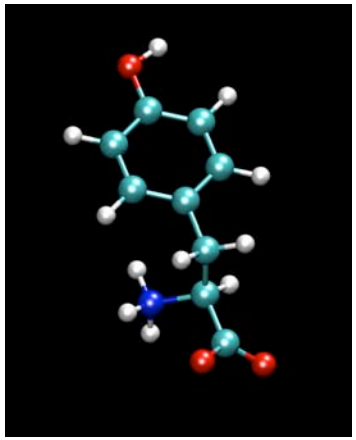
Lecture 1a

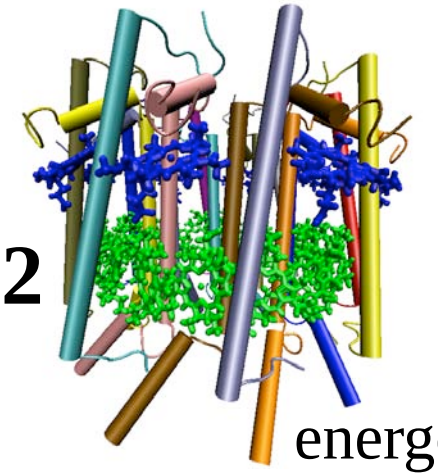
# Introduction to Protein Structures - Molecular Graphics Tool



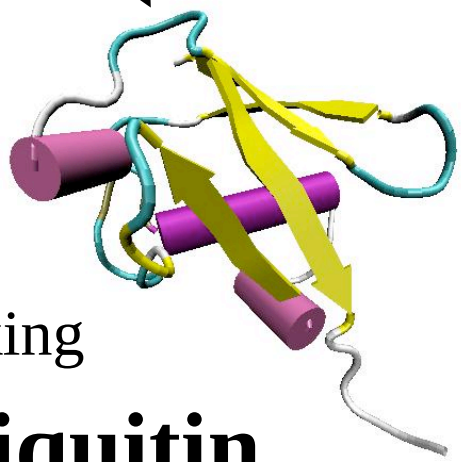
*amino acid*

*tyrosine*

**LH2**

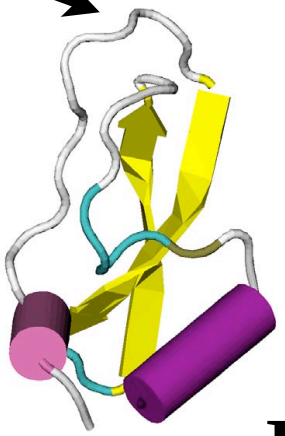


energetics



traficking

**Ubiquitin**

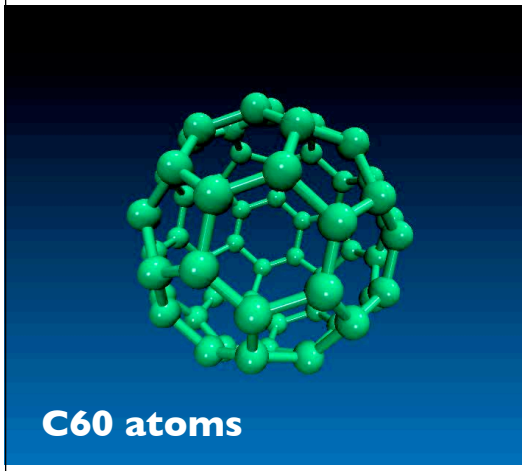


enzymatic control

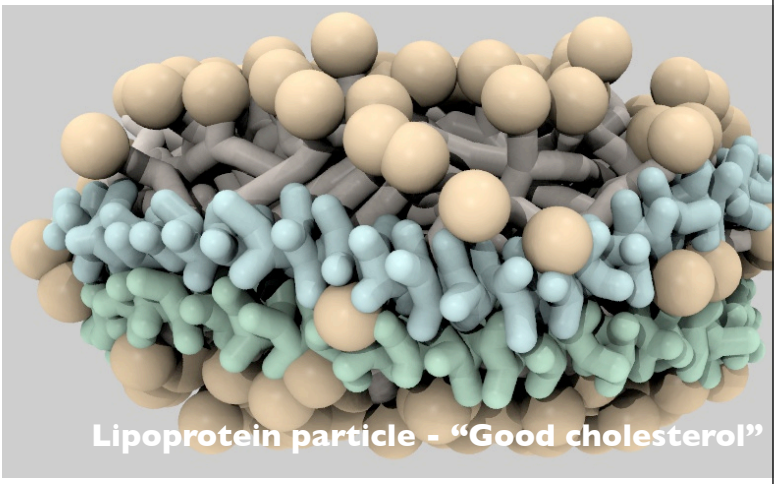
**BPTI**

# VMD - a "Tool to Think"

Carl Woese




**Atomic, CG, Particle:**
Coordinates, Trajectories,
Energies, Forces,
Secondary Structure, …

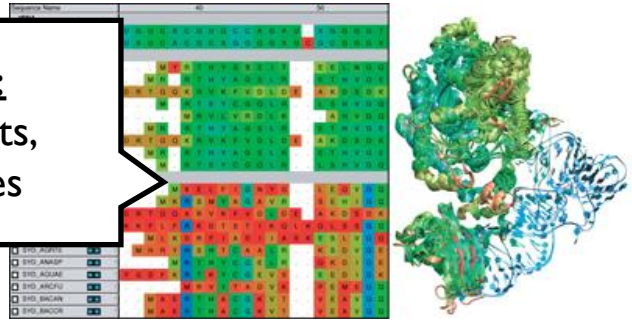



**Graphics, Geometry**

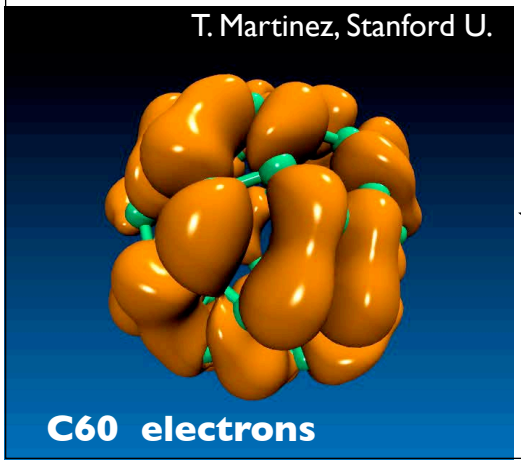
**Annotations**

**VMD**

**Sequence Data:**
Multiple Alignments,
Phylogenetic Trees



T. Martinez, Stanford U.

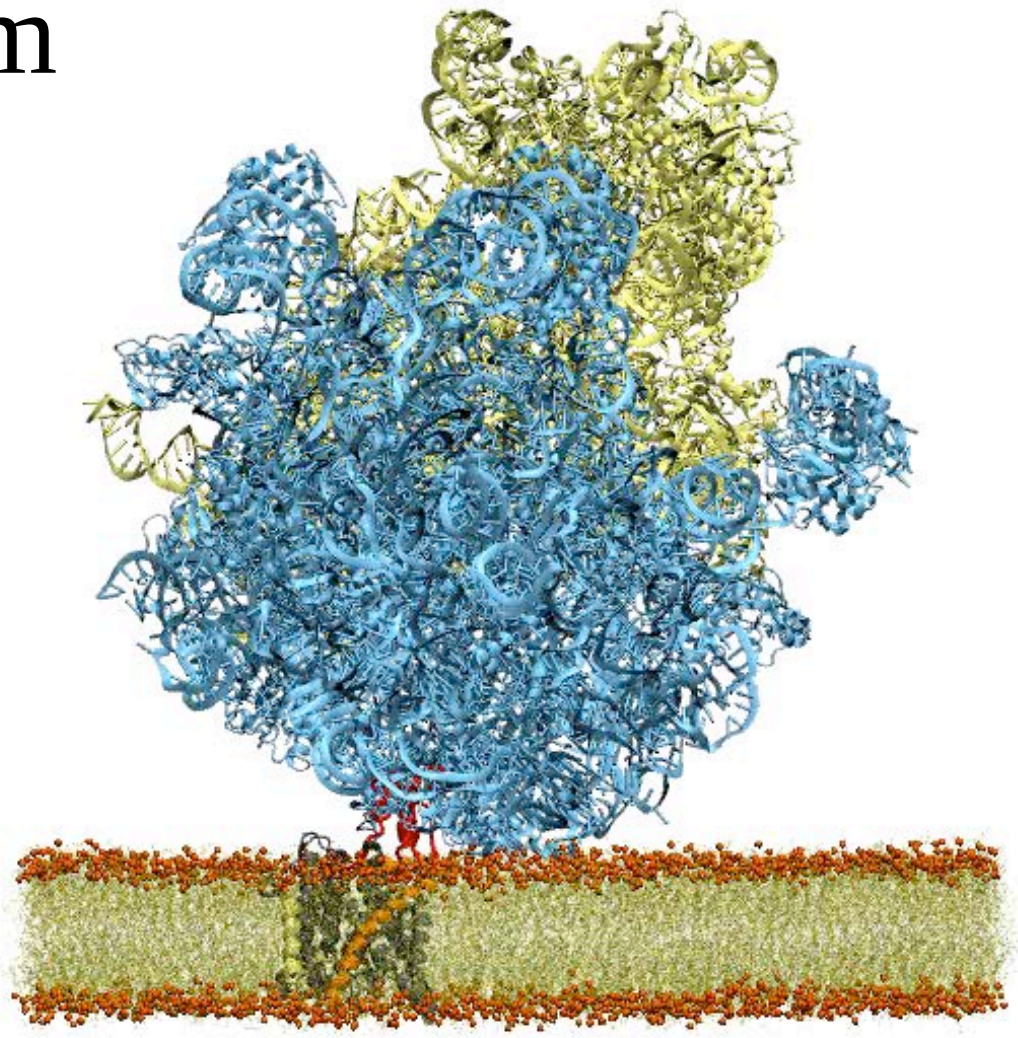



**Volumetric Data:**
Density maps,
Electron orbitals,
Electrostatic potential, …


*120,000 registered users*

# Highlights of the VMD Molecular Graphics Program

- > 120,000 registered users
- Platforms:
  - Unix / Linux
  - Windows
  - MacOS X
- Display of large biomolecules and simulation trajectories
- Sequence browsing and structure highlighting
- Multiple sequence - structure analysis
- User-extensible scripting interfaces for analysis and customization



***The program is used today more for preparation and analysis of modeling than for graphics.***

# Key Features of VMD

- General 3-D molecular visualization with extensive drawing and coloring methods
- Extensive atom selection syntax for choosing subsets of atoms for display
- Visualization of dynamic molecular data
- Visualization of volumetric data
- Supports all major molecular data file formats
- No limits on the number of molecules or trajectory frames, except available memory
- Rendering high-resolution, publication-quality molecule images
- Movie making capability
- Extensions to the Tcl/Python scripting languages
- Extensible source code written in C and C++
- Building and preparing systems for molecular dynamics simulations
- Interactive molecular dynamics simulations
- Molecular analysis commands

# VMD Analysis Features

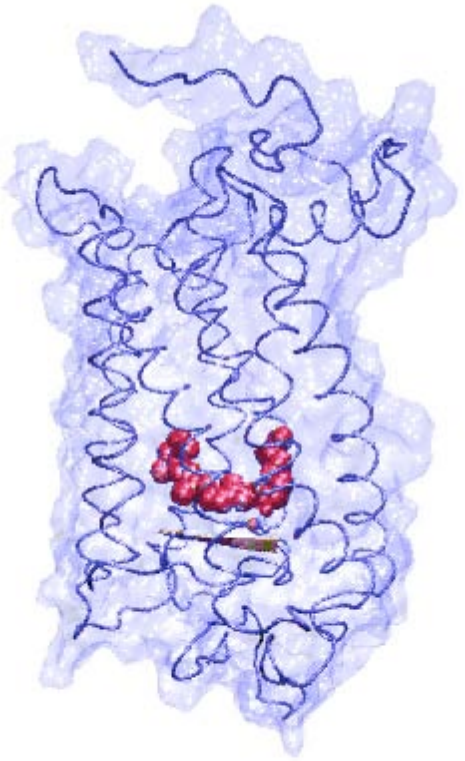
- User extensible scripting language for automating complex analysis tasks
- Powerful atom selection language
  - Select atoms by name, type, many other properties
  - Topological selections: rings, bond connectivity, etc
  - Select atoms by volumetric field, e.g. electrostatic field, electron density, etc.
  - Set per-atom properties for selected atoms, user-defined property fields, etc.

