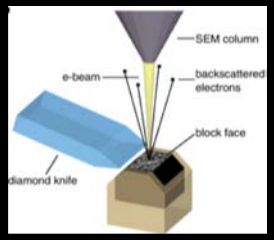
Serial Section EM
Resin-embedded samples





Serial Block EM
Resin-embedded tissues

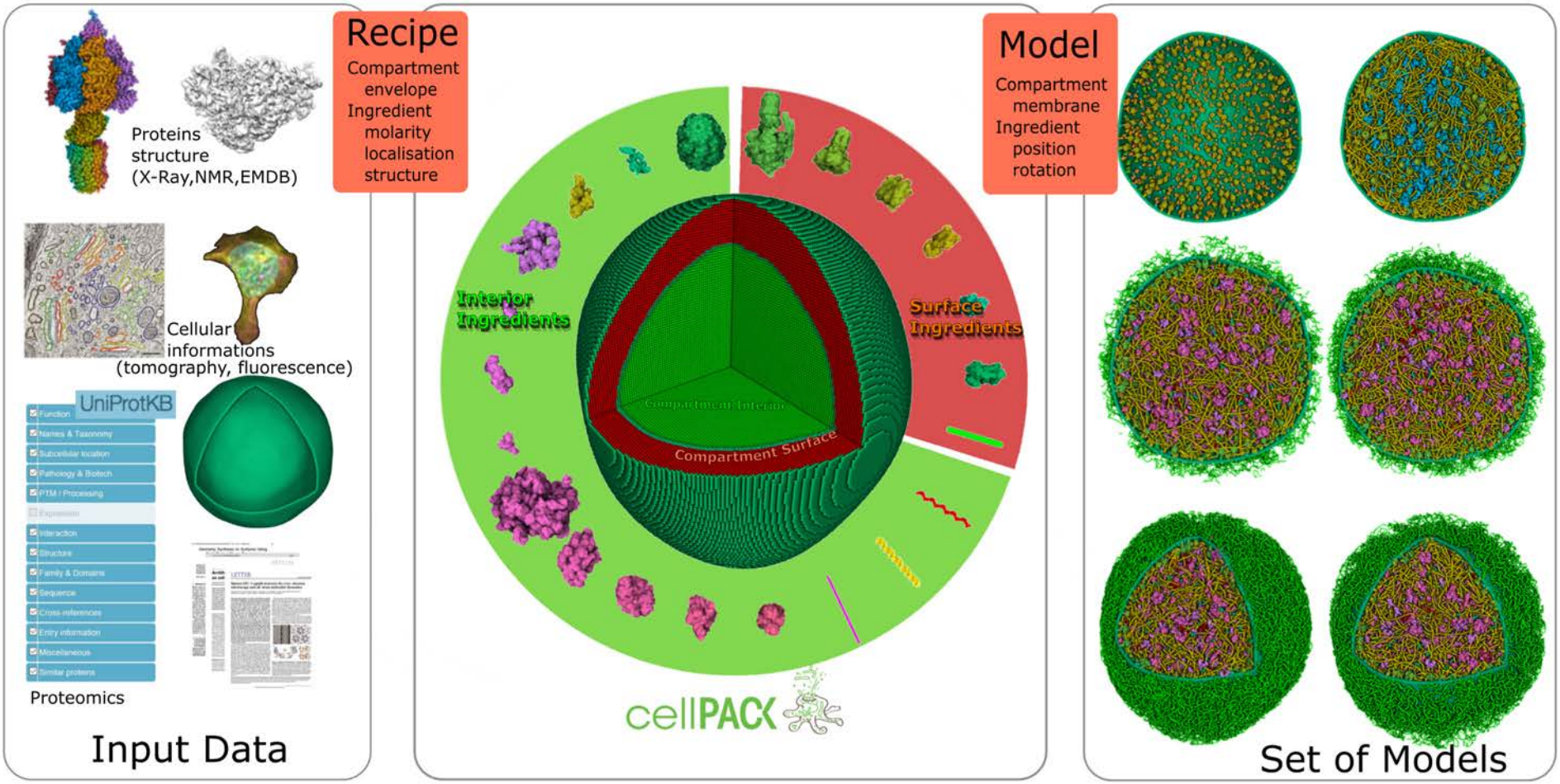




Routine dataset is 1.2 trillion pixels

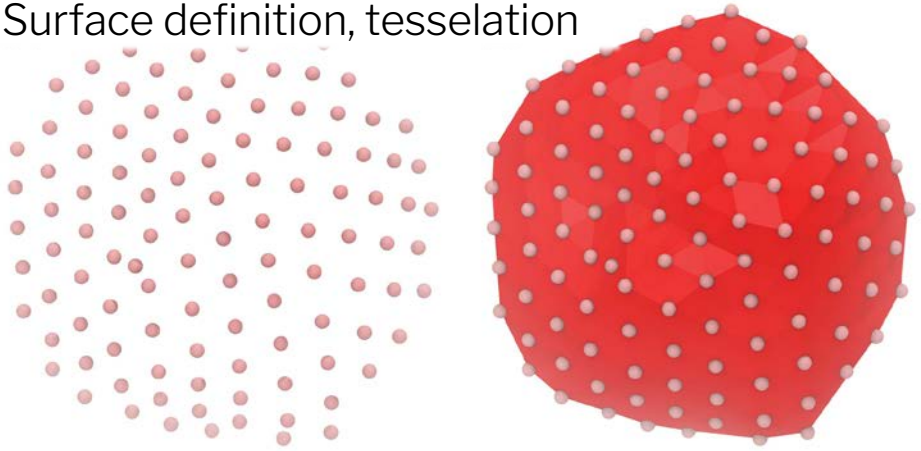
- 100,000's of structures in a single dataset

