# Using AWS EC2 GPU Instances for Computational Microscopy and Biomolecular Simulation

John E. Stone
Theoretical and Computational Biophysics Group
Beckman Institute for Advanced Science and Technology
University of Illinois at Urbana-Champaign

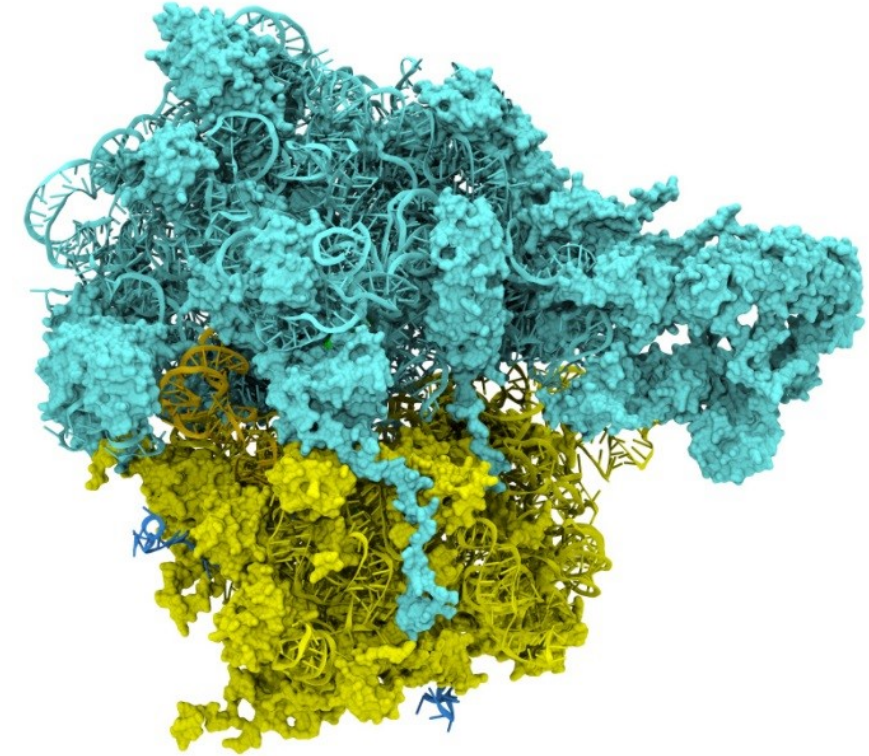
**http://www.ks.uiuc.edu/Research/cloud/**

1:00-1:25, Amazon Booth, Supercomputing 2018 Exhibition
Dallas, TX, Wednesday, November 14th, 2018

