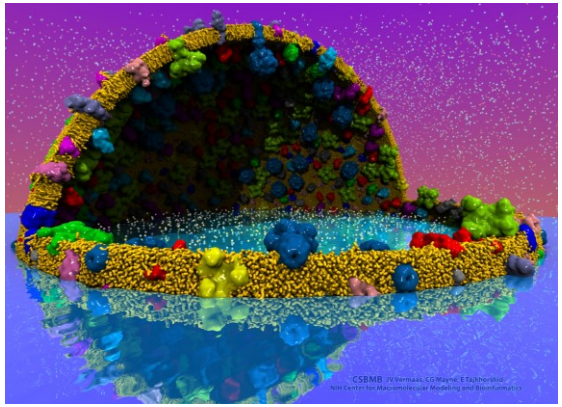
Interactive HPC Requirements, Challenges, and Solutions for Cutting Edge Molecular Simulation Science Campaigns

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/vmd/

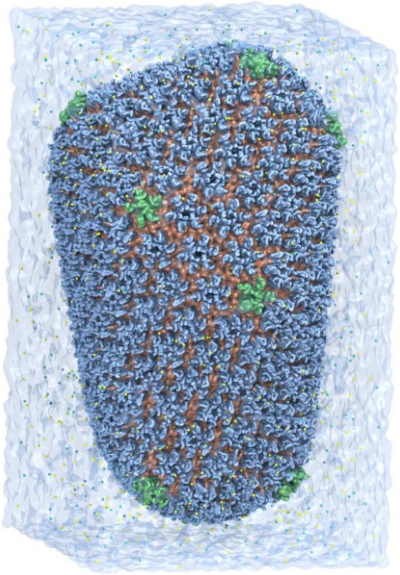
First Workshop on Interactive High-Performance Computing, ISC 2018
14:10-14:35, Alabaster 2, Marriot Hotel,
Frankfurt, Germany, Thursday June 28th, 2018

VMD – “Visual Molecular Dynamics”

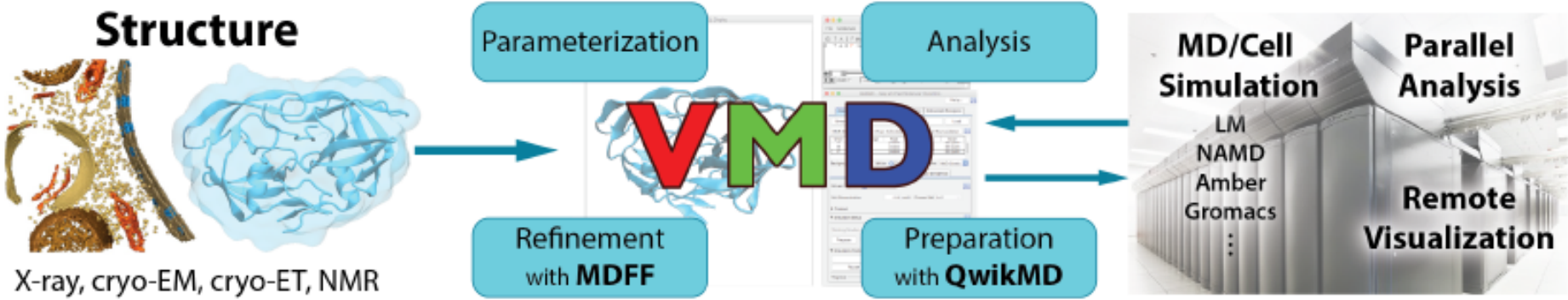
- Visualization and analysis of:
  - Molecular dynamics simulations
  - Lattice cell simulations
  - Quantum chemistry calculations
  - Sequence information
- User extensible scripting and plugins
- http://www.ks.uiuc.edu/Research/vmd/



