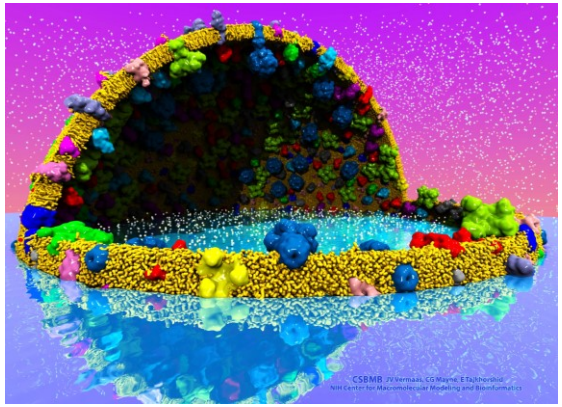
ATPESC Track 4: Scalable Molecular Visualization and Analysis Tools in VMD

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/

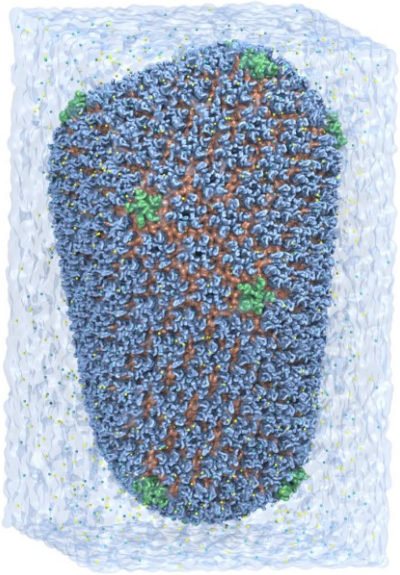
Argonne Training Program on Exascale Computing (ATPESC)
10:15am-11:00pm,
Monday August 9th, 2021

VMD – “Visual Molecular Dynamics”

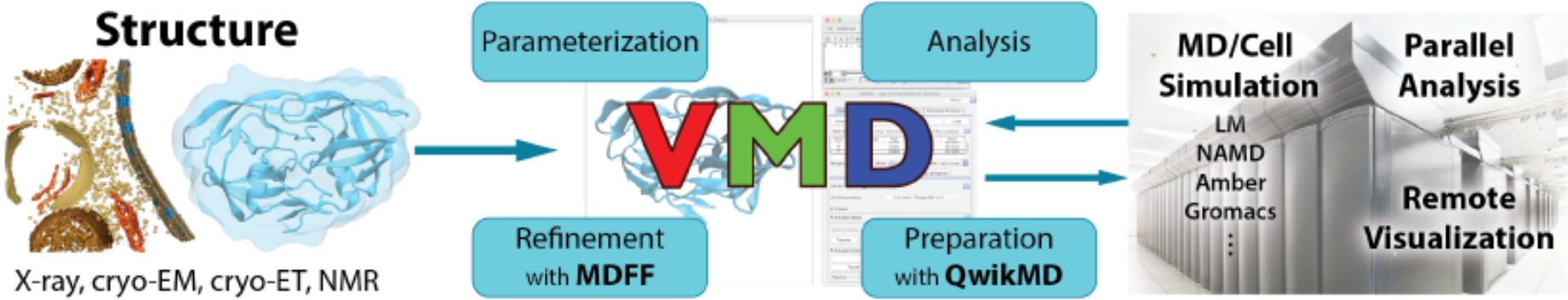
- 100,000 active users worldwide
- Visualization and analysis of:
  - Molecular dynamics simulations
  - Lattice cell simulations
  - Quantum chemistry calculations
  - Cryo-EM densities, volumetric data
- User extensible scripting and plugins
- http://www.ks.uiuc.edu/Research/vmd/



Cell-Scale Modeling



MD Simulation



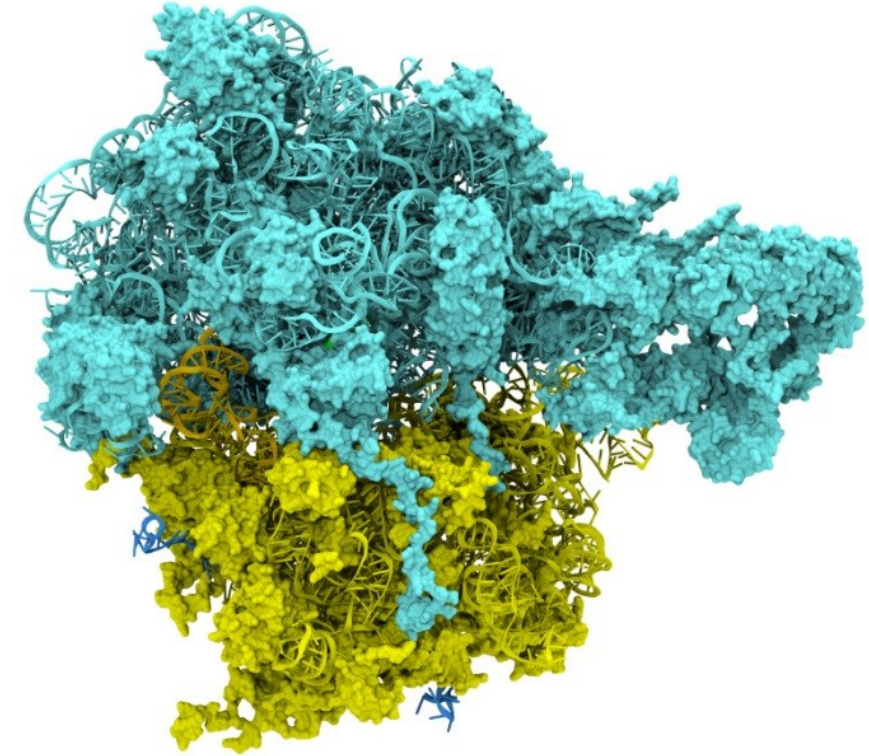
VMD Hands-On Tutorials

- http://www.ks.uiuc.edu/Training/Tutorials/#vmd
  - Main VMD tutorial
  - QwikMD simulation preparation and analysis plugin
  - VMD images and movies tutorial
  - Structure check
  - VMD quantum chemistry visualization tutorial
  - Visualization and analysis of CPMD data with VMD
  - Parameterizing small molecules using ffTK

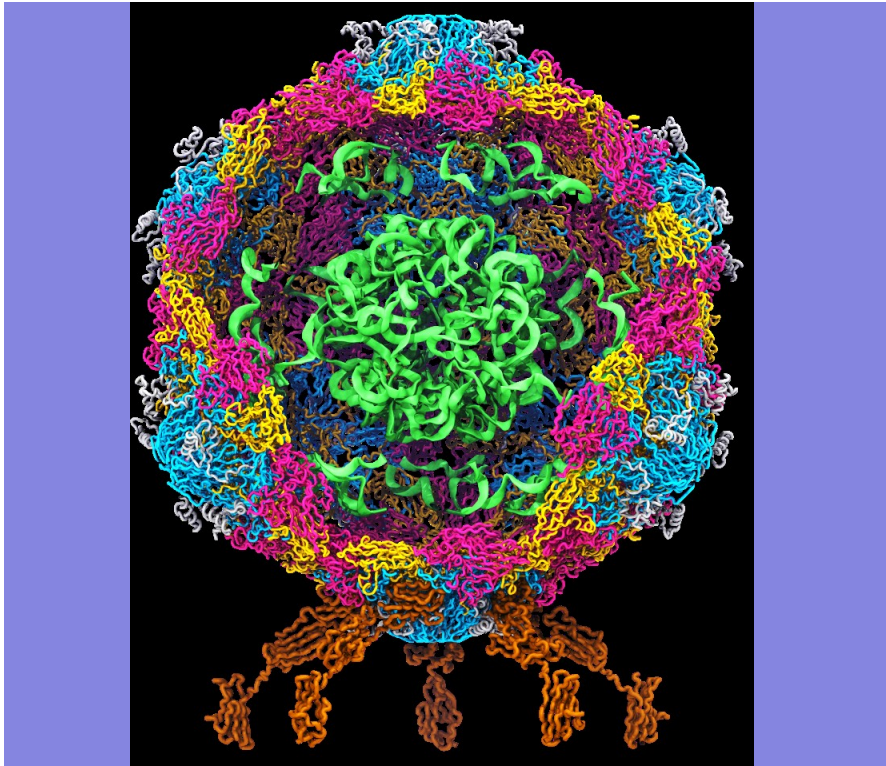
Goal: A Computational Microscope

Study the molecular machines in living cells

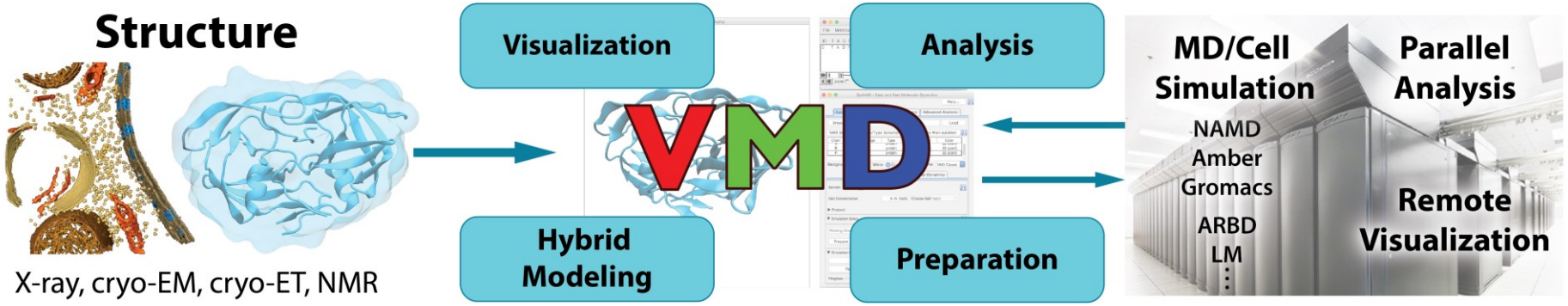
Ribosome: target for antibiotics



Poliovirus



VMD: Building A Next Generation Modeling Platform



- Provide tools for simulation preparation, visualization, and analysis
  - Reach cell-scale modeling w/ all-atom MD, coarse grained, Lattice Microbes
  - Improved performance, visual fidelity, exploit advanced technologies (GPUs, VR HMDs)
- Enable hybrid modeling and computational electron microscopy
  - Load, filter, process, interpret, visualize multi-modal structural information
- Connect key software tools to enable state-of-the-art simulations
  - Support new data types, file formats, software interfaces
- Openness, extensibility, and interoperability are VMD hallmarks
  - Reusable algorithms made available in NAMD, for other tools

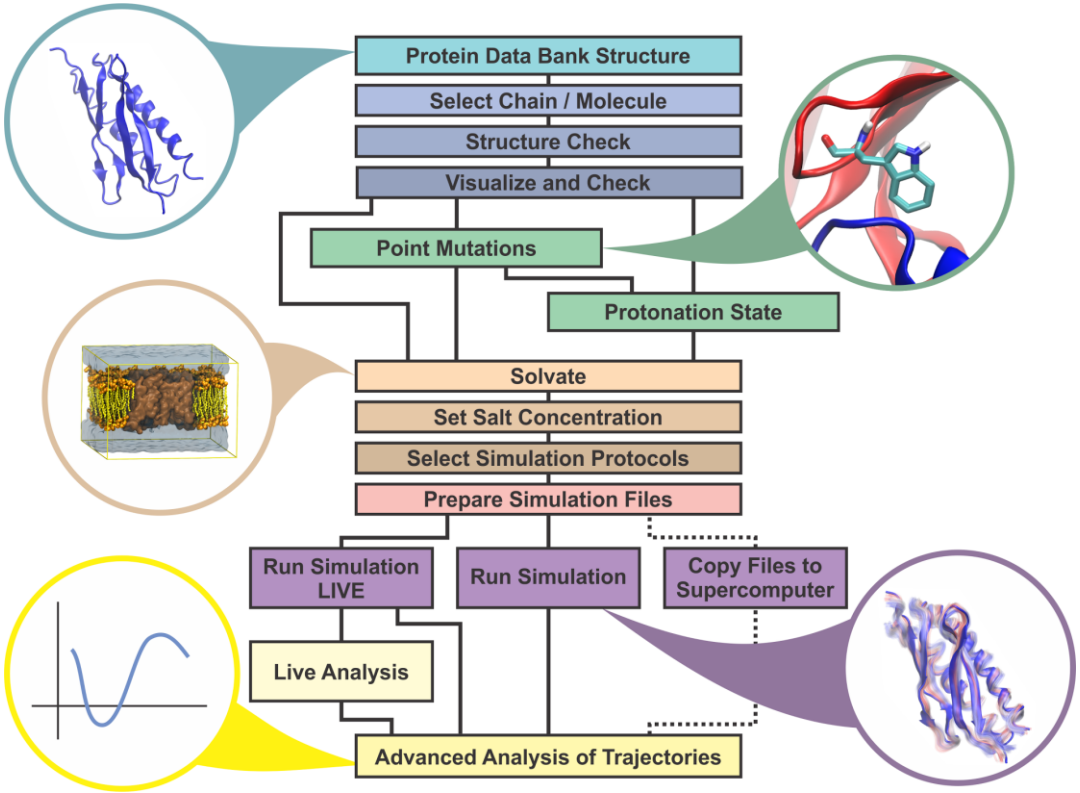
QwikMD: Guided MD Simulation and Training

Smooths initial learning curve (non-expert users)

Speed up tedious simulation preparation tasks (expert users)

Reproducibility: detailed log of all steps

Interactive preparation, simulation, and analysis



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

Selected VMD Plugins: Center Developed, and User Developed

Analysis

- APBSRun
- CatDCD
- Contact Map
- GofRGUI
- HeatMapper
- ILSTools
- IRSpecGUI
- MultiSeq
- NAMD Energy
- NAMD Plot
- NetworkView
- NMWiz
- ParseFEP
- PBCTools
- PMEpot
- PropKa GUI
- RamaPlot
- RMSD Tool
- RMSD Trajectory Tool
- RMSD Visualizer Tool
- Salt Bridges
- Sequence Viewer
- Symmetry Tool
- Timeline
- TorsionPlot
- VolMap

Modeling

- AutoIonize
- AutoPSF
- Chirality
- Cionize
- Cispeptide
- CGTools
- Dowser
- ffTK
- Inorganic Builder
- MDFF
- Membrane
- Merge Structs
- Molefacture
- Mutator
- Nanotube
- Psfgen
- RESPTool
- RNAView
- Solvate
- SSRestraints
- Topotools

Visualization

- Clipping Plane Tool
- Clone Rep
- DemoMaster
- Dipole Watcher
- Intersurf
- Navigate
- NavFly
- MultiMolAnim
- Color Scale Bar
- Remote
- Palette Tool
- ViewChangeRender
- ViewMaster
- Virtual DNA Viewer
- VMD Movie Maker

Simulation

- AlaScan
- AutoIMD
- IMDMenu
- NAMD GUI
- NAMD Server
- QMTool

Collaboration

- Remote Control

Data Import and Plotting

- Data Import
- Multiplot
- PDBTool
- MultiText

Externally Hosted Plugins and Extensions

- Check sidechains
- MultiMSMS
- Interactive Essential Dynamics
- Mead Ionize
- Clustering Tool
- iTrajComp
- Swap RMSD
- Intervor
- SurfVol
- vmdICE

75 MolFile I/O Plugins:
structure, trajectory, sequence, and density map

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

VMD Visualization Concepts

VMD Approach to Visualization

- Molecular scene is composed of “graphical representations”
- Each representation encapsulates a group of selected atoms, a drawing style, coloring style, and other parameters
- Representations are independent of each other, can be toggled on/off easily, allowing molecular scenes to be built-up incrementally
- VMD atom selection language is shared with its analytical and scripting interfaces

Selection, Filtering

- Most viz tools allow interactive visual picking, menu-driven selections of structure components to display or operate on
- VMD also extensively uses a text-based selection language (think google):