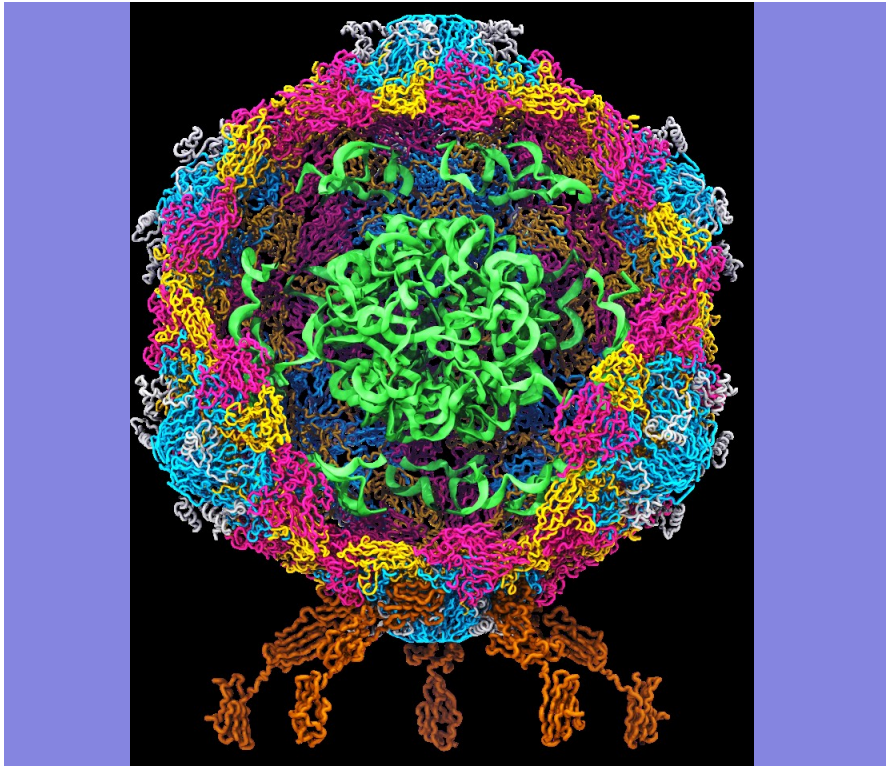
# Goal: A Computational Microscope

Study the molecular machines in living cells

Ribosome: target for antibiotics

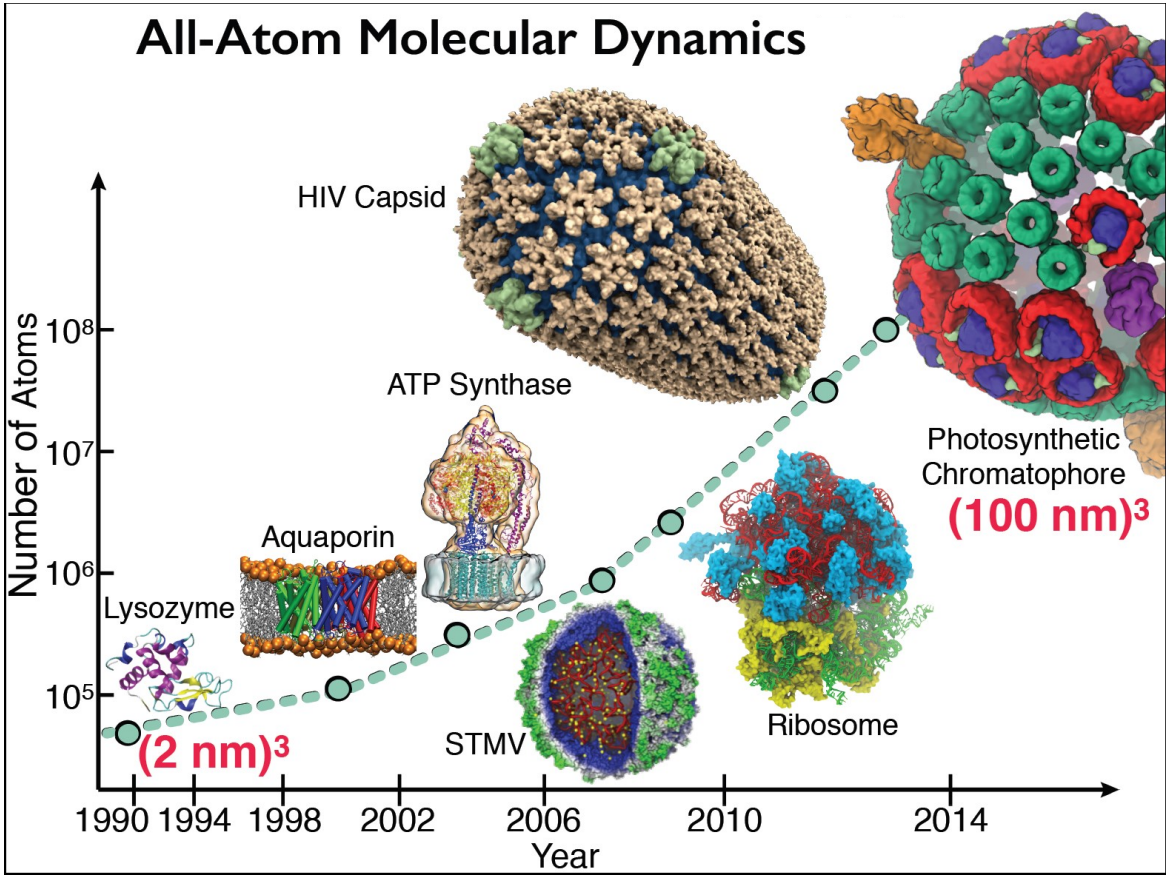


Poliovirus



# Goal: A Computational Microscope

Study the molecular machines in living cells

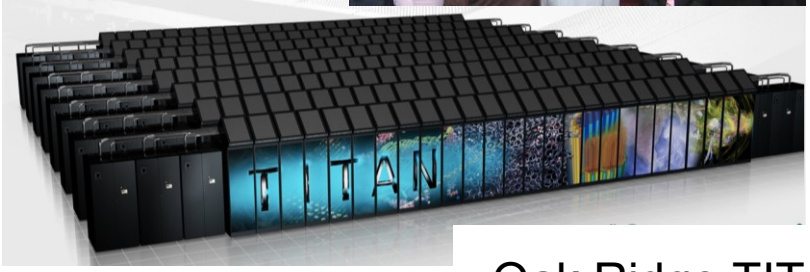


# NAMD Highlights

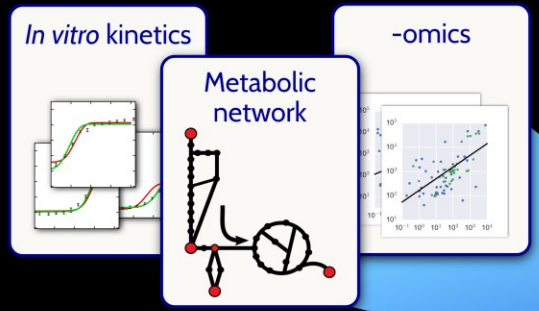
- **Parallel Molecular Dynamics**
- **Over 12,600 citations of NAMD**
- One program available on all platforms.
  - Desktops and laptops – setup and testing
  - Linux clusters – affordable local workhorses
  - Supercomputers – free allocations on XSEDE
  - Blue Waters – sustained petaflop/s performance
  - GPUs – from desktop to supercomputer
- User knowledge is preserved across platforms.
  - Run any simulation on **any number of cores**.
- Available free of charge to all.

**http://www.ks.uiuc.edu/Research/namd/**


Hands-On Workshops

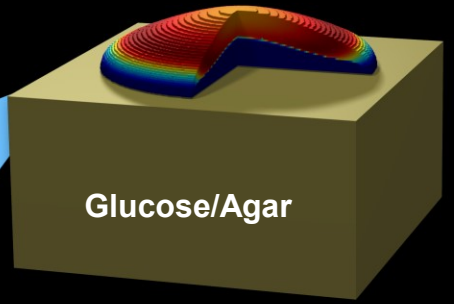


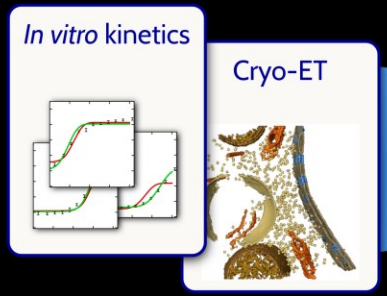

Oak Ridge TITAN



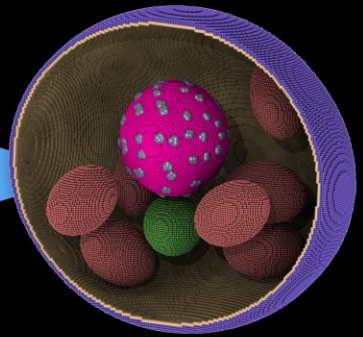

- In
imaging
- Scriptable in Py

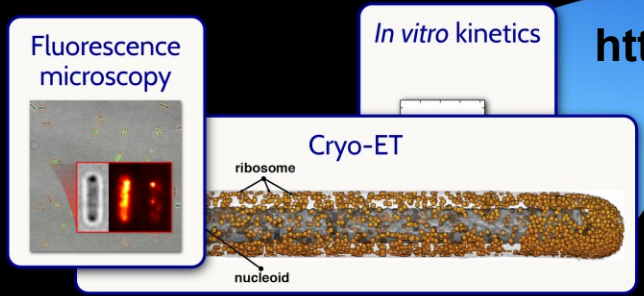
ental



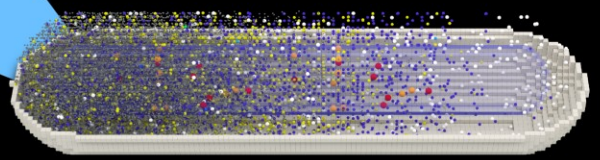







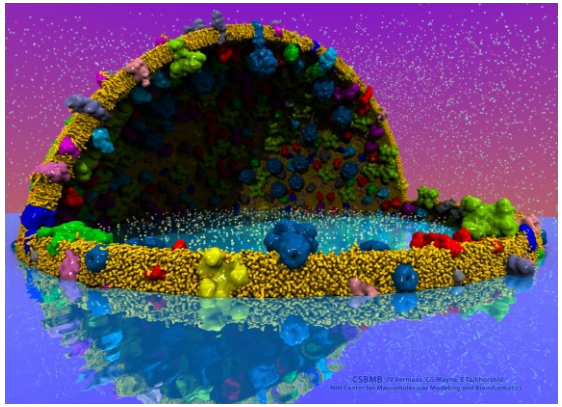



http://ww is.edu/sc

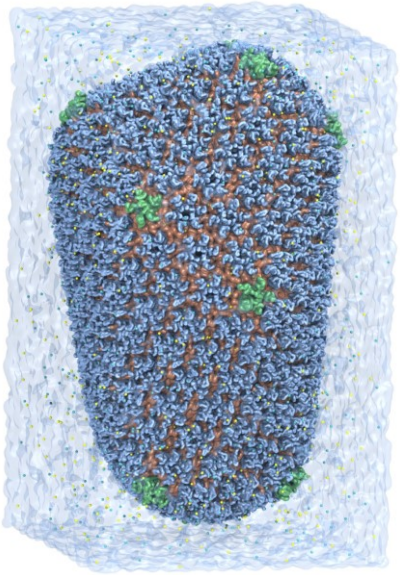


# VMD – “Visual Molecular Dynamics”

- Visualization and analysis of:
  - Molecular dynamics simulations
  - Lattice cell simulations
  - Quantum chemistry calculations
  - Sequence information
- User extensible scripting and plugins

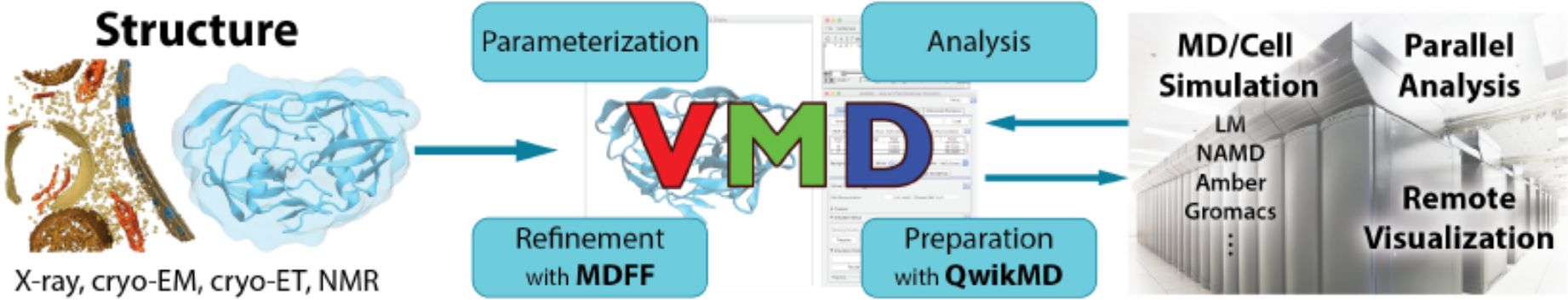


Cell-Scale Modeling



MD Simulation

**http://www.ks.uiuc.edu/Research/vmd/**



# Making Our Research Tools Easily Accessible

- Cloud based deployment
  - Full virtual machines (known as “AMI” in Amazon terminology)
  - Amazon AWS EC2 GPU-accelerated instances: **http://www.ks.uiuc.edu/Research/cloud/**


- Docker “container” images available in NVIDIA NGC registry
  - Users obtain Docker images via registry, download and run on the laptop, workstation, cloud, or supercomputer of their choosing
  - **https://ngc.nvidia.com/registry/**
  - **https://ngc.nvidia.com/registry/hpc-vmd**