Cell-Scale Modeling

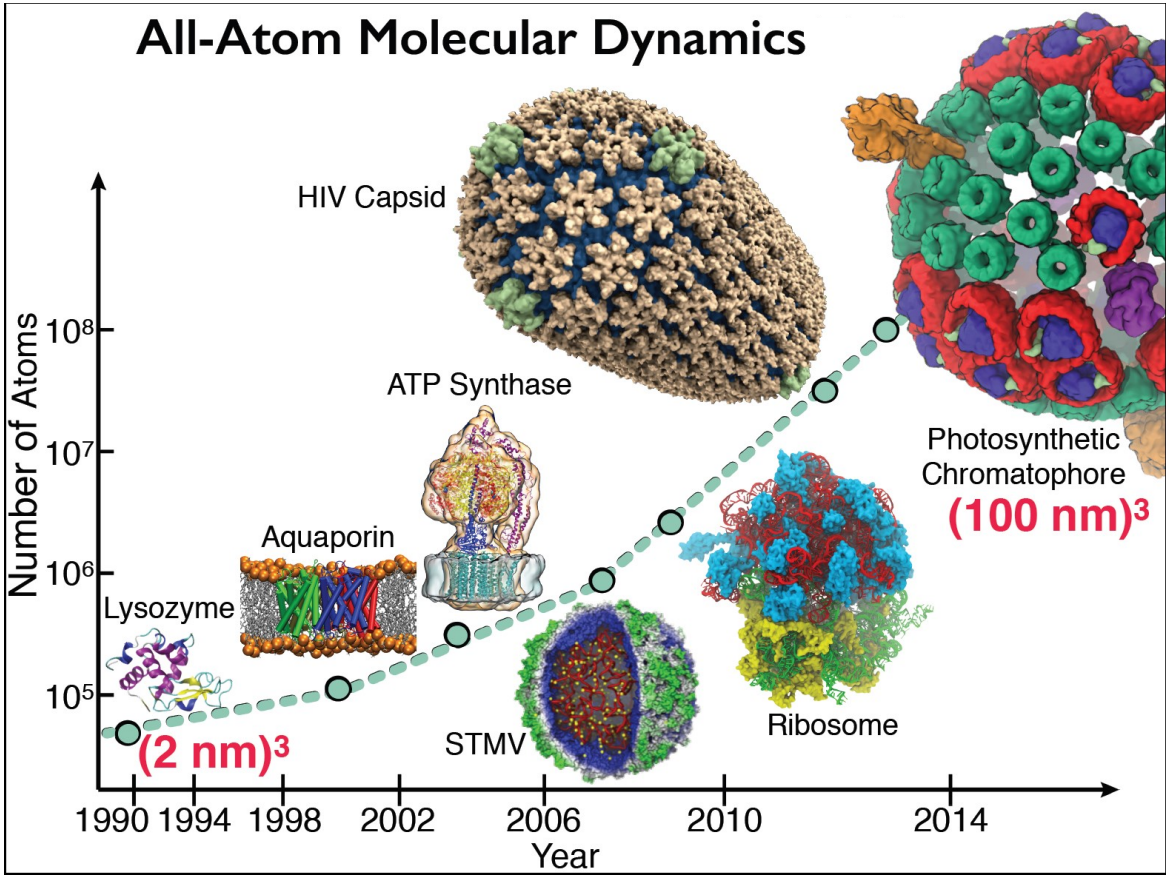


MD Simulation



Goal: A Computational Microscope

Study the molecular machines in living cells



How Can Interactivity in HPC Benefit Scientific Productivity

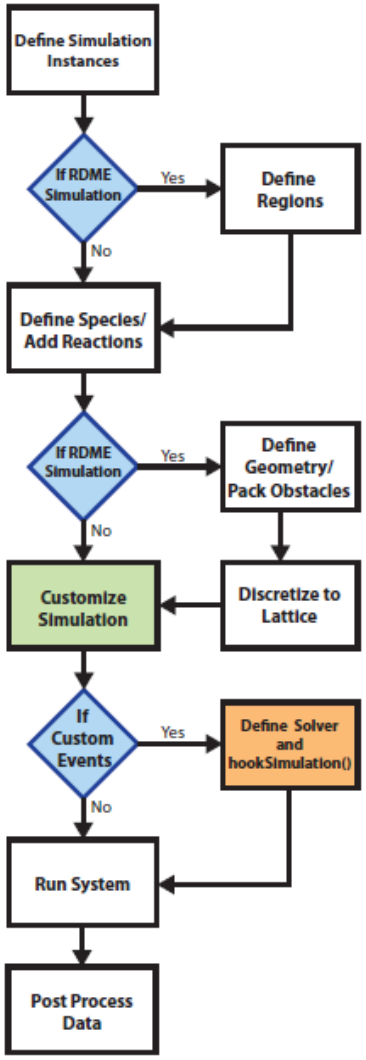
- Historically, MD simulation performance was the major limiting factor on large-scale biomolecular simulation productivity, maybe 95% factor
- 2000s: Batch MD sim performance a gradually shrinking factor affecting overall scientific productivity
- **2012 onward: with petascale supercomputers, "all of the other tasks" now gate scientific productivity as much as simulation throughput does**
- **An extremely simplified outline:**
  - **Months** of science team effort go into **preparing large biomolecular simulations** w/ experimental structures before MD sims begin
  - **Early phases of simulation campaign,** hundreds of modeling problems must be identified and addressed, **interactivity has a huge impact on productivity**
  - Batch simulation commences at scale…
  - **Analytical work begins, interactivity is again critical to science productivity**
  - Science teams make key decisions about outcomes, potentially conducting further simulations, developing manuscripts using analyses, figures, and movies of dynamics of systems under study

What Do Molecular Scientists Need?

- Interactive resources for **human-intensive** science team activities:
  - **Science campaign "bullpen" interactive compute/viz**
  - **Interactive supercomputing "front end" w/ large memory capacities (persistent memory NVDIMMs) dense GPU counts**
  - **On-demand batch queues, exploit suspend-resume of other jobs**
- Workflows composed from **arbitrary science tools**, including **non-traditional HPC tools** (e.g. CryoEM and X-Ray crystallography pkgs)
- M**eso-scale interactive calculations** performed **"at a mouse click"**
  - Support early-phase simulation bring-up activities
  - provide rapid-turnaround of results in interactive modeling sessions to correct flaws in molecular models, simulation parameters, etc.

Broader Motivations for Interactive HPC

- Ongoing push for greater **reproducibility**
  - **Sharing of data** to permit further/alternative analyses by others
  - Large data difficult to “share” w/ non-HPC environments
  - **Remote access to interactive, graphically-driven science apps, eliminate roadblocks for non-HPC-expert scientists**
- **Interactive “lab notebook” environments**
  - Very popular among data scientists
  - Encapsulate data, math, analysis, visualization
  - Mathematica, **Jupyter**
  - **HPC apps can be driven directly by Jupyter notebook**
  - **Jobs run remotely on cloud, cluster, supercomputer…**
- To reach the scale of modern HPC workloads, we need software infrastructure and HPC center policies that facilitate interactivity and modern workflow technologies



jupyter Tut1.2-StochBimol Last Checkpoint: 6 hours ago (unsaved changes)

File Edit View Insert Cell Kernel Help Python 3

Markdown CellToolbar

Tutorial 1.2 - Stochastic Solution of a Bimolecular Reaction

Here we examine a stochastic version of Tutorial 1.1.

In Python you "import" libraries to be able to use their functionality. The first several lines import certain functionality including certain operating system functions (os), standard numeric capabilities that are much like Matlab (numpy) and plotting capabilities (matplotlib). There lines are boiler-plate code for most pyLM scripts.

In order to use pyLM we need to import several libraries. The first is pyLM proper (pyLM). The second is a library with a number of functions such as `nm()`, `micron()`, `ms()`, `microsecond()`, etc. that allow cleaner definition of units. Finally, we import the `pyLM` standard library of functionality `pySTDLM`, which contains standard plotting and post-processing commands.

In [ ]:

```
# Import Standard Python Libraries
import os
import numpy as np
#import matplotlib.pyplot as plt


# Import pyLM Libraries
from pyLM import *
from pyLM.units import *
from pySTDLM import *
from pySTDLM.PostProcessing import *

# Enable plotting inline in the Jupyter notebook
%matplotlib inline
```

Hit Shift+Enter to execute cell

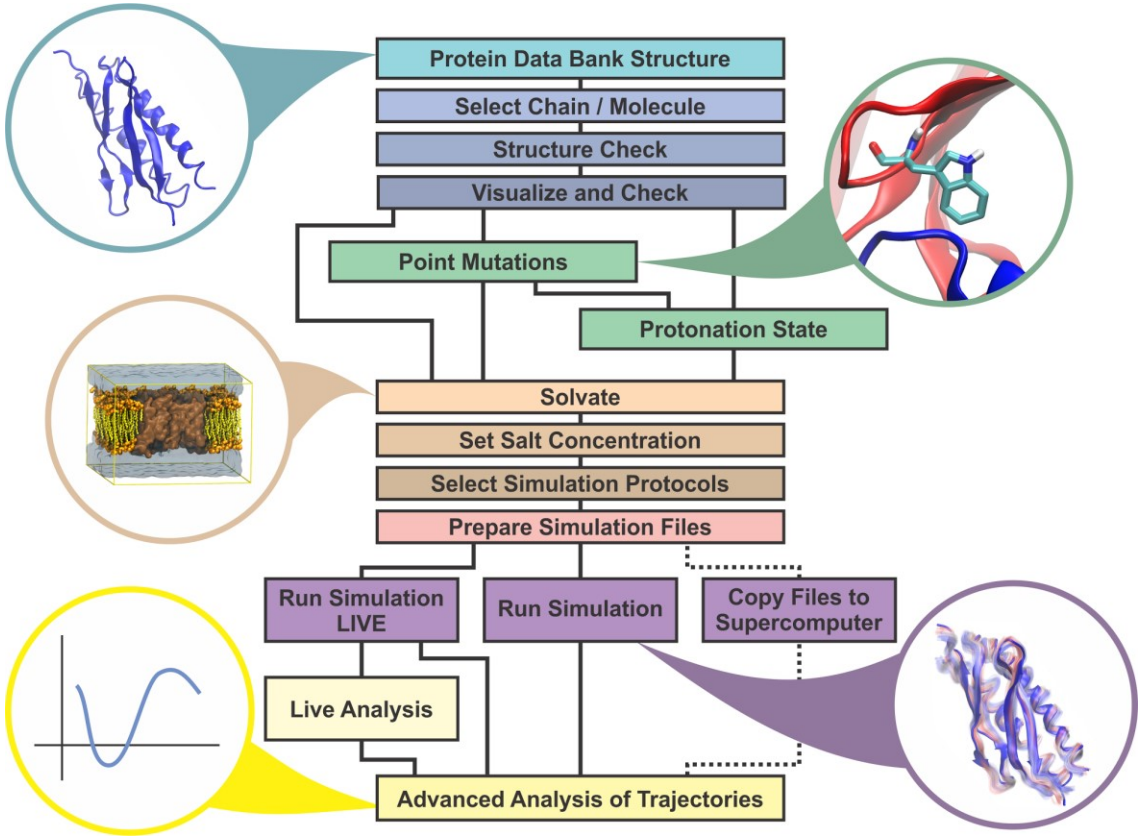
QwikMD: Interactive, Guided MD Simulation