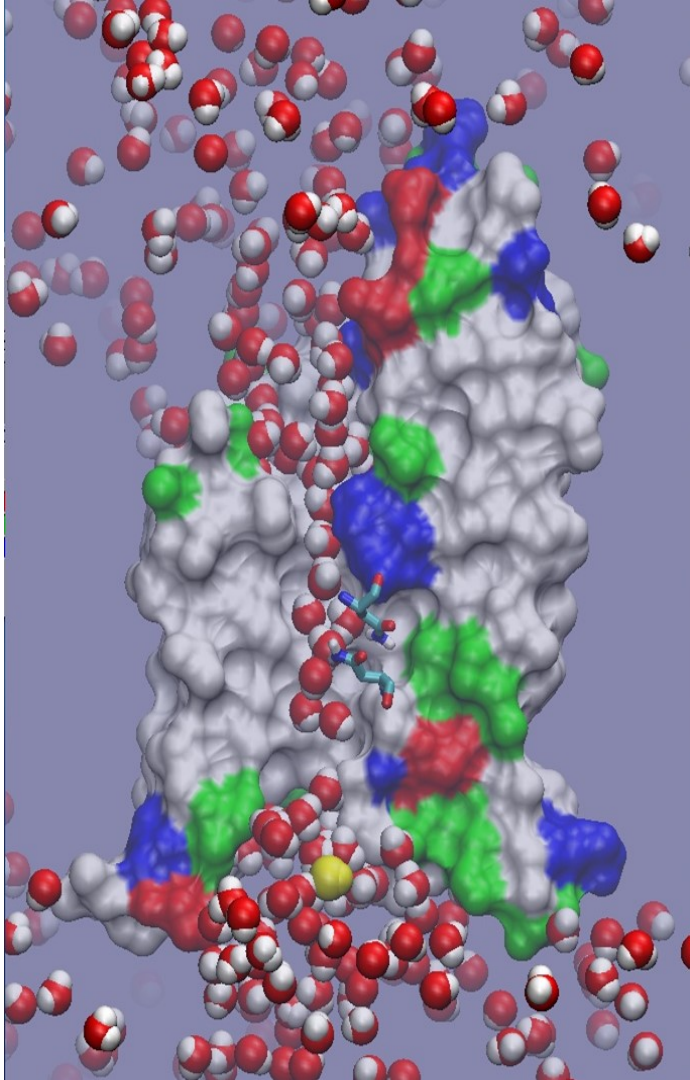
  water within 10 of protein and z > 0

  nucleic or protein or ions

  segname BR

  name “C.*”

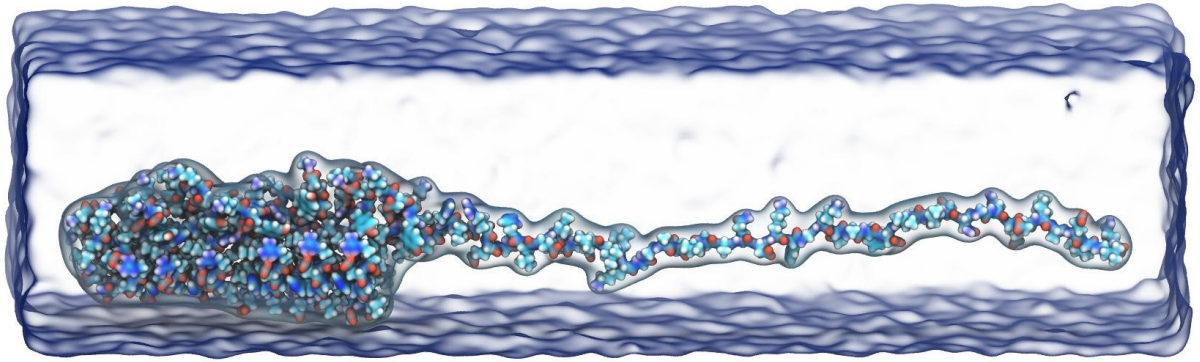
  – Allows selection on user-defined data fields
  – ***Promotes synergy between interactive and scripting interfaces, visualization and quantitative analysis tasks***
  – Works well with huge time-varying structures



Structure Visualization

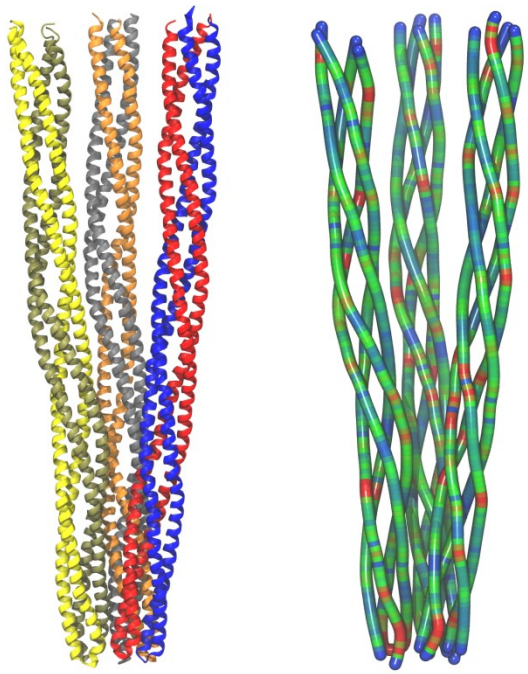
Molecular representations provide different levels of abstraction, atomic detail vs. higher level organizational information

- Atoms, VdW spheres, bonds, ball-stick, …
- Coarse-grained “beads”
- Ribbons, secondary structure, “cartoon” reps, RNA/DNA
- Molecular surfaces
- Molecular orbitals (quantum chemistry)



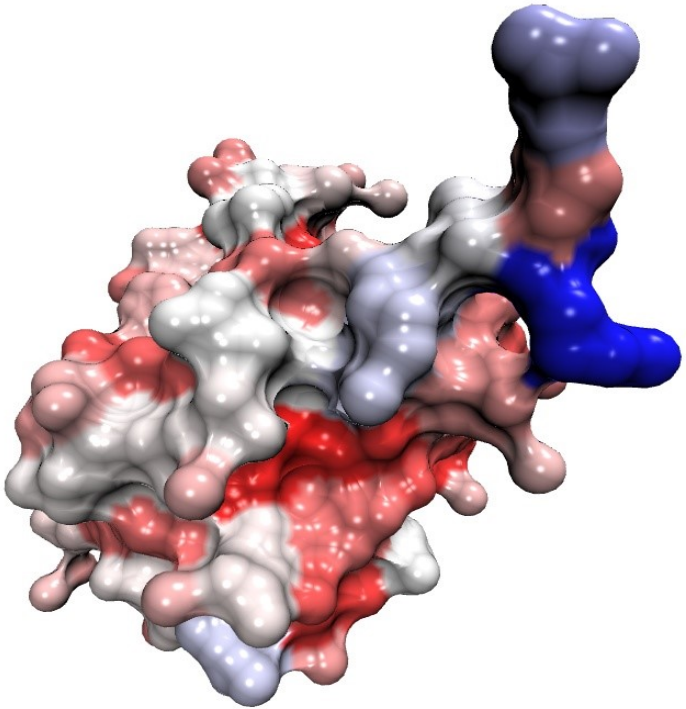
Computed Properties

- Smoothing of thermal noise
- Secondary structure
- Hydrogen bonds, salt bridges
- Forces, energies, stress, strain
- Time averaging of electrostatic fields, occupancy maps
- Quality-of-fit cross correlation with cryo-EM density maps
- Normal modes, principal component analysis, essential dynamics
- Cluster simulation trajectory timesteps by structural similarity

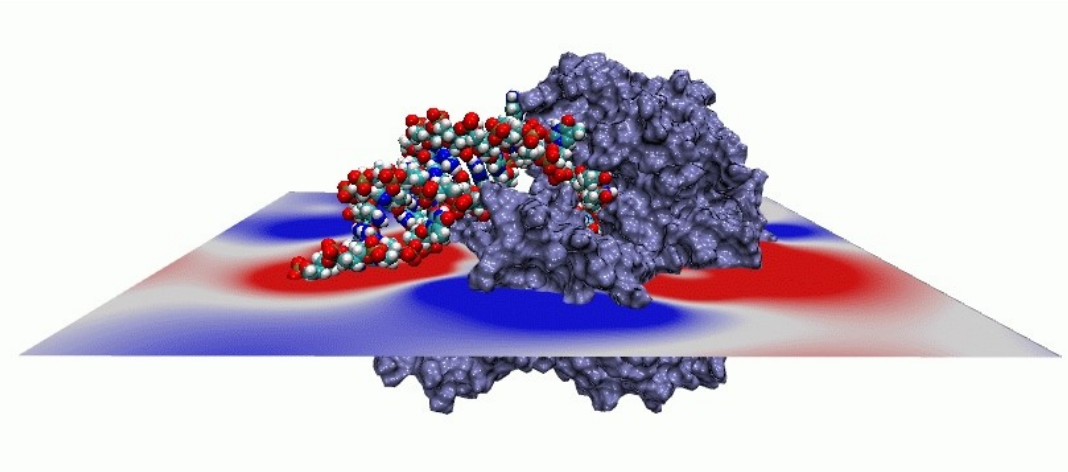


Chemoreceptor trimer-of-dimers analysis with Bendix plugin in VMD

Display of Computed Properties on Structures

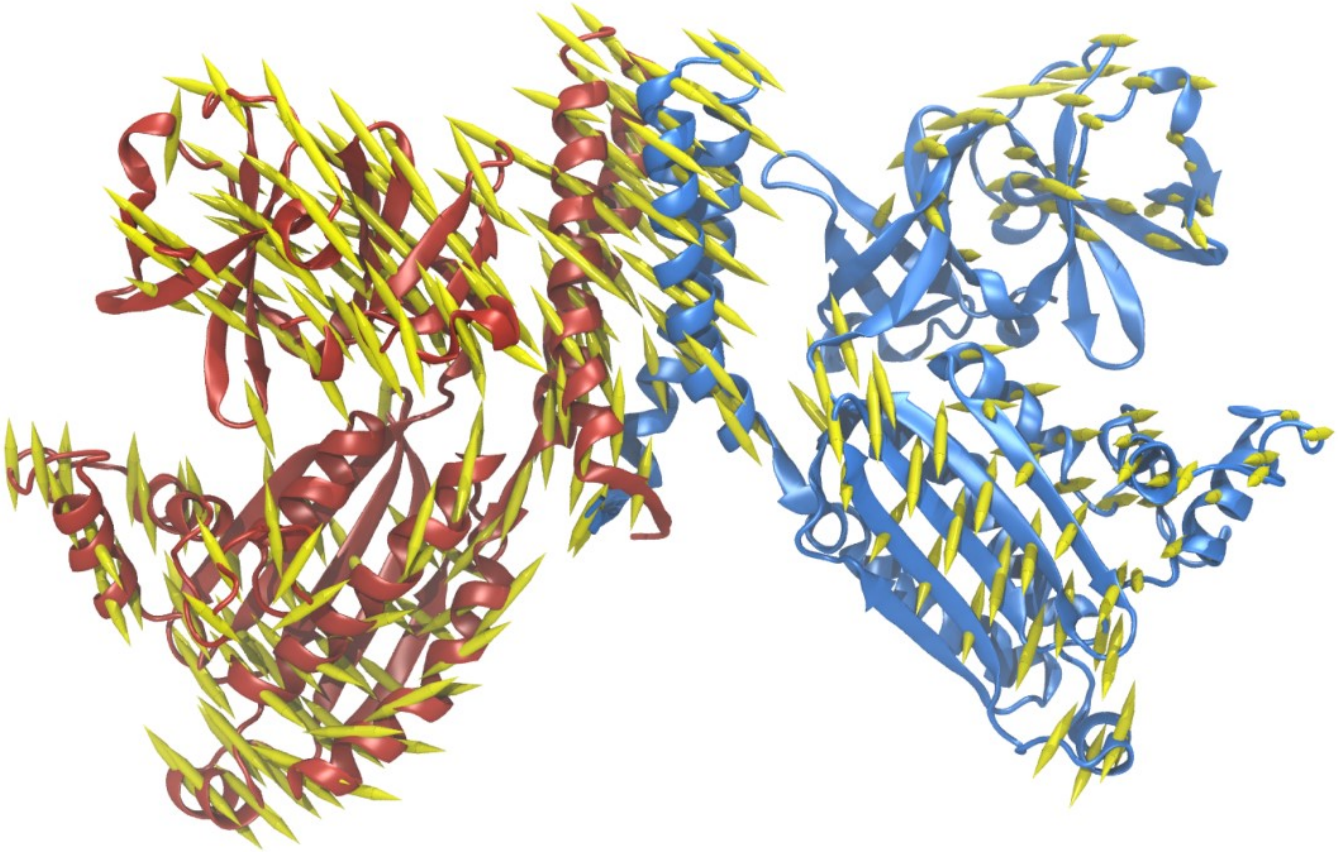


Per-residue solvent-accessible surface area of Ubiquitin



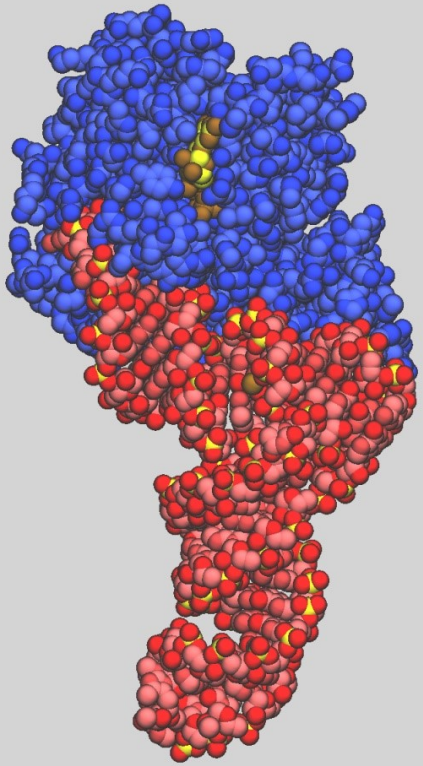
PME electrostatic potential contour for a helicase on a volumetric slice plane

CheA kinase PCA: first principal component porcupine plot

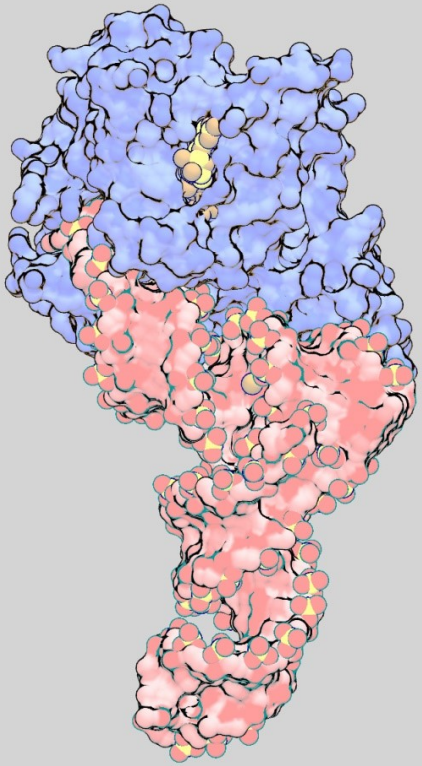


VMD Shading Comparison: EF-Tu

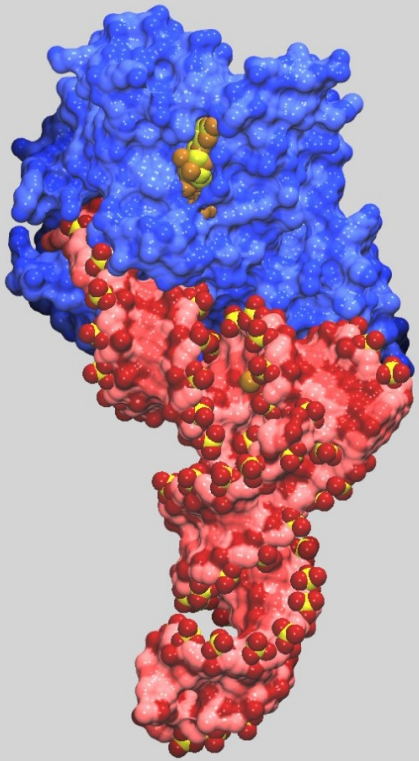
Outline Shader



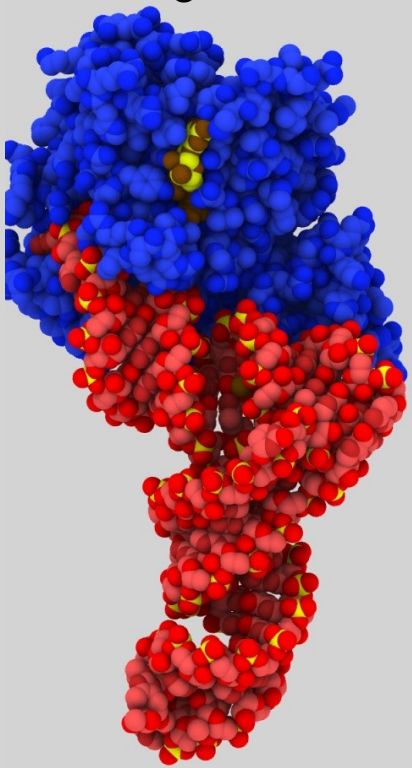
“Goodsell” Shader



Glossy Shader



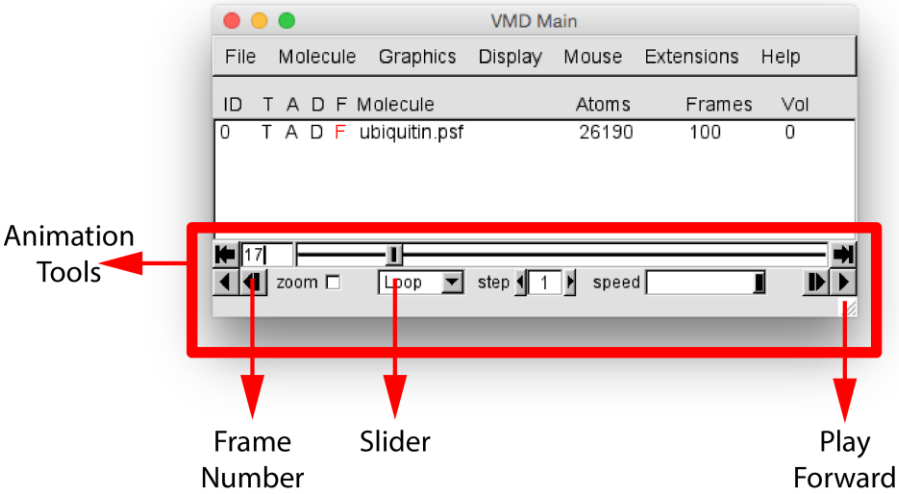
Ray Tracing: Ambient Occlusion, Shadowing