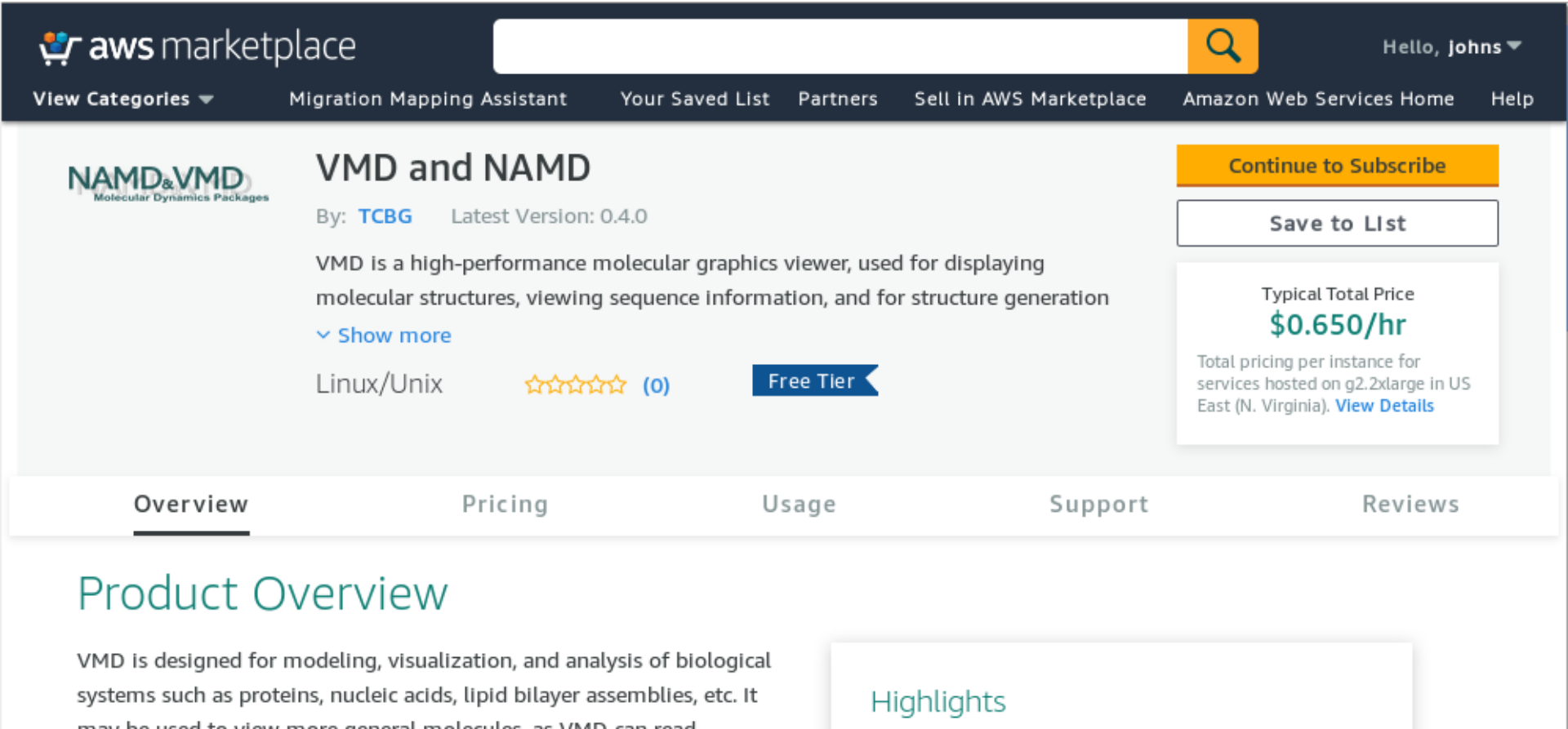
**Research articles incorporating use of Amazon AWS EC2:**

**Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps.** Abhishek Singharoy, Ivan Teo, Ryan McGreevy, John E. Stone, Jianhua Zhao, and Klaus Schulten. *eLife*, 10.7554/eLife.16105, 2016. (66 pages).

**QwikMD-integrative molecular dynamics toolkit for novices and experts.** Joao V. Ribeiro, Rafael C. Bernardi, Till Rudack, John E. Stone, James C. Phillips, Peter L. Freddolino, and Klaus Schulten. *Scientific Reports*, 6:26536, 2016.

**High performance molecular visualization: In-situ and parallel rendering with EGL.** John E. Stone, Peter Messmer, Robert Sisneros, and Klaus Schulten. *2016 IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW)*, pp. 1014-1023, 2016.

# Easy to Launch: AWS EC2 Marketplace



# NAMD+VMD AWS EC2 AMIs

**Production AMI:**

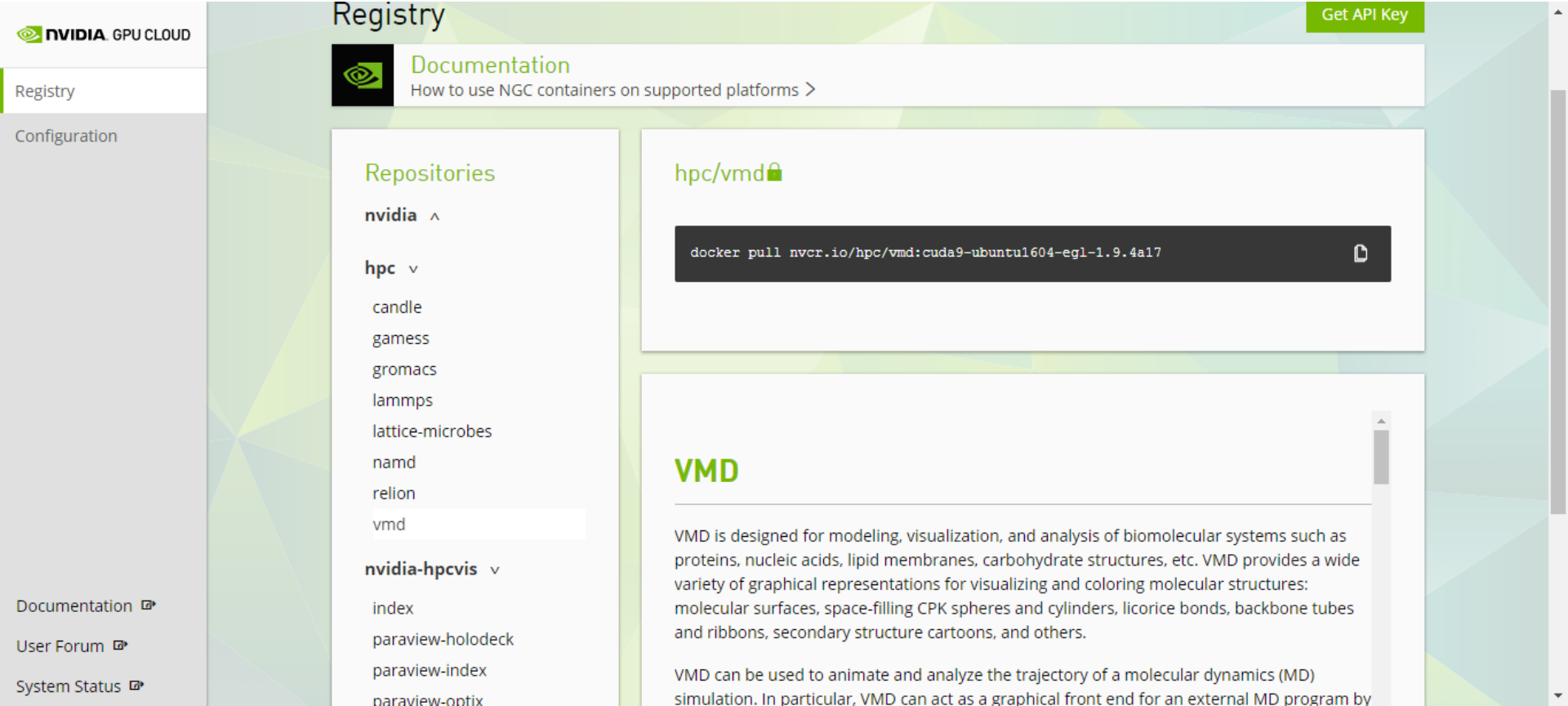
- (ami-a01125df) VMD-NAMD-VNC-R1.9.4.1, 64-bit Ubuntu Linux, EBS storage, HVM, created July 10, 2018

**Pre-release or developmental AMI:**

- (ami-ac604ed3) 64-bit Centos 7.5 Linux, EBS storage, HVM, created July 6, 2018
- This is an experimental image using Centos and **DCV** for increased remote visualization performance and smoother interaction. This image will only run on g3 instance types.