Interactive preparation, simulation, and analysis

Smooths initial learning curve (non-expert users)

Speed up tedious simulation preparation tasks (expert users)

Reproducibility: detailed log of all steps



Ribeiro, et al. QwikMD - Integrative Molecular Dynamics Toolkit for Novices and Experts. ***Sci. Rep.*** 6, 26536, 2016

Making Our Research Tools and Science Workflows Easily Accessible

- 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**

- 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/**

Clusters, Supercomputers

Workstations, Servers, Cloud

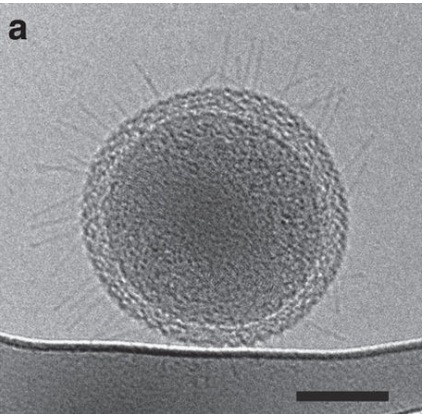
Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. Abhishek Singharoy, Ivan Teo, Ryan McGreevy, John E. Stone, Jianhua Zhao, and Klaus Schulten. *eLife*, 10.7554/eLife.16105, 2016. (66 pages).

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.

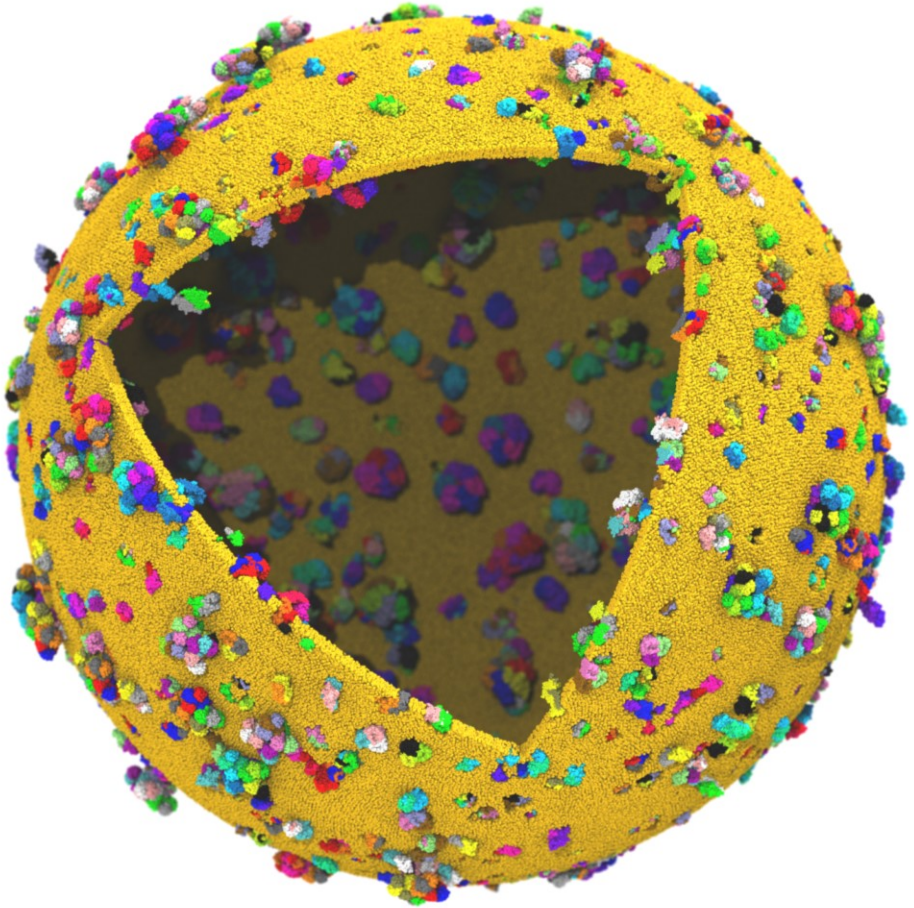
High performance molecular visualization: In-situ and parallel rendering with EGL. John E. Stone, Peter Messmer, Robert Sisneros, and Klaus Schulten. *2016 IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW)*, pp. 1014-1023, 2016.

Next Generation: Simulating a Proto-Cell

- Emulate aspects of the *Mycoplasma mycoides* bacterium
- 200nm diameter
- ~1 billion atoms w/ solvent
- ~1400 proteins in membrane

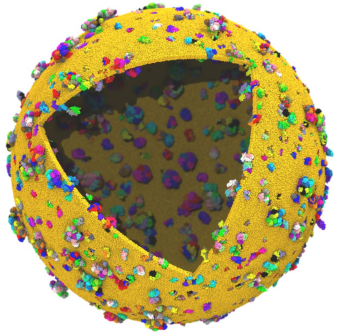


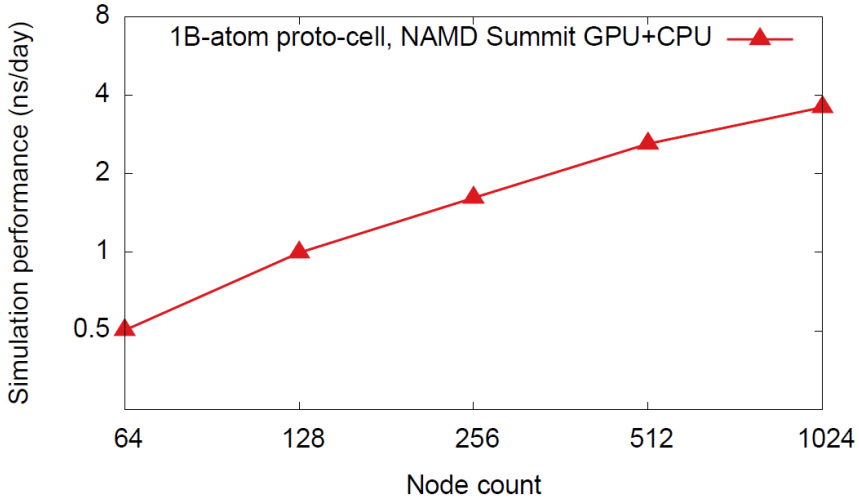
Cryo-ET image of ultra-small bacteria (scale bar 100nm) Luef et al. Nature Comm., 6:6372, 2015.