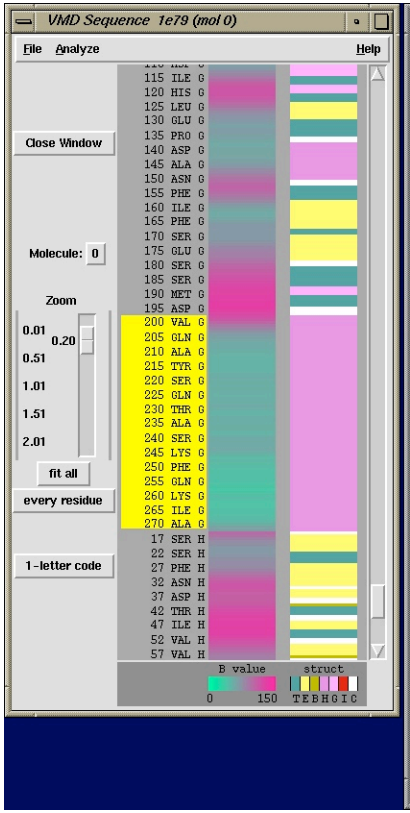
# VMD Analysis Features

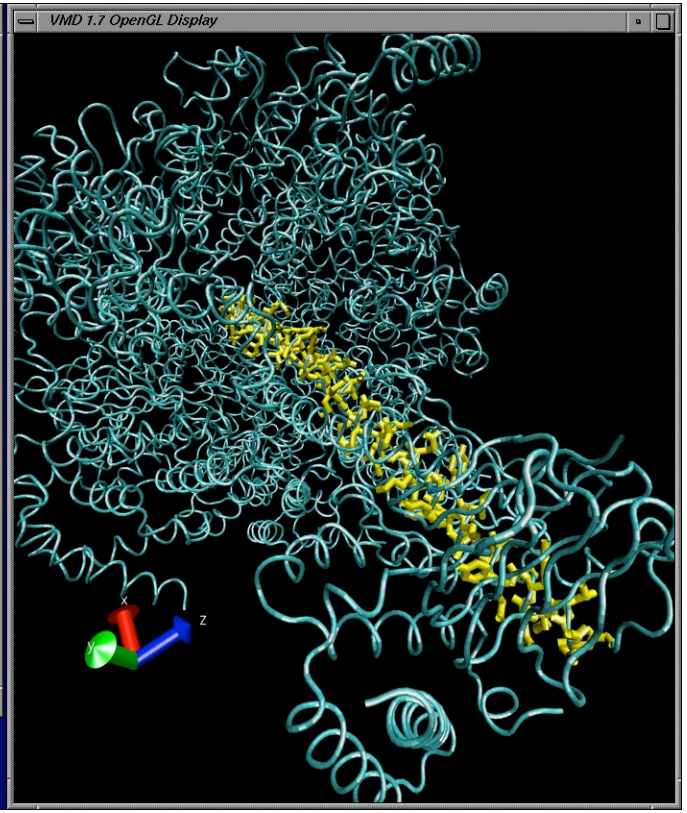
- VMD “measure” analysis commands:
  - Compute RMSD, RMSF between selections
  - Compute RMS alignment transform between selections
  - Find atoms on the surface of a selection
  - Compute COM, axes of inertia, moments of inertia, dipole moment, radius of gyration
  - Compute molecular symmetry for a selection
  - Compute radial distribution functions
  - Compute energies (bond/angle/dihedral/…)
- VMD “volmap” volumetric analysis commands:
  - Compute time-averaged electrostatic potential maps
  - Compute time-averaged occupancy/density/… maps

# Molecular Graphics Perspective of Protein Structure and Function

see tutorial at http://www.ks.uiuc.edu/Training/Tutorials/





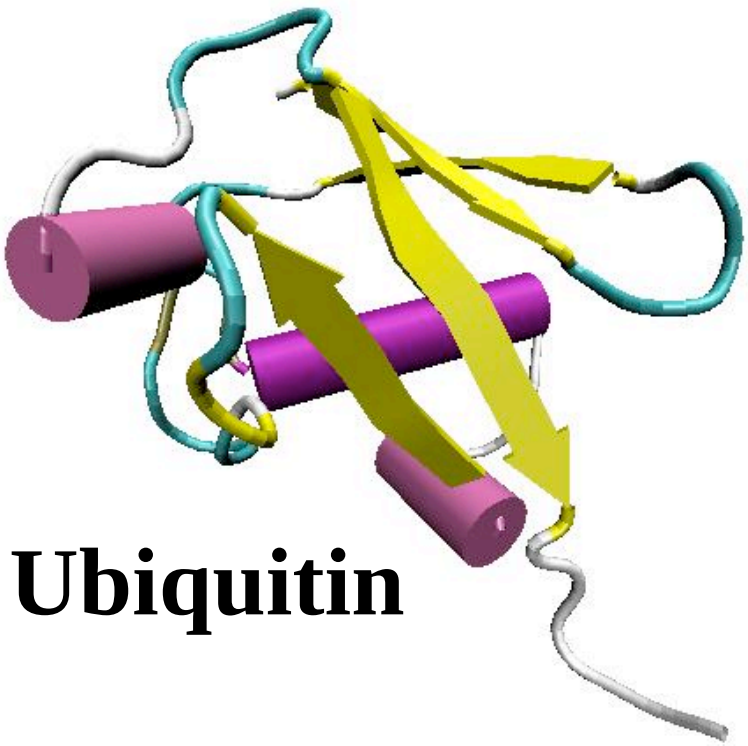




animation sequence structure

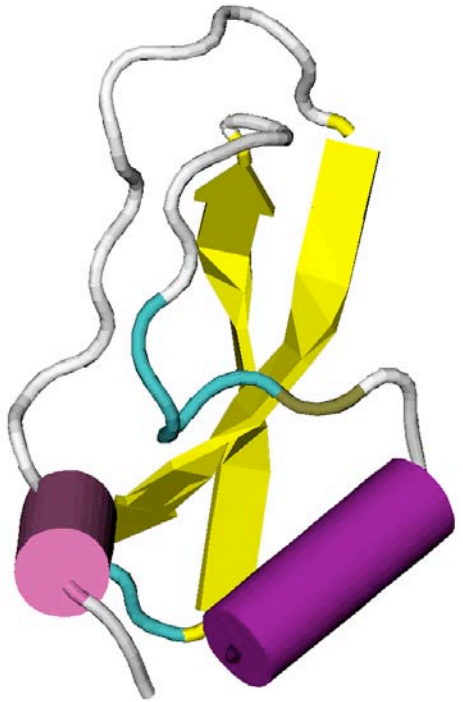
# Focus on two proteins

## Ubiquitin (used in VMD Tutorial)

## Bovine Pancreatic Trypsin Inhibitor (BPTI, available as a case study, www.ks.uiuc.edu)



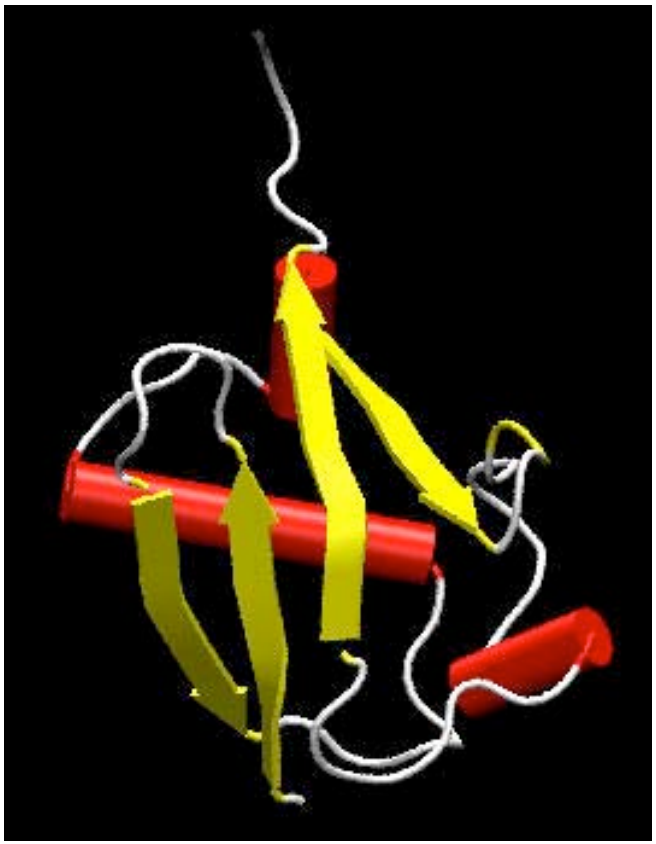
Ubiquitin



BPTI

# Ubiquitin

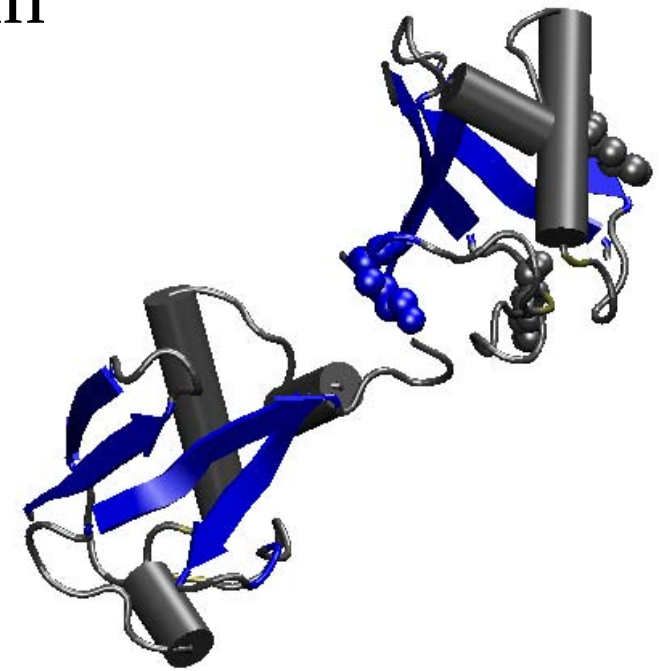
- 76 amino acids
- highly conserved
- covalently attaches to proteins and tags them for degradation
- other cell traficking



• Glycine at C-terminal attaches to the Lysine on the protein by an isopeptide bond.

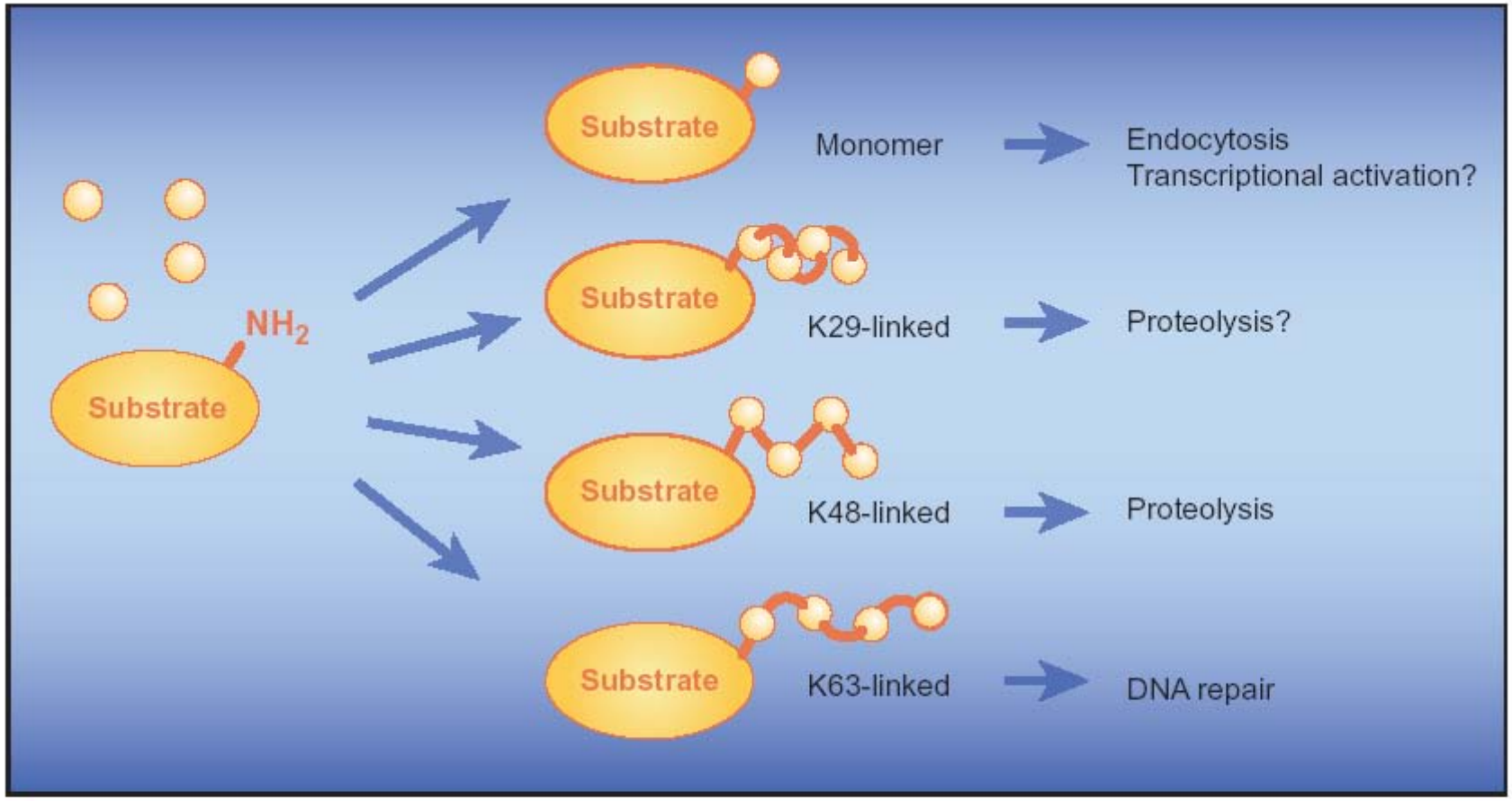
• it can attach to other ubiquitin molecules and make a polyubiquitin chain.