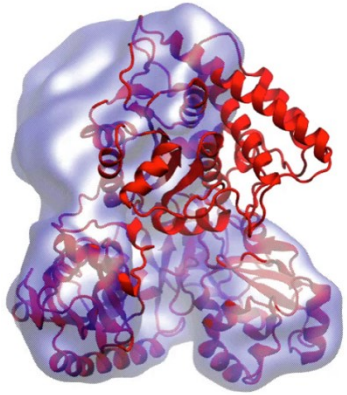
# VMD / NAMD / LM, NGC Containers



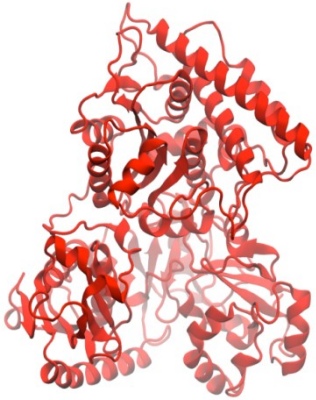
# Molecular Dynamics Flexible Fitting (MDFF)

X-ray crystallography → MDFF ← Electron microscopy

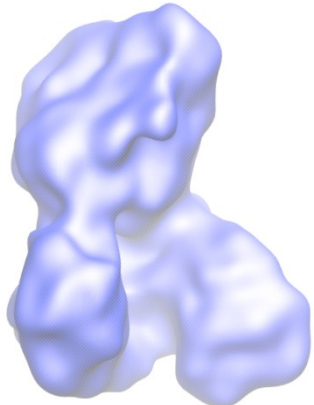
APS at Argonne



FEI microscope

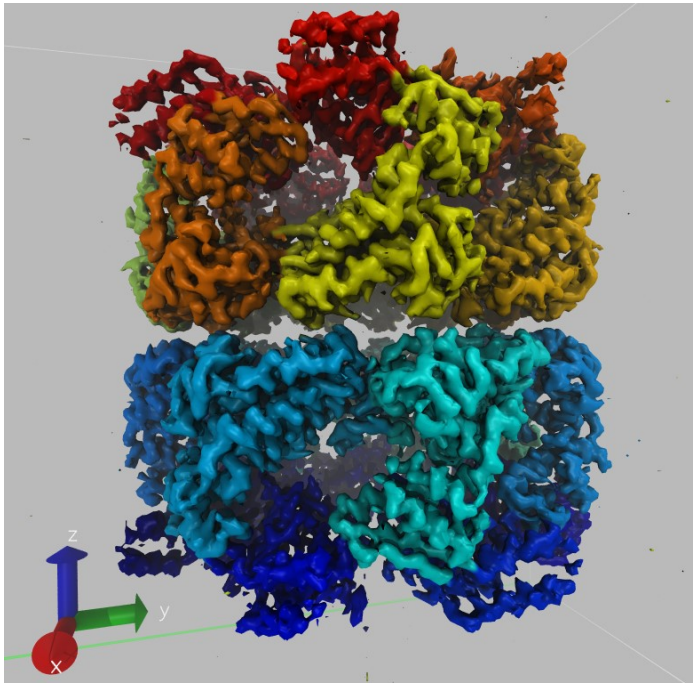



ORNL Titan

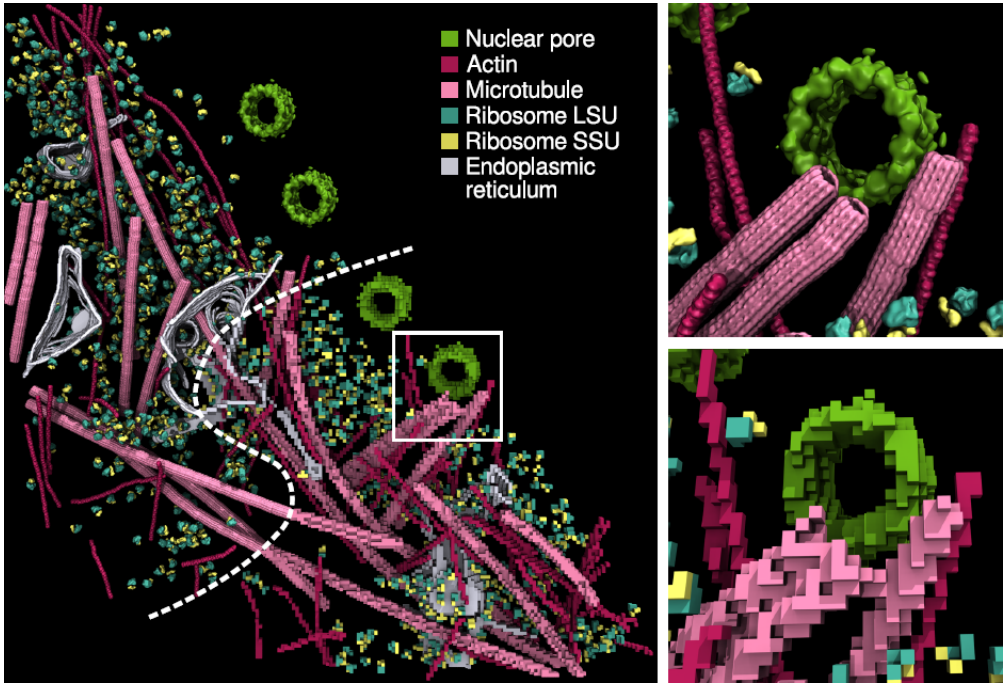


**Molecular dynamics-based refinement and validation for sub-5Å cryo-electron microscopy maps**. A. Singharoy, I. Teo, R. McGreevy, J. E. Stone, J. Zhao, and K. Schulten. eLife 2016;10.7554/eLife.16105

# Density Map Segmentation




**VMD GPU-accelerated density map segmentation of GroEL**




**Earnest, et al. J. Physical Chemistry B, 121(15): 3871-3881, 2017.**

# Molecular Dynamics Flexible Fitting - Theory

Two terms are added to the MD potential

$$U_{total} = U_{MD} + U_{EM} + U_{SS}$$

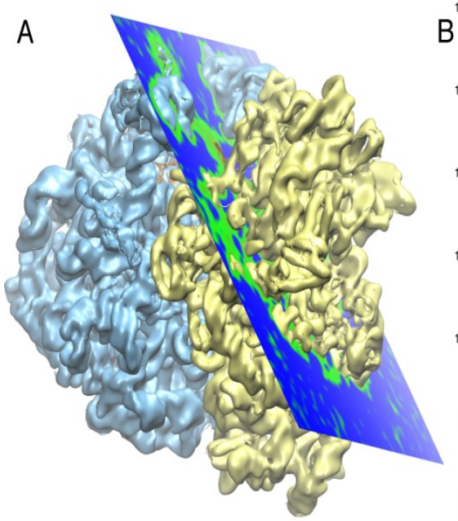
An external potential derived from the EM map is defined on a grid as

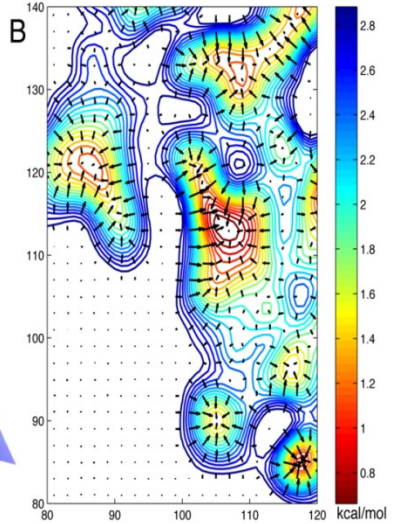
$$U_{EM}(\mathbf{R}) = \sum_j w_j V_{EM}(\mathbf{r}_j)$$

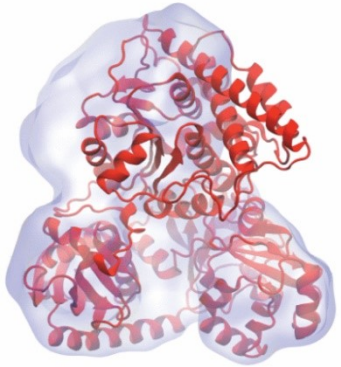
$$V_{EM}(\mathbf{r}) = \begin{cases} \xi \left(1 - \frac{\Phi(\mathbf{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right) & \text{if } \Phi(\mathbf{r}) \geq \Phi_{thr}, \\ \xi & \text{if } \Phi(\mathbf{r}) < \Phi_{thr}. \end{cases}$$

A mass-weighted force is then applied to each atom

$$\mathbf{f}_i^{EM} = -\nabla U_{EM}(\mathbf{R}) = -w_i \partial V_{EM}(\mathbf{r}_i)/\partial r_i$$



# MDFF on the Cloud Costs Less than a Cup of Coffee

MDFF and ReMDFF (Resolution Exchange) require many CPU/GPU cores but little compute time, making making them good candidates for cloud compı