Proto-Cell Data Challenges

- **1B-atom proto-cell requires nodes with more than TB RAM to build complete model…**
- **1B-atom proto-cell binary structure file: 63GB**
- **Trajectory frame atomic coordinates: 12GB, 1.2TB/ns of simulation (1 frame per 10ps)**
- **Routine modeling and visualization tasks are a big challenge at this scale**
  - **Models contain thousands of atomic-detail components** that must work together in harmony
  - **Exploit persistent memory technologies** to enable “instant on” operation on massive cell-scale models – eliminate several minutes of startup during analysis/visualization of known structure
  - **Sparse output of results at multiple timescales** will help ameliorate visualization and analysis I/O
  - **Data quantization, compression, APIs like ZFP**





VMD Petascale Visualization and Analysis

- **Analyze/visualize large trajectories too large to transfer off-site:**
  - User-defined parallel analysis operations, data types
  - Parallel rendering, movie making
- Supports GPU-accelerated Cray XK7 nodes for both visualization and analysis:
  - **GPU accelerated trajectory analysis w/ CUDA**
  - **OpenGL and GPU ray tracing for visualization and movie rendering**
- Parallel I/O rates up to **275 GB/sec** on 8192 Cray XE6 nodes – can read in **231 TB in 15 minutes!**

Parallel VMD currently available on:

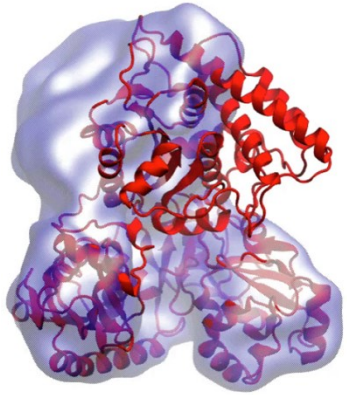
ORNL Titan, NCSA Blue Waters, Indiana Big Red II, CSCS Piz Daint, and similar systems

NCSA Blue Waters Hybrid Cray XE6 / XK7
22,640 XE6 dual-Opteron CPU nodes
4,224 XK7 nodes w/ Telsa K20X GPUs

Molecular Dynamics Flexible Fitting (MDFF)

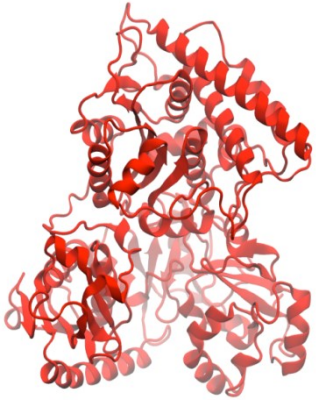
X-ray crystallography → MDFF ← Electron microscopy

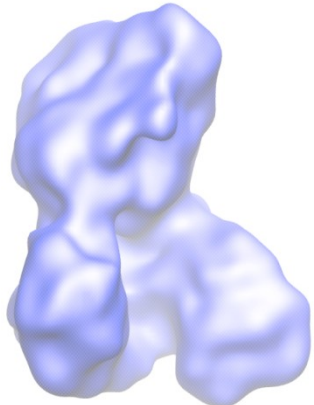
APS at Argonne



FEI microscope

ORNL Titan

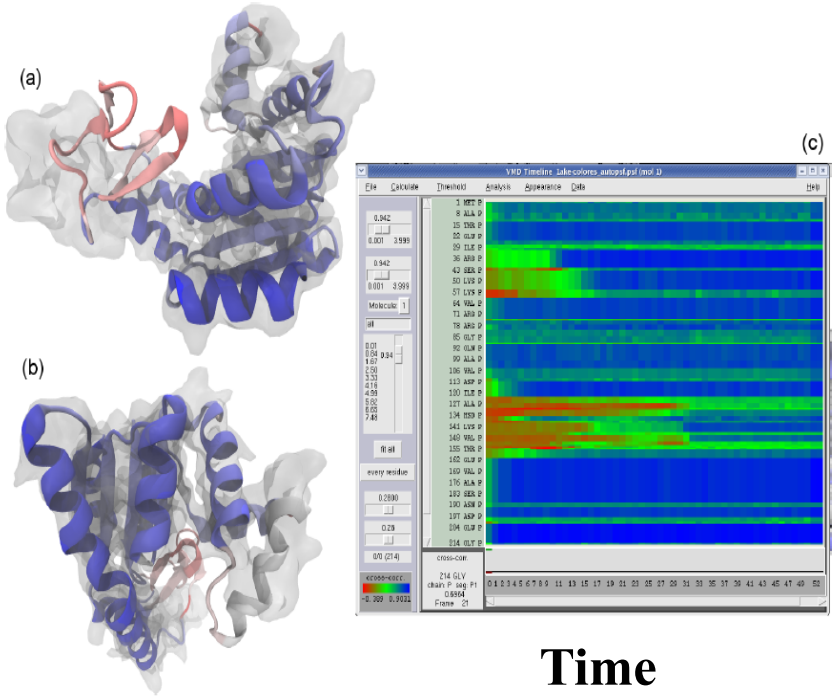




Molecular dynamics-based refinement and validation for sub-5Å cryo-electron microscopy maps. A. Singharoy, I. Teo, R. McGreevy, J. E. Stone, J. Zhao, and K. Schulten. eLife 2016;10.7554/eLife.16105

Interactive and Parallel Analysis

- New graphical interfaces for batch and interactive exploration, calculation
  - User interactions drive analysis focus with progressive refinement of details
  - Interactive in-situ analysis of running simulations
- Enabled by GPU acceleration, parallel computing on desktops, clouds, clusters, and supercomputers

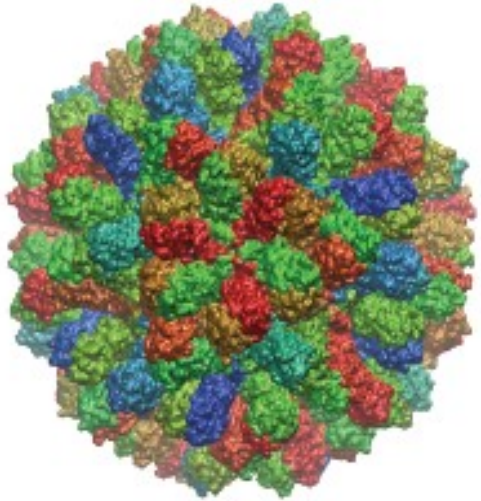


MDFF Cross Correlation Analysis

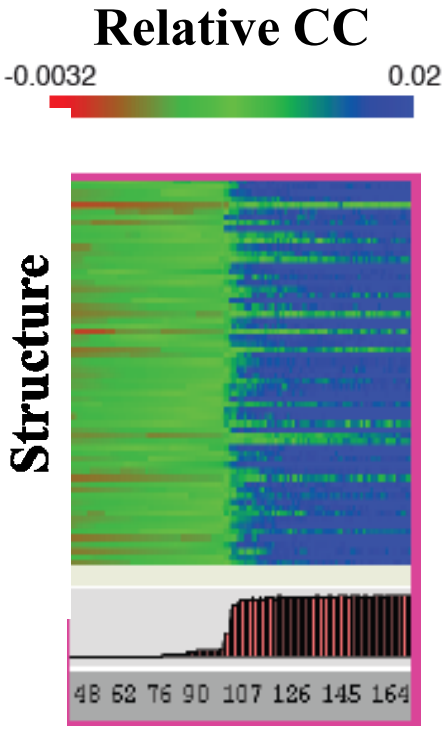
Regions with poor fit **Regions with good fit**

Parallel MDFF Cross Correlation Analysis on Cray XK7

Rabbit Hemorrhagic Disease Virus (RHDV)	
Traj. frames	10,000
Structure component selections	720
Single-node XK7 (projected)	336 hours (14 days)
128-node XK7	3.2 hours 105x speedup
2048-node XK7	19.5 minutes 1035x speedup



RHDV colored by relative CC



Calculation of 7M CCs would take **5 years** using serial CPU algorithm!