There are 7 conserved lysine residues in ubiquitin.



Two ubiquitins attached together through LYS 48. LYS 63 and LYS 29 are also shown there.

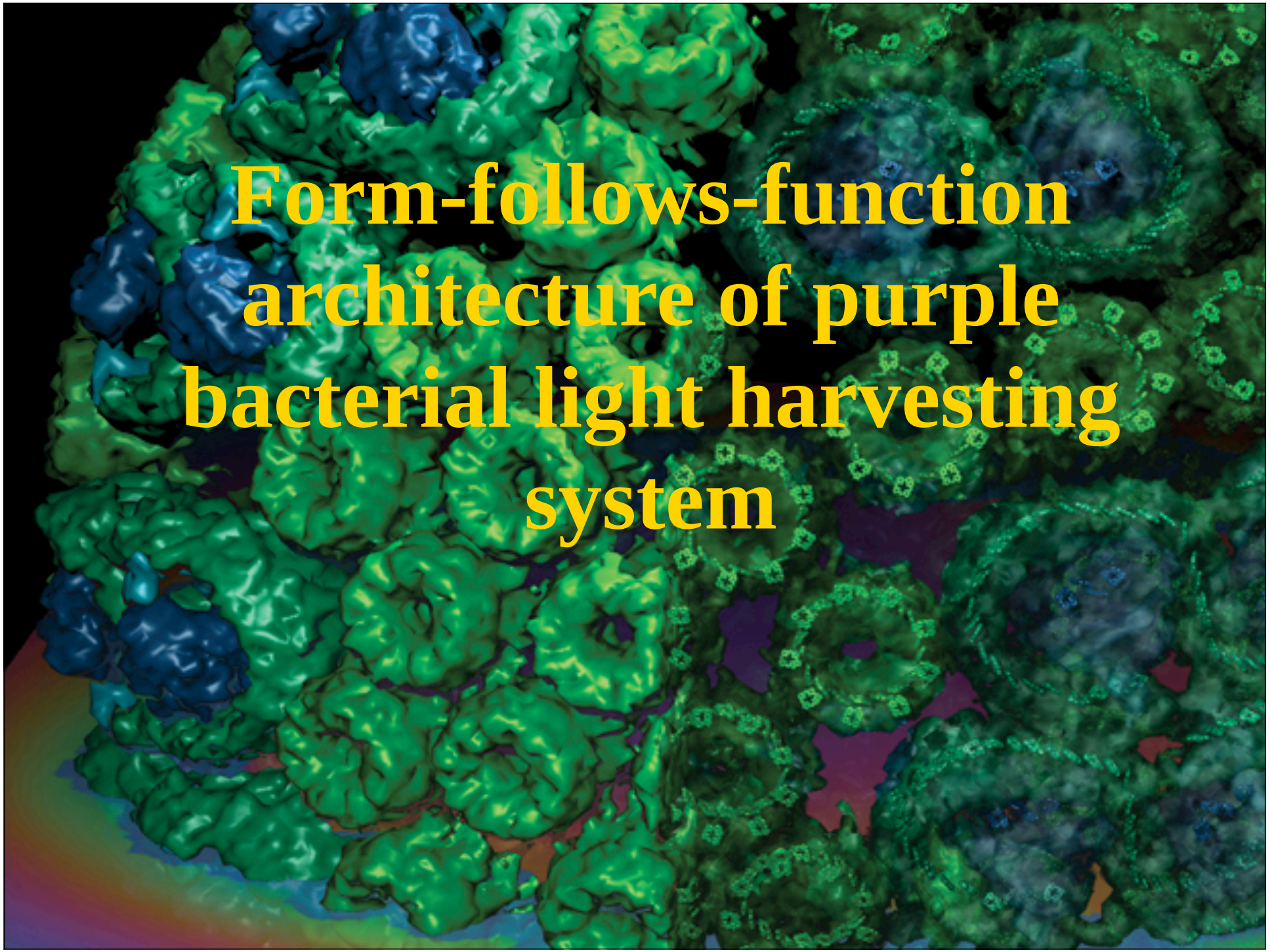
# Multi-ubiquitylation targets destination of proteins



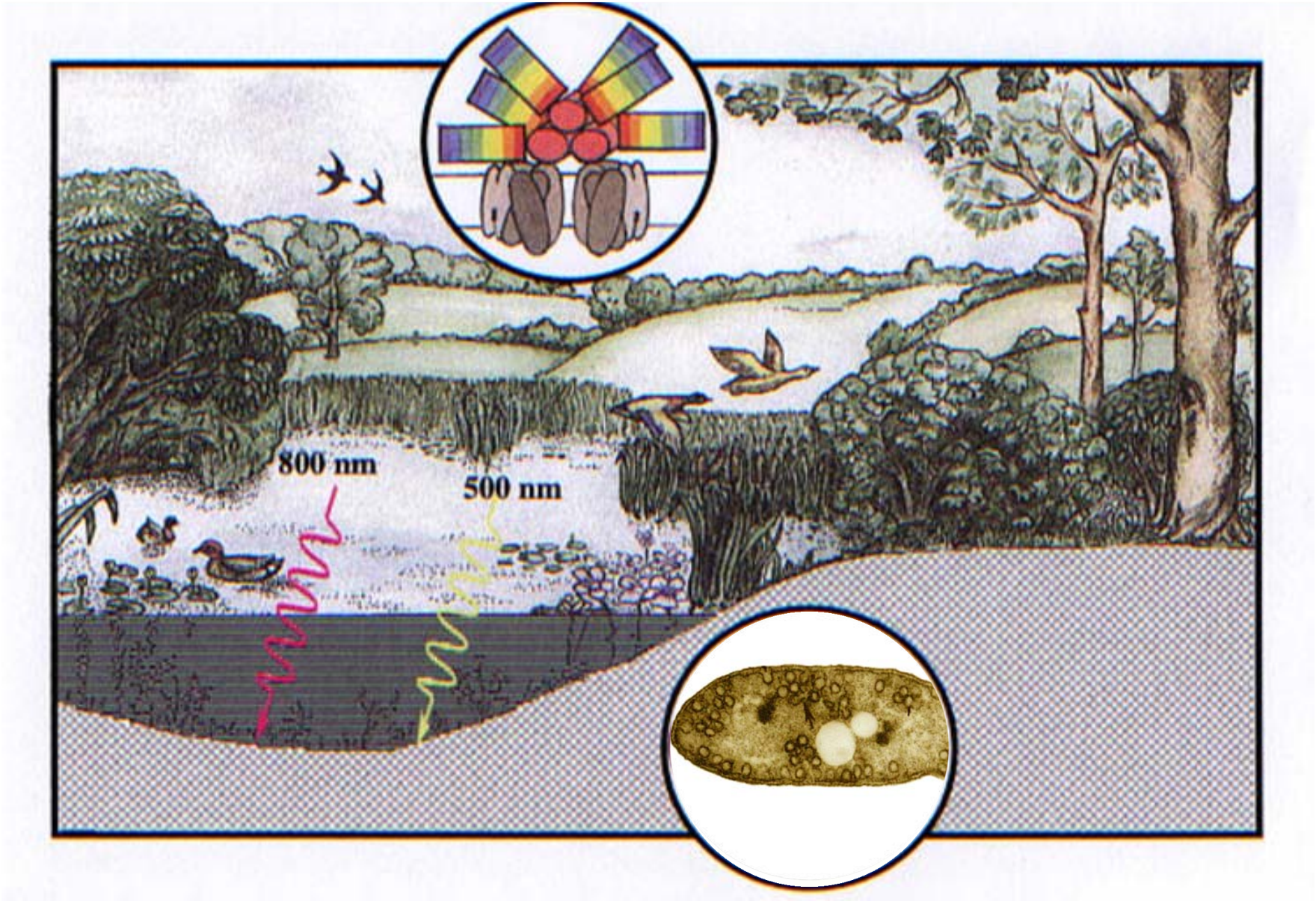

**Multifaceted.** Ubiquitin can attach to its various substrate proteins, either singly or in chains, and that in turn might determine what effect the ubiquitination has. (K29, K48, and K63 refer to the particular lysine amino acid used to link the ubiquitins to each other.)

Marx, J., Ubiquitin lives up its name, Science 297, 1792-1794 (2002)