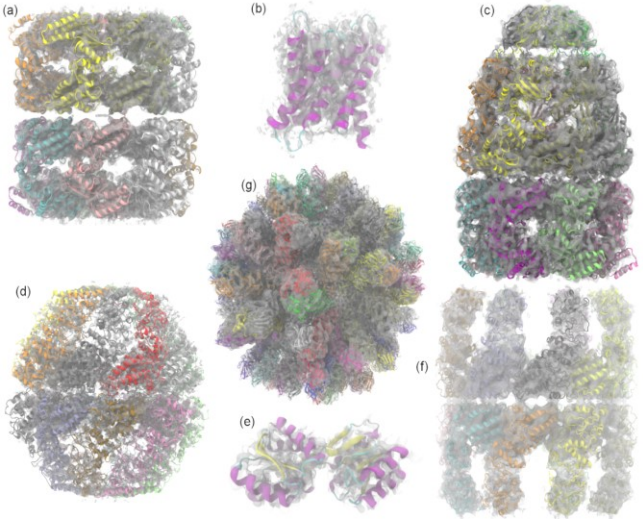
Singharoy, *et al*. eLife 2016

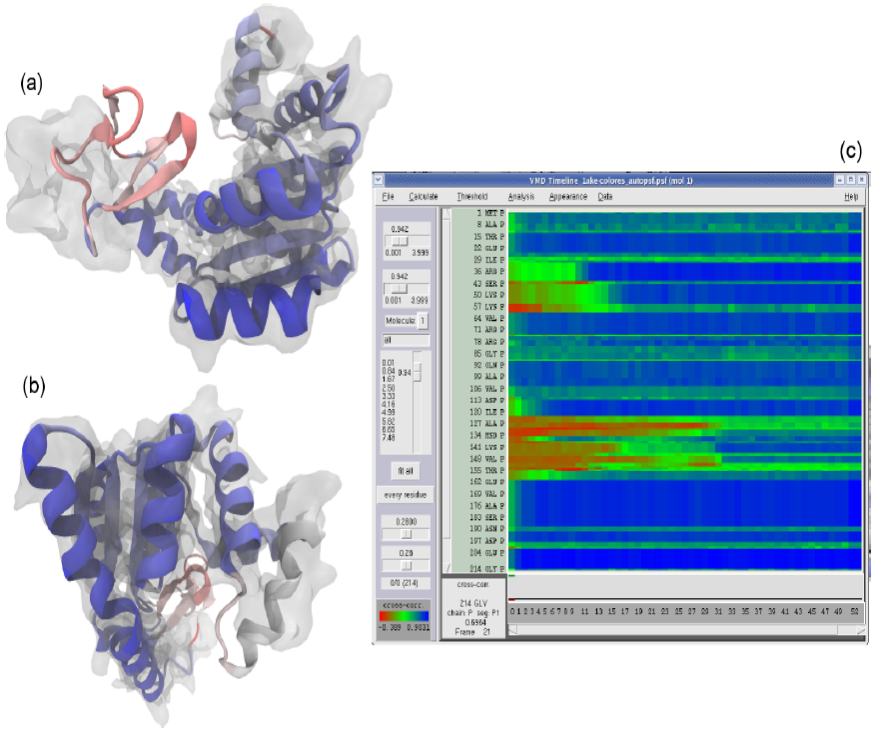
| Molecule | Instance | Performance (ns/day) | Time (hours) | Simulation Cost / ns ($) |
|---|---|---|---|---|
| Adenylate Kinase | p3.2xlarge | 112 | 0.2 | 0.67 |
| Acetyl-CoA Synthase | p3.2xlarge | 82 | 0.3 | 0.89 |
| J1 Nitrilase | p3.2xlarge | 5 | 4.8 | 14.6 |


Cloud computing allows researchers to focus on the scientific challenges of their project without having to worry about local availability and administration of suitable computer hardware, and installation or compilation of software.

# Evaluating Quality-of-Fit for Structures Solved by Hybrid Fitting Methods

Compute Pearson correlation to evaluate quality-of-fit between a reference cryo-EM density map and a **simulated density map** from an **all-atom structure**.




**MDFF Cross Correlation Timeline**

**Regions with poor fit** **Regions with good fit**

# VMD Tesla V100 Cross Correlation Performance

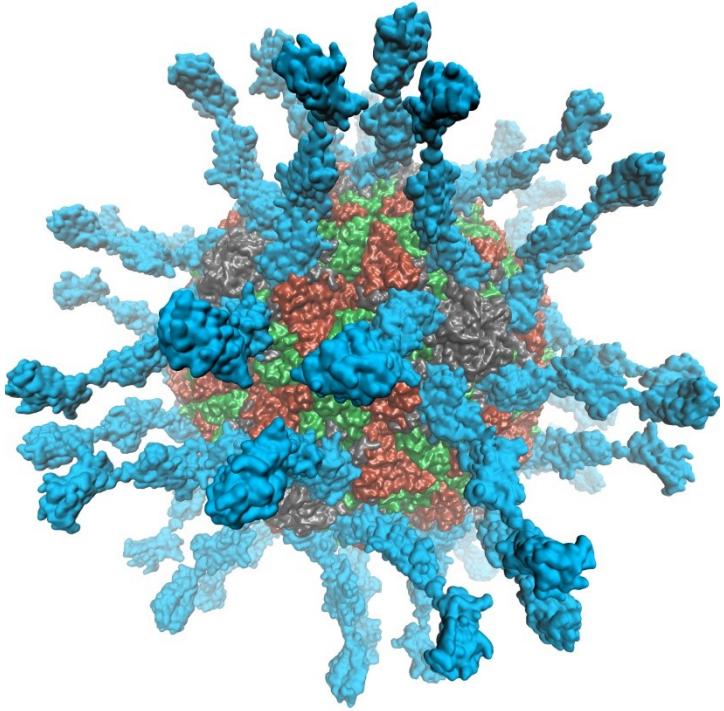
## Rabbit Hemorrhagic Disease Virus: 702K atoms, 6.5Å resolution
## Volta GPU architecture almost 2x faster than previous gen Pascal:

| Application and Hardware platform | Runtime, Speedup vs. Chimera, | | VMD+GPU |
|---|---|---|---|
| Chimera Xeon E5-2687W (2 socket) [1] | 15.860s, | 1x | |
| VMD-CUDA Intel Xeon E5-2687W + 1x Quadro K6000 [1,2] | 0.458s, | 35x | **1.0x** |
| **VMD-CUDA Intel Xeon E5-2698v3 + 1x Tesla P100** | **0.090s,** | 176x | **5.1x** |
| **VMD-CUDA Intel Xeon E5-2697Av4 + 1x Tesla V100** | **0.050s,** | 317x | **9.2x** |

**[1] GPU-Accelerated Analysis and Visualization of Large Structures Solved by Molecular Dynamics Flexible Fitting.** J. E. Stone, R. McGreevy, B. Isralewitz, and K. Schulten. Faraday Discussions 169:265-283, 2014.
**[2] Early Experiences Porting the NAMD and VMD Molecular Simulation and Analysis Software to GPU-Accelerated OpenPOWER Platforms.** J. E. Stone, A.-P. Hynninen, J. C. Phillips, K. Schulten. International Workshop on OpenPOWER for HPC (IWOPH'16), LNCS 9945, pp. 188-206, 2016.

# VMD supports EGL for in-situ and parallel rendering on Amazon EC2

- **No windowing system dependency**
- Easily deploy parallel VMD builds supporting off-screen rendering
- Maintains 100% of VMD OpenGL shaders and rendering features

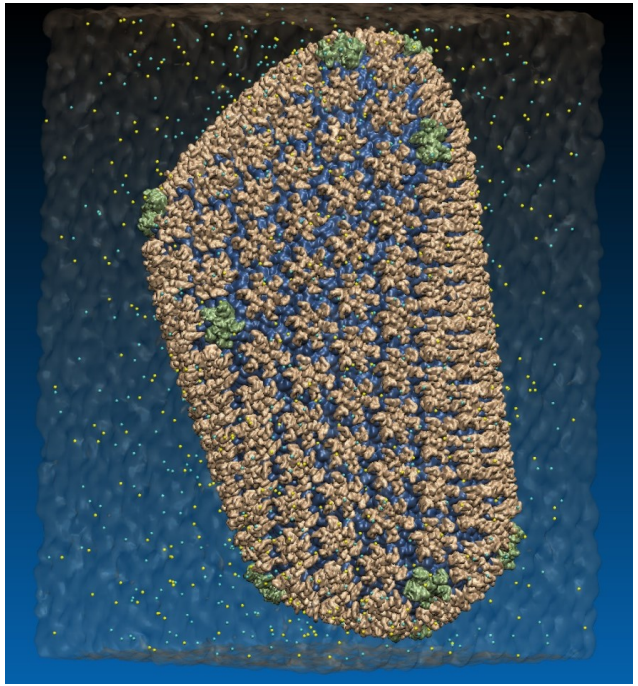


**Poliovirus**

# VMD EGL Performance on Amazon EC2 Cloud

| MPI Ranks | EC2 "G2.8xlarge" GPU Instances | HIV-1 movie rendering time (sec), (I/O %) 3840x2160 resolution |
|---|---|---|
| 1 | 1 | 626s (10% I/O) |
| 2 | 1 | 347s (19% I/O) |
| 4 | 1 | 221s (31% I/O) |
| 8 | 2 | 141s (46% I/O) |
| 16 | 4 | 107s (64% I/O) |
| 32 | 8 | 90s (76% I/O) |