Stone et al., Faraday Discuss., 169:265-283, 2014.

VMD Tesla V100 Cross Correlation Performance

Rabbit Hemorrhagic Disease Virus: 702K atoms, 6.5Å resolution

VMD on Volta GPUs now **~9x faster** than Kepler GPUs

37 Summit nodes ~= 2048 BW XK7 nodes:

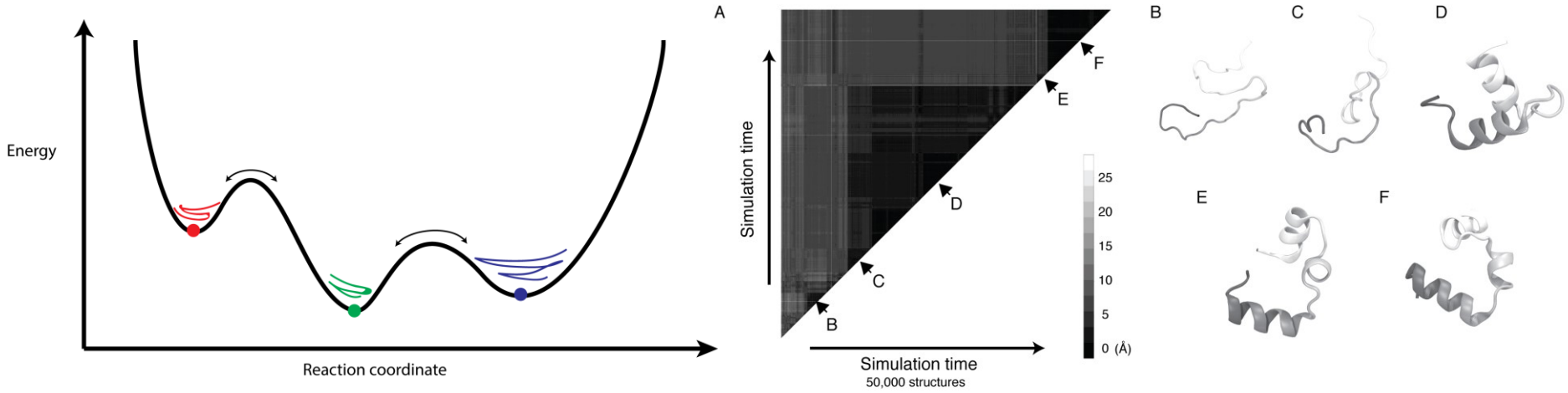
In the realm of "Interactive HPC" now!

Application and Hardware platform	**Runtime, Speedup vs. Chimera,**		**VMD+GPU**
VMD-CUDA Intel Xeon E5-2687W + 1x Quadro K6000 [1,2]	0.458s,	35x	**1.0x**
VMD-CUDA Intel Xeon E5-2697Av4 + 1x Tesla V100	**0.050s,**	317x	**9.2x**
VMD-CUDA IBM Power9 "Newell" + 1x Tesla V100	**0.049s,**	**323x**	**9.3x**

[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.

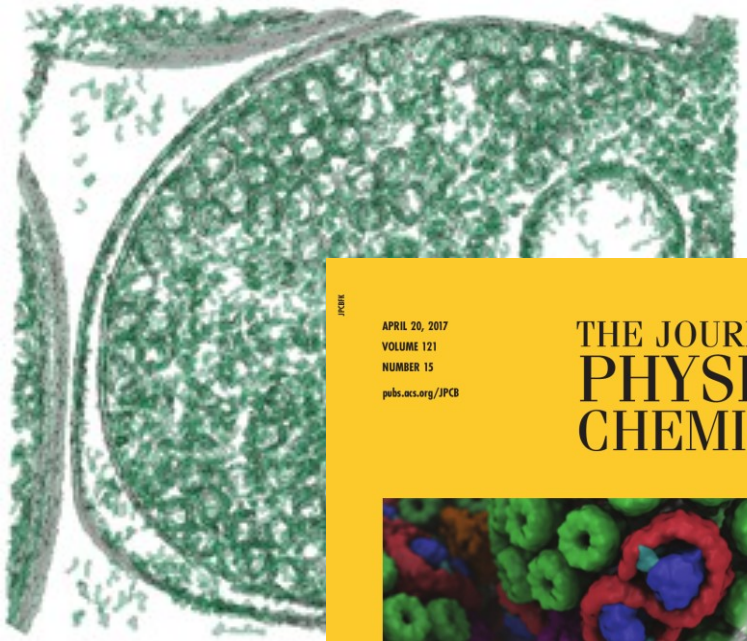
[2] Early Experiences Porting the NAMD and VMD Molecular Simulation and Analysis Software to GPU-Accelerated OpenPOWER Platforms. J. E. Stone, A.-P. Hynninen, J. C. Phillips, K. Schulten. International Workshop on OpenPOWER for HPC (IWOPH'16), LNCS 9945, pp. 188-206, 2016.

Clustering Analysis of Molecular Dynamics Trajectories

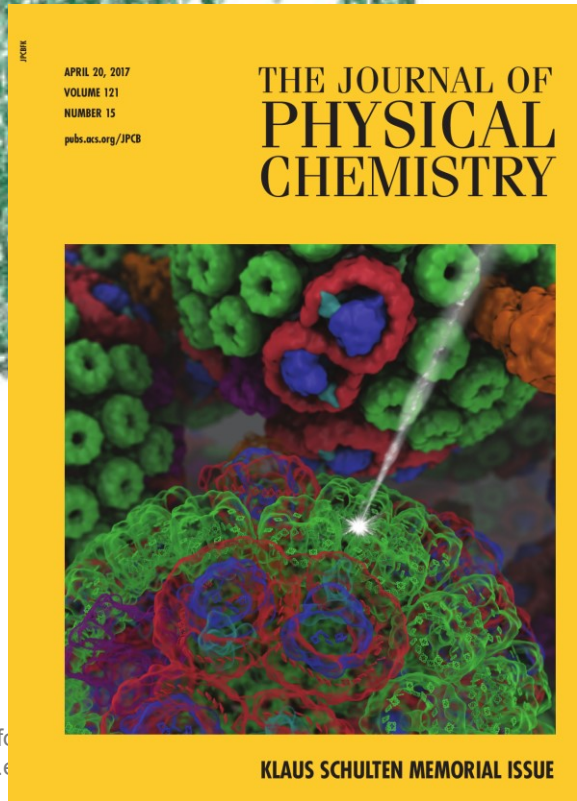


GPU-Accelerated Molecular Dynamics Clustering Analysis with OpenACC. J.E. Stone, J.R. Perilla, C. K. Cassidy, and K. Schulten. In, Robert Farber, ed., Parallel Programming with OpenACC, Morgan Kaufmann, Chapter 11, pp. 215-240, 2016.