Cell-centered, data-centric modeling framework

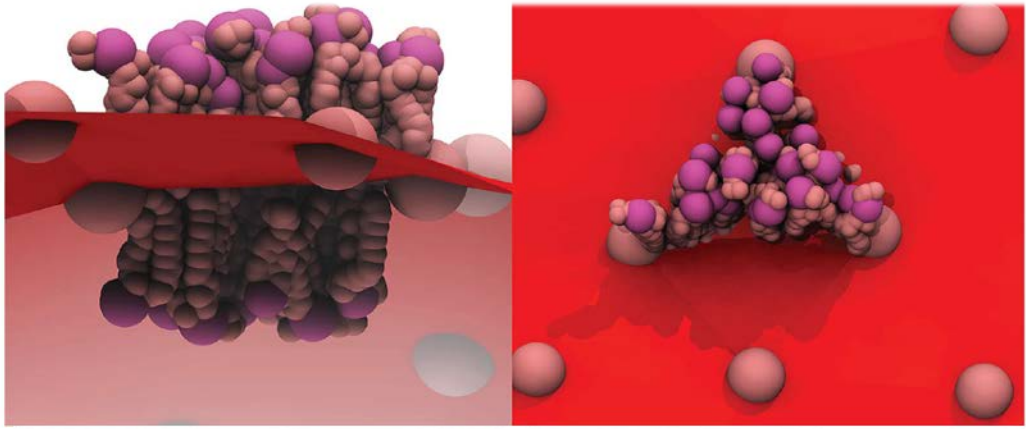


Lipid bilayers with realistic geometries

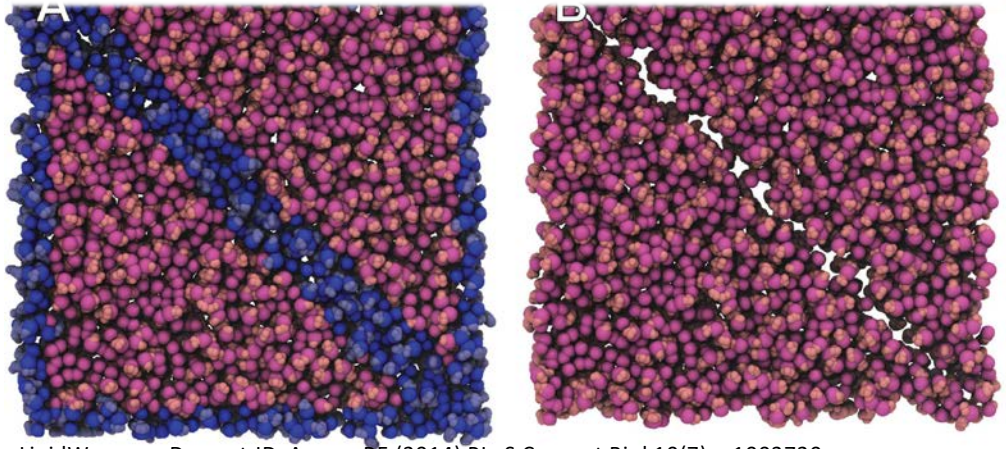
Surface definition, tesselation



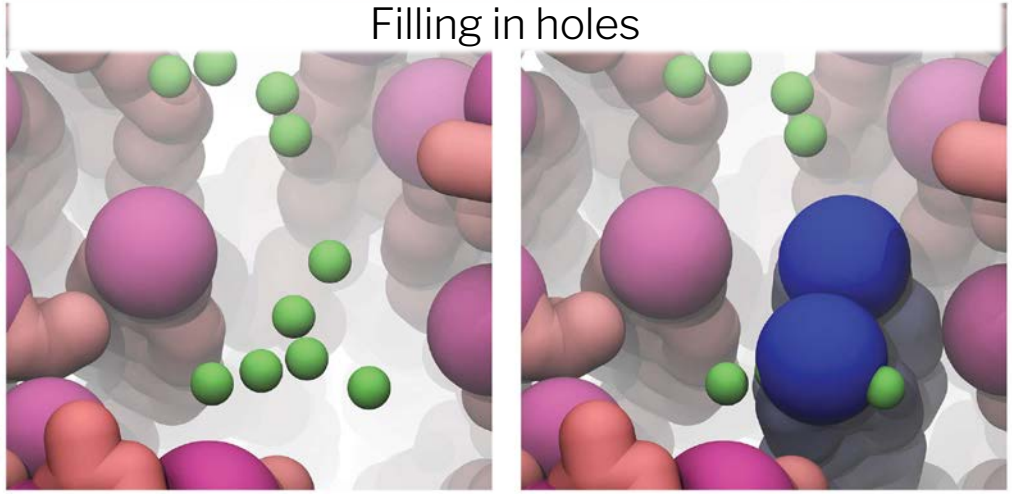
Lipid triangles positioned



Remove clashing lipids

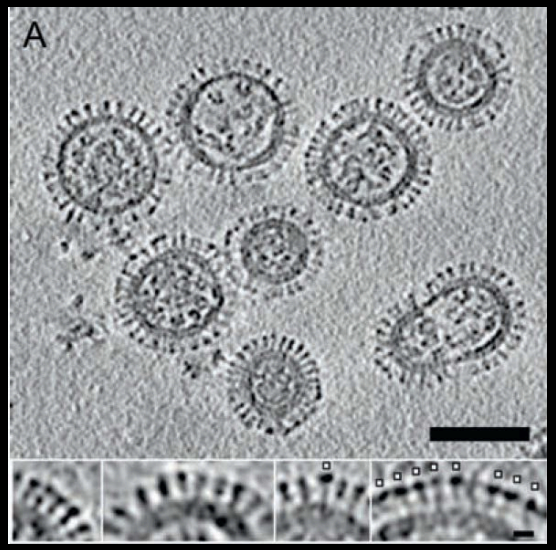


Filling in holes



LipidWrapper: Durrant JD, Amaro RE (2014) PLoS Comput Biol 10(7): e1003720.

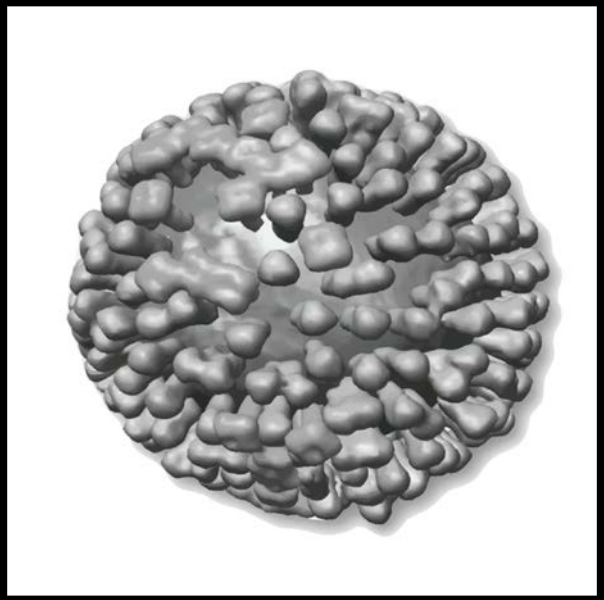
Moving from single protein to whole virus