Trajectory Analysis and Visualization

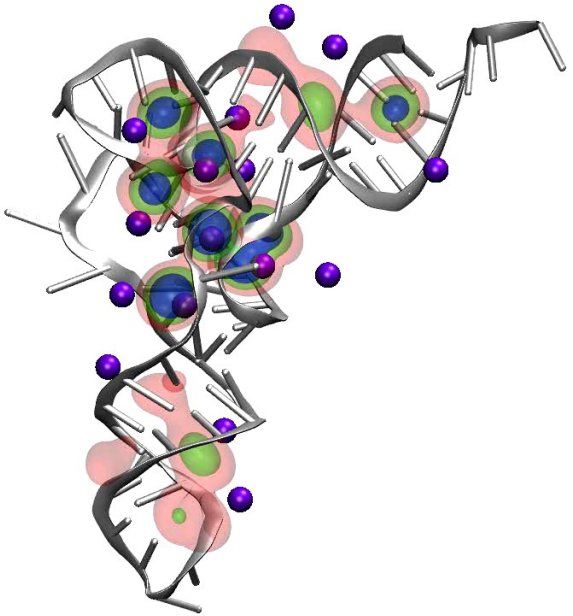
Visualization of MD Trajectories

- Allow researchers study trajectories by analyzing force profiles, energies, structural changes, etc.
- **Visualization selections, graphics, structure properties, colors can all be recomputed for each trajectory timestep!**



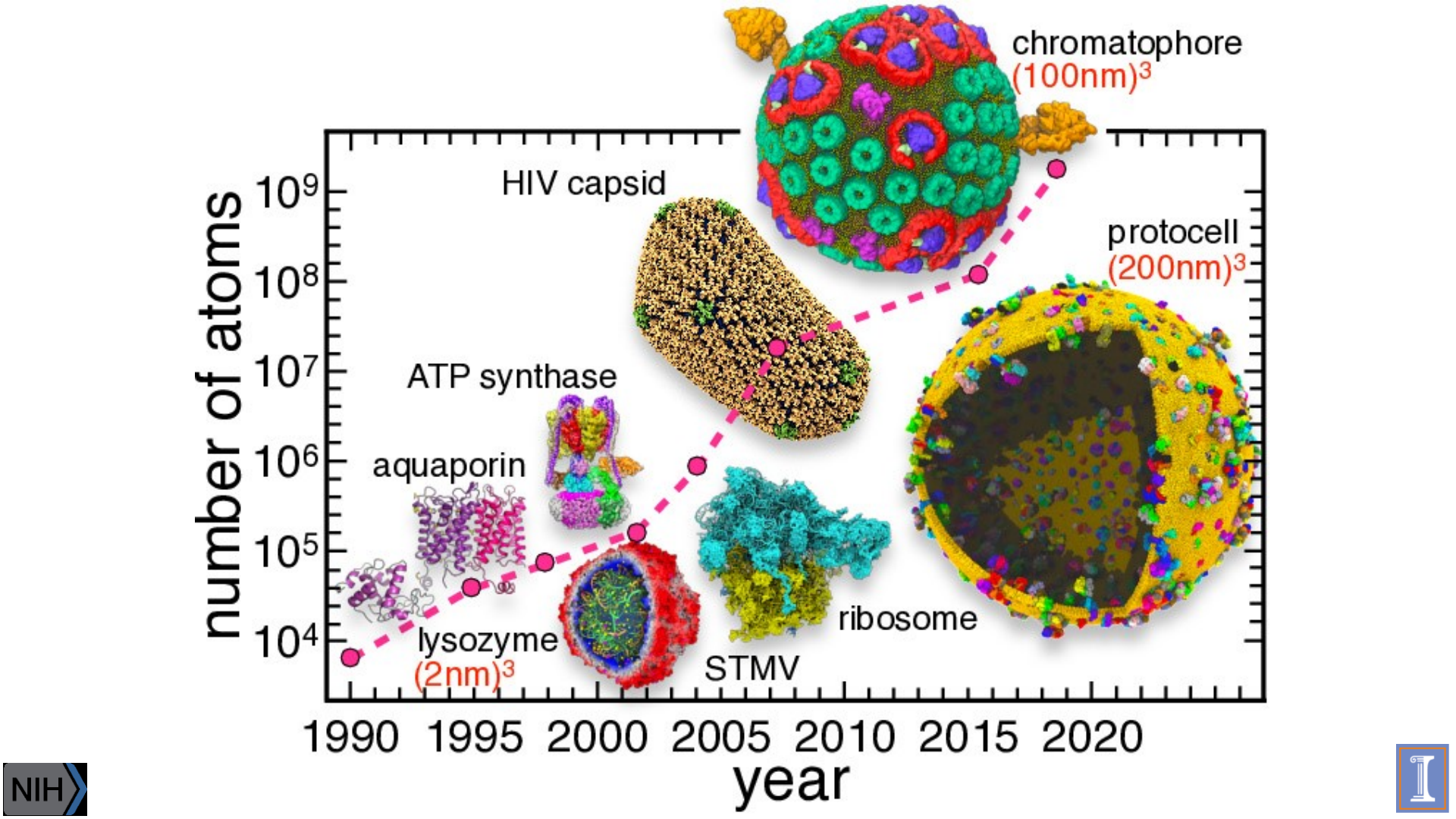
Time-Averaged Volumetric Properties

- Compute density, distance, occupancy, potential maps for a frame or averaged over a trajectory
- Example: display binding sites for diffusively bound ions as probability density isosurfaces



tRNA magnesium ion occupancy: VMD volmap plugin

Large System Analysis and Visualization


chromatophore
(100nm)3
HIV capsid
protocell
(200nm)3
ATP synthase
aquaporin
ribosome
lysozyme
(2nm)3
STMV
number of atoms
10^9
10^8
10^7
10^6
10^5
10^4
1990
1995
2000
2005
2010
2015
2020
year

VMD Petascale Visualization and Analysis

- Combination of growing system sizes and timescales of simulation trajectories poses a major data size challenge for molecular visualization and analysis
- Parallel I/O rates up to **275 GB/sec** on 8192 Cray XE6 nodes – can read in **231 TB in 15 minutes!**
- Analyze/visualize large trajectories **too large to transfer off-site:**
  - User-defined parallel analysis operations, data types
  - Parallel rendering, movie making
- Supports GPU-accelerated compute nodes for both visualization and analysis tasks:
  - GPU accelerated trajectory analysis w/ CUDA
  - OpenGL and GPU ray tracing for visualization and movie rendering

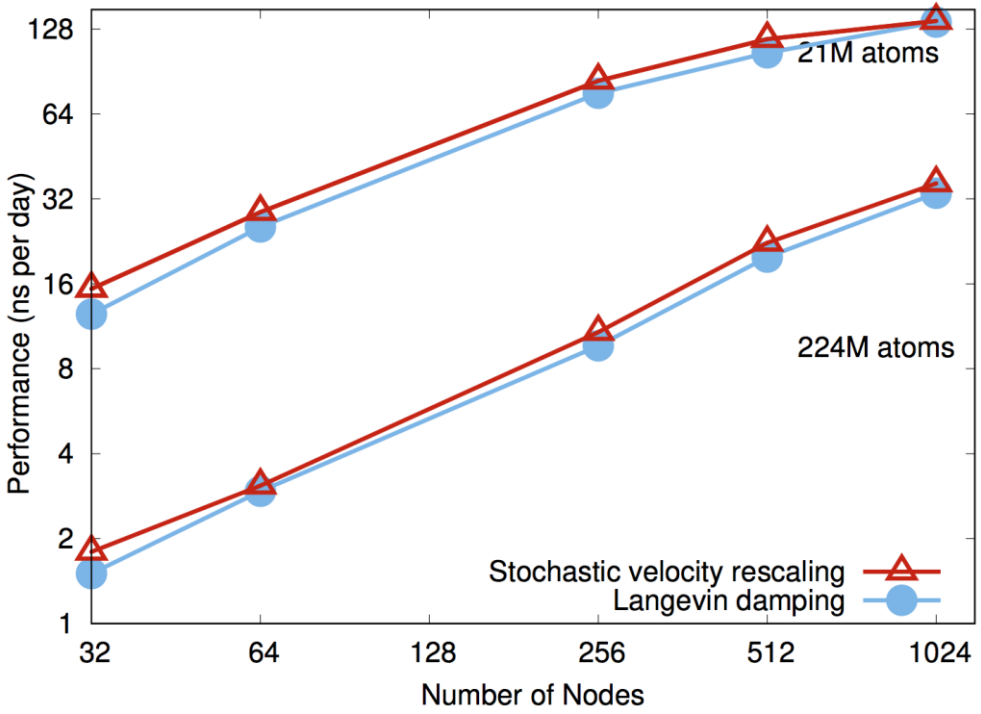
NCSA Blue Waters Hybrid Cray XE6 / XK7
22,640 XE6 dual-Opteron CPU nodes
4,224 XK7 nodes w/ Telsa K20X GPUs

Parallel VMD currently available on: ORNL Summit and Titan, NCSA Blue Waters, IU Big Red II, CSCS Piz Daint, many similar systems

Petascale Molecular Dynamics I/O and Storage Challenges

- NAMD simulations can produce up to **10TB/day @ 1024 nodes (~20%) of ORNL Summit**, more as optimizations raise NAMD performance further
- Petascale science campaigns require **months of simulation runs**
- **Long-term storage of large-fractional petabytes impractical**
- **Historical "download output files for analysis and visualization" approach is a non-starter at this scale**
- Demands visualization and analysis operate on the data **in-place** on the HPC system, **whether *post-hoc, in-transit, or in-situ***
- Analyses must **identify salient features of structure, dynamics**, cull data that don't contribute to biomolecular processes of interest

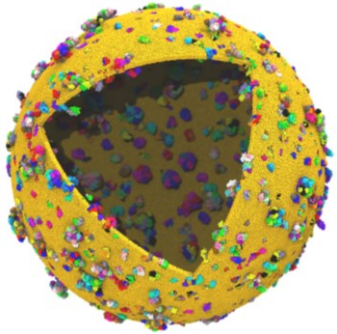
NAMD on Summit, May 2018

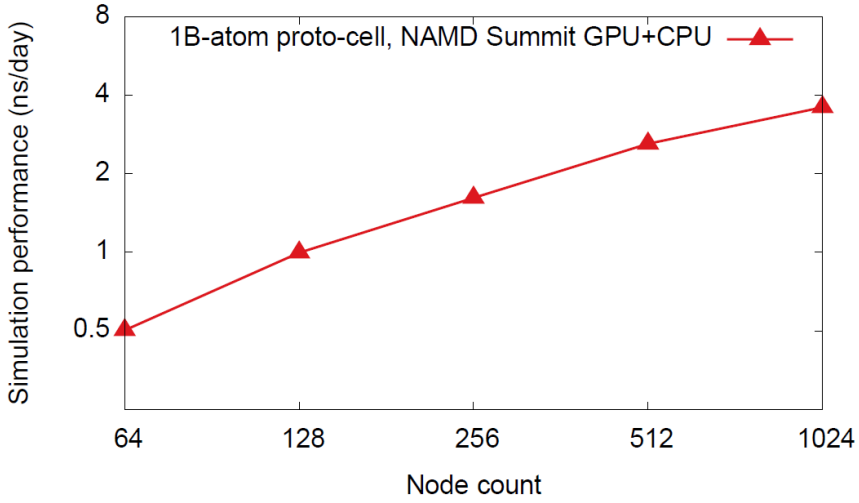


NAMD simulations can generate up to 10TB of output per day on 20% of Summit

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**





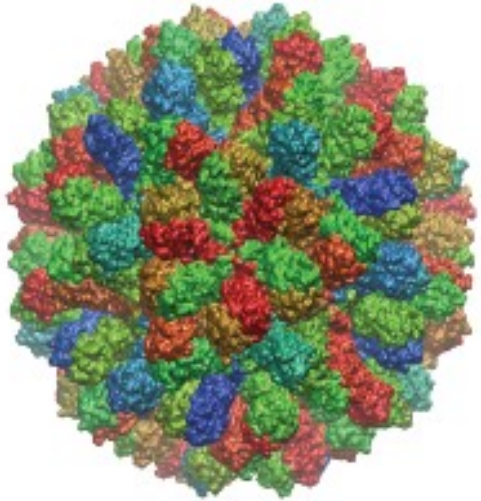
Running VMD on Distributed Memory Clouds, Clusters, and Supercomputers with MPI

Using VMD MPI Builds

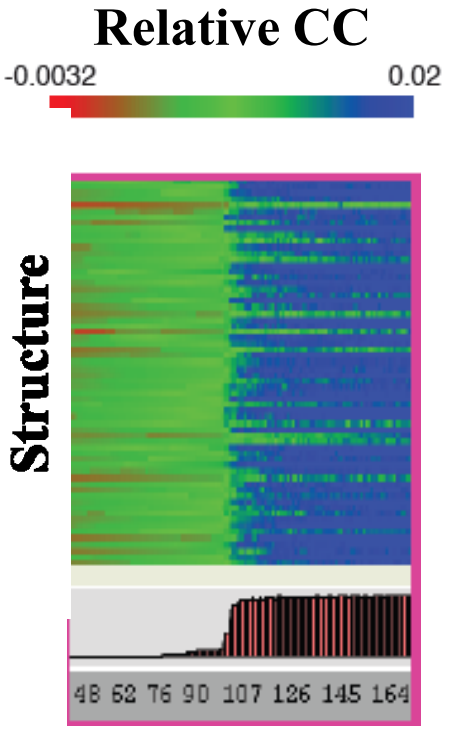
- **See “Running VMD on Supercomputers” section in the VMD tutorial**
- Run one MPI rank (VMD process) per compute node
  - Each rank uses all CPU cores and all GPU accelerators they find
  - GPUs are shared between graphics/visualization and computing tasks
- Parallel VMD Scripting APIs:
  - “parallel” script subcommands, e.g., “parallel for ...”:
    - “nodename”, “noderank”, “nodecount”, “barrier”, “allgather”, “allreduce”, “for”
  - Available even in non-MPI builds, for single-node script compatibility
- Work scheduling:
  - Intra-node work scheduled via multithreading, CUDA, etc.
  - Inter-node distributed memory work scheduling of user analysis/viz scripts is provided by built-in load balancing implementation for VMD

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



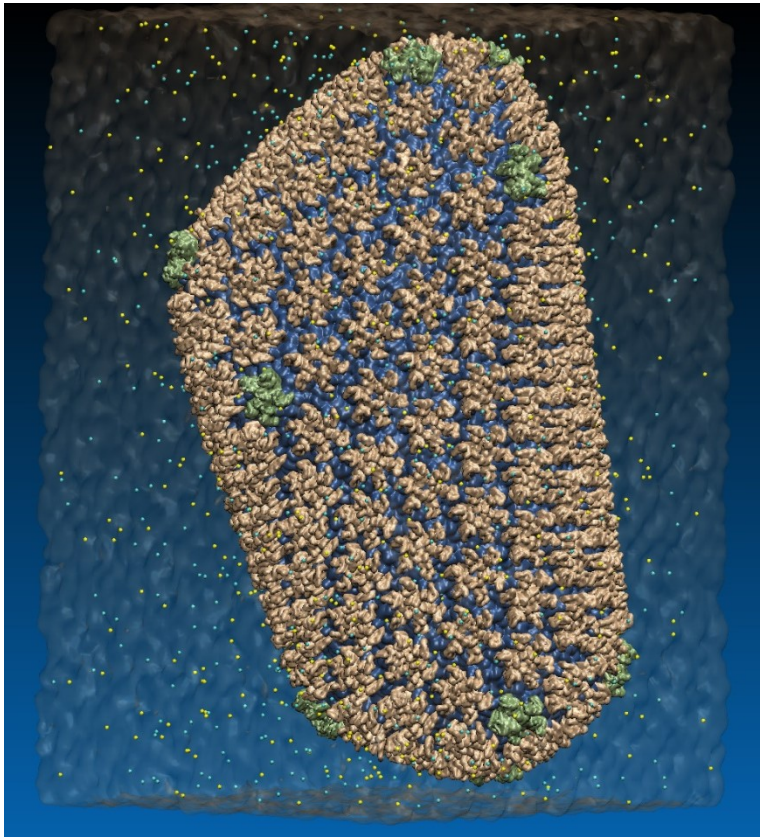
Calculation of 7M CCs would take **5 years** using serial CPU algorithm!