Interactive Remote Visualization



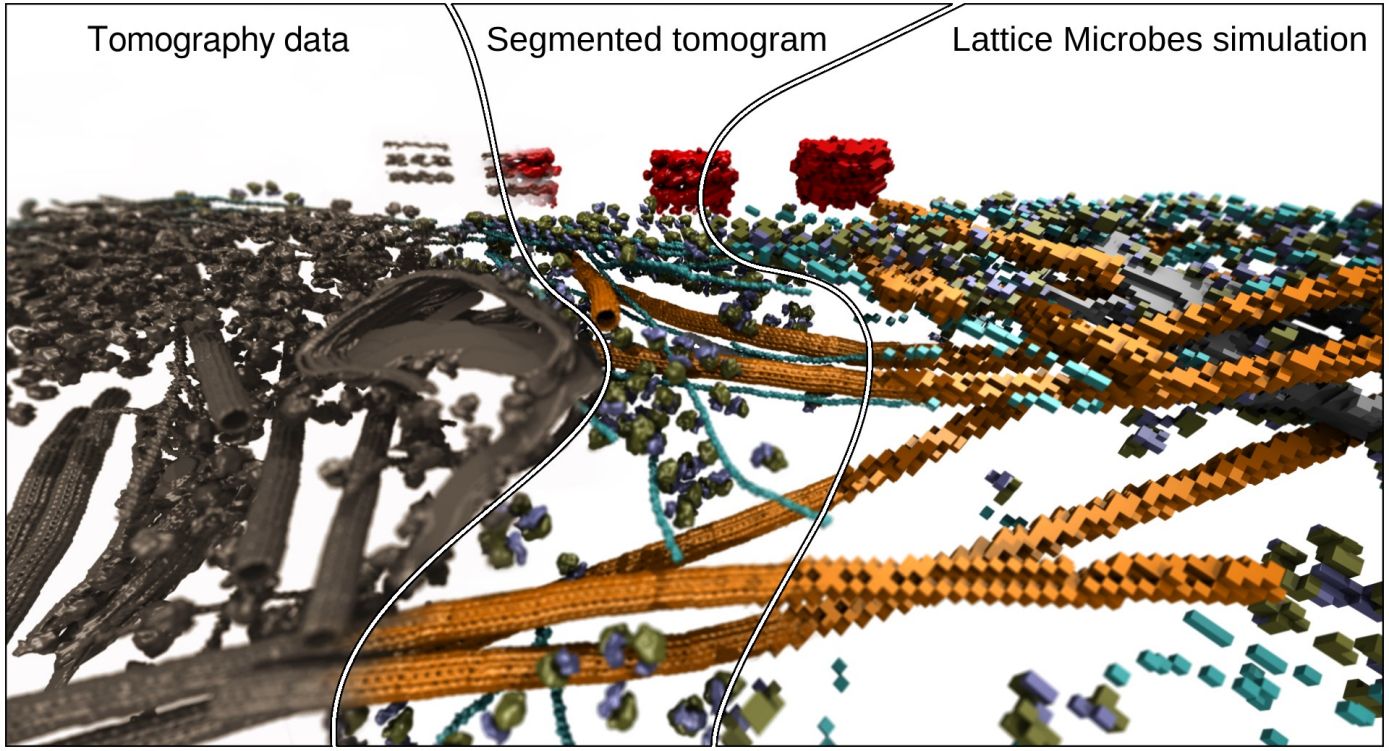
- **Enable access to massive data sets**
- **A Grander Goal: A Thousand-Fold Increase In Human Capabilities.** B. Shneiderman. Educom Review, 32, 6 (Nov/Dec 1997), 4-10.
  - Continuous visual display of status
  - Rapid, incremental, and reversible actions with 100ms updates
  - **Visual Information Seeking mantra: "Overview first, zoom and filter, then details-on-demand"**



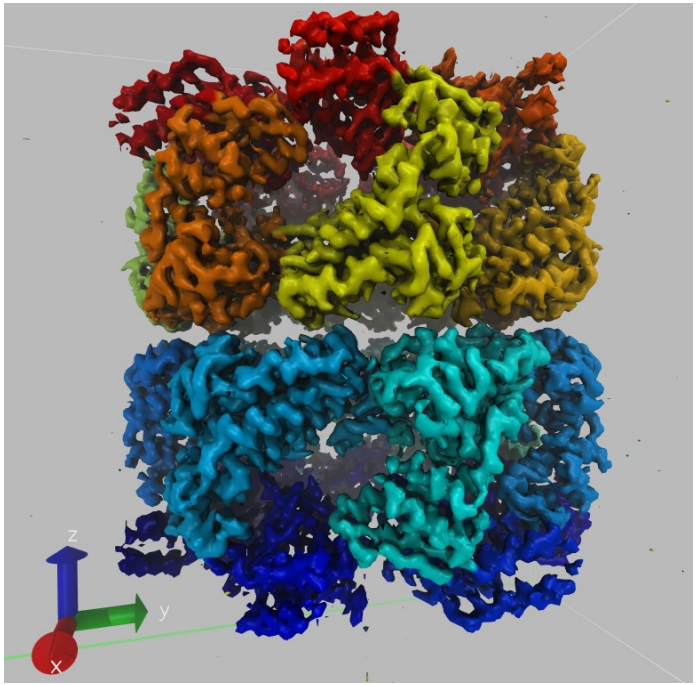
Cryo-EM / Cryo-ET Density Map Segmentation

Evaluate 3-D volumetric electron density maps and segment them, to identify key structural components

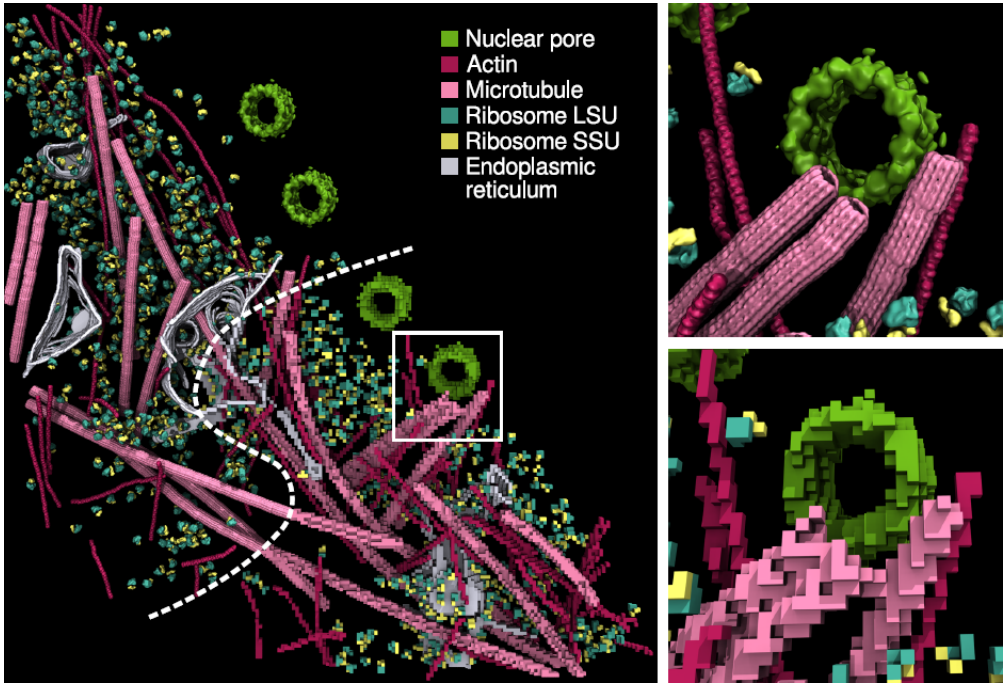
Index/label components so they can be referred to, colored, analyzed, and simulated…