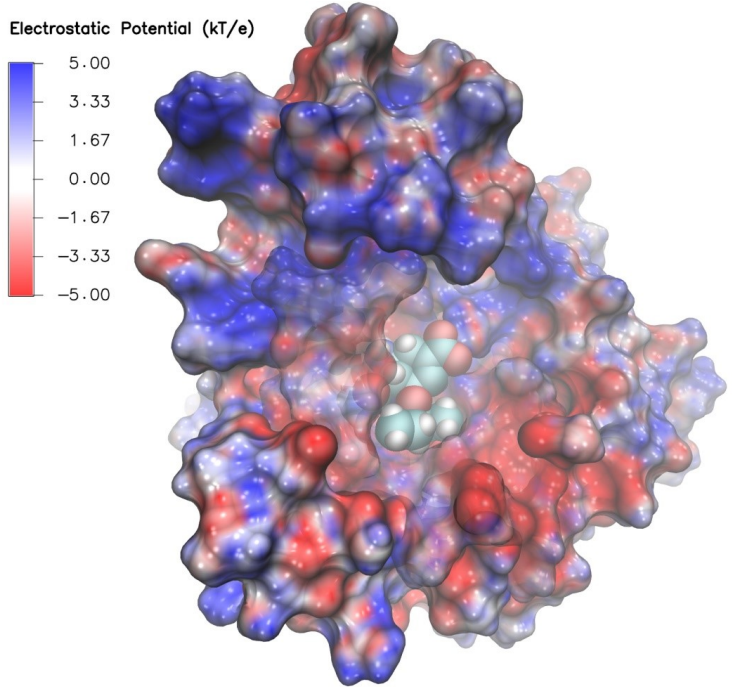
**Performance at 32 nodes reaches ~48 frames per second**

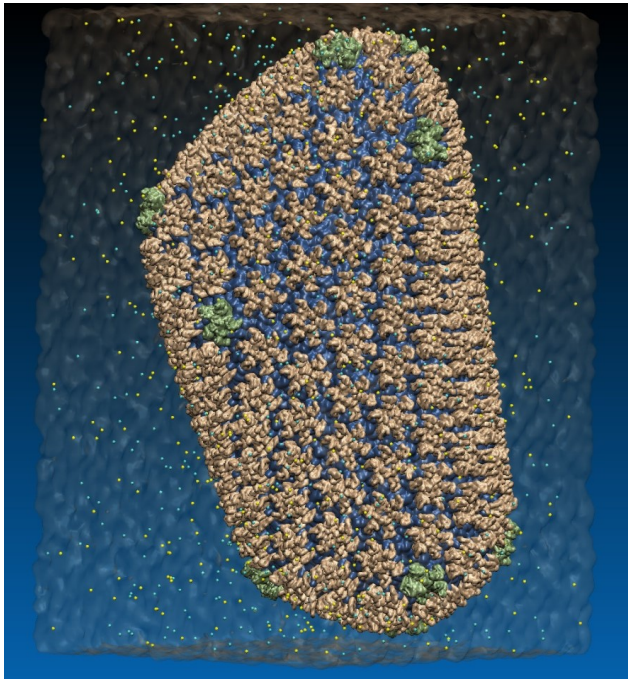


**64M atom HIV-1 capsid simulation rendered via EGL**

**High Performance Molecular Visualization: In-Situ and Parallel Rendering with EGL.** J. E. Stone, P. Messmer, R. Sisneros, and K. Schulten. High Performance Data Analysis and Visualization Workshop, IEEE IPDPSW, 2016.

# VMD EGL Rendering: Supports full VMD GLSL shading features
# Vulkan support coming soon...



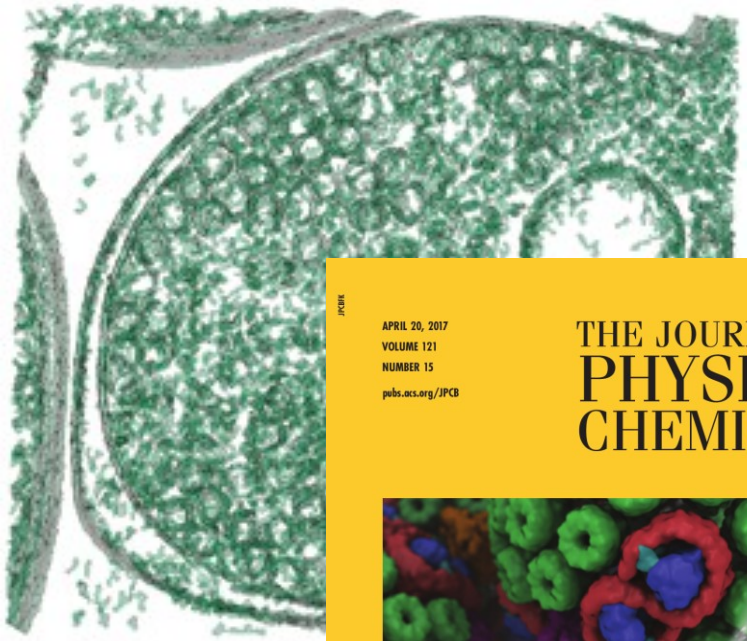
**Swine Flu A/H1N1 neuraminidase bound to Tamiflu**

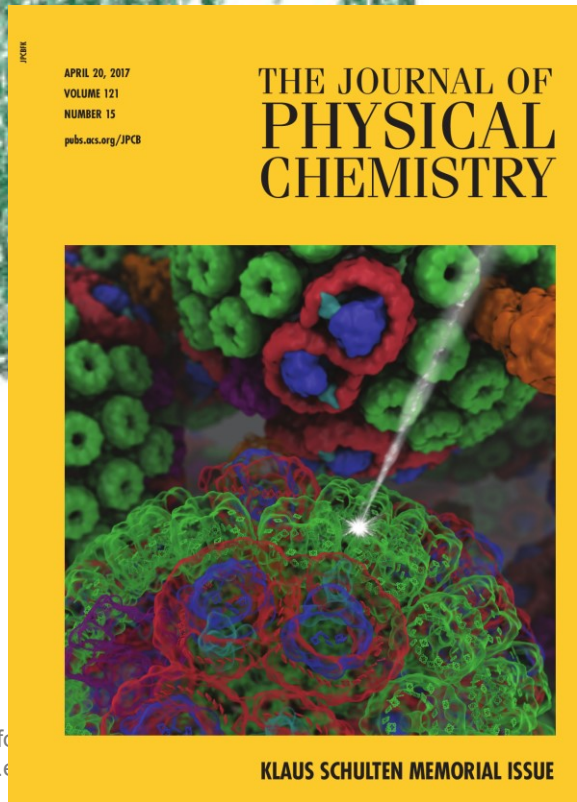
**64M atom HIV-1 capsid simulation**

**High Performance Molecular Visualization: In-Situ and Parallel Rendering with EGL.**
J. E. Stone, P. Messmer, R. Sisneros, and K. Schulten. High Performance Data Analysis and Visualization Workshop, IEEE IPDPSW, pp. 1014-1023, 2016.

NIH

# *NEW:* Cloud-Based Interactive Remote Visualization

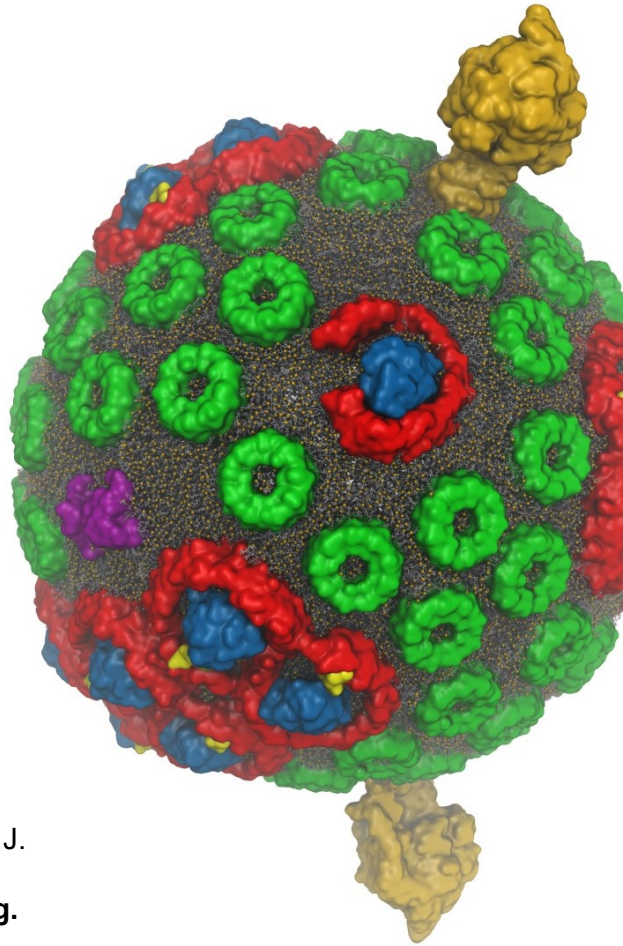





- Built-into VMD itself
- Enable access to massive data sets
- Uses GPU H.264 / HEVC hardware accelerated video encode/decode
- Supports interactive remote visualizations (both rasterization and ray tracing)
- Development ongoing, expected in next major VMD release, in 1H 2019…