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

VMD 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 FPS



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. *2016 IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW)*, pp. 1014-1023, 2016.

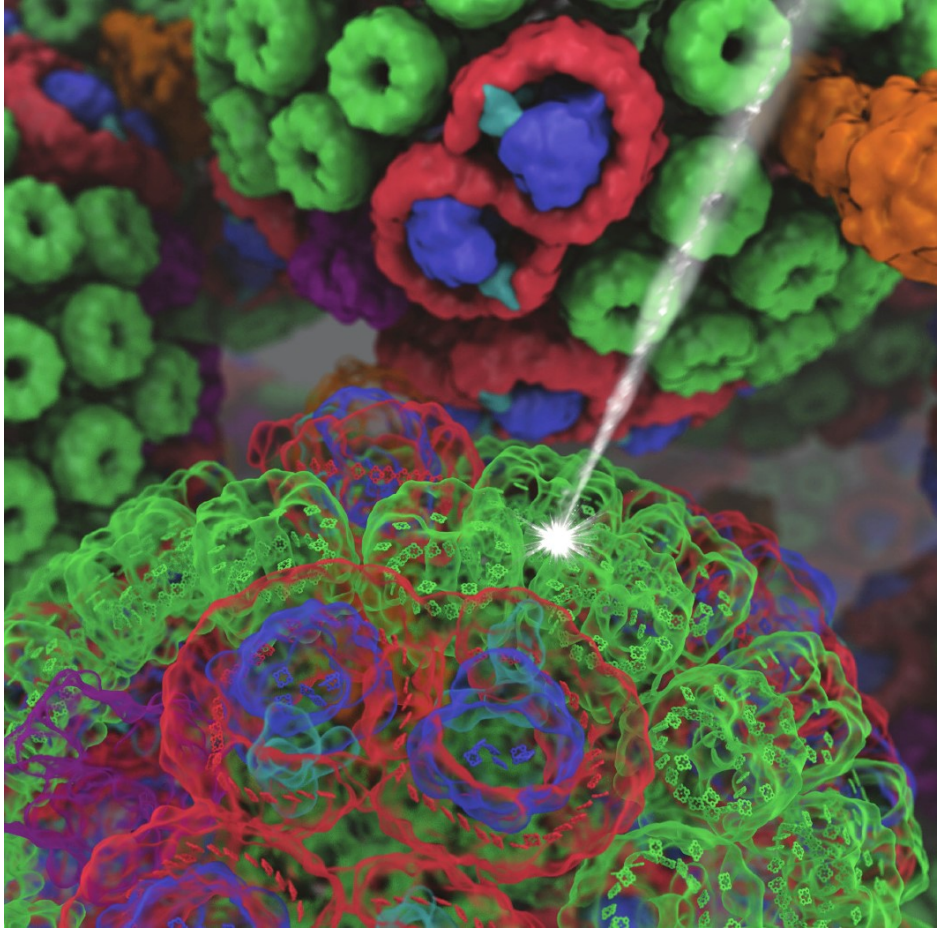
Trajectory I/O on Parallel Filesystems

- Don’t use old/inefficient trajectory file formats if they can be avoided
- DO use file formats like DCD, NetCDF, JS, that permit efficient strided reads
- DO enable file striping for large trajectory files (more I/O parallelism)
- DO write simulation outputs to multiple files (more I/O parallelism)
- DO use VM page-aligned file formats that permit kernel-bypass direct-I/Os, and support for technologies like GPU-Direct Storage
- **VMD has read trajectories at up to 71GB/sec from JS files on DGX-2 dense multi-GPU nodes with GDS, w/ Weka I/O**

High Fidelity Ray Tracing for Interactive and Cinematic Visualization

High Fidelity Ray Tracing

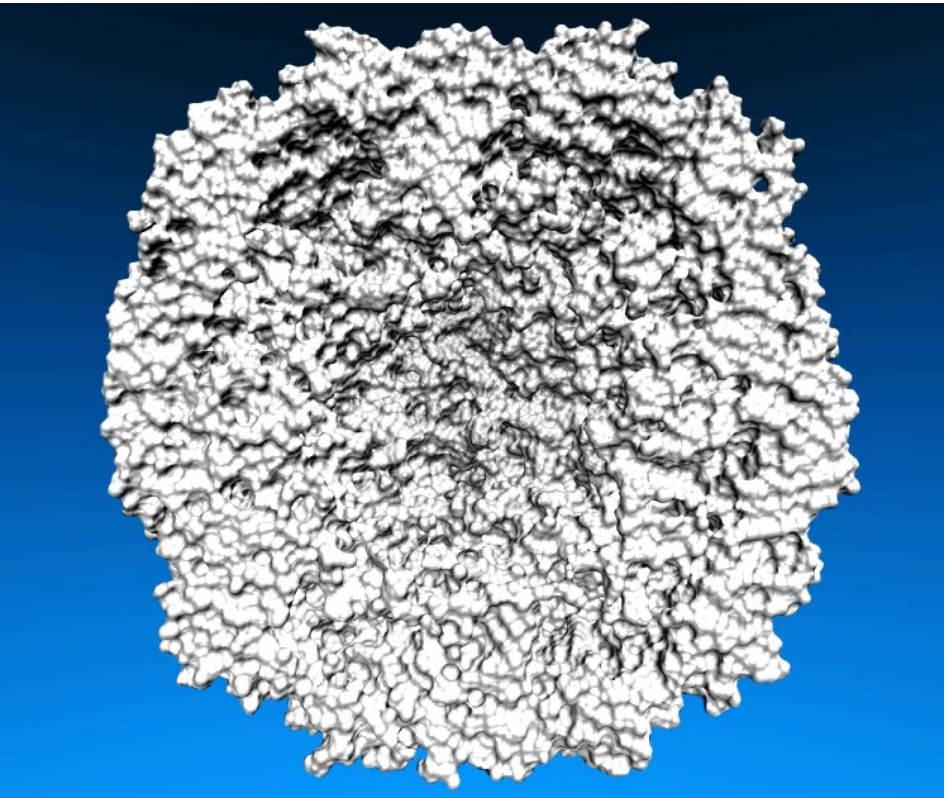
- Advanced rendering techniques save scientists time, produce images that are easier to interpret
- Ambient Occlusion, Depth of Field, high quality transparency, instancing, ….
- **Interactive RT** on laptops, desk, cloud, and **remote supercomputers**
- **Large-scale MPI parallel rendering:** in situ or post hoc visualization tasks
- Stereoscopic panorama and full-dome projections
- **Omnidirectional VR: YouTube, HMDs**
- Built-in ray tracing engines:
  - **Tachyon:** cross-platform RT
  - **NVIDIA OptiX:** GPU RTX-accelerated
  - **Intel OSPRay:** CPU x86-optimized



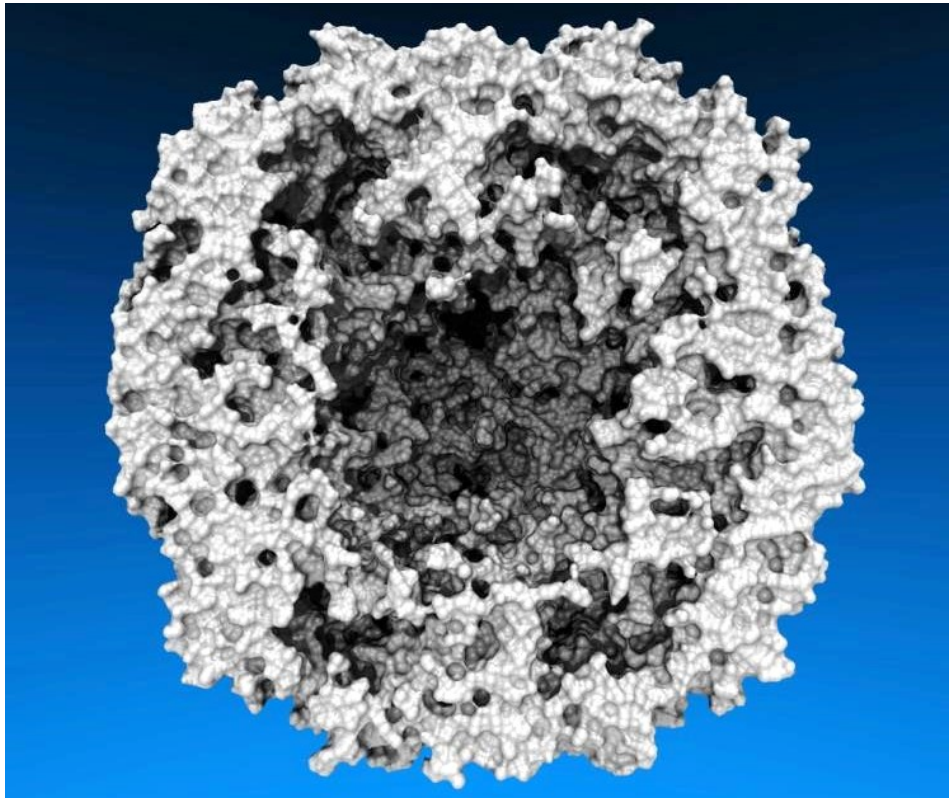
VMD/OptiX all-atom Chromatophore

Lighting Comparison, STMV Capsid

Two lights, no shadows



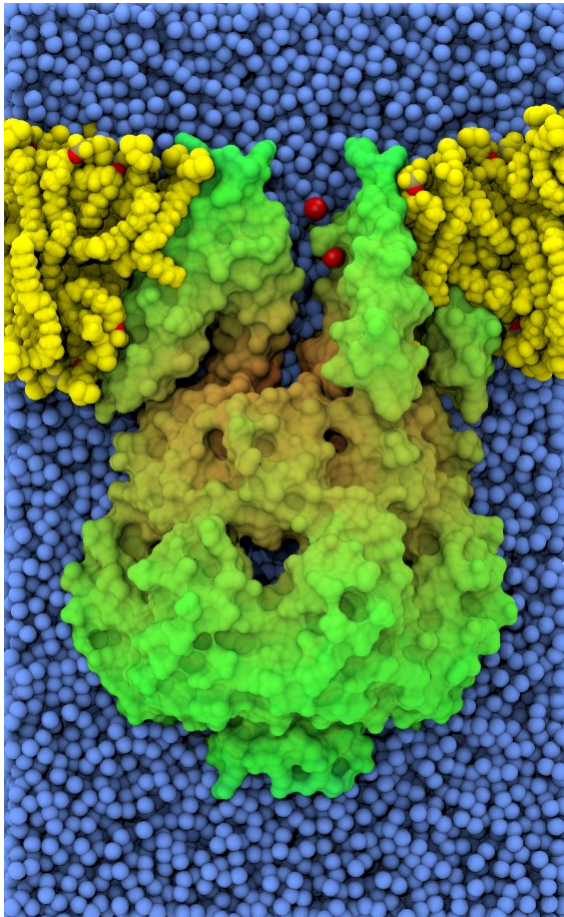
Ambient occlusion + two lights, 144 AO rays/hit



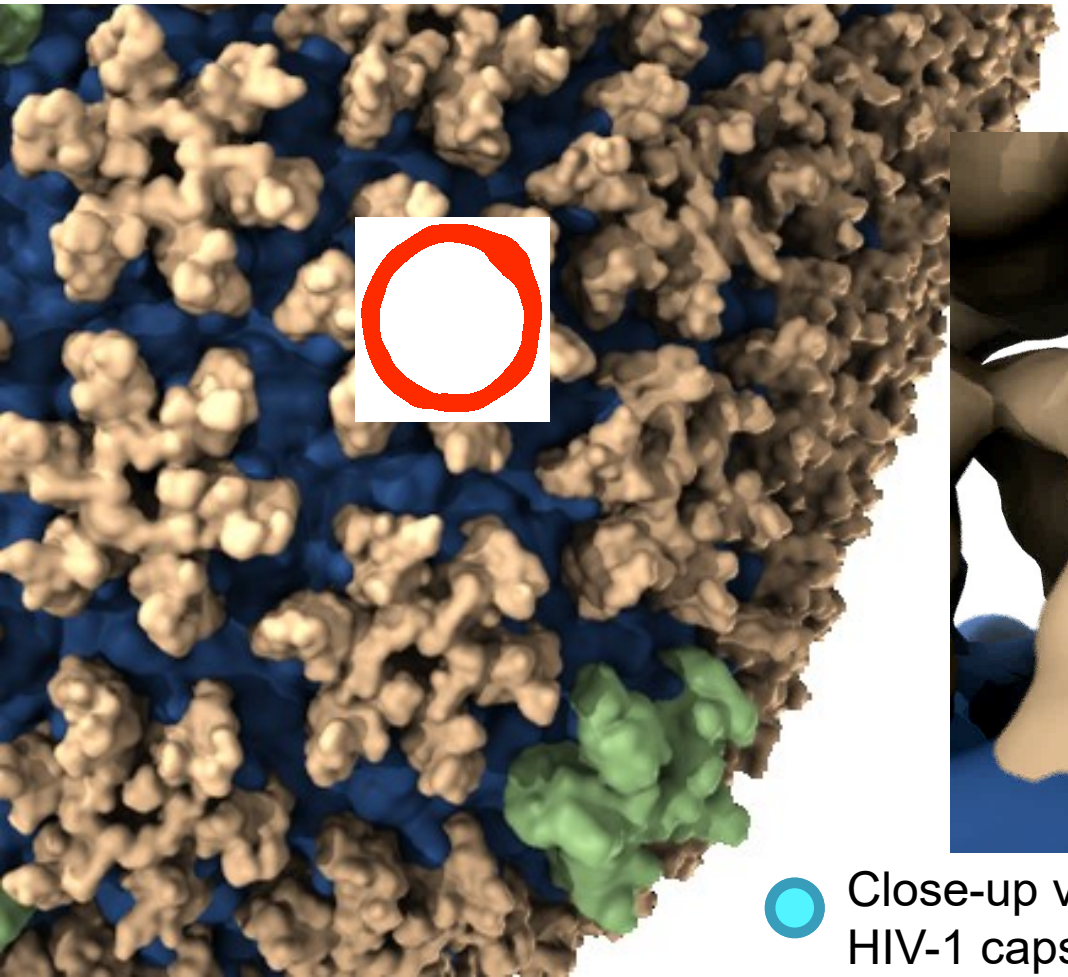
Geometrically Complex Scenes

Ray tracing techniques well matched to molecular viz. needs:

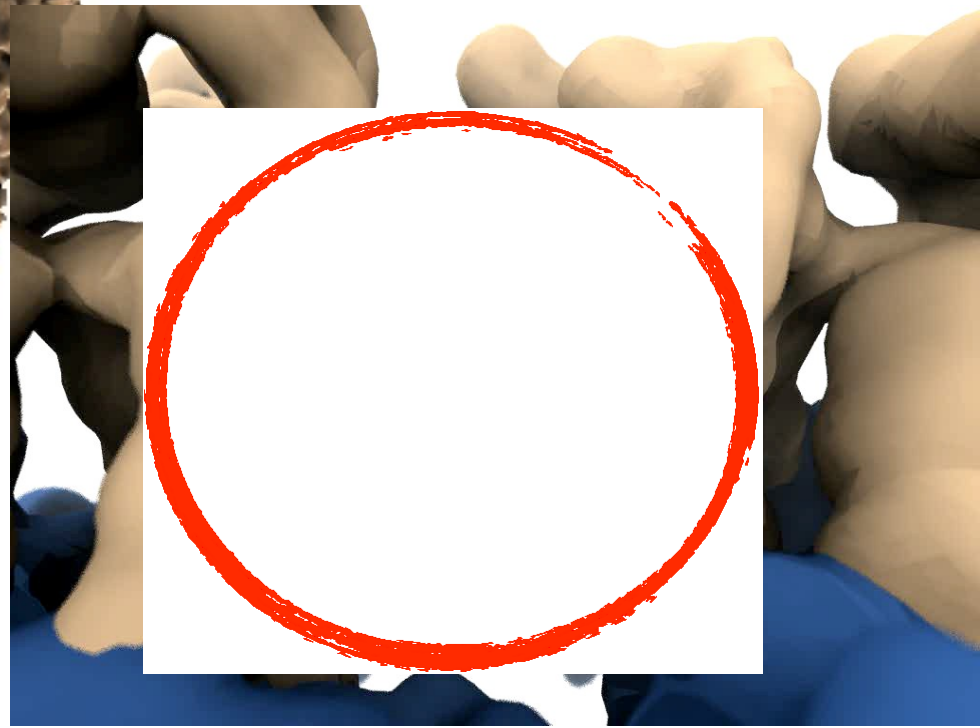
- Curved geometry, e.g. spheres, cylinders, toroidal patches, easily supported
- Greatly reduced memory footprint vs. polygonalization
- Runtime scales only moderately with increasing geometric complexity
- Occlusion culling is “free”, RT acceleration algorithms do this and much more



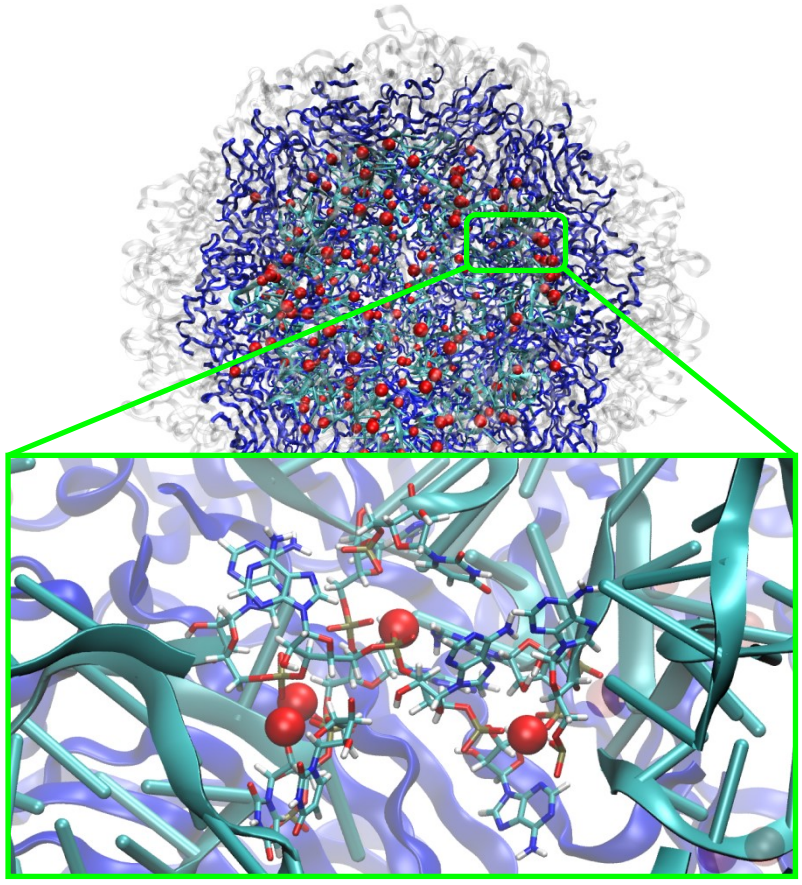
Goal: Intuitive interactive viz. in crowded molecular complexes

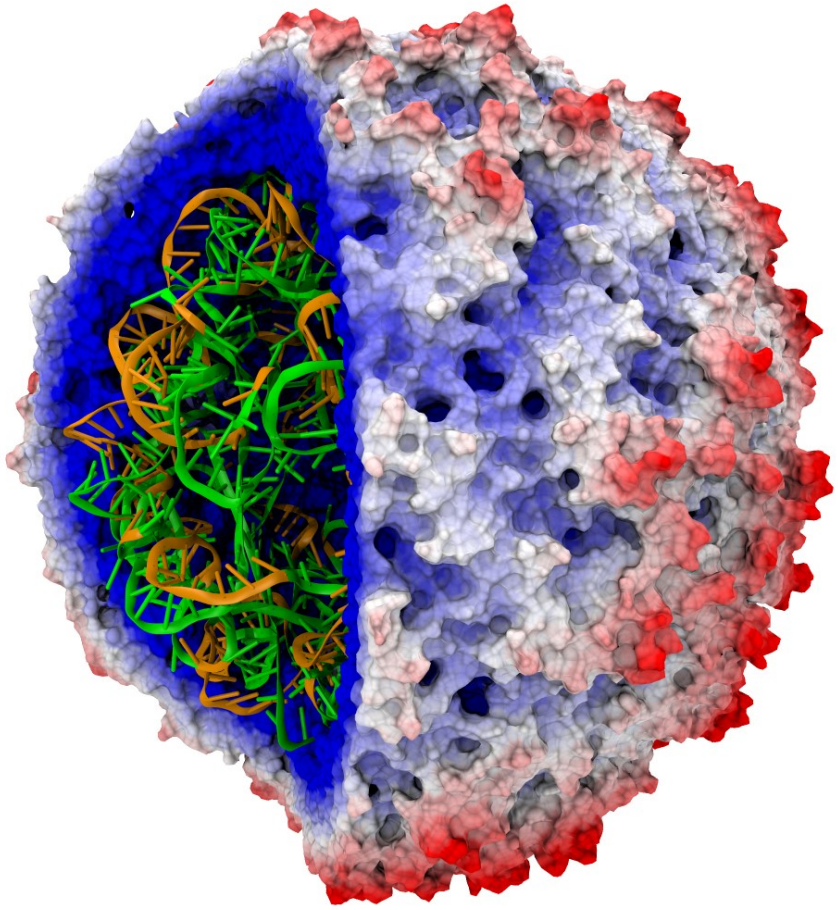


Results from 64 M atom, 1 μs sim!



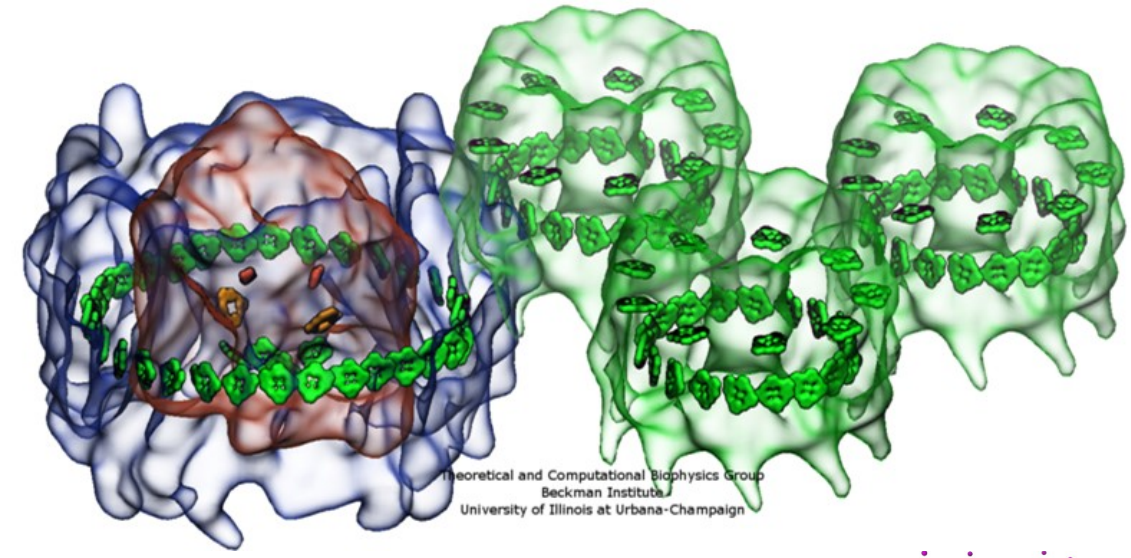
- Close-up view of chloride ions permeating through HIV-1 capsid hexameric centers

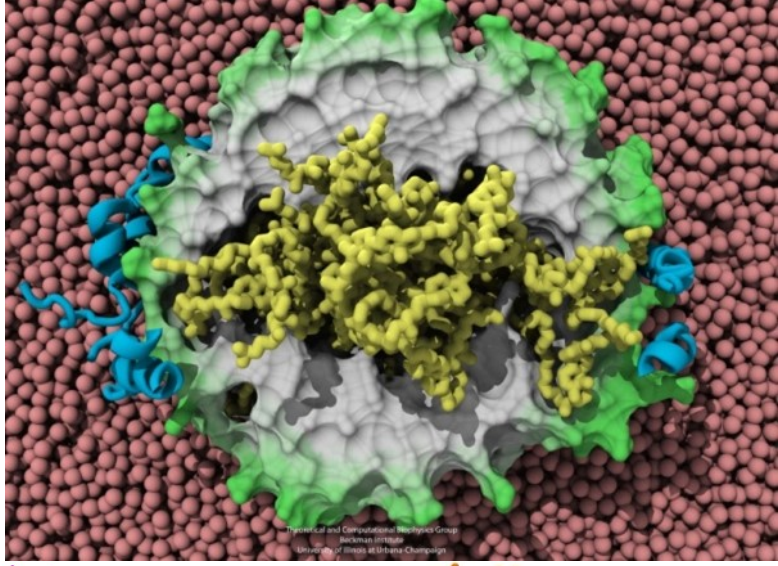


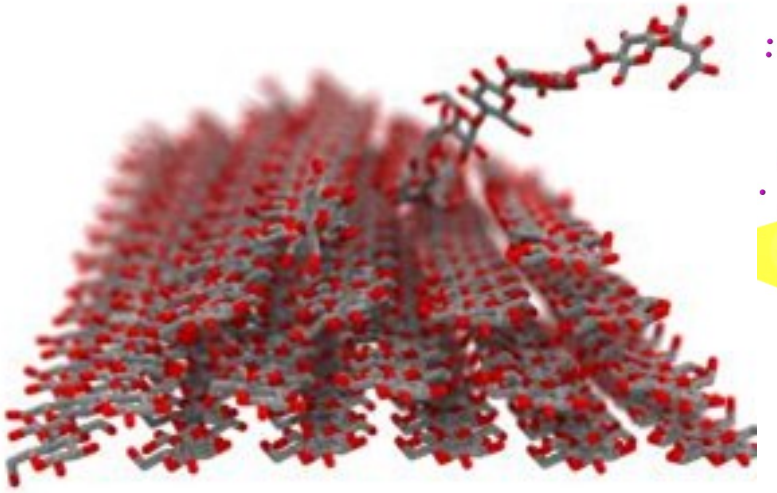


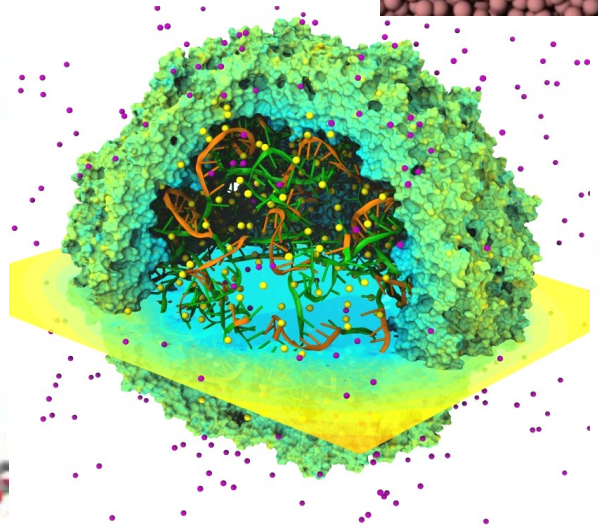
Satellite Tobacco Mosaic Virus

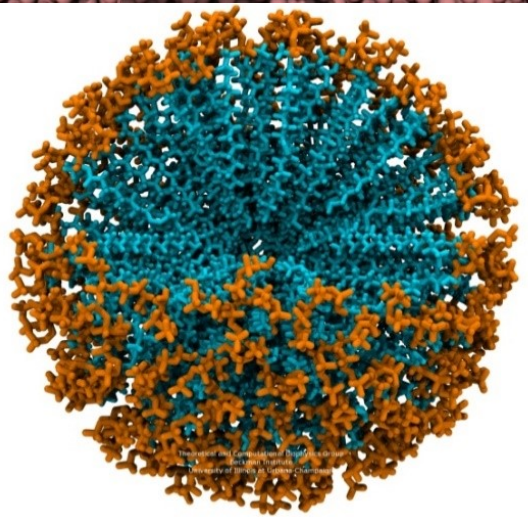
Biomedical Technology Research Center for Macromolecular Modeling and Bioinformatics
Beckman Institute, University of Illinois at Urbana-Champaign - www.ks.uiuc.edu