Density Map Segmentation



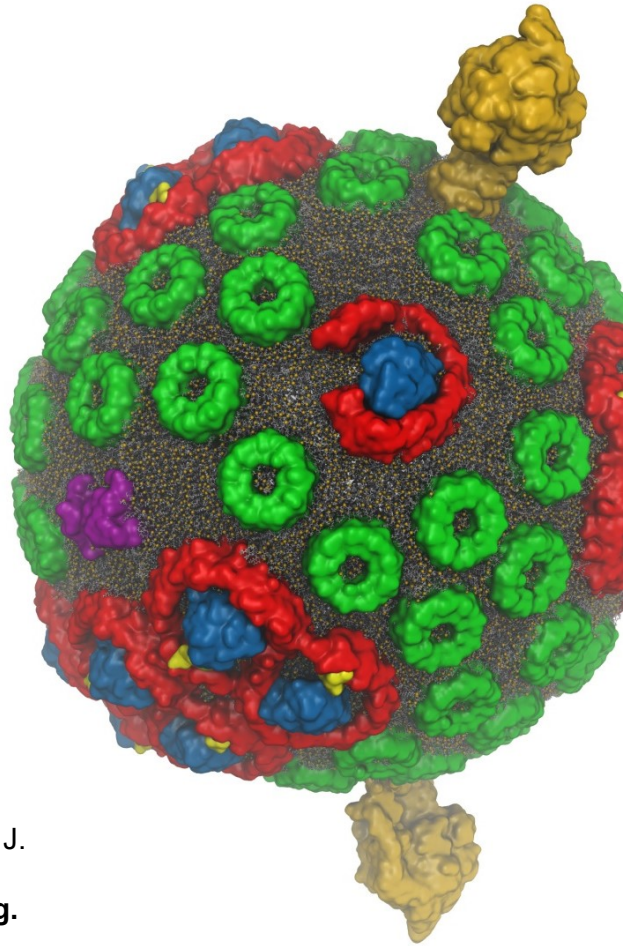
VMD GPU-accelerated density map segmentation of GroEL



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

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.

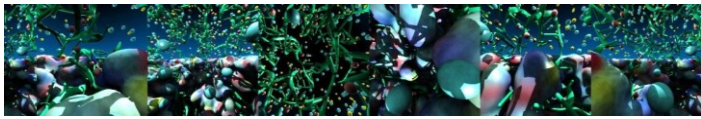
Chemical Visualization of Human Pathogens: the Retroviral Capsids. J. R. Perilla, B.-C. Goh, J. E. Stone, and K. Schulten. SC'15 Visualization and Data Analytics Showcase, 2015.

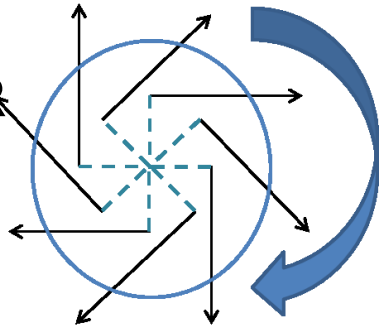
Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing. J. E. Stone et al., J. Parallel Computing, 55:17-27, 2016.

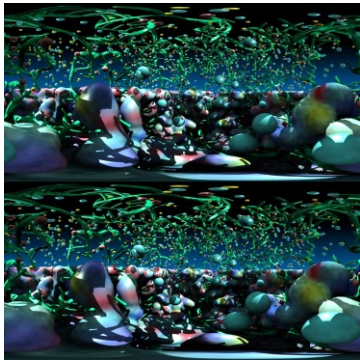
Immersive Molecular Visualization with Omnidirectional Stereoscopic Ray Tracing and Remote Rendering J. E. Stone, W. R. Sherman, and K. HPDAV, IPDPSW, pp. 1048-1057, 2016.

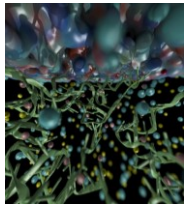
Interactive RT For High-Fidelity Immersive Viz.

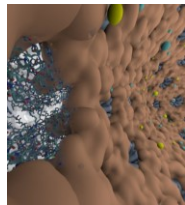
- **Permit immersive interaction, while maintaining high visual fidelity**
- **Ray trace stereoscopic 360° views**
- **H.264, H.265 video streaming to remote VR HMDs**
- Stereo spheremaps or cubemaps allow very high-frame-rate interactive OpenGL display