# VMD Interactive Ray Tracing

- **Exploit computational power to improve rendering of the structural details of biomolecular complexes**
- **Remote visualization tasks on very large macromolecular complexes**
- High fidelity shading, shadows, AO lighting, depth of field, …



**VMD/OptiX GPU Ray Tracing of all-atom Chromatophore w/ lipids.**


**GPU-Accelerated Molecular Visualization on Petascale Supercomputing Platforms.** J. E. Stone, K. L. Vandivort, and K. Schulten. UltraVis'13, pp. 6:1-6:8, 2013.

**Visualization of Energy Conversion Processes in a Light Harvesting Organelle at Atomic Detail.** M. Sener, et al. SC'14 Visualization and Data Analytics Showcase, 2014.

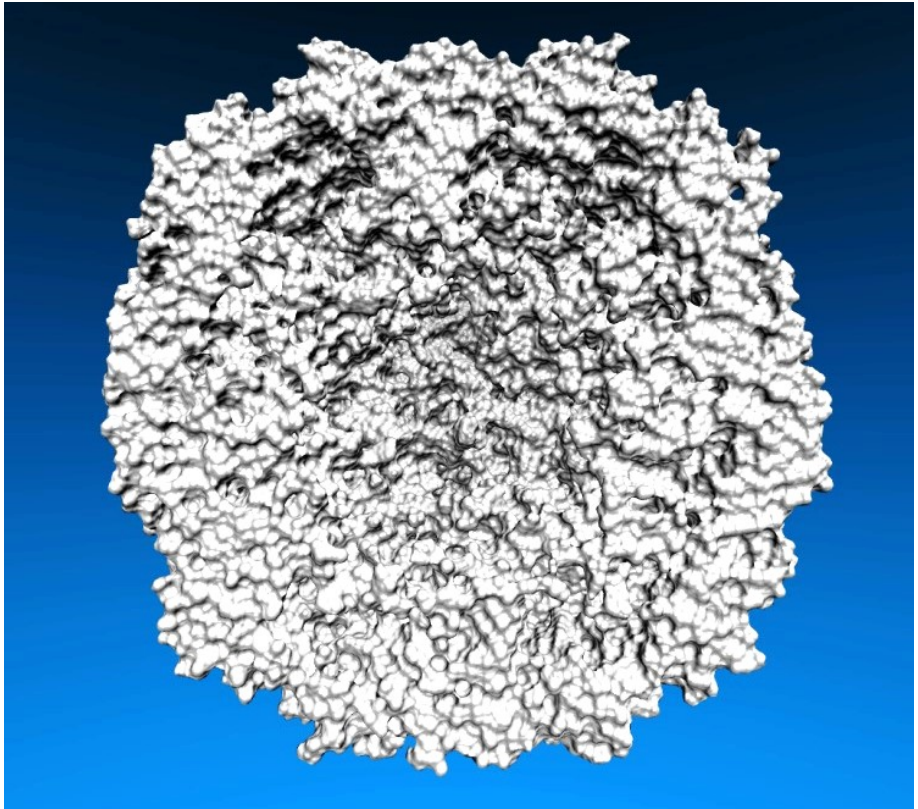
**Chemical Visualization of Human Pathogens: the Retroviral Capsids.** J. R. Perilla, B.-C. Goh, J. E. Stone, and K. Schulten. SC'15 Visualization and Data Analytics Showcase, 2015.

**Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing.** J. E. Stone et al., J. Parallel Computing, 55:17-27, 2016.

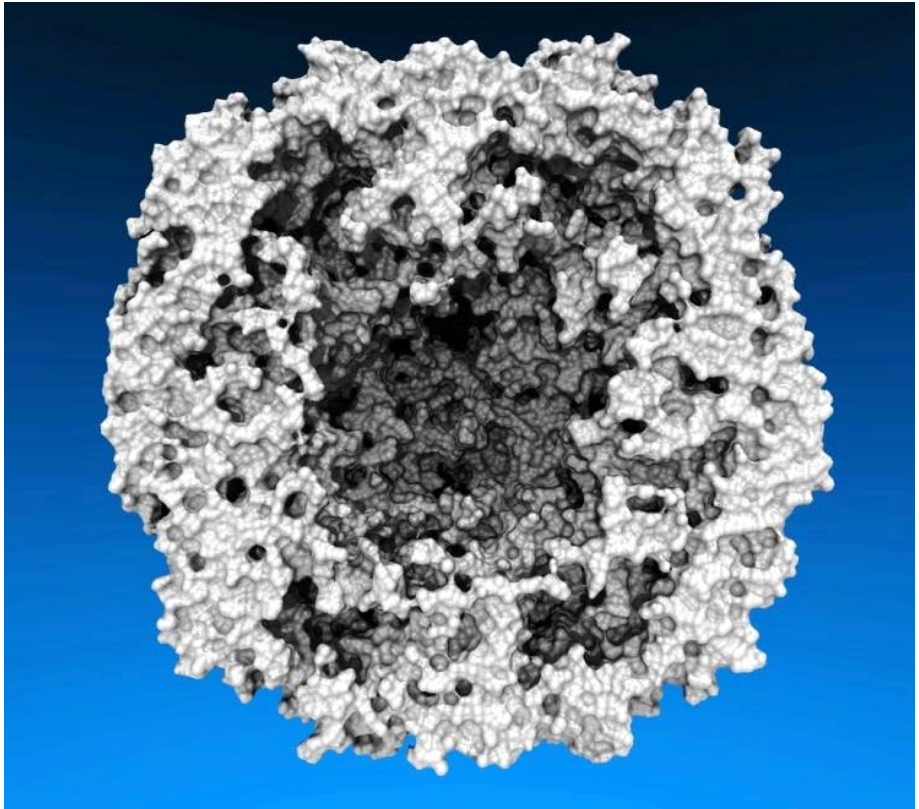
**Immersive Molecular Visualization with Omnidirectional Stereoscopic Ray Tracing and Remote Rendering** J. E. Stone, W. R. Sherman, and K. HPDAV, IPDPSW, pp. 1048-1057, 2016.

# *Interactive* Ray Tracing, Lighting Comparison: STMV Capsid

**Two lights, no shadows (e.g. as used by OpenGL)**

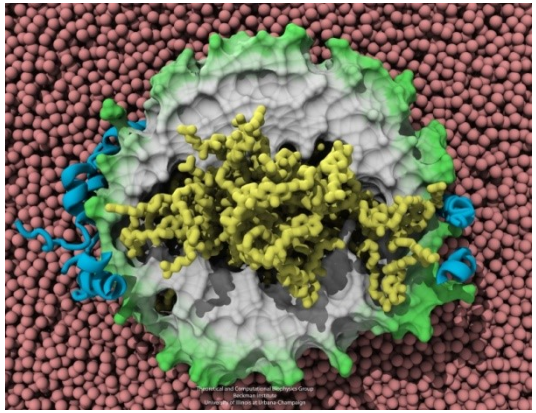


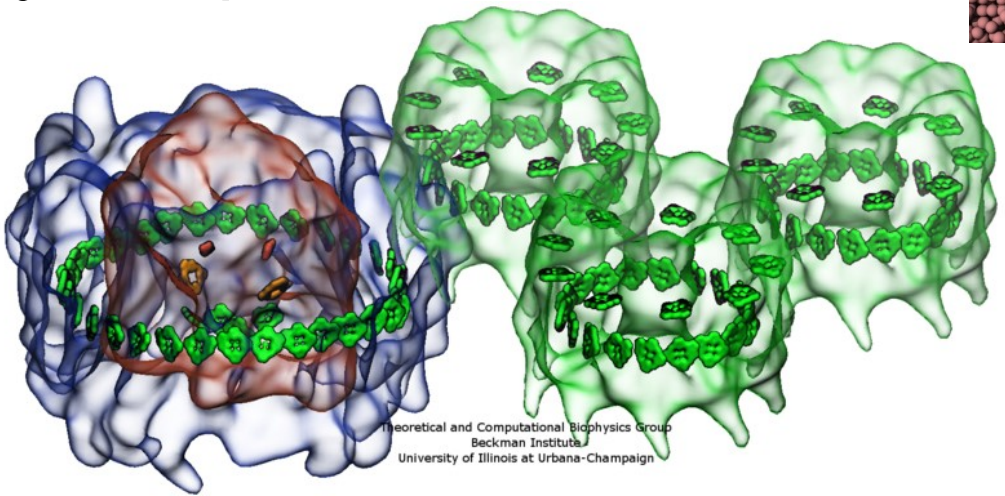
**Ambient occlusion lighting and shadows w/ RT**