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


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



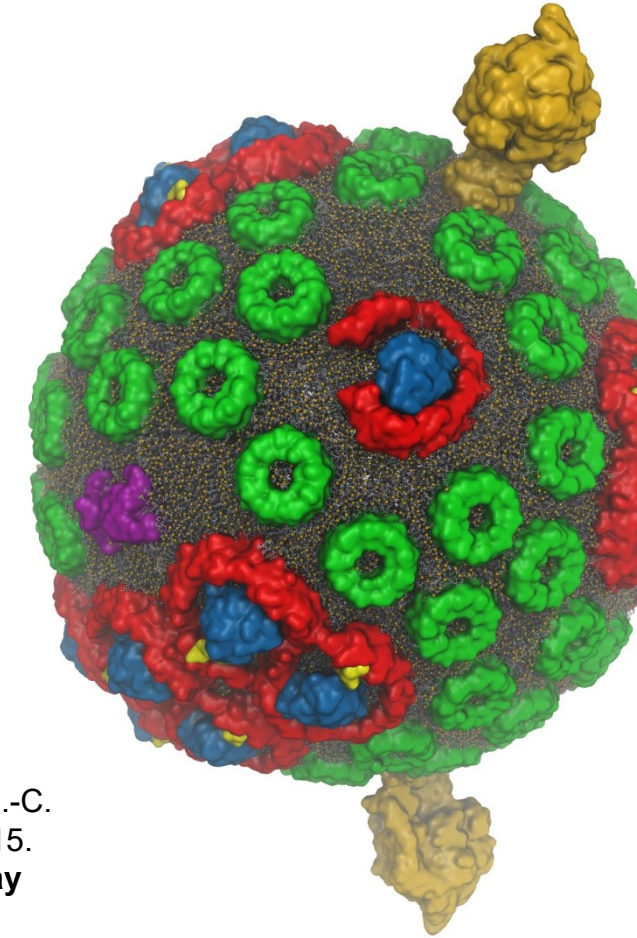




VMD w/ OptiX RTX Ray Tracing

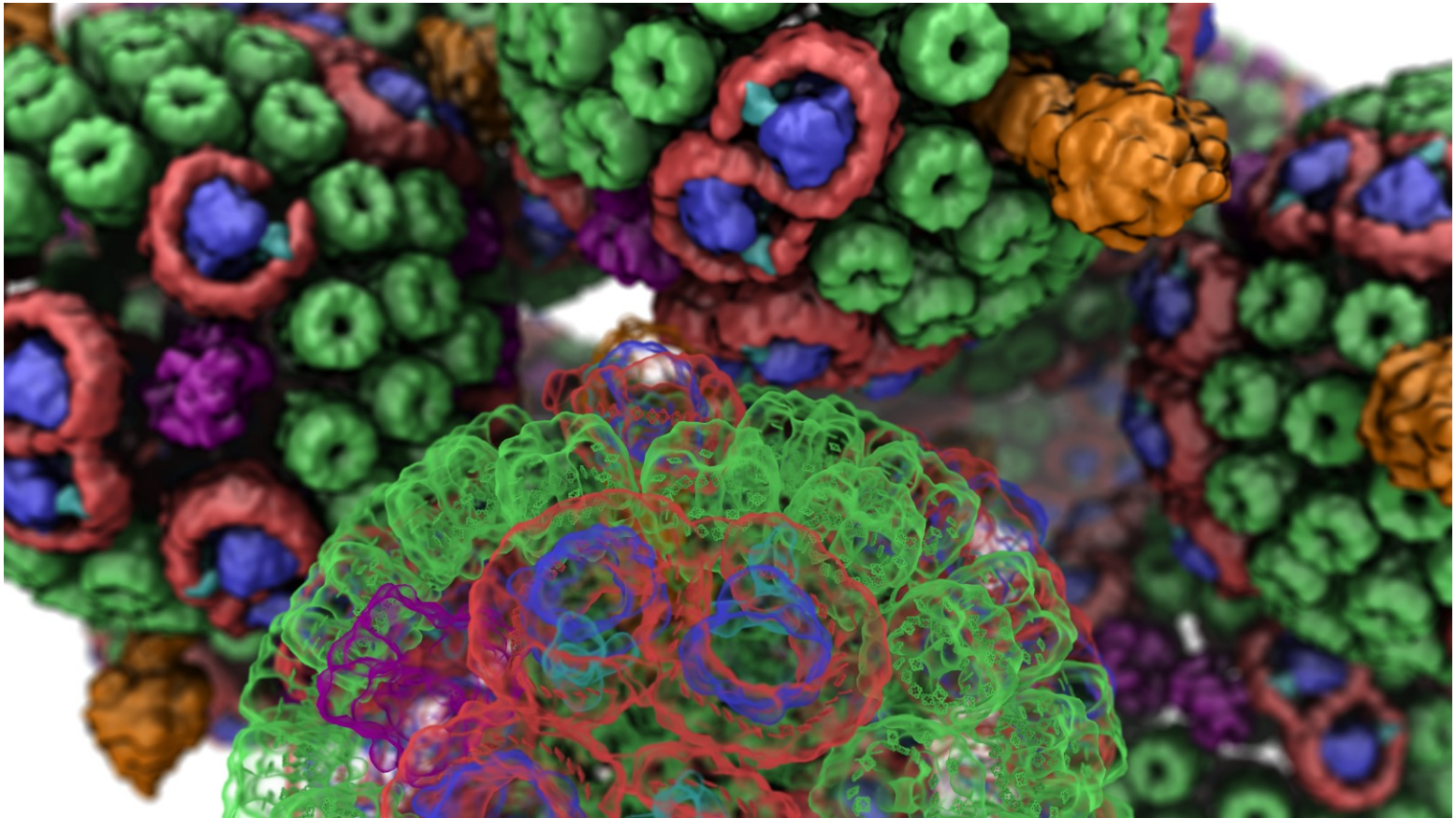
- Interactive RT on laptops, desktops, and cloud
- Large-scale parallel rendering: in situ or post hoc visualization
- **Remote ray tracing with NvPipe video streaming**
- Stereoscopic panoramic and full-dome projections
- **Omnidirectional VR for YouTube, VR HMDs**
- VMD+OptiX NGC container: https://ngc.nvidia.com/registry/
- GPU memory sharing via NVLink
- **In-progress:**
  Denoising: faster turnaround w/ AO, DoF, etc

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.

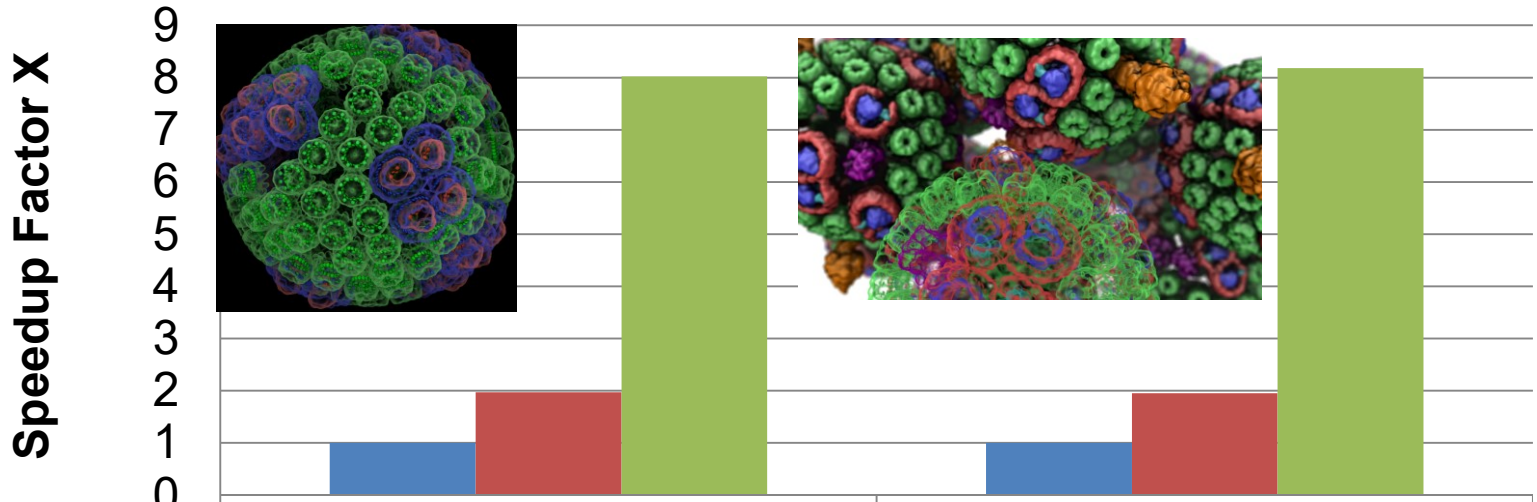


VMD/OptiX GPU Ray Tracing of all-atom Chromatophore w/ lipids.

VMD/OptiX RTX Acceleration



VMD OptiX RT performance on Quadro RTX 6000

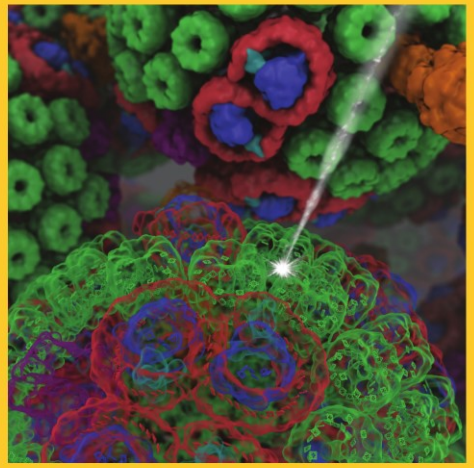


	Chromatophore @ 4Kx4K	Chrom Cell, 512x DoF @ 1080p
Quadro GV100	1	1
2x Quadro GV100	1.97	1.95
Quadro RTX 6000	8.02	8.18

JPCBFK

APRIL 20, 2017
VOLUME 121
NUMBER 15
pubs.acs.org/JPCB

THE JOURNAL OF PHYSICAL CHEMISTRY B



KLAUS SCHULTEN MEMORIAL ISSUE

ACS Publications
Most Trusted. Most Cited. Most Read.

www.acs.org

Faraday Discussions

Volume: 169

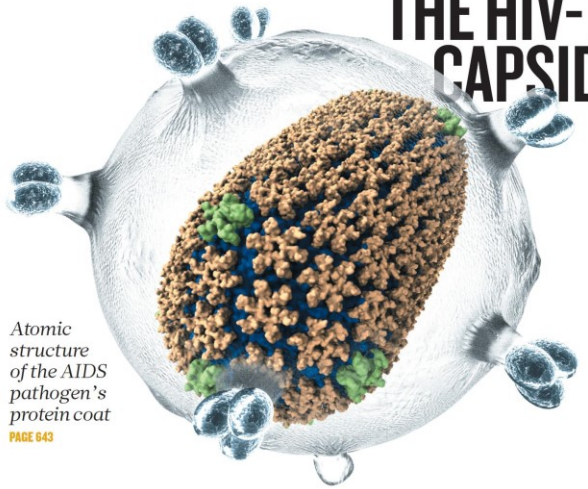
Molecular Simulations and Visualization

ROYAL SOCIETY OF CHEMISTRY

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

THE HIV-1 CAPSID



Atomic structure of the AIDS pathogen's protein coat

PAGE 643

COSMOLOGY
THE FIRST LIGHT
In pursuit of the most distant galaxies
PAGE 554

CITATION
CROSSING THE BORDERS
International collaborations make the most impact
PAGE 557

ANTICANCER DRUGS
A SITTING TARGET
An indirect hit on 'undruggable' KRAS protein
PAGES 577 & 638

NATURE.COM/NATURE
30 May 2013

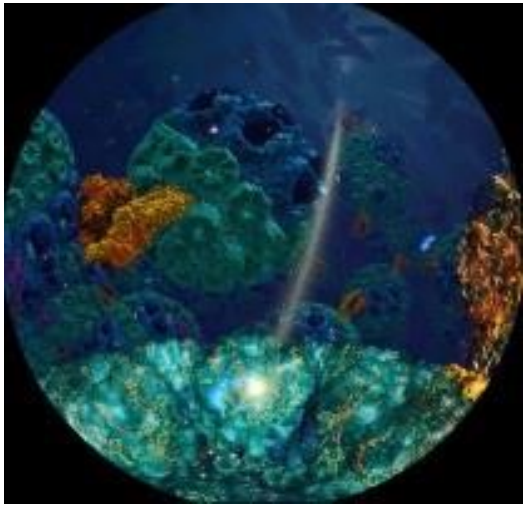
VMD Application Examples and Implementation Details

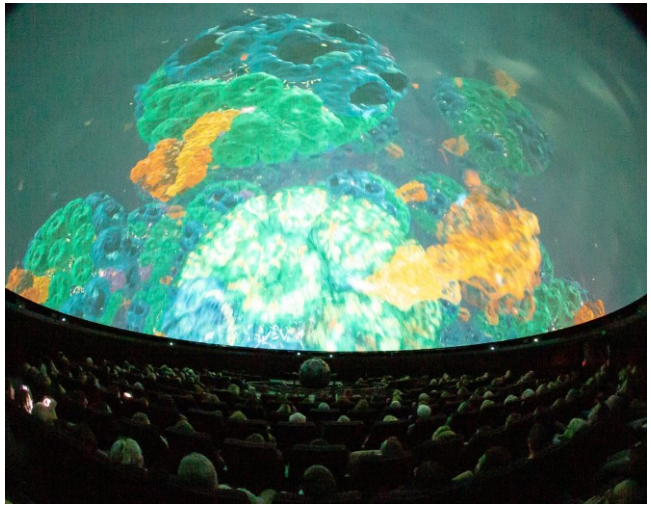
VMD Cinematic Molecular Visualization and Rendering: “Birth of Planet Earth” Fulldome Show

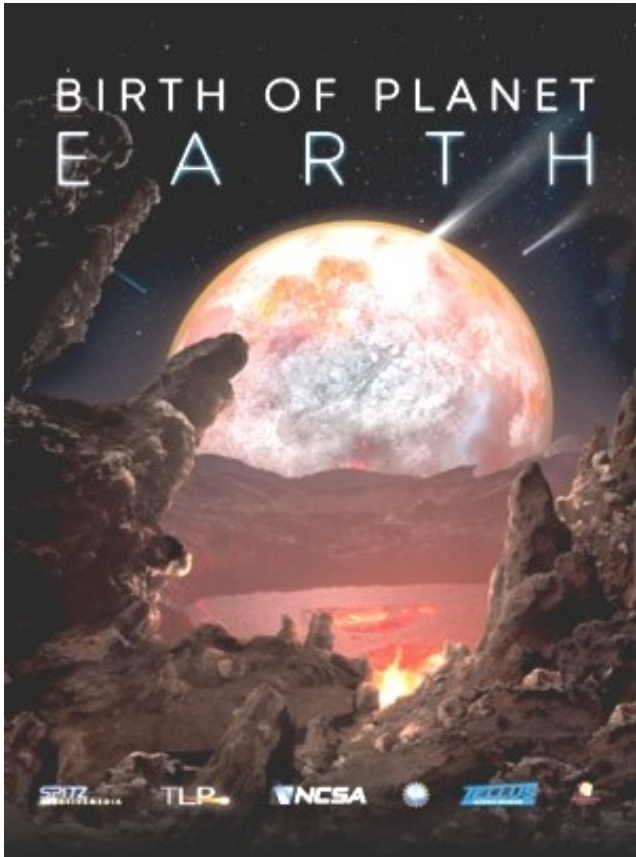
https://www.youtube.com/watch?v=NTgAok6n7I4

VMD RTX Performance Gains for “BoPE” Content:

- BoPE production used Quadro M6000 (Maxwell) GPUs
- BoPE w/ OptiX 6.5 on Quadro RTX 6000 **up to 15x faster!**



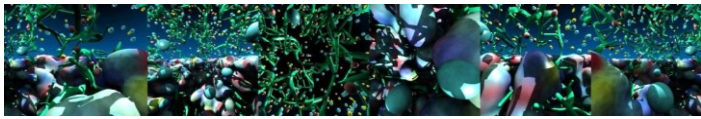


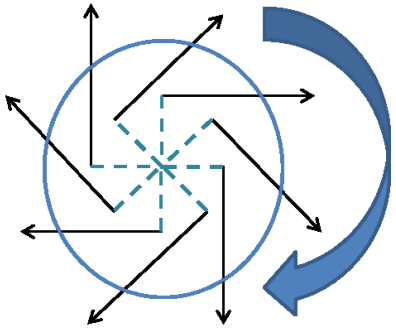


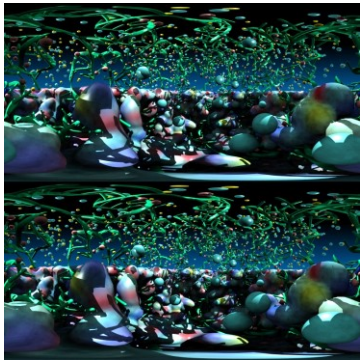
Multiscale modeling and cinematic visualization of photosynthetic energy conversion processes from electronic to cell scales.
M. Sener, S. Levy, J. E. Stone, et al., J. Parallel Computing, 2021.