Alasdair Steven, NIH

Fully Atomic Reconstructions

PyMolecule
LipidWrapper
CellPACK



- Improved sense of the physical arrangement of biological entities in complex biological milieu
- Enables simultaneous study of multiple components
- Mesoscale molecular models as a platform for other simulation approaches (e.g., Brownian dynamics, Mcell, lattice boltzmann MD)

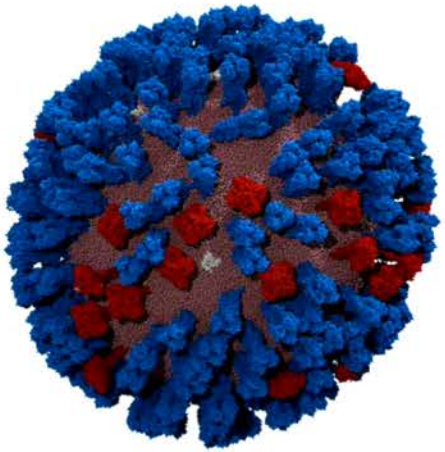
... leads us to new avenues of investigation, not possible on the single protein scale

Johnson et al, Nature Methods (2014),; Durrant & Amaro, PLOS Comp Bio (2014).

In-silico pandemic 2009 H1N1 Influenza Virus

~160 millions atoms

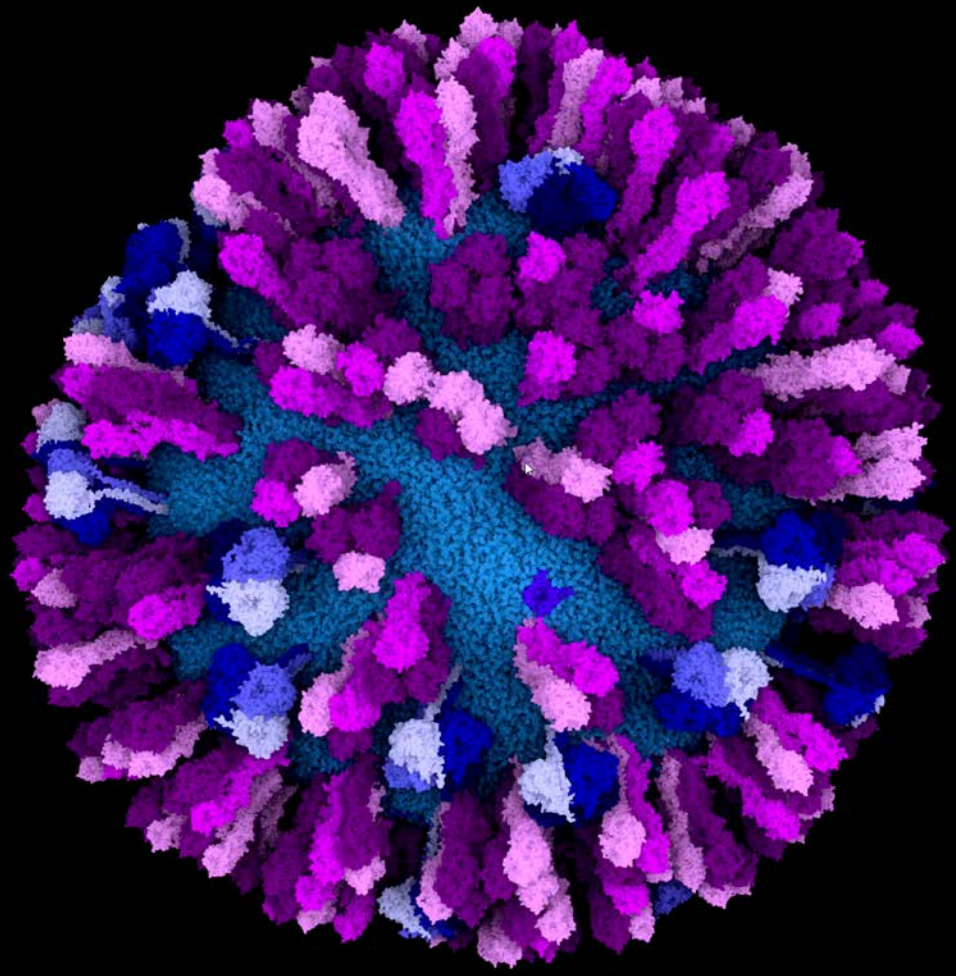
Explicit water
(115 nm x 120 nm x 116 nm)