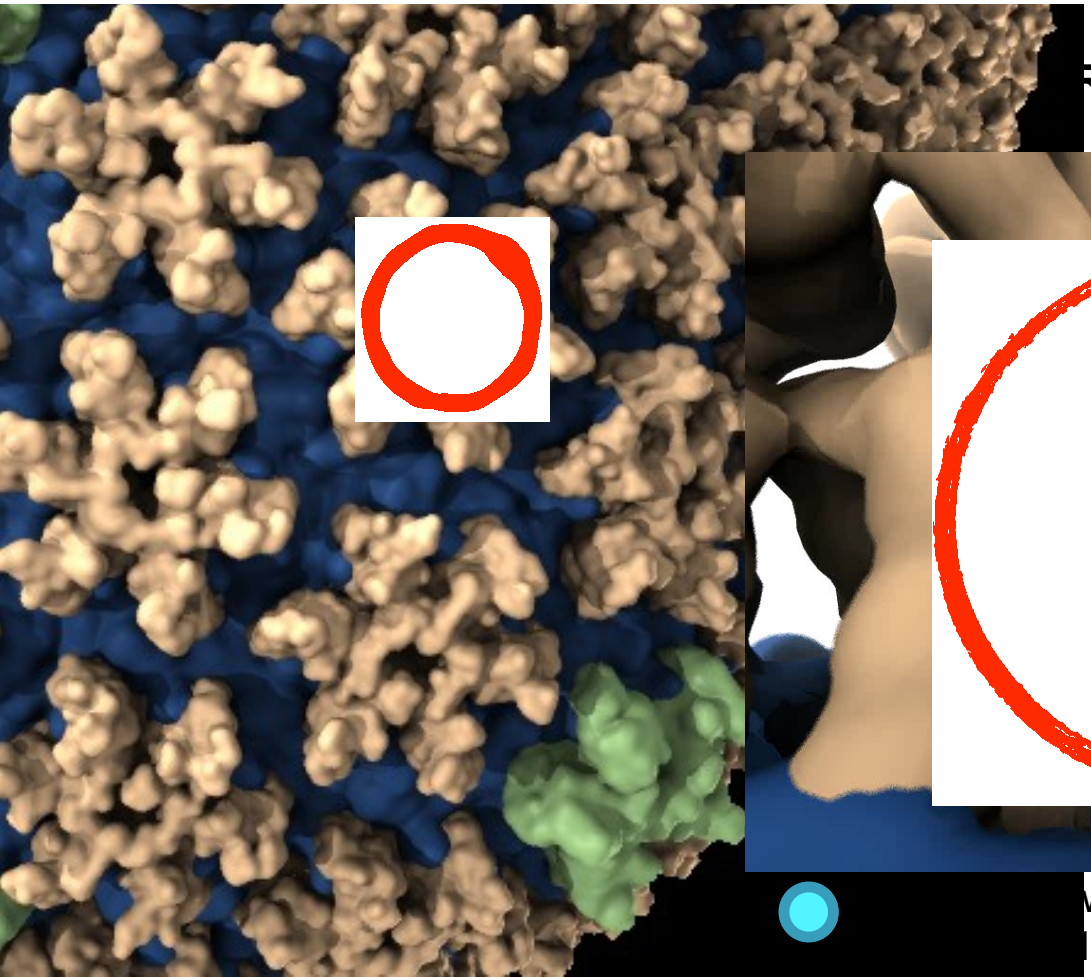




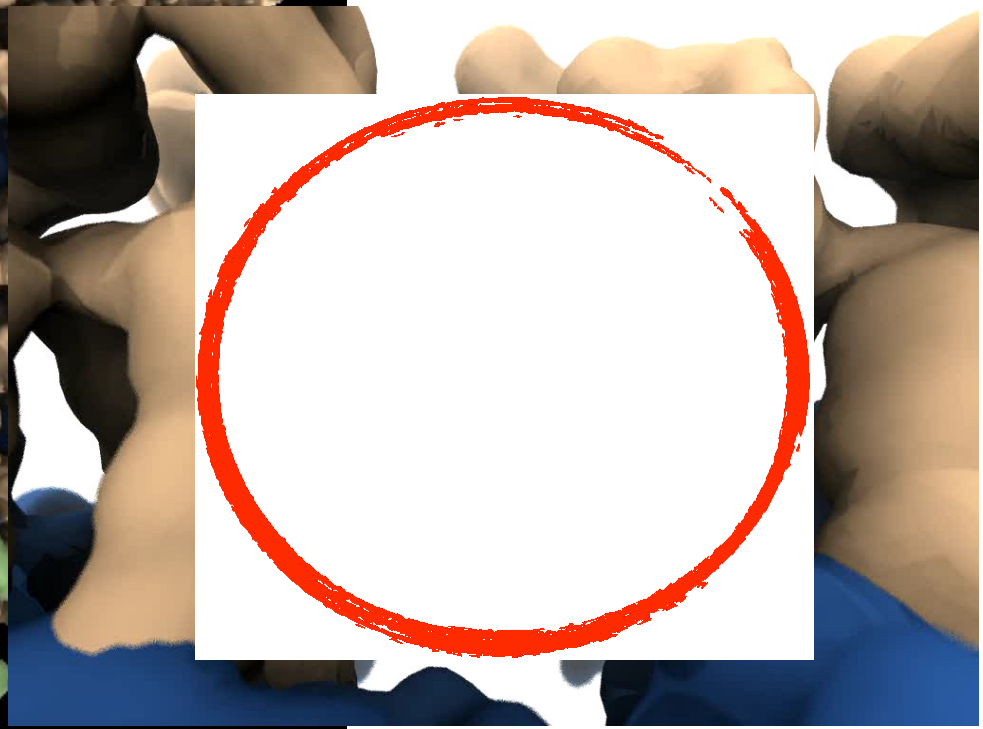
Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing. J. E. Stone, et al. J. Parallel Computing, 55:17-27, 2016.

Immersive Molecular Visualization with Omnidirectional Stereoscopic Ray Tracing and Remote Rendering. J. E. Stone, W. R. Sherman, and K. Schulten. High Performance Data Analysis and Visualization Workshop, IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 1048-1057, 2016.

Goal: Intuitive interactive viz. in crowded molecular complexes



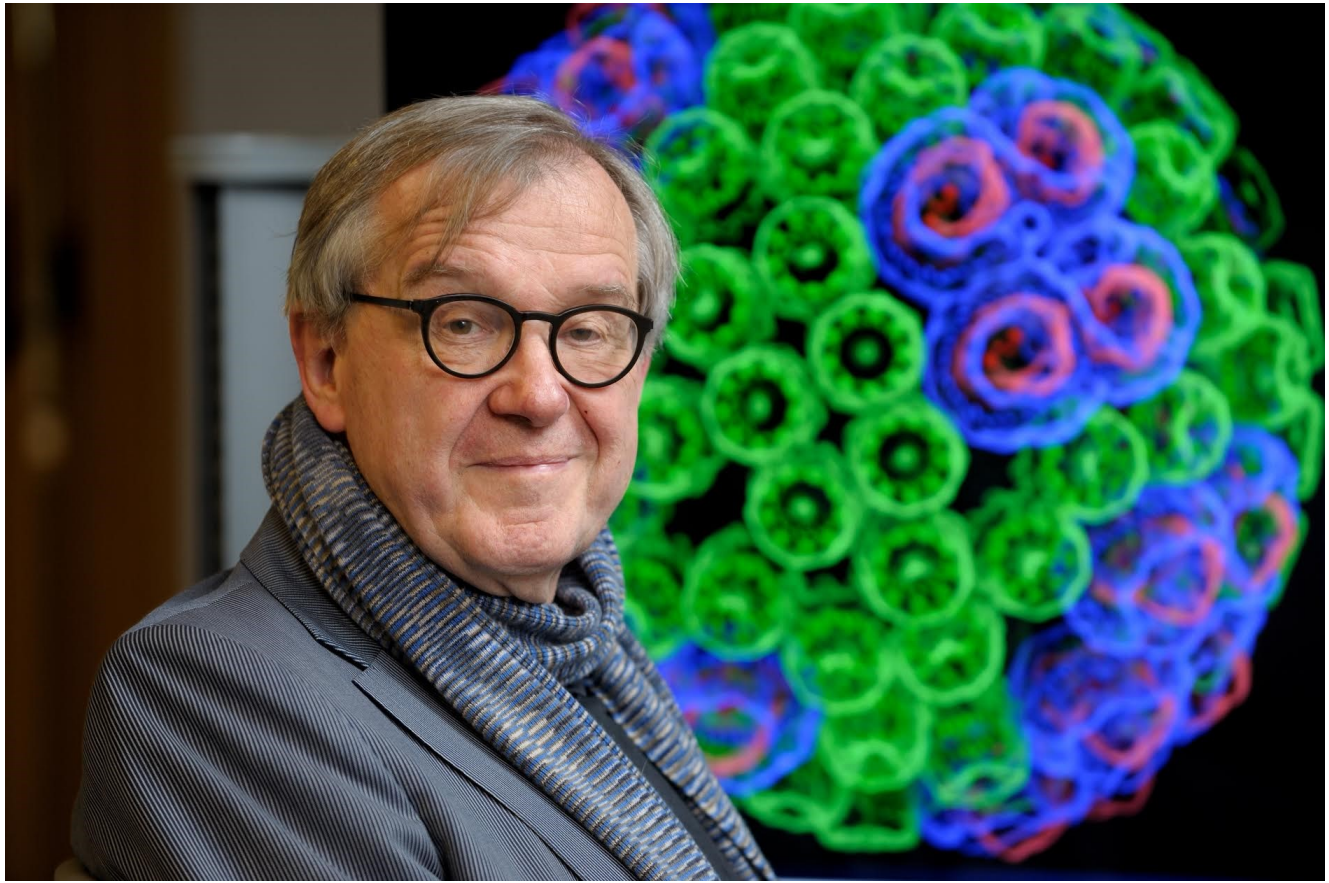
Results from 64 M atom, 1 μs sim!



v of chloride ions permeating
capsid hexameric centers

Acknowledgements

- Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
- 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