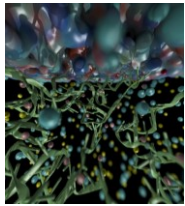
Omnidirectional Stereoscopic Ray Tracing

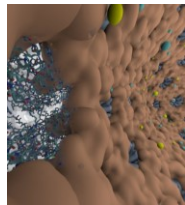
- **Ray trace 360° images and movies for Desk and VR HMDs: Oculus, Vive, Cardboard**
- Stereo spheremaps or cubemaps allow very high-frame-rate interactive OpenGL display
- **AO lighting, depth of field**, shadows, transparency, curved geometry, …
- **Summit 6x Tesla V100 GPU nodes:**
  - **Render many omni-stereo viewpoints, no acceleration structure rebuilds, tens of frames/sec per-node!**
  - **OptiX multi-GPU rendering, NVLink compositing and data distribution, etc…**
  - **Future: AI for warping between views**







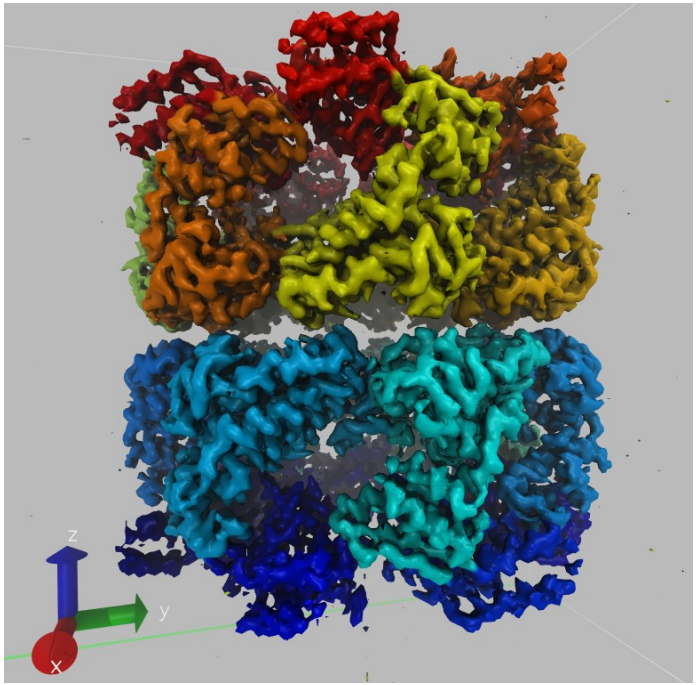




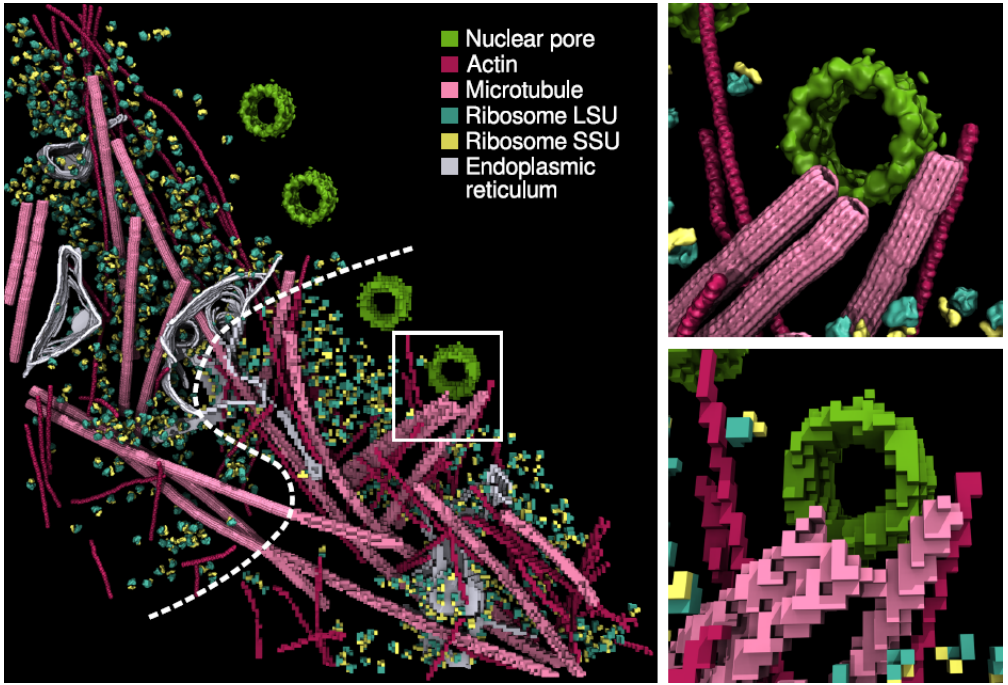
Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing. J. E. Stone, et al. J. Parallel Computing, 55:17-27, 2016.

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.

Density Map Segmentation



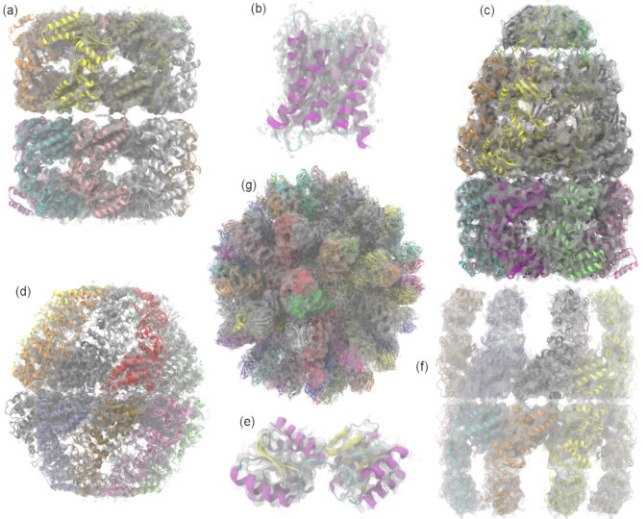
VMD GPU-accelerated density map segmentation of GroEL

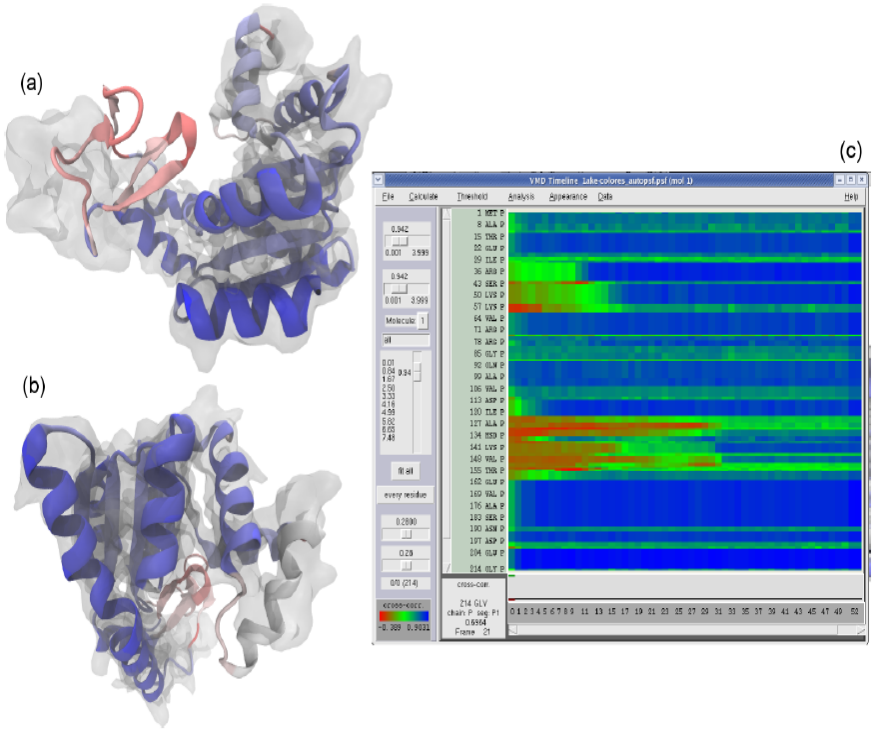


Earnest, et al. J. Physical Chemistry B, 121(15): 3871-3881, 2017.

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**

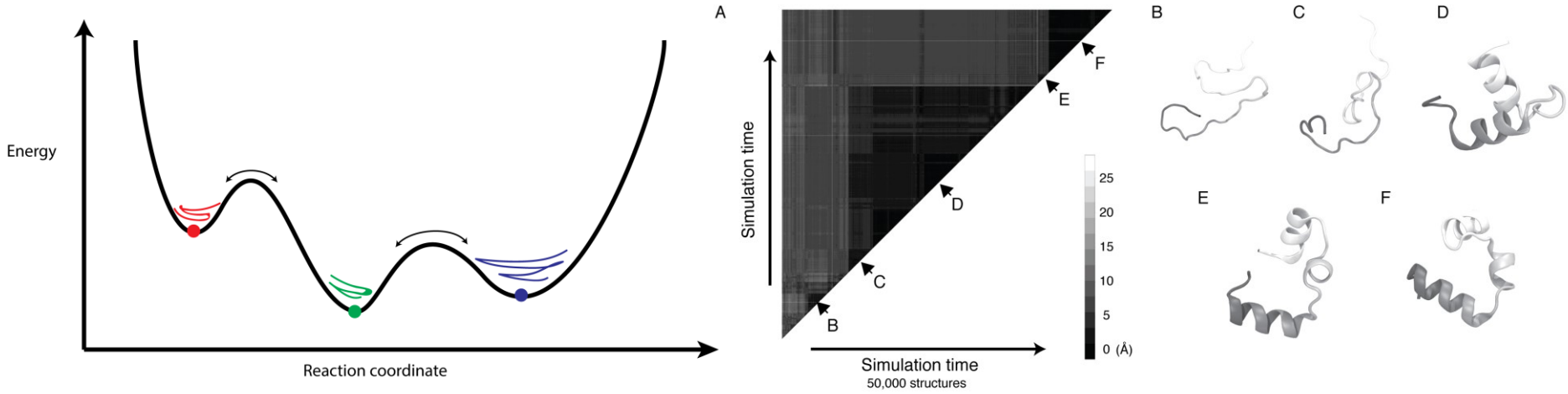
VMD Tesla V100 Cross Correlation Performance

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 IBM Power8 + 1x Tesla K40 [2]	0.488s,	32x	**0.9x**
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 IBM Power8 "Minsky" + 1x Tesla P100	**0.080s,**	198x	**5.7x**
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:
Requires I/O+Memory for All-Pairs of Trajectory Frames

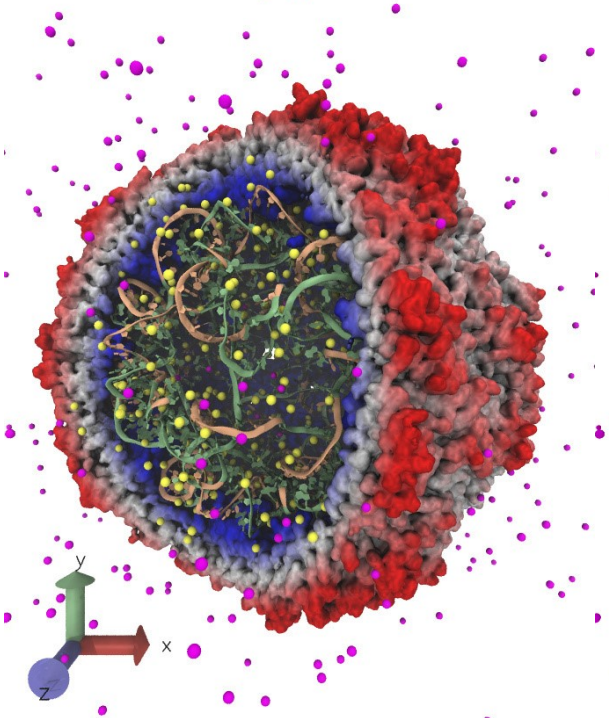


GPU-Accelerated Molecular Dynamics Clustering Analysis with OpenACC. J.E. Stone, J.R. Perilla, C. K. Cassidy, and K. Schulten. In, Robert Farber, ed., Parallel Programming with OpenACC, Morgan Kaufmann, Chapter 11, pp. 215-240, 2016.

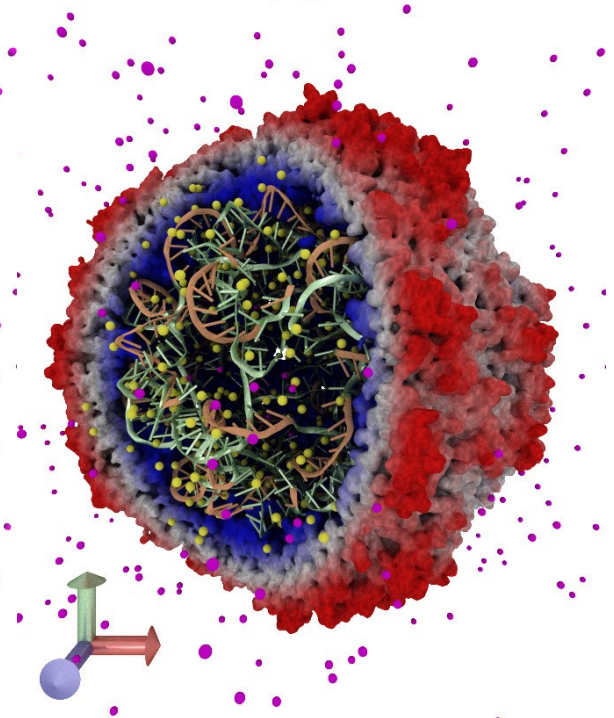
VMD Examples from In-Progress ANARI Renderers