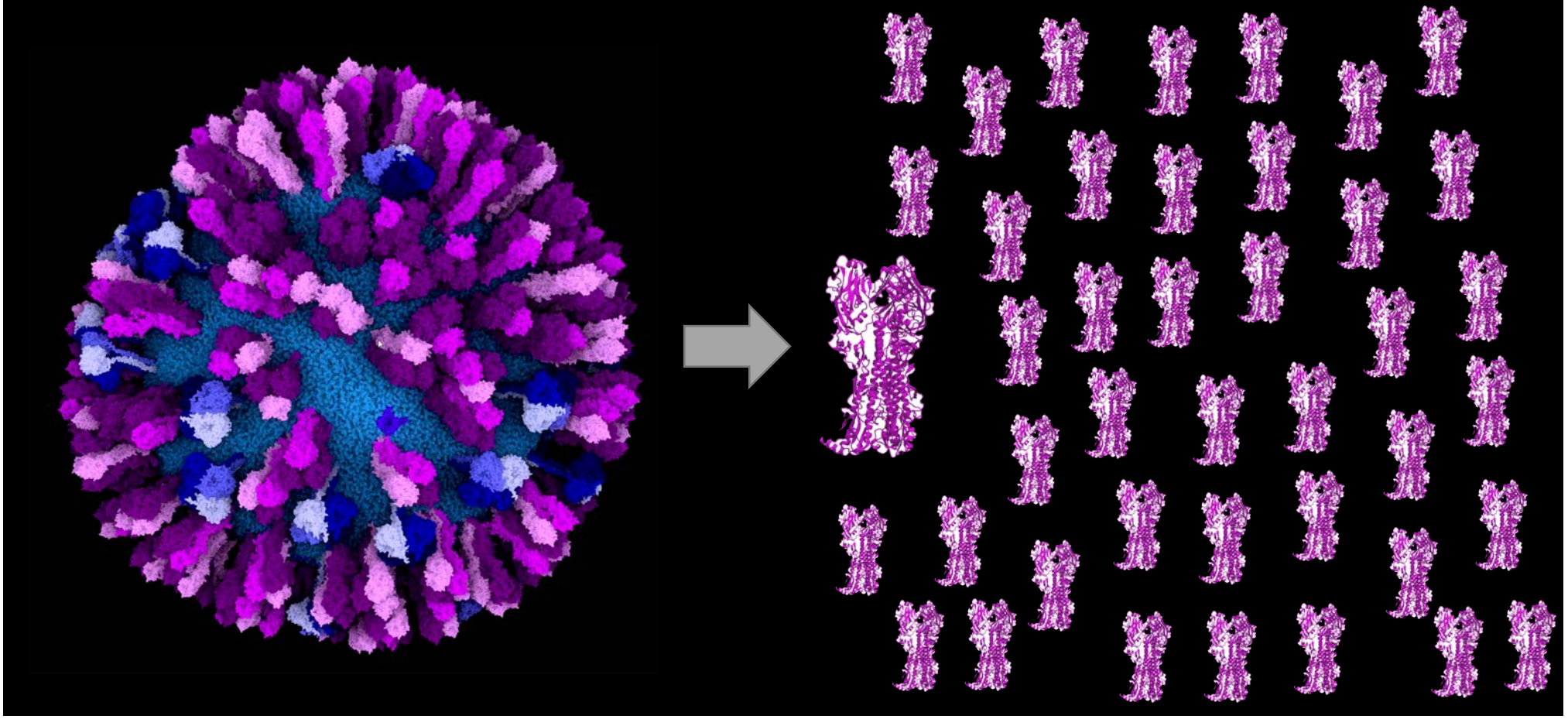
~110 nm diameter

708 HA monomers
(236 HA trimers)
120 NA monomers
(30 NA tetramers)
11 M2

~ 160 million all-atom MD simulation with NAMD on Blue Waters, viz'd with CellView

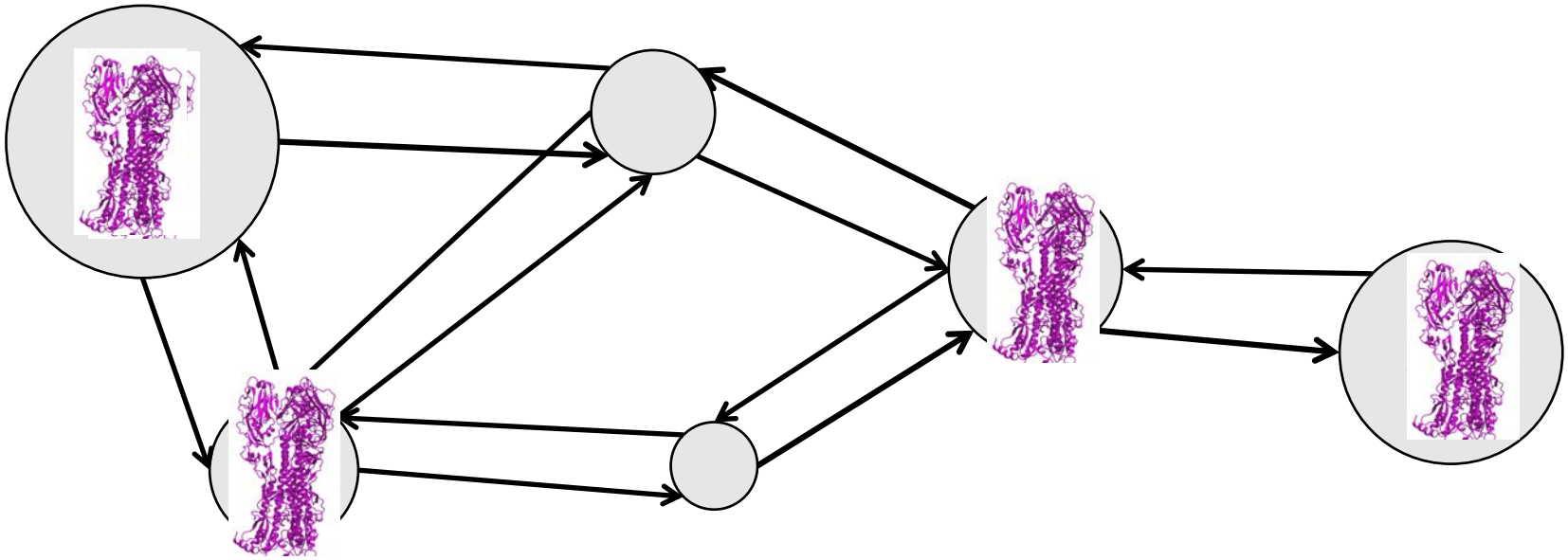


Protein Dynamics from Viral Simulations




moovly

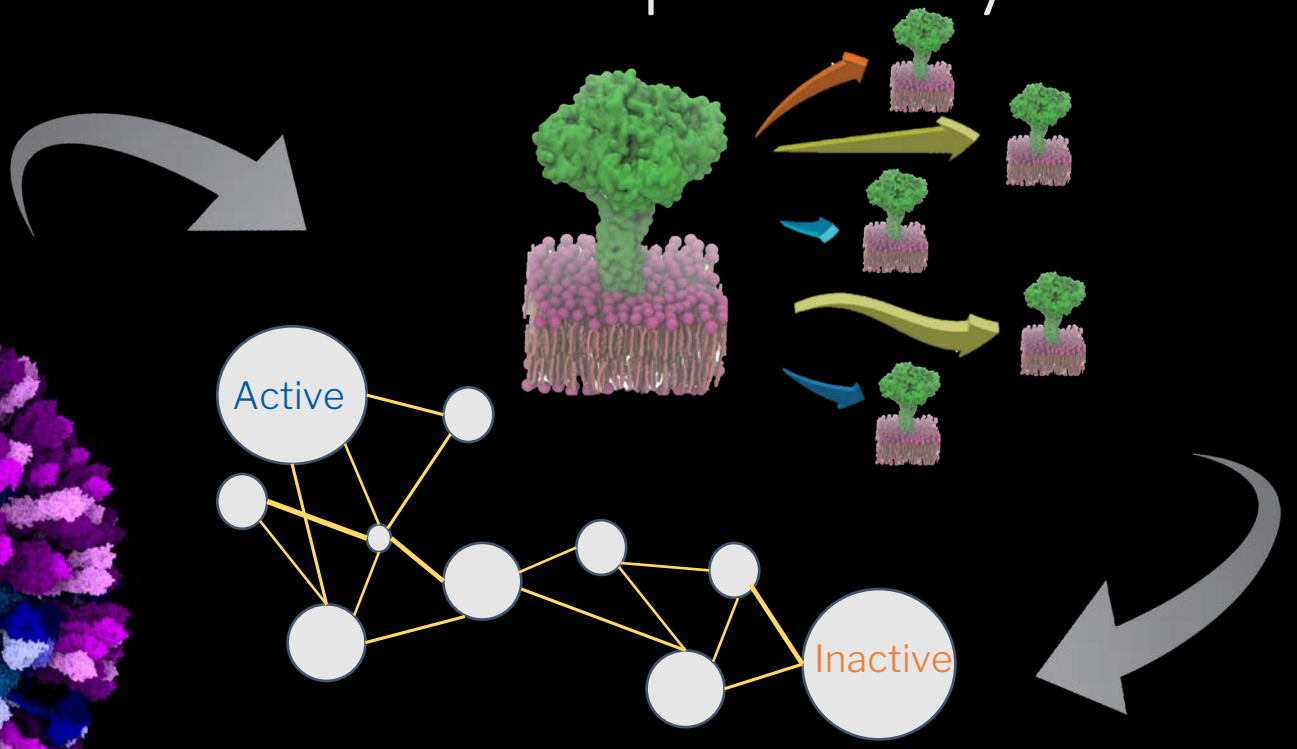
Cell-Scale Markov State Models of Protein Dynamics