# VMD session with ubiquitin

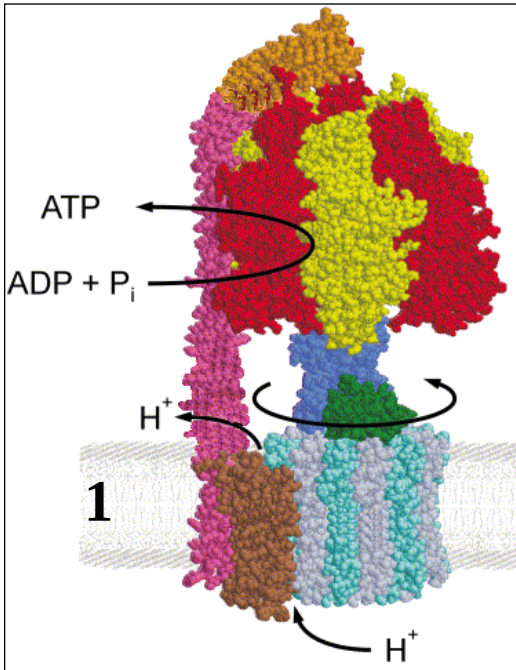

Form-follows-function
architecture of purple
bacterial light harvesting
system

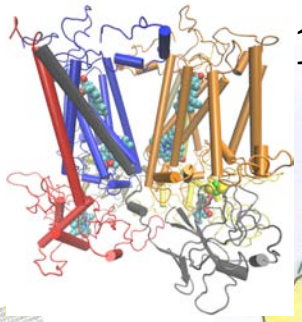
# Habitats of Photosynthetic Life Forms




purple bacterium

# The proteins that make up the chromatophore of photosynthetic bacteria

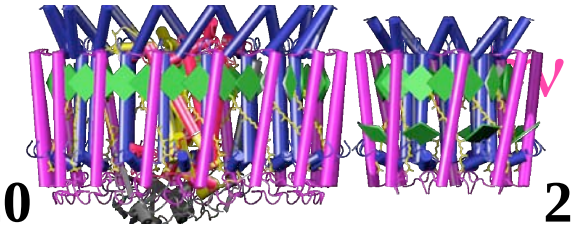


10

1

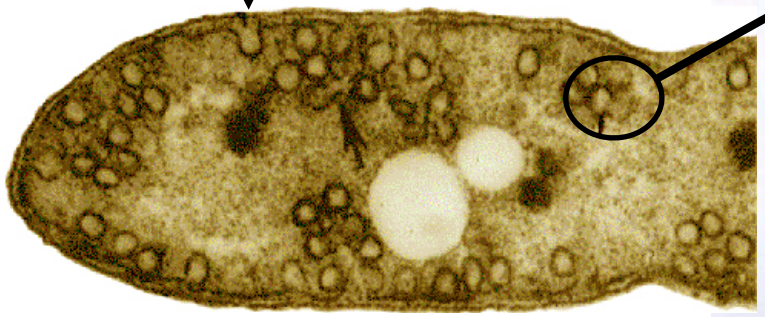
5



20

200

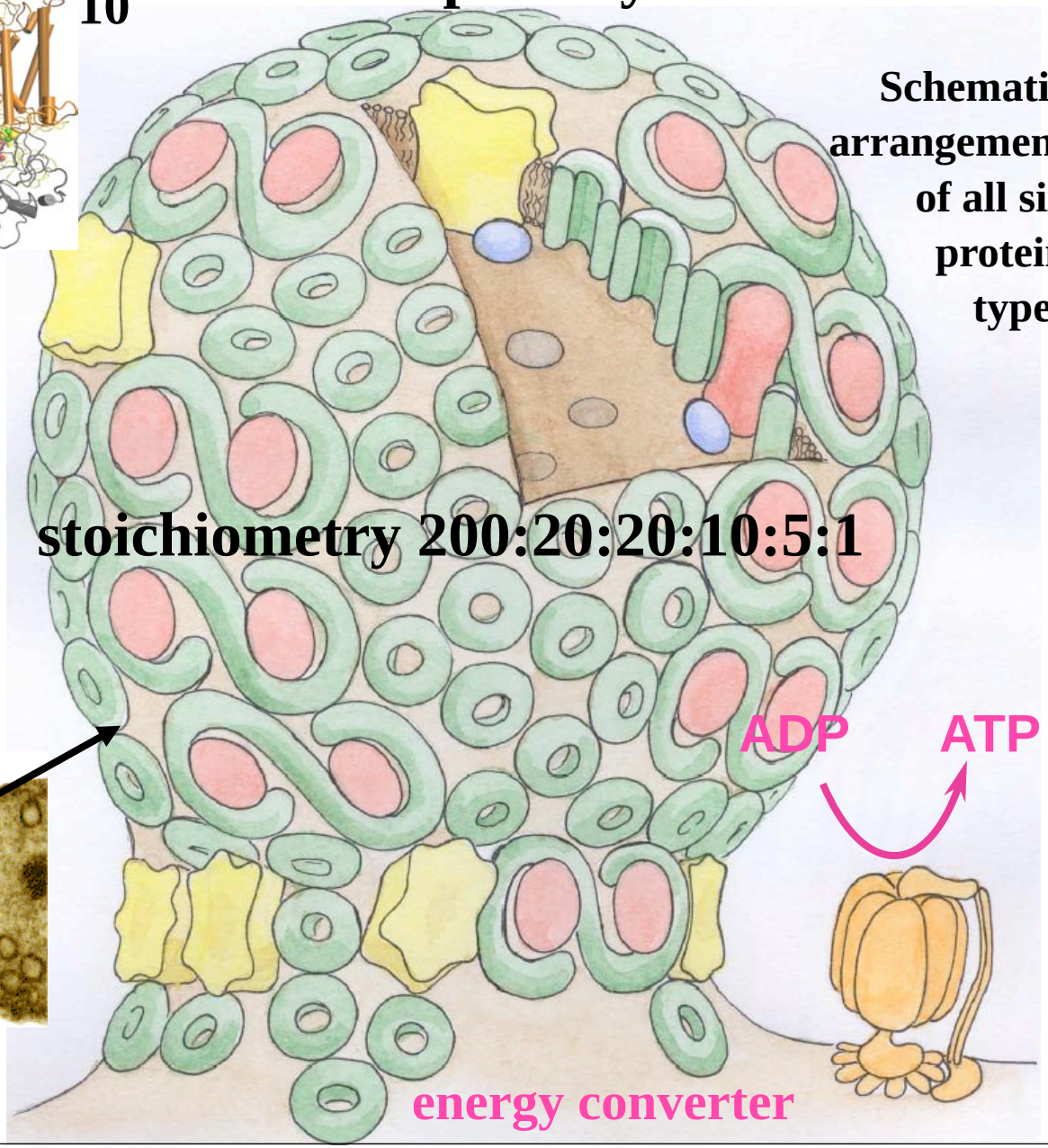
photosynthetic unit



purple bacterium cell

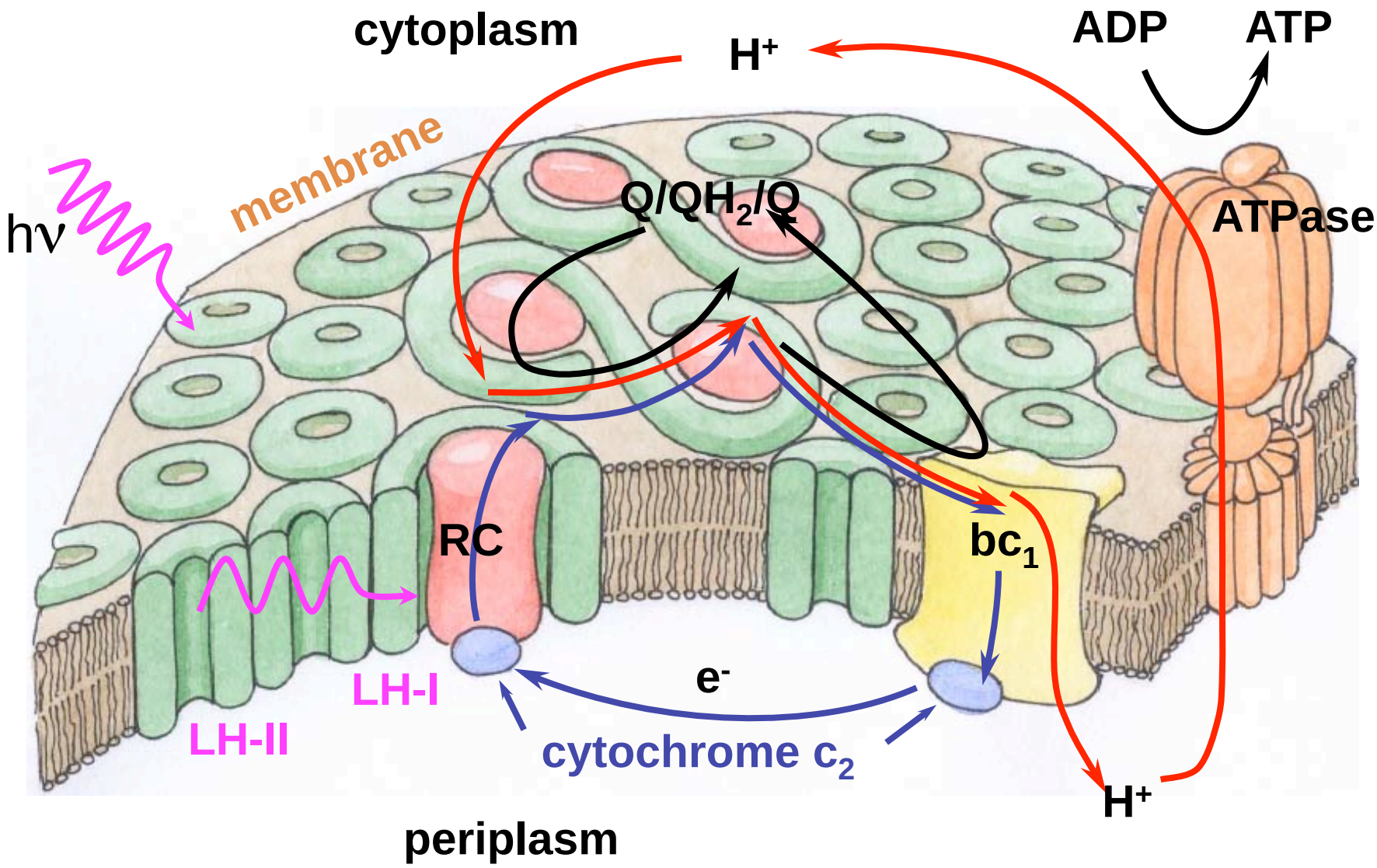
Schematic arrangement of all six protein types