https://www.khronos.org/anari

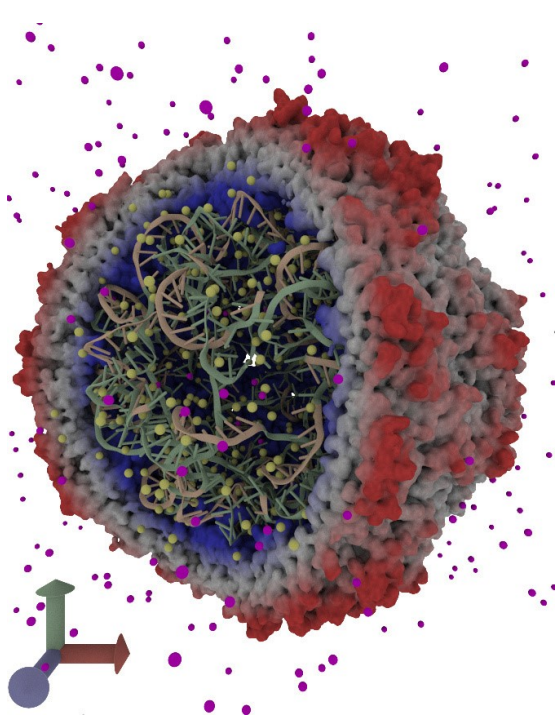
Tachyon Ray Tracer



OSPRay Path Tracer



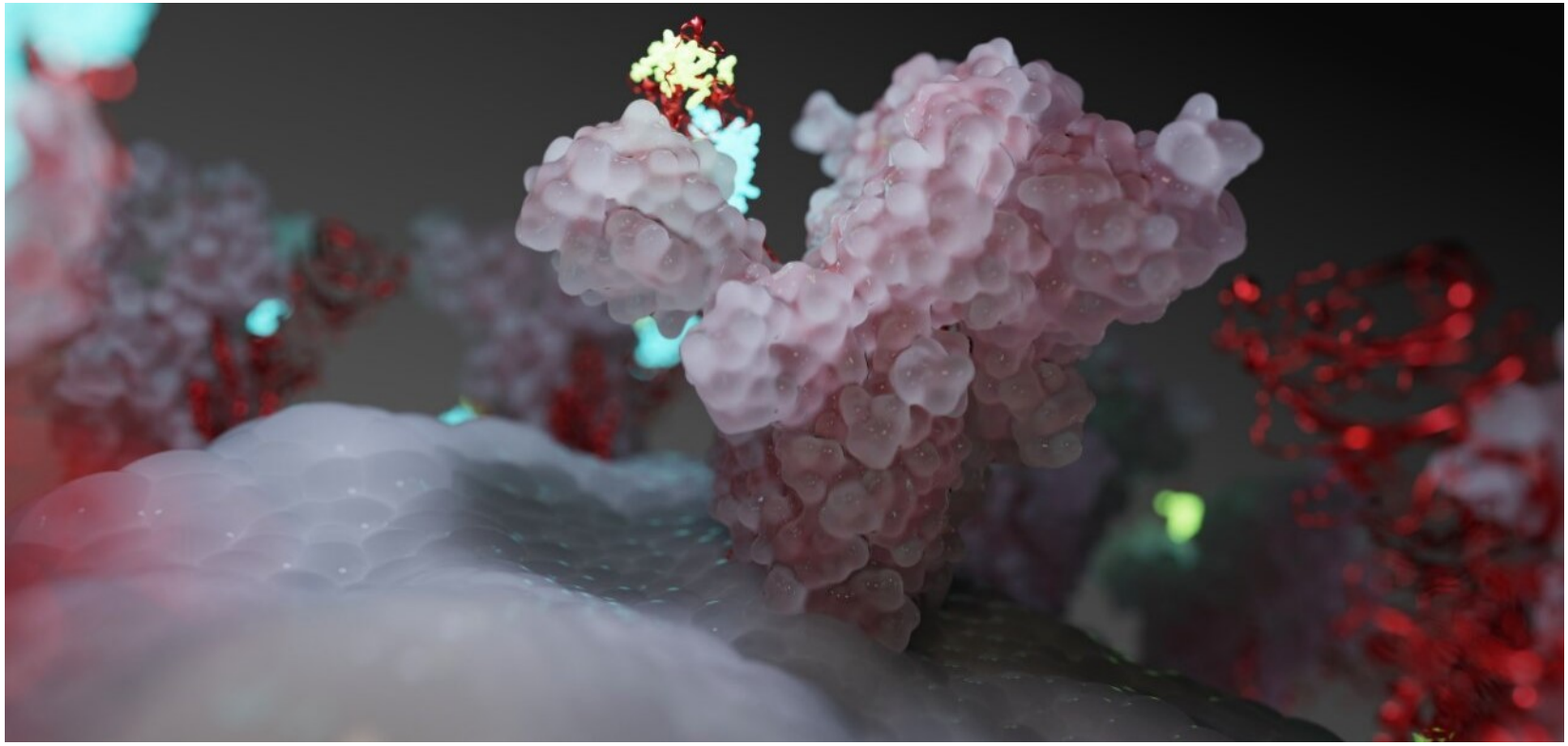
OptiX Path Tracer



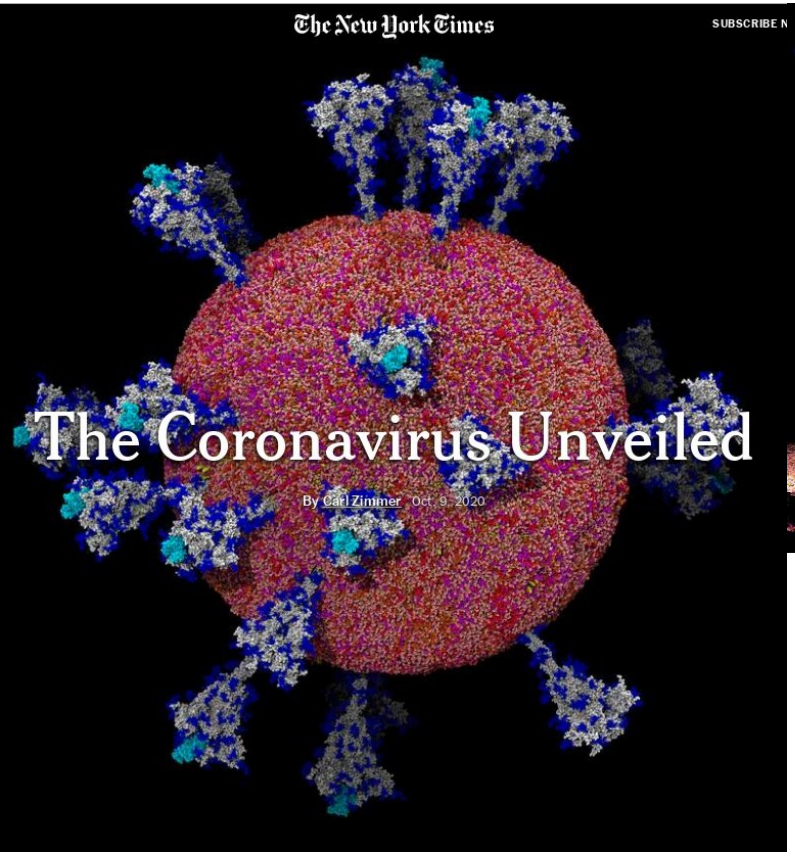
This work is licensed under a Creative Commons Attribution 4.0 International License

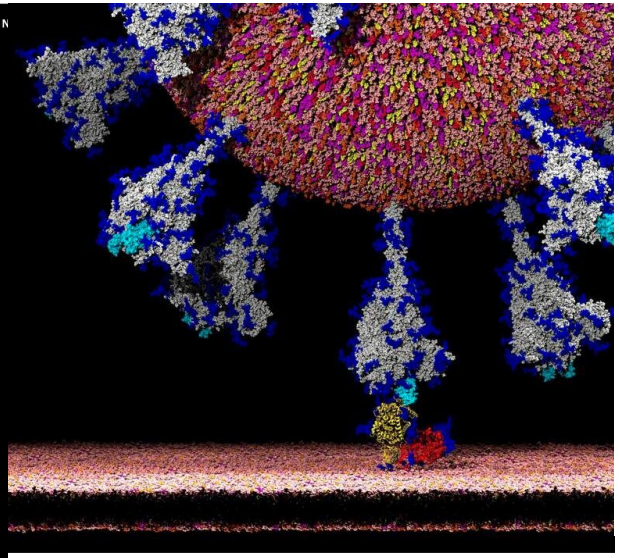
VMD+Folding@Home w/ NVIDIA Omniverse+Maya

- https://blogs.nvidia.com/blog/2020/10/07/foldingathome-omniverse-coronavirus/
- Movie (YouTube): https://youtu.be/Y9N_lmvwnUI

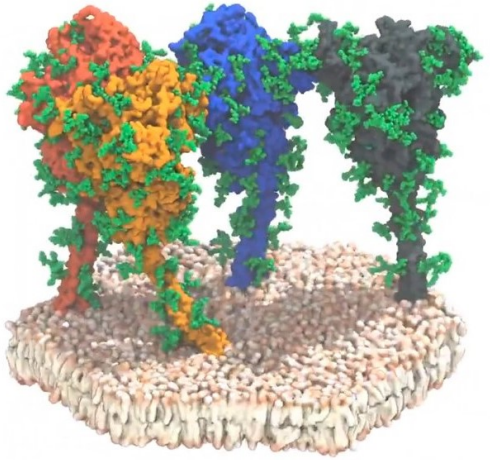


https://www.nytimes.com/interactive/2020/health/coronavirus-unveiled.html





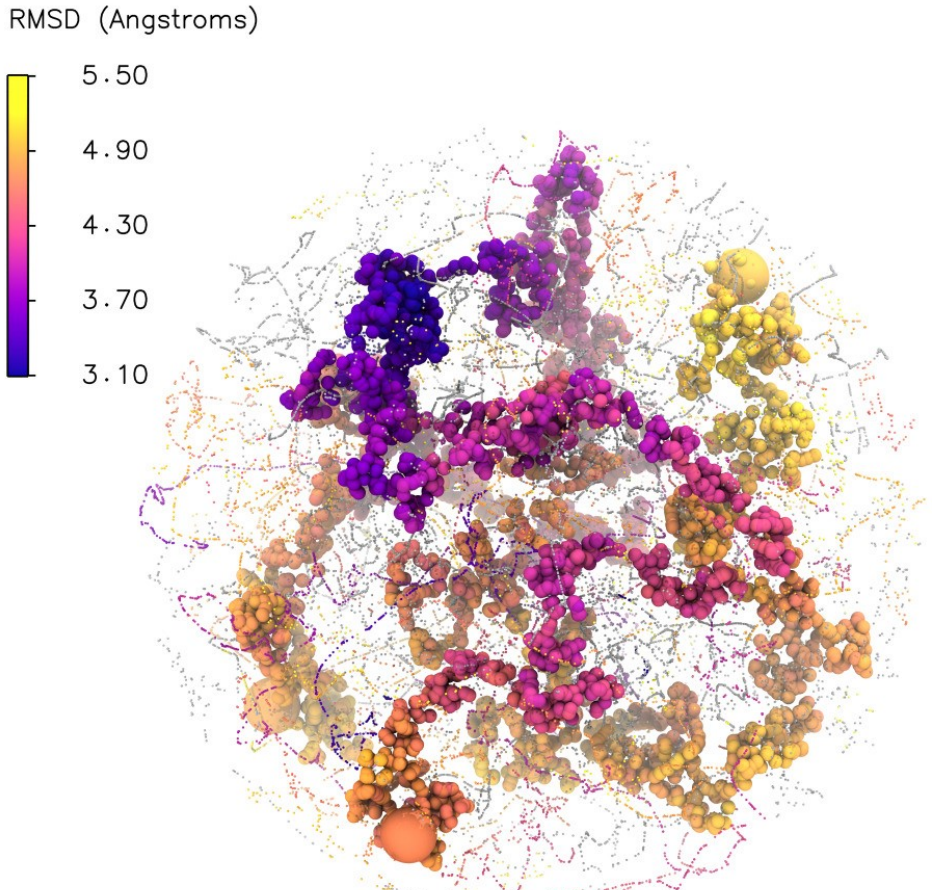
Latching on to an ACE2 receptor, in yellow, allows the coronavirus to enter human cells. Lorenzo Casalino, Amaro Lab, U.C. San Diego.



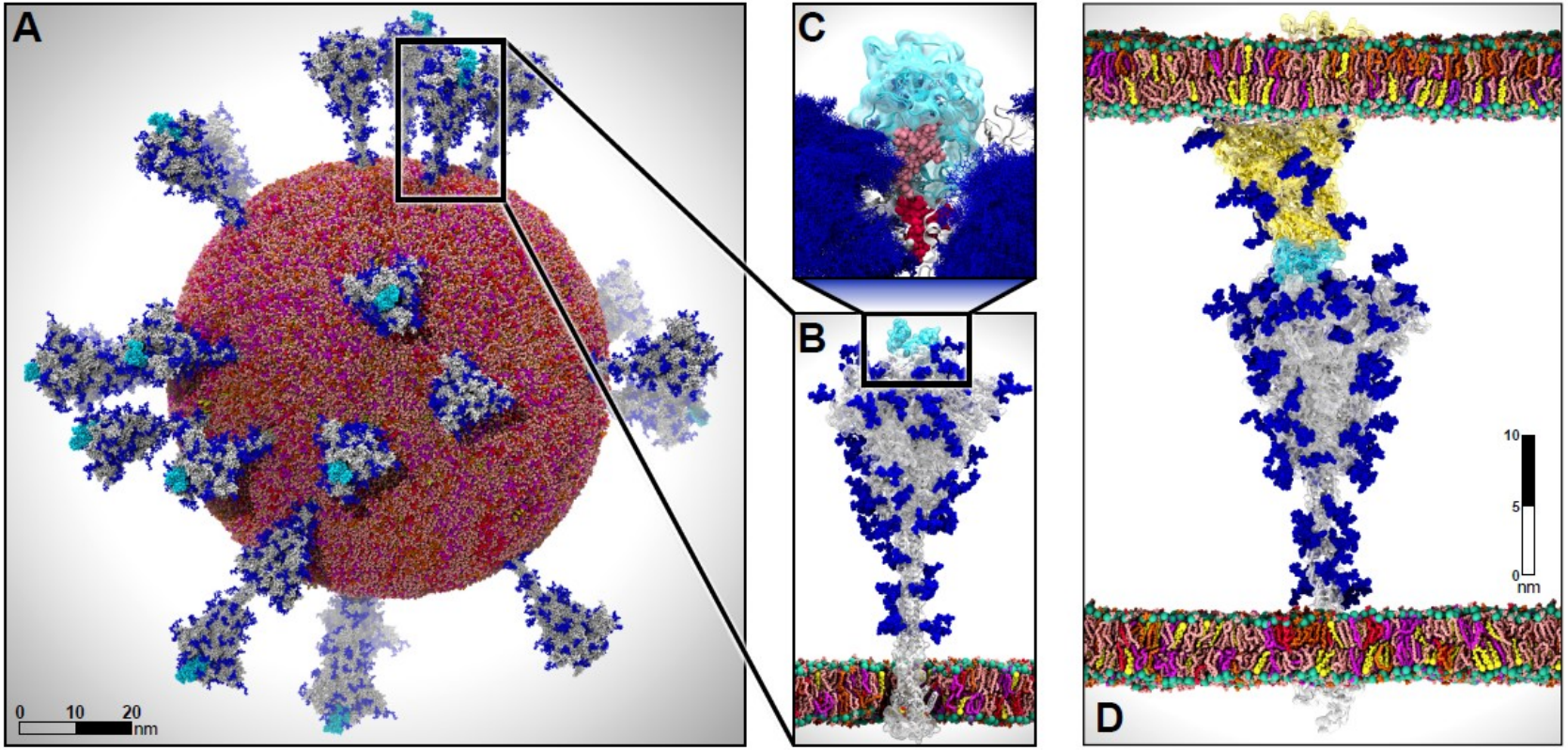
A simulation of four spike proteins, each bending on three hinges. Sören von Bülow, Mateusz Sikora and Gerhard Hummer, Max Planck Institute of Biophysics

Visualizing High-Dimensional Data:

t-SNE plot of SARS-CoV2 Simulation Campaign



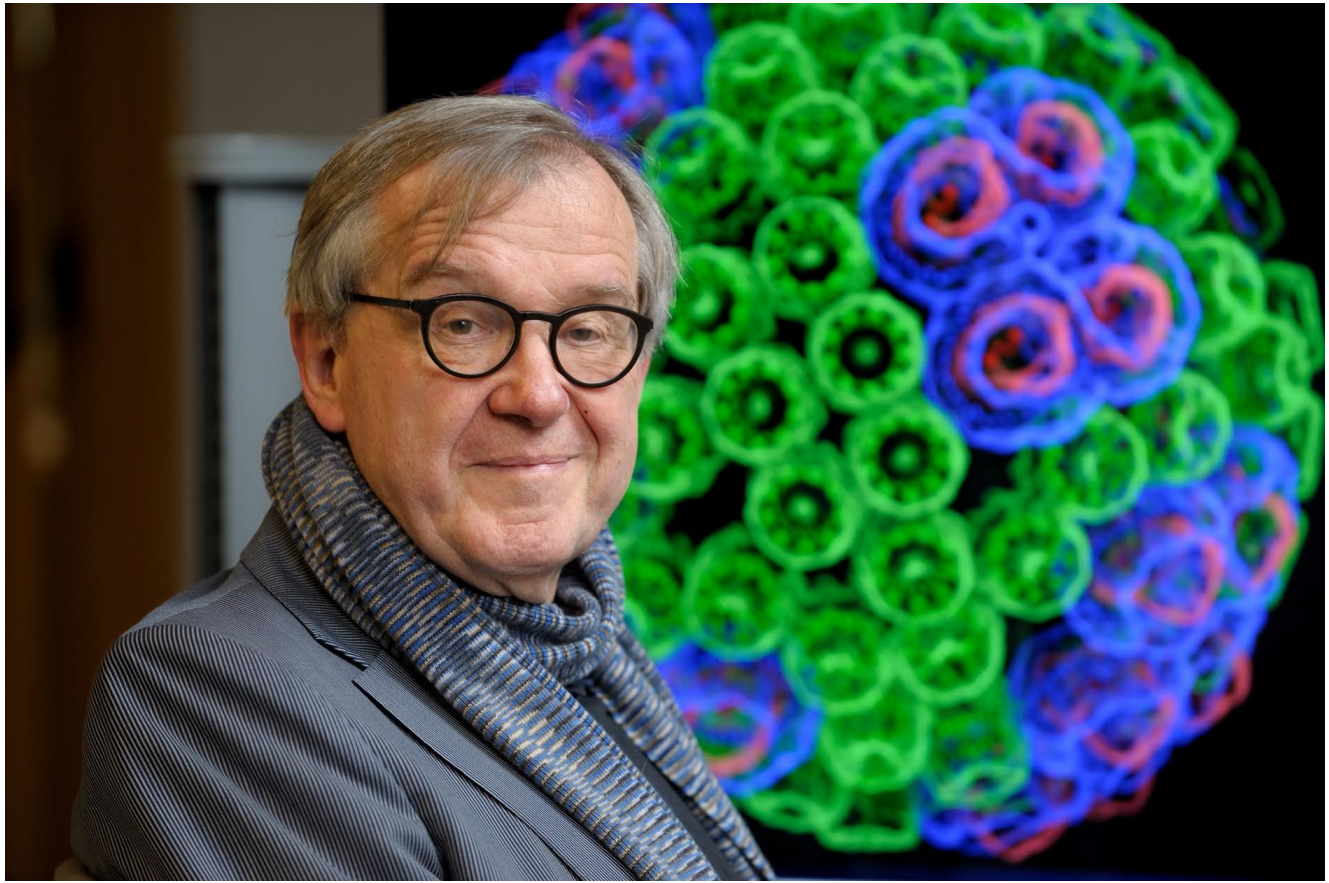
AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics.
L. Casalino, A. Dommer, Z. Gaieb, et al., IJHPCA, 2021.
https://dx.doi.org/10.1177/10943420211006452



AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics.
L. Casalino, A. Dommer, Z. Gaieb, et al., IJHPCA, 2021.
https://dx.doi.org/10.1177/10943420211006452

Acknowledgements

- Theoretical and Computational Biophysics Group, University of Illinois at Urbana-Champaign
- CUDA Center of Excellence, University of Illinois at Urbana-Champaign
- NVIDIA CUDA and OptiX teams
- Intel OSPRay team
- Funding:
  - NIH support: P41-GM104601
  - DOE INCITE, ORNL Summit
  - 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