- Markov state models define metastable states and transitions between states
- Allows extraction of long timescale dynamics from many short timescale simulations

B. E. Husic, V. S. Pande, *J. Am. Chem. Soc.* (2018).; J. D. Chodera, F. Noé, *Curr. Opin. Struct. Biol.* (2014).

Cell-scale Markov state models of protein dynamics

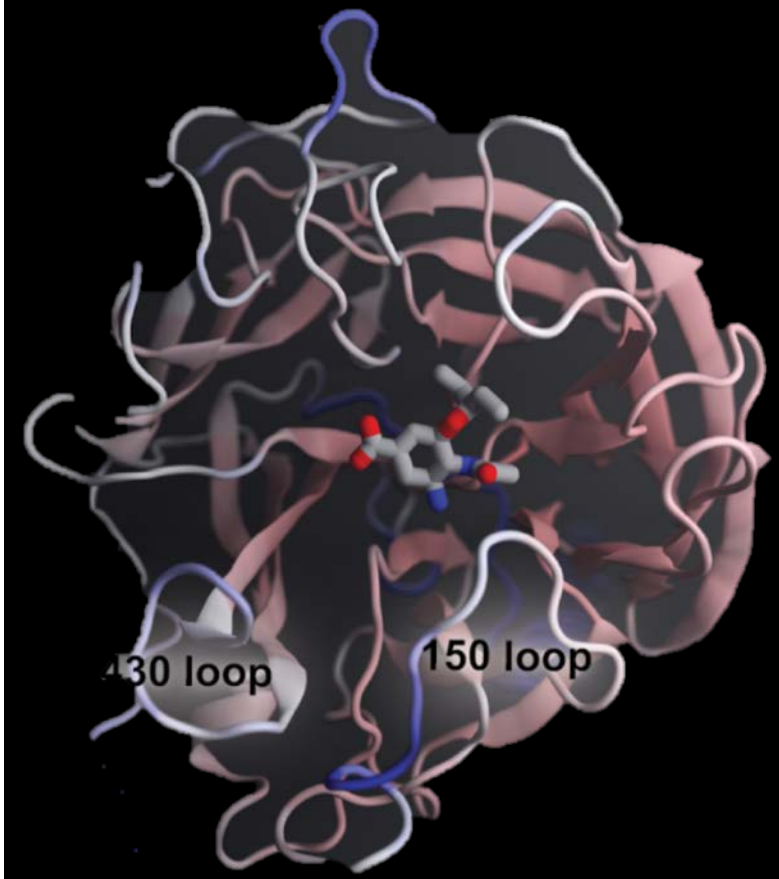


Markov state models define metastable states and transitions between states

Allows one to extract long timescale dynamics from many short timescale simulations

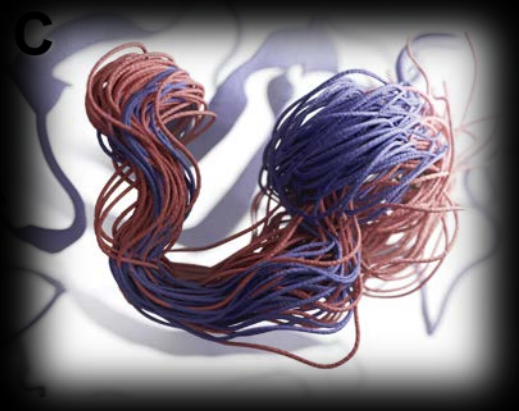
Swope, Pande, Schutte, Noe…

MSMs characterize loop dynamics & druggable pockets



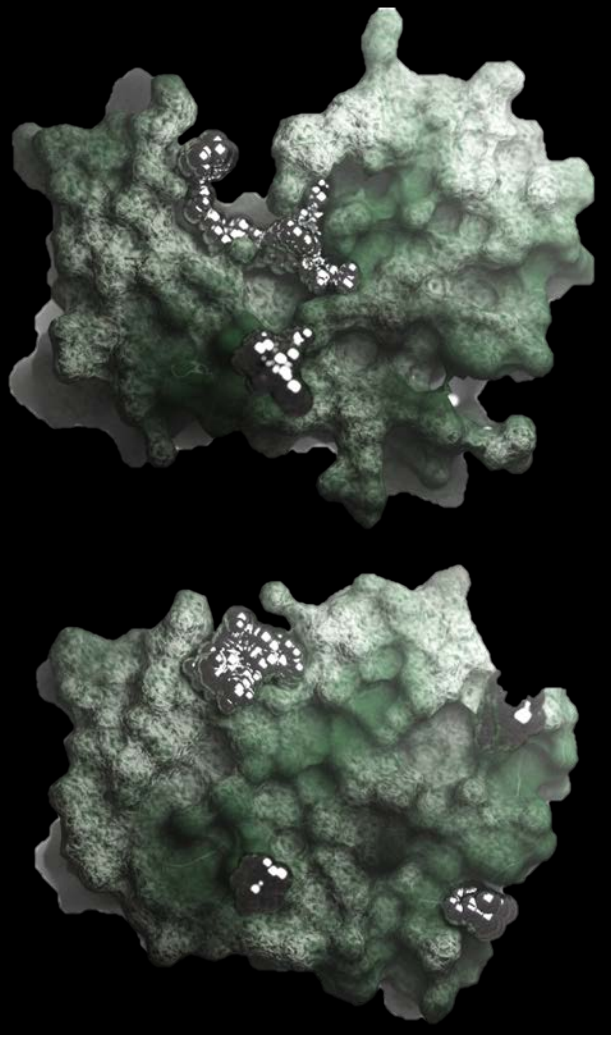
Virion has 30 NAs, 236 HAs
Enough sampling to make a Markov state model (MSM) of NA loop dynamics