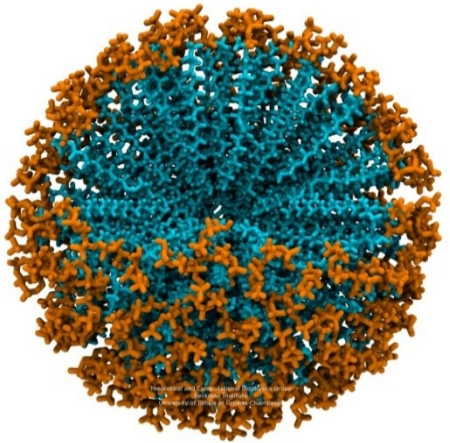


# Benefits of Advanced Lighting and Shading Techniques

- Exploit visual intuition
- Spend computer time in exchange for scientists' time, make images that are more easily interpreted

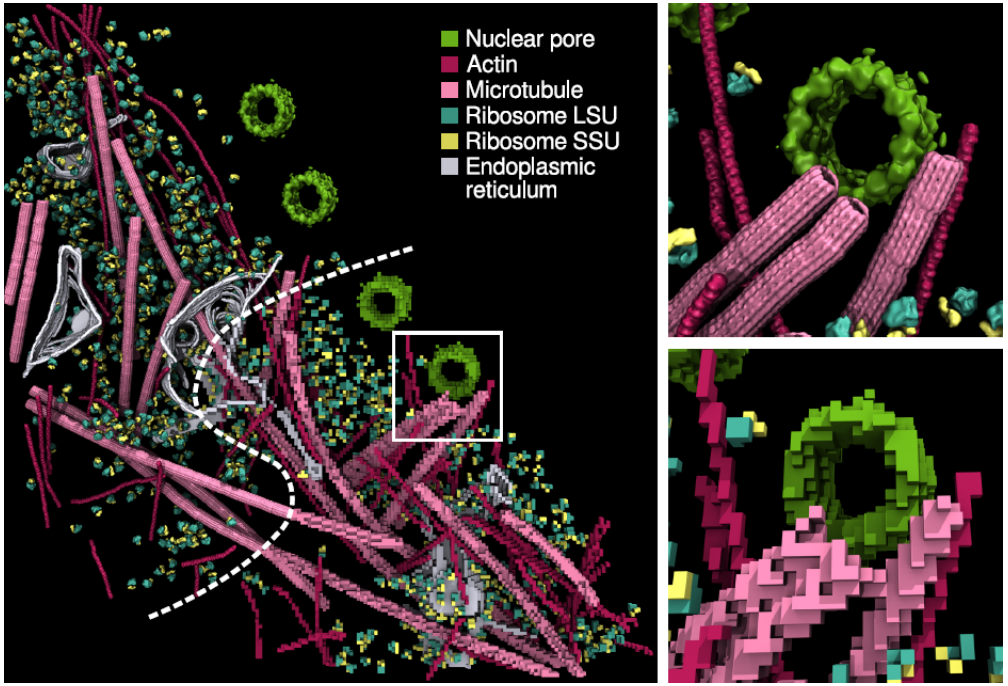




# Interactive Ray Tracing of Cells

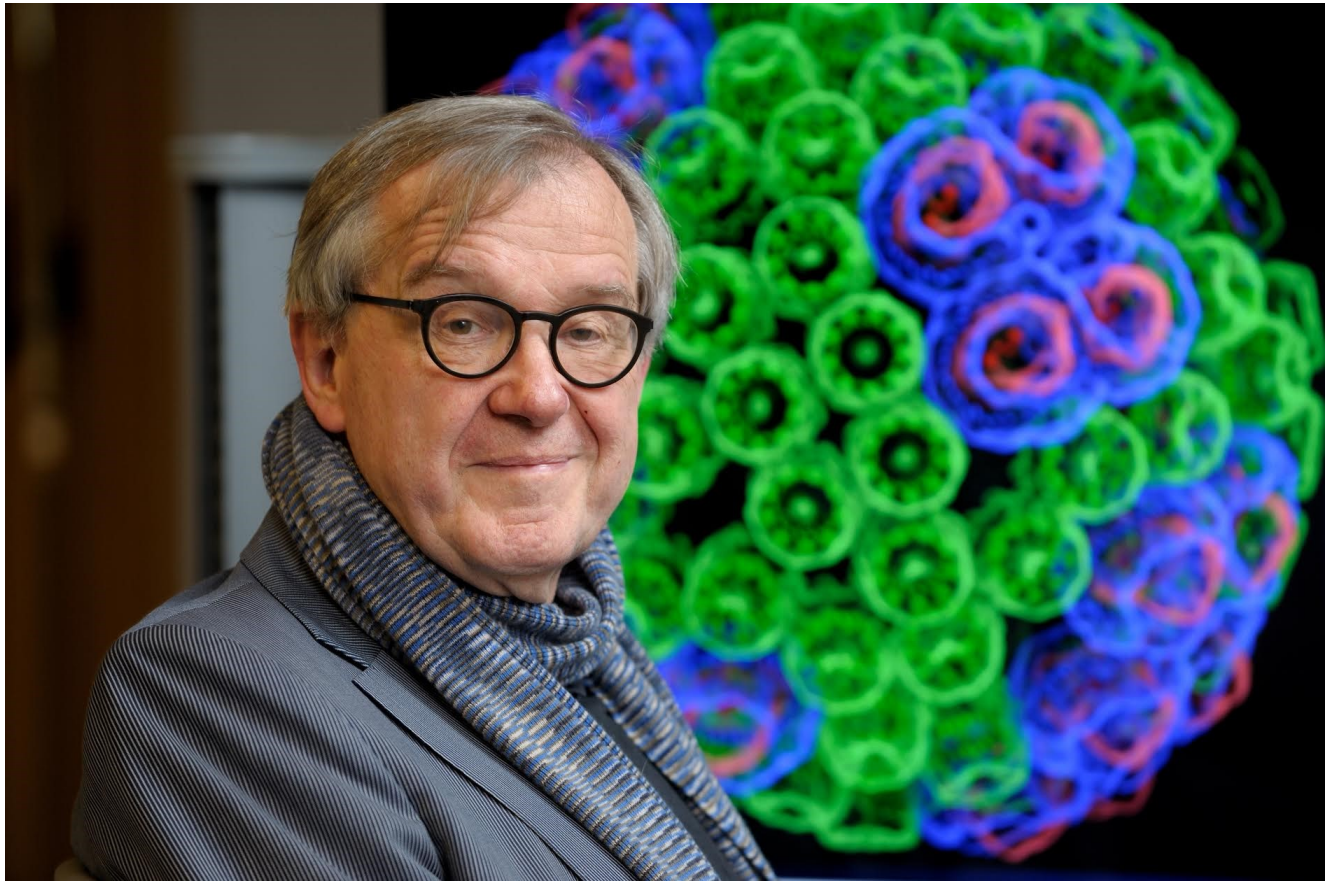
- High resolution cellular tomograms, **billions of voxels**
- Even isosurface or lattice site graphical representations involve ~100M geometric primitives
- Amazon P3 instance types
- **Tesla V100 GPUs benefit from OptiX support for NVLink and distribution of scene data across multiple GPUs**




Earnest, et al. J. Physical Chemistry B, 121(15): 3871-3881, 2017.

# Acknowledgements


- Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
- NVIDIA CUDA and OptiX teams
- Funding:
  - NIH support: P41GM104601
  - DOE INCITE, ORNL Titan: DE-AC05-00OR22725
  - NSF Blue Waters:
    NSF OCI 07-25070, PRAC “The Computational Microscope”,
    ACI-1238993, ACI-1440026



*“When I was a young man, my goal was to look with mathematical and computational means at the inside of cells, one atom at a time, to decipher how living systems work. That is what I strived for and I never deflected from this goal.” – Klaus Schulten*