stoichiometry 200:20:20:10:5:1

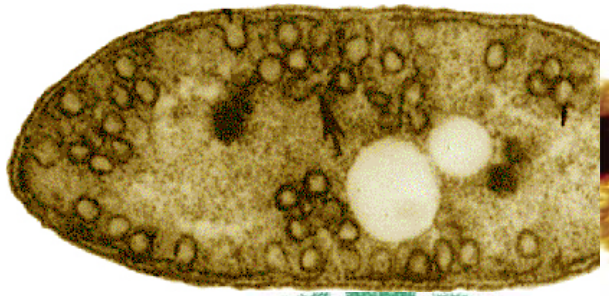


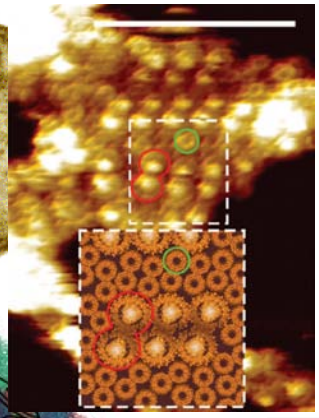
# Chromatophore of Purple Bacteria

*(section of the chromatophore membrane)*

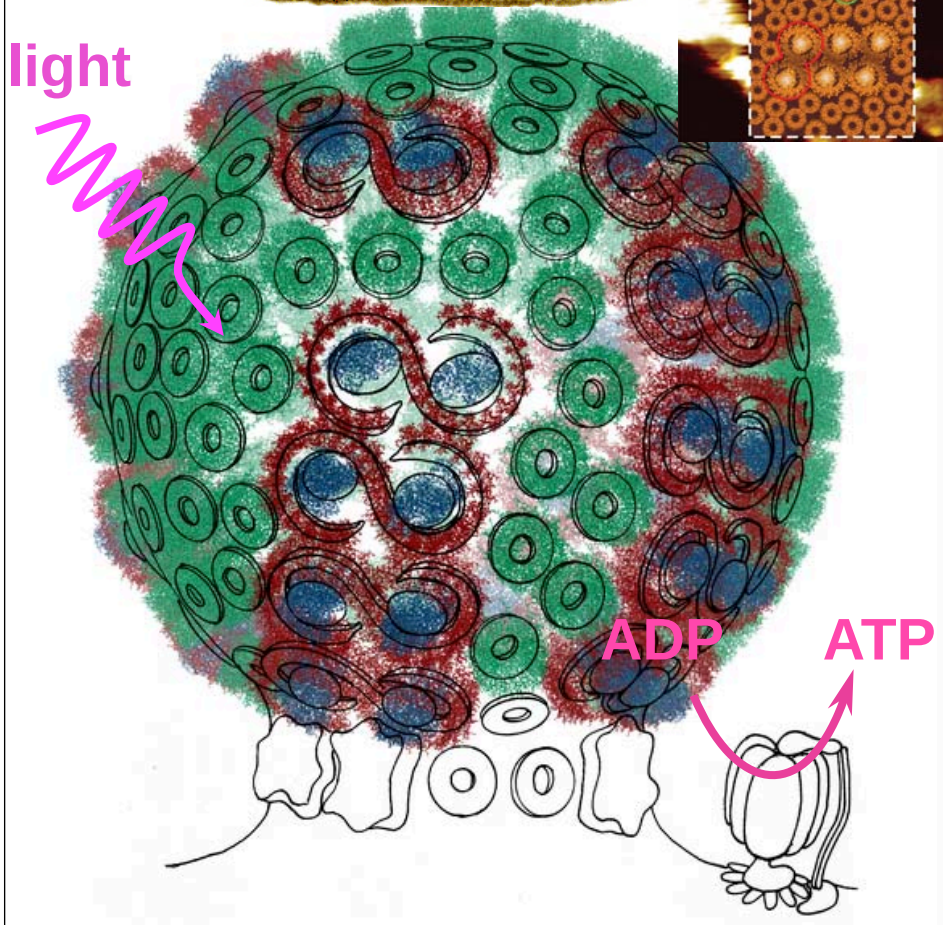


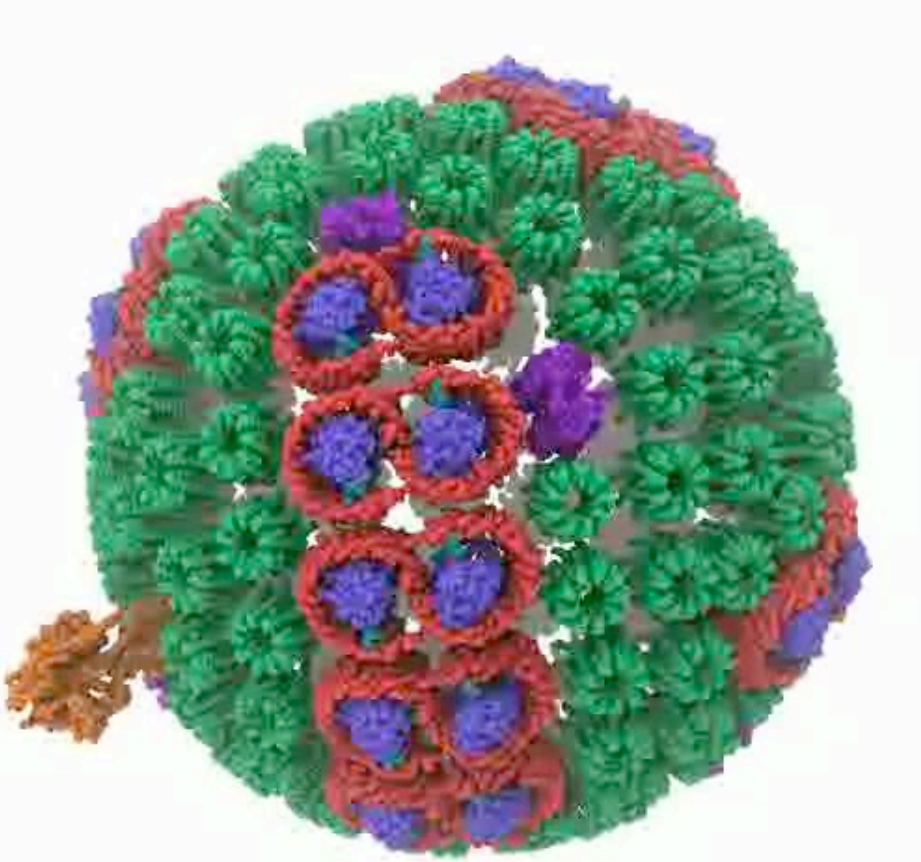
# Knowing the Atomic Level Structure





of the chromatophore, one can systematically describe its physical mechanism

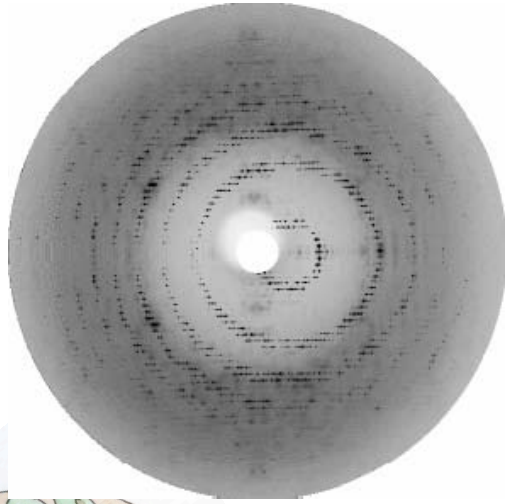


M. Sener, J. Olsen, N. Hunter, and K. Schulten. *PNAS*, **104:** 15723-15728, 2007

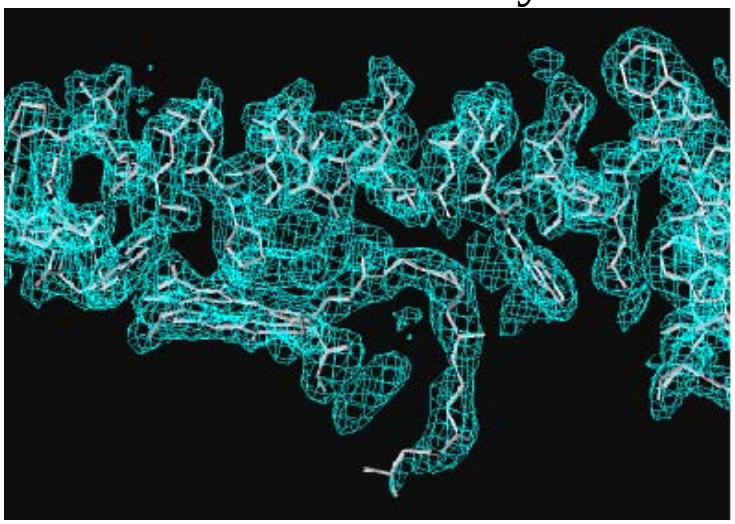
# Structure of LH 2 of *Rs. molischianum*

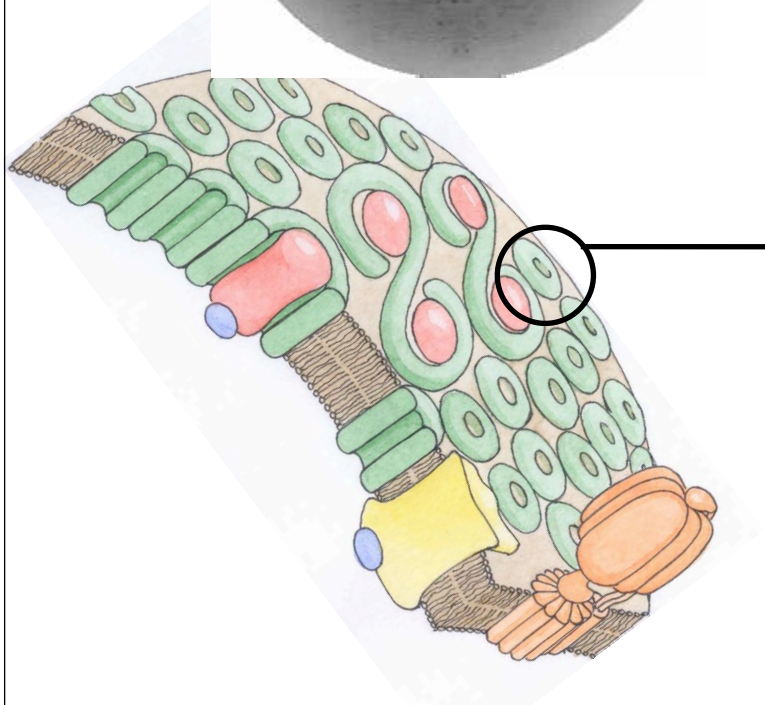
crystallographic diffraction pattern

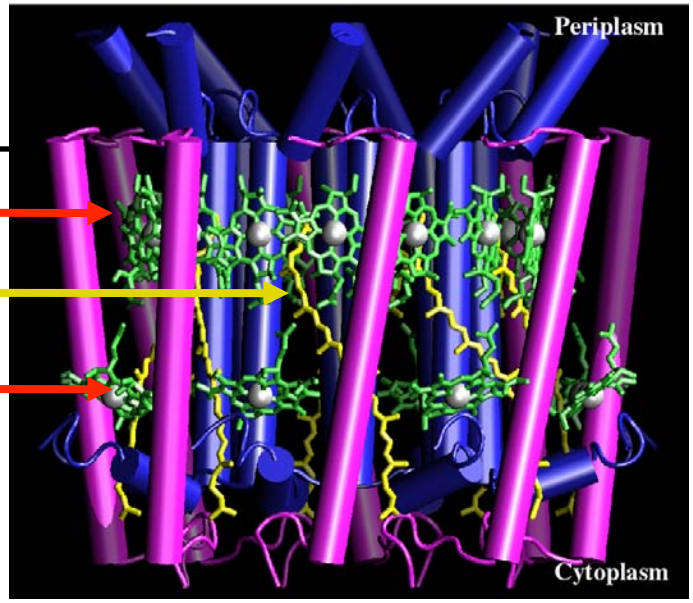


molecular replacement through modeling

electron density








**B850 band**

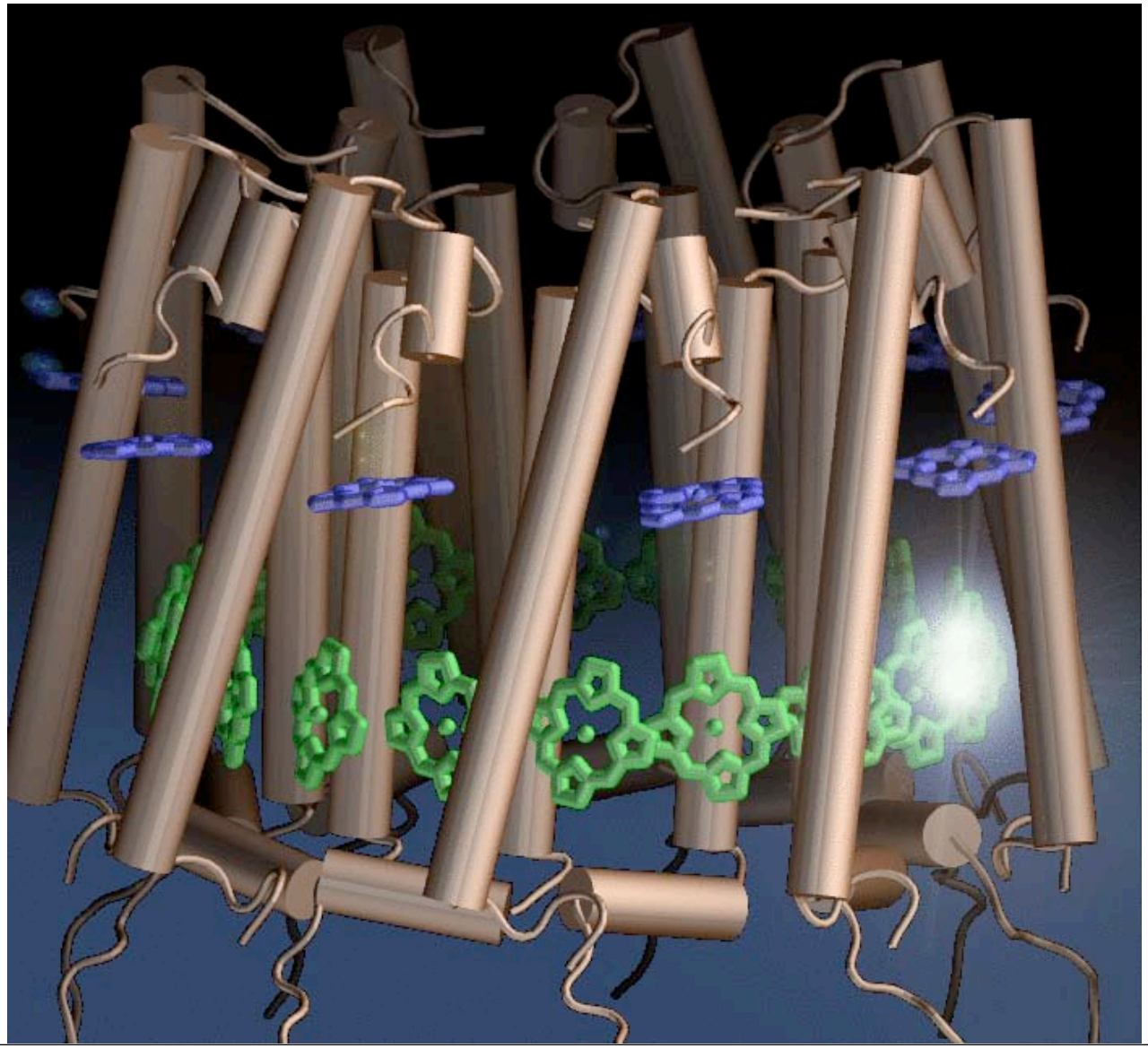
**B500 band**

**B800 band**

**optical spectrum**

Koepke et al., Structure, **4**, 581 (1996)

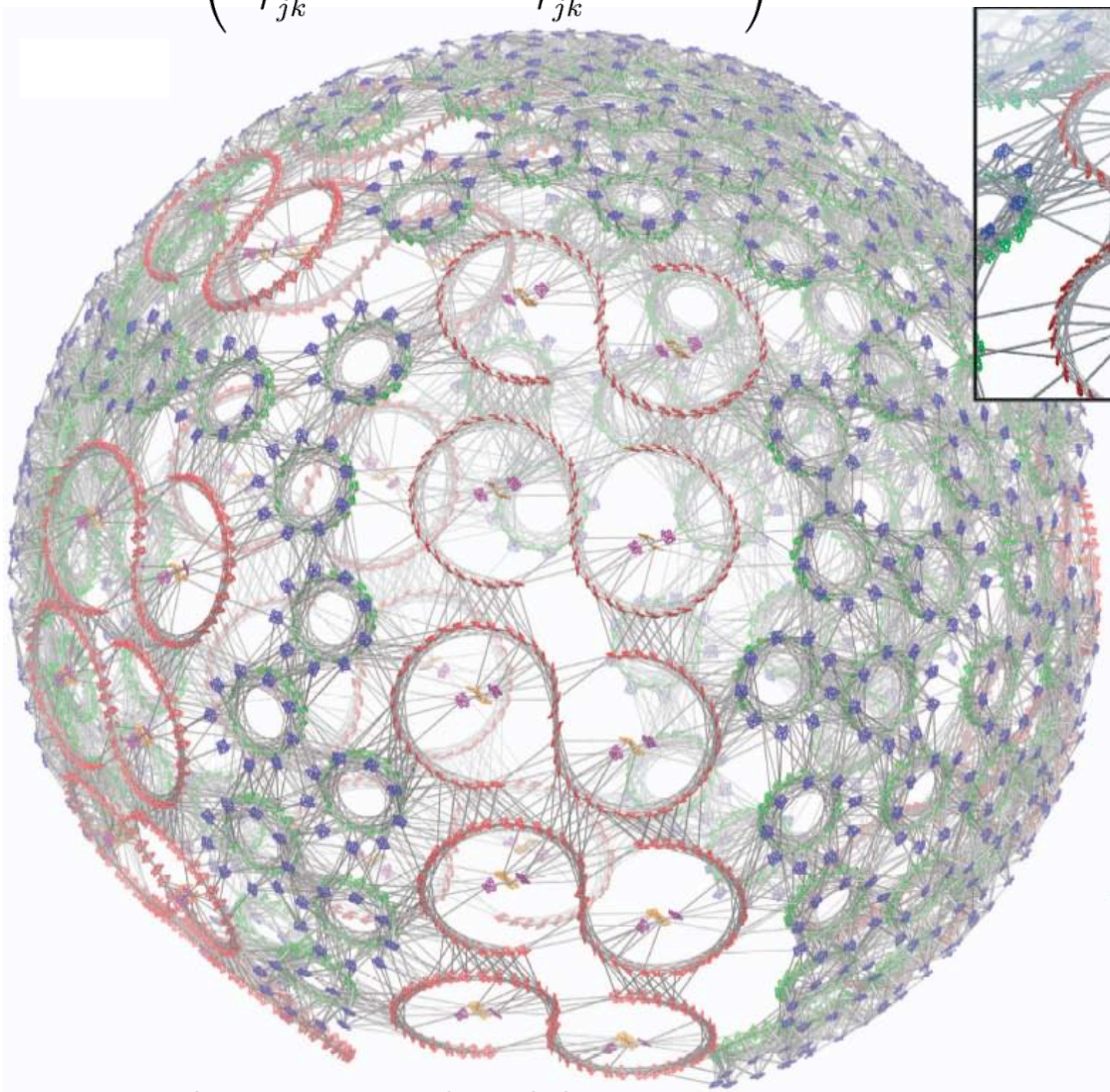
# Excitonic dynamics in LH2

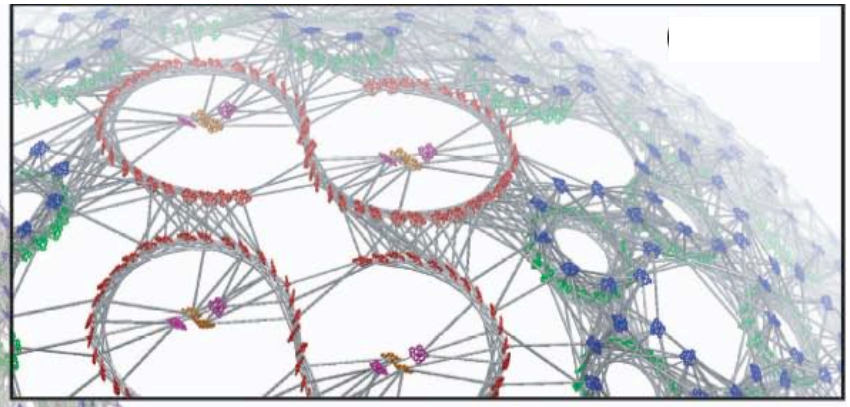


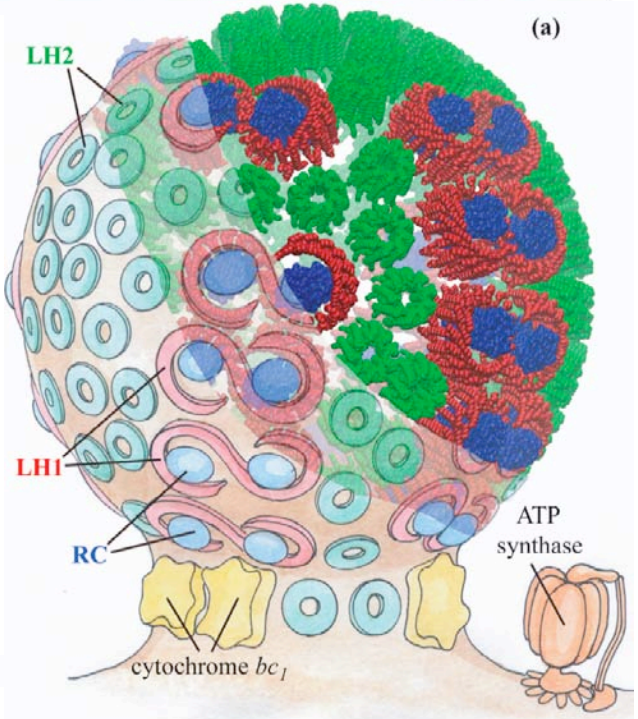
# The “Physics” of Light Harvesting in the Chromatophore

## Calculated Energy Transfer Rates Determine Optimal Placement of Proteins in Chromatophore

$$W_{jk} = C\left(\frac{\vec{d}_j \cdot \vec{d}_k}{{r_{jk}}^3} - \frac{3(\vec{r}_{jk} \cdot \vec{d}_j)\,(\vec{r}_{jk} \cdot \vec{d}_k)}{{r_{jk}}^5}\right)$$ links: induced dipole - induced dipole interaction








M. Sener, J. Olsen, N. Hunter, and K. Schulten. *PNAS*, **104:** 15723-15728, 2007

# VMD session with LH2

# VMD – "Visual Molecular Dynamics"
# http://www.ks.uiuc.edu/Research/vmd/

**Visualization and analysis of:**

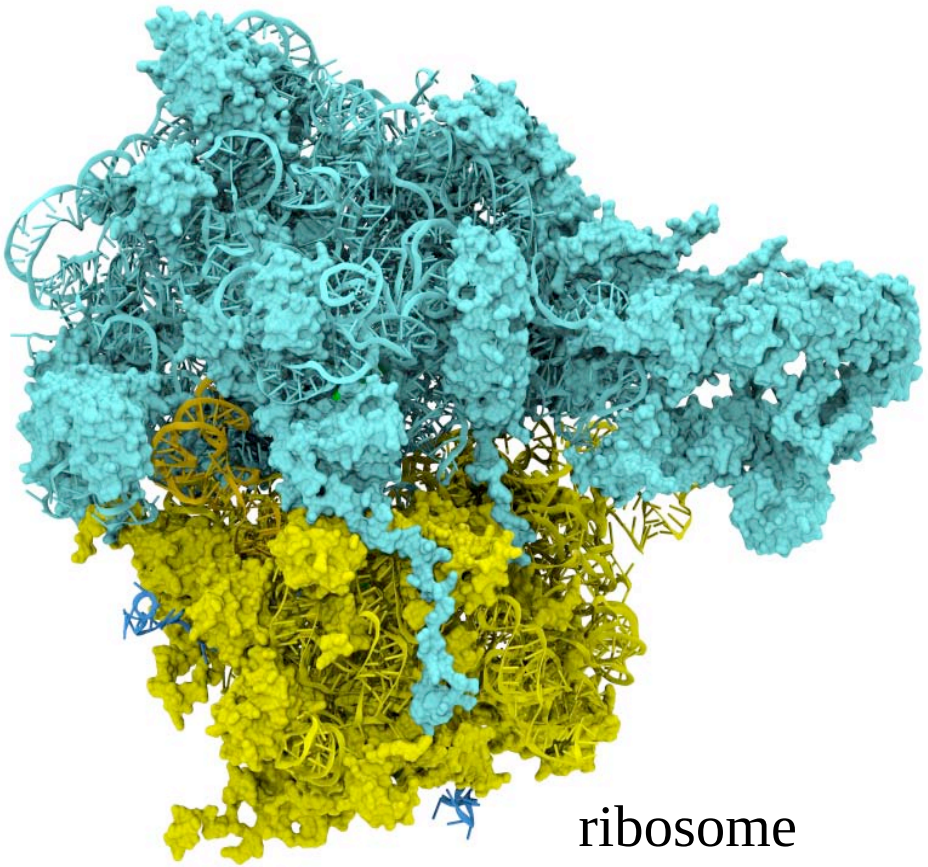
- Molecular dynamics and quantum chemistry simulations
- Sequence data
- Cryo-electron microscopy maps

**User-extensible with built-in scripting and many plugins**

**Supports very large data sets, batch mode analysis on clusters**

**Takes advantage of advanced technological opportunities:**

- High quality interactive display, batch mode rendering, movie making
- Supports multi-core processors, GPUs



ribosome

# VMD – "Visual Molecular Dynamics"
# http://www.ks.uiuc.edu/Research/vmd/

**Visualization and analysis of:**

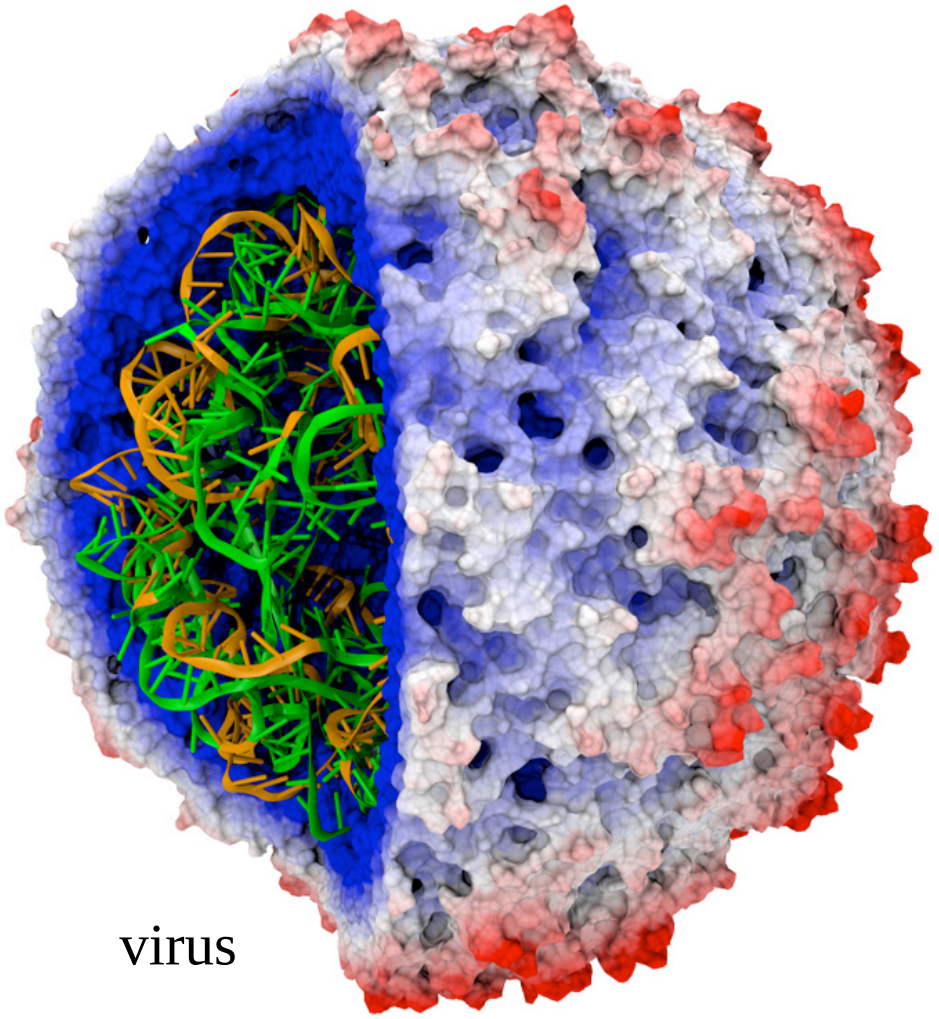
- Molecular dynamics and quantum chemistry simulations
- Sequence data
- Cryo-electron microscopy maps

**User-extensible with built-in scripting and many plugins**

**Supports very large data sets, batch mode analysis on clusters**

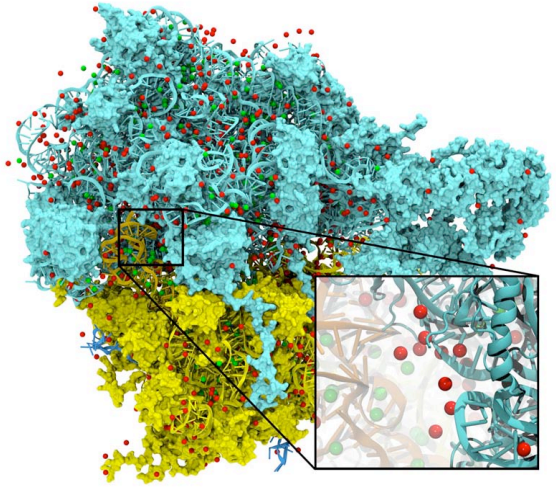
**Takes advantage of advanced technological opportunities:**

- High quality interactive display, batch mode rendering, movie making
- Supports multi-core processors, GPUs

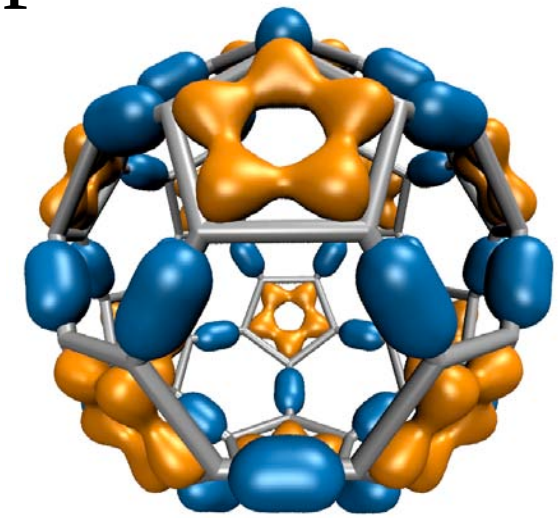


virus

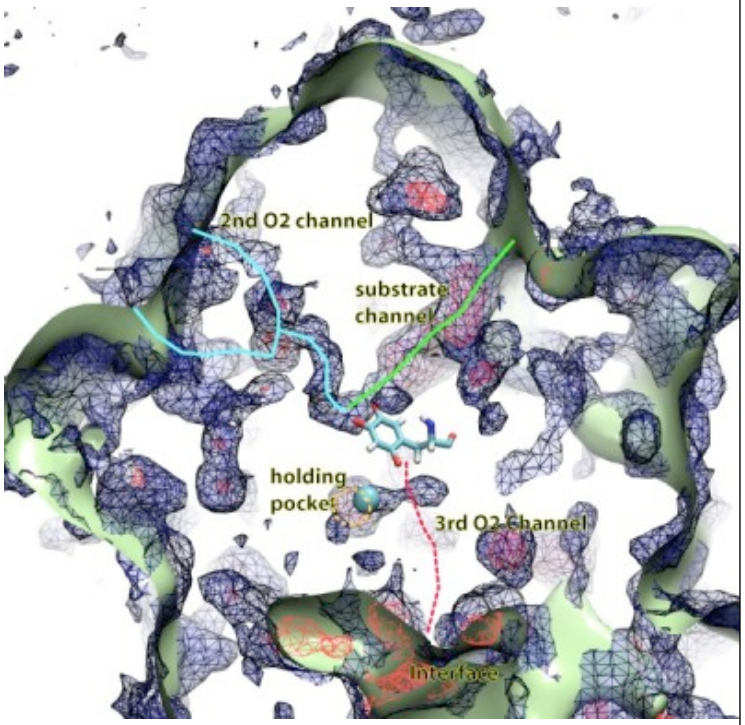
# CUDA+OpenCL Acceleration in VMD



Electrostatic field calculation,
Multilevel Summation Method
20x to 44x faster

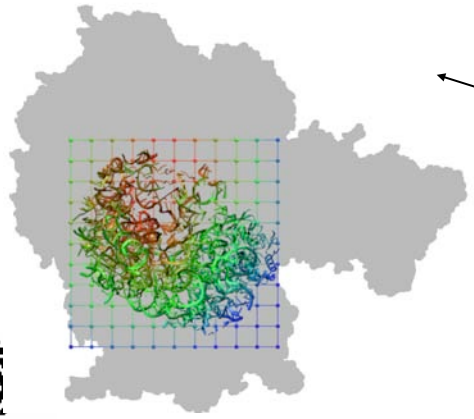


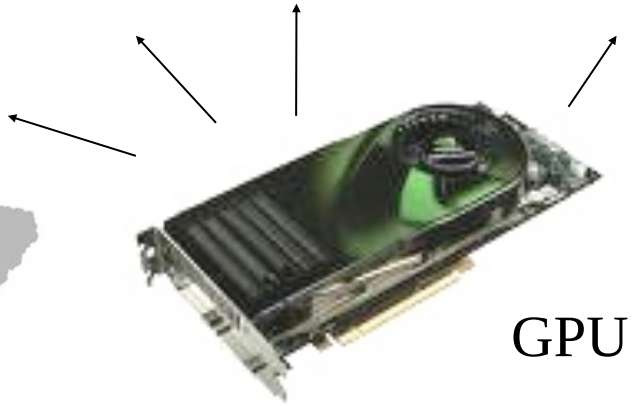
Molecular orbital
calculation and display
100x to 120x faster




Imaging of gas migration pathways in proteins with Implicit Ligand Sampling (ILS) algorithm

20x to 30x faster



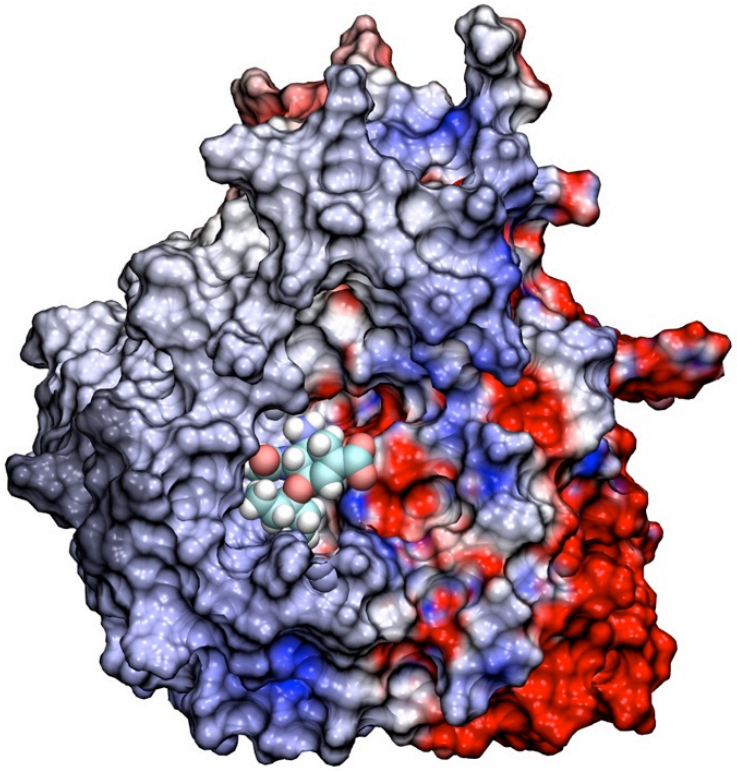


GPU

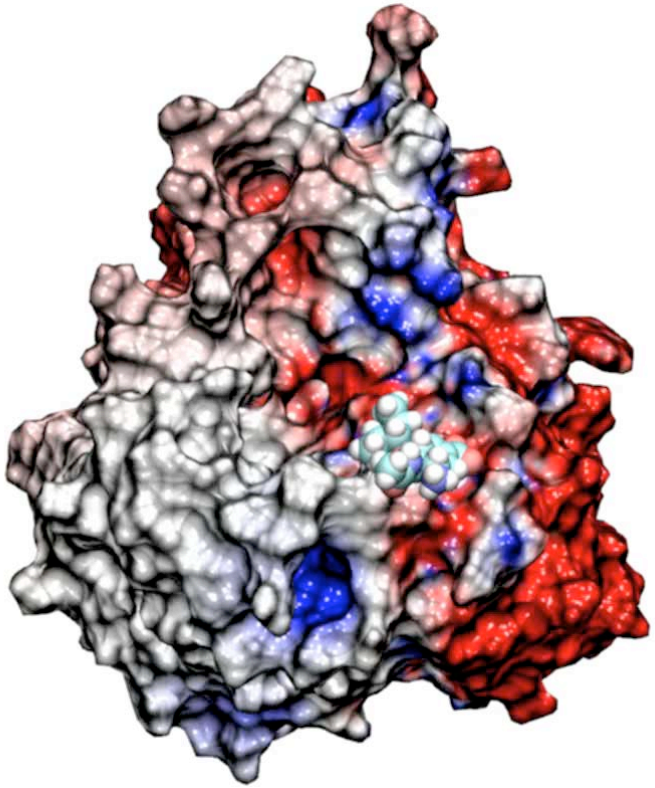
National Center for Research Resources

# Swine Flu Neuraminidase Electrostatics

## Mean electrostatic field needed to identify drug binding pathway



Time-averaged electrostatic field of H1N1 neuraminidase calculated from VMD Multilevel Summation Tool:
**Compution with NVIDIA Tesla C1060 is 20x faster than computation on a single CPU core**

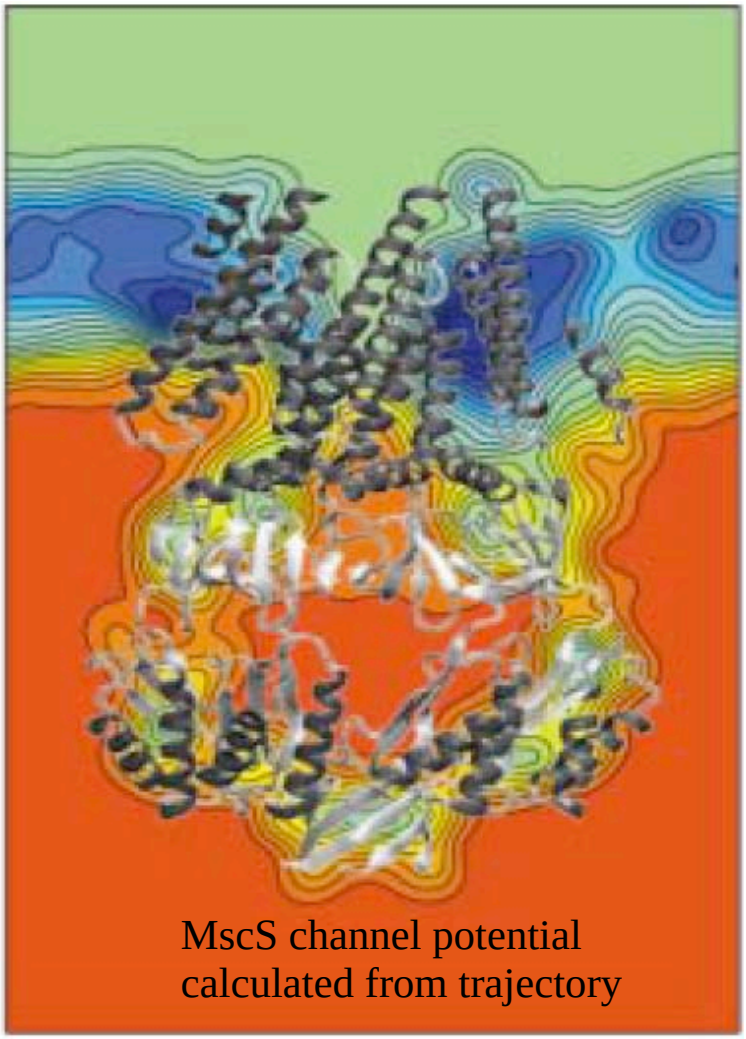


Movie of drug (tamiflu) binding

# Electrostatic Potential Maps

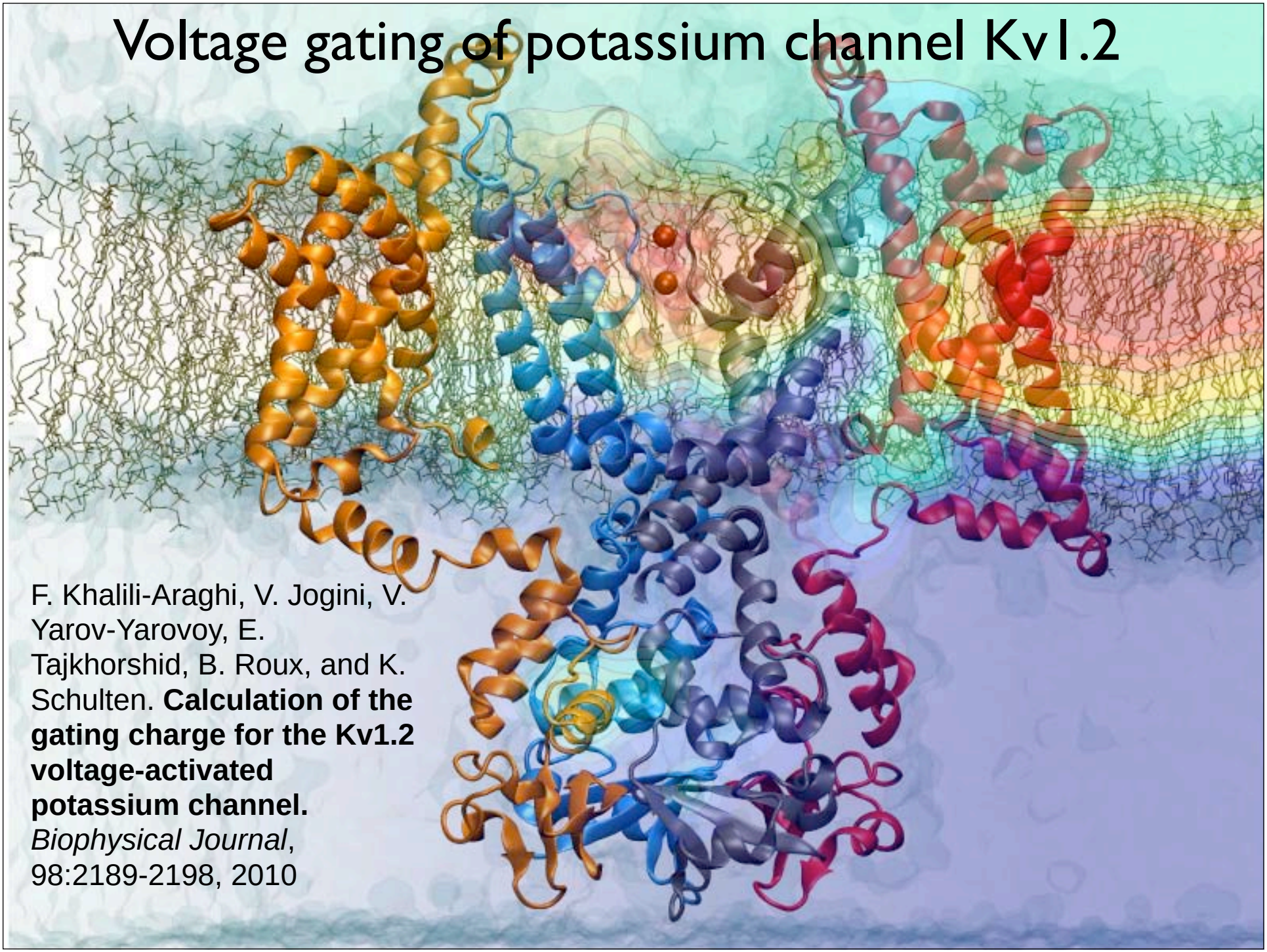
New VMD features made possible through GPU computing

- Electrostatic potentials evaluated on 3-D lattice
- Applications include:
  - Ion placement for structure building
  - Time-averaged potentials for simulation
  - Visualization and analysis



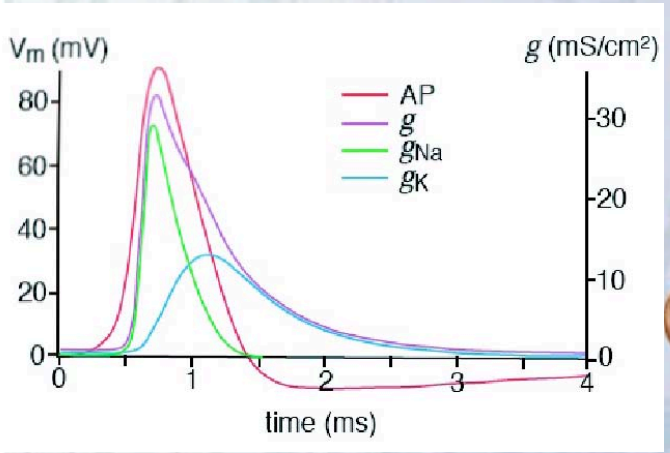
MscS channel potential calculated from trajectory

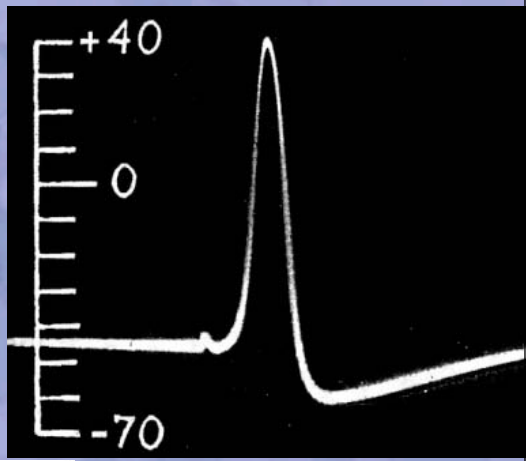
# Voltage gating of potassium channel Kv1.2



F. Khalili-Araghi, V. Jogini, V. Yarov-Yarovoy, E. Tajkhorshid, B. Roux, and K. Schulten. **Calculation of the gating charge for the Kv1.2 voltage-activated potassium channel.** *Biophysical Journal*, 98:2189-2198, 2010

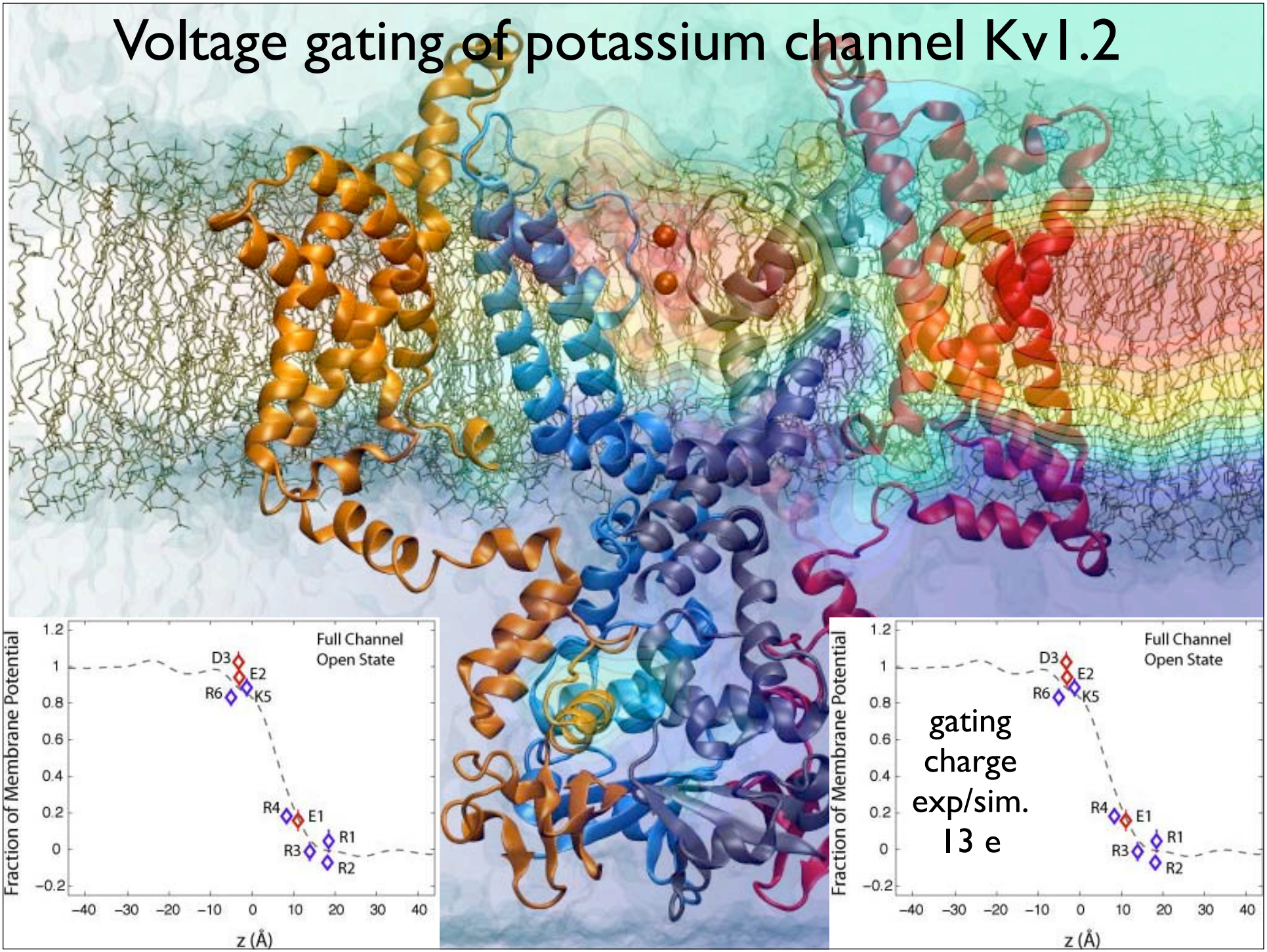
# Voltage gating of potassium channel Kv1.2



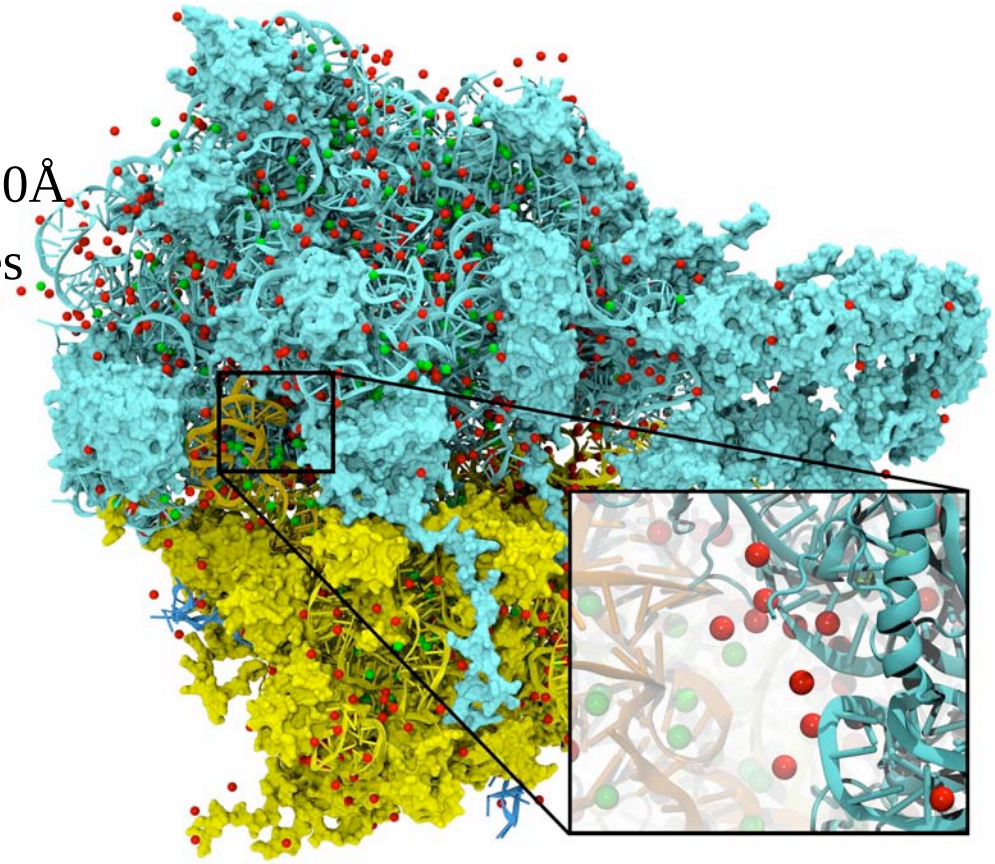

$$I = \overline{g_{Na}} m^3 h (V - E_{Na}) + \overline{g_K} n^4 (V - E_K) + g_{leak} (V - E_{leak})$$

# Voltage gating of potassium channel Kv1.2



# Time-averaged Electrostatic Potential Calculation for the Ribosome with VMD

- Direct Coulomb summation ~580,000 atoms
  - Lattice spacing 1.0Å, padding 10Å
  - Time-average from 1,000 frames
- 3 GPUs: 49 hours
- 3 CPUs: 0.23 years (est.)

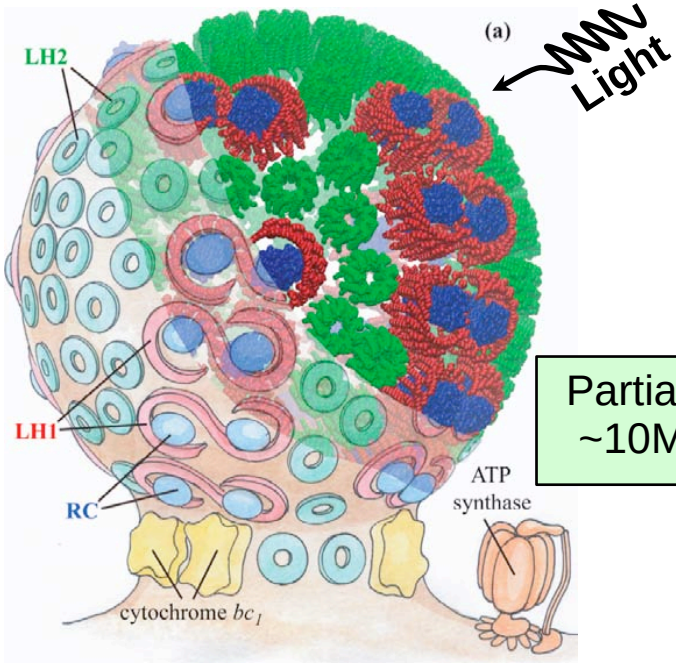


**This was one of our early results, using the multi-GPU direct Coulomb summation algorithm, showing the benefit it gave at the time. Now that we have MSM (multilevel summation) we would get much faster performance since it is a linear-time algorithm, but we haven't yet re-run these tests using MSM.**