2-state Macrostate model
open/closed

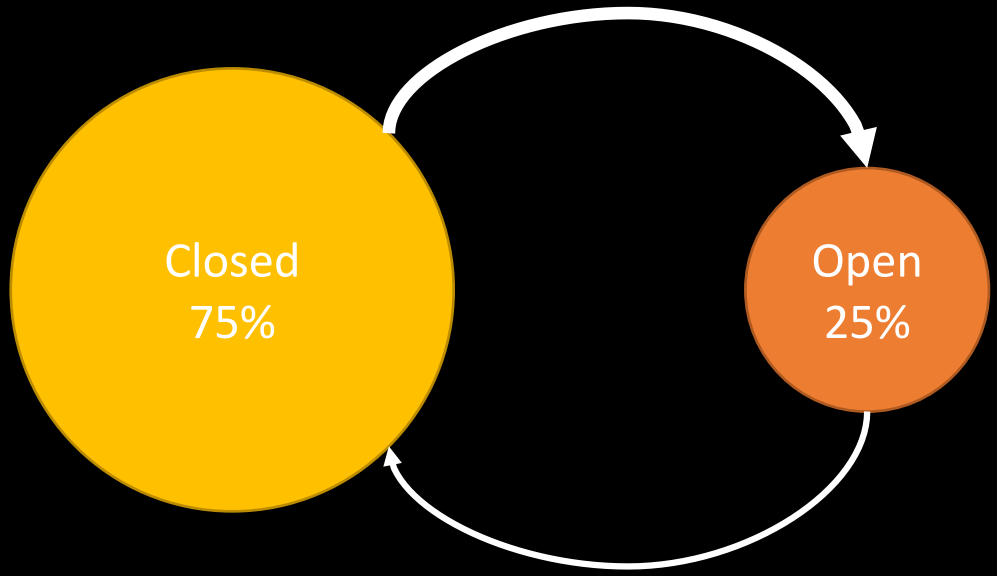


MFPT for the 150-loop:

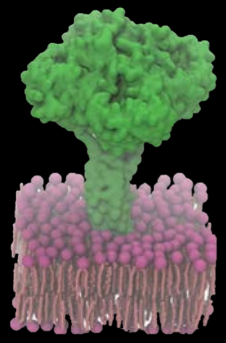
- *open to closed* 52.9ns
- *closed to open* 198.4 ns



Comparison to single glycoprotein simulations

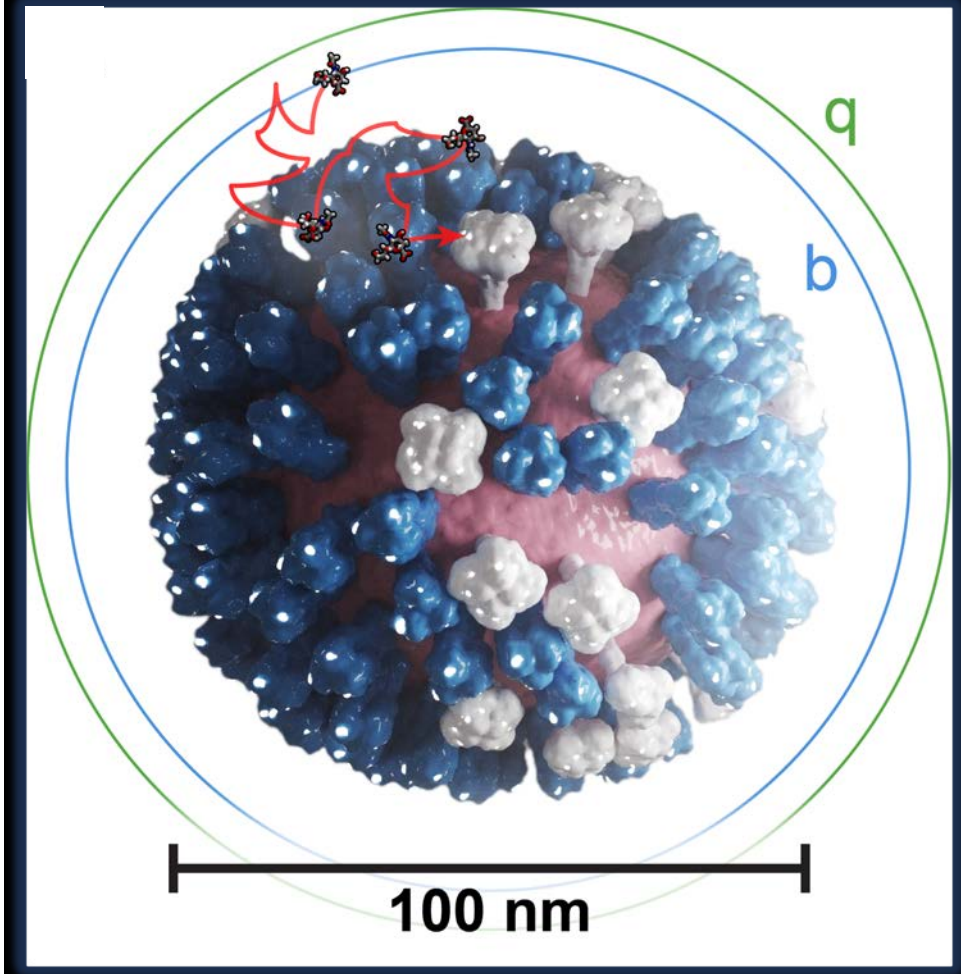


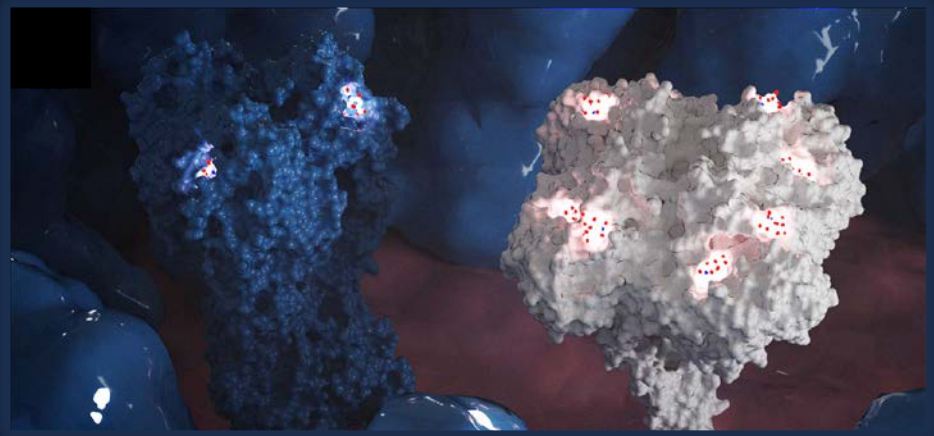
- Two MSMs, one each for viral coat and single glycoprotein simulations
- Same feature selection (residues 146 to 156) and lag time (7.5 ns)
- Both independently converged according to CK tests
- Same relative populations, and similar MFPT ratios, but different absolute rates

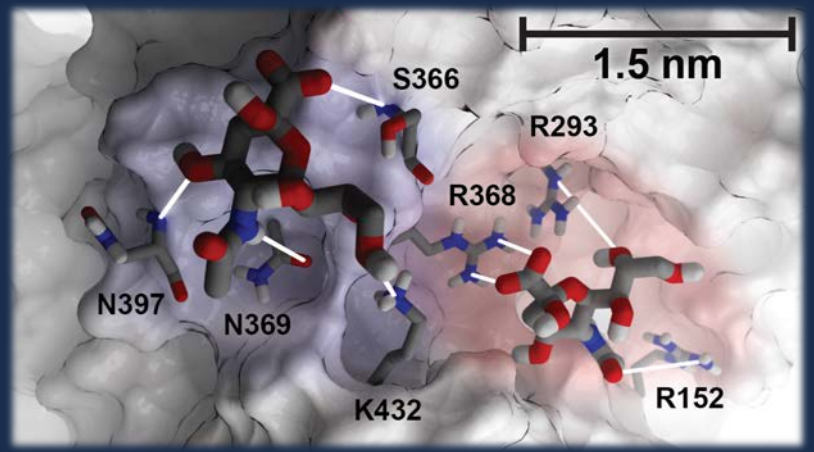


	Mean First Passage Time (ns)	
	Single Glycoprotein	Viral Coat
Open to Closed	390	164
Closed to Open	1000	520
MFPT Ratio	*2.6*	*3.2*

Molecular and Brownian Dynamics at Cellular Scales



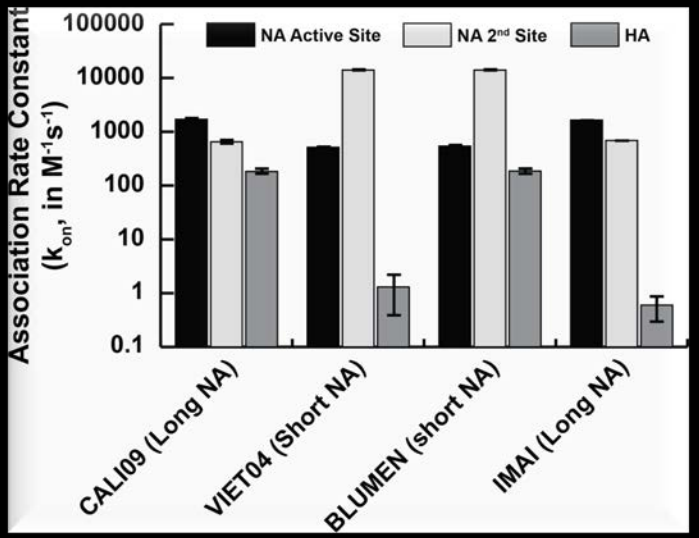


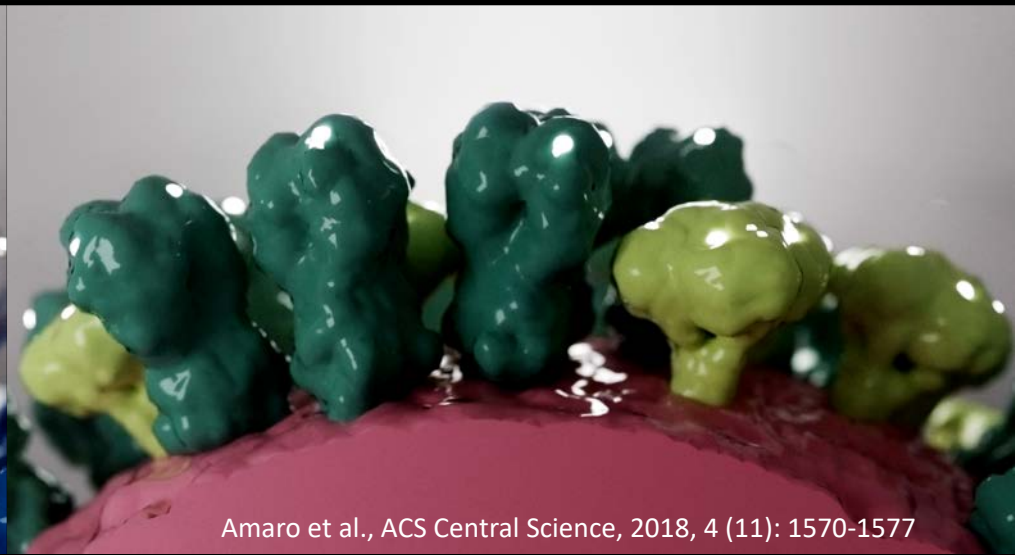


Unexpected role of the NA secondary site

Created 4 virions, ~15 million atoms each
Including reassorted strains, "gain of function" strain:

	HA	NA	NA stalk length
CALI09 (wt)	CALI09	CALI09	long
VIET04 (wt)	VIET04	VIET04	short
BLUMEN	CALI09	VIET04	short
IMAI	VIET04	CALI09	long



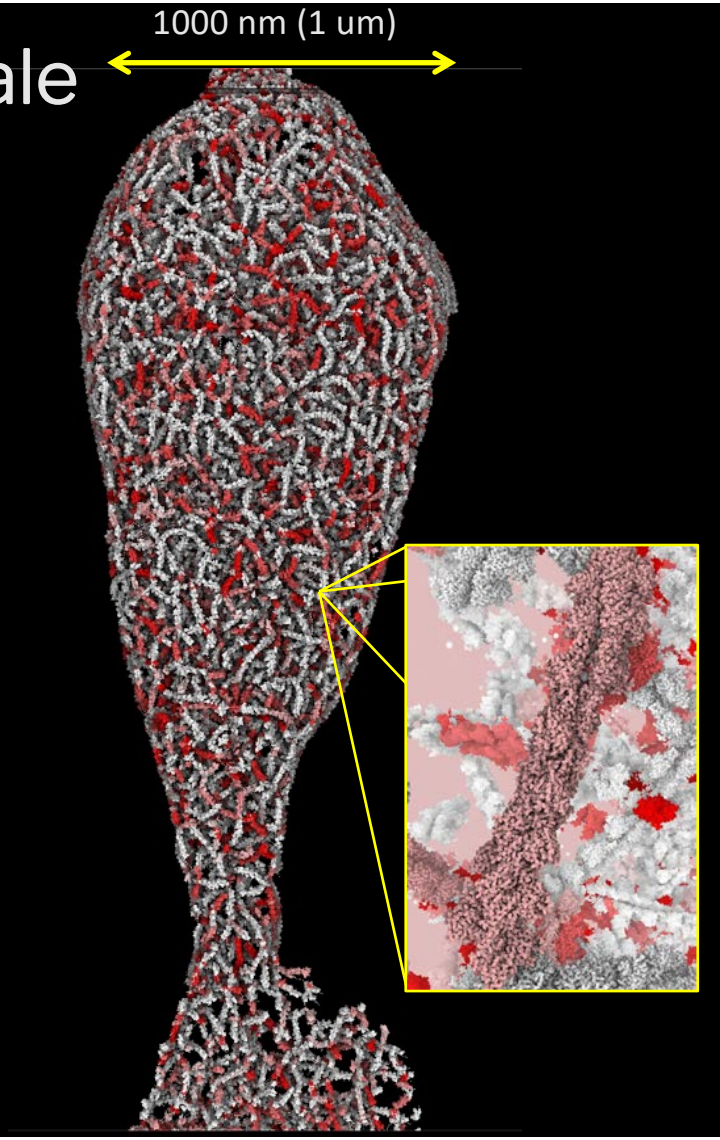


Amaro et al., ACS Central Science, 2018, 4 (11): 1570-1577

Molecular simulation at the mesoscale

10 nm

100 nm



Biophysics is ready for exascale!

http://amarolab.ucsd.edu

Acknowledgements

http://nbcr.ucsd.edu

Jacob Durrant

Lorenzo Casalino

Sarah Kochanek

Lane Votapka

Frank Noe, Freie Univ Berlin
Gary Huber, UCSD

COMP ACS Division of Computers in Chemistry

NBCR

NSF

NIH National Institute of General Medical Sciences
Basic Discoveries for Better Health

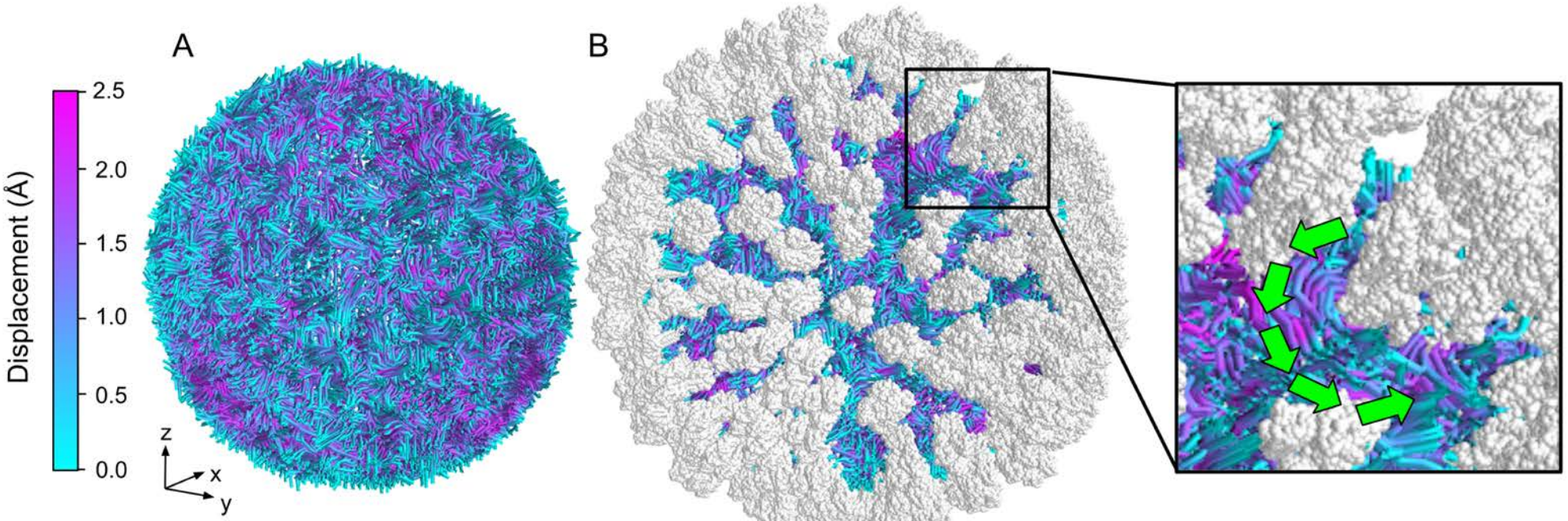
NIH Director's New Innovator Award

Theoretical and Computational Biophysics Group
NIH Resource for Macromolecular Modeling and Bioinformatics
University of Illinois at Urbana-Champaign

Hellman Fellows

Compute the Cure

NIH National Institute of Allergy and Infectious Diseases
Leading research to understand, treat, and prevent infectious, immunologic, and allergic diseases.


A
B
Displacement (Å)
2.5
2.0
1.5
1.0
0.5
0.0
z
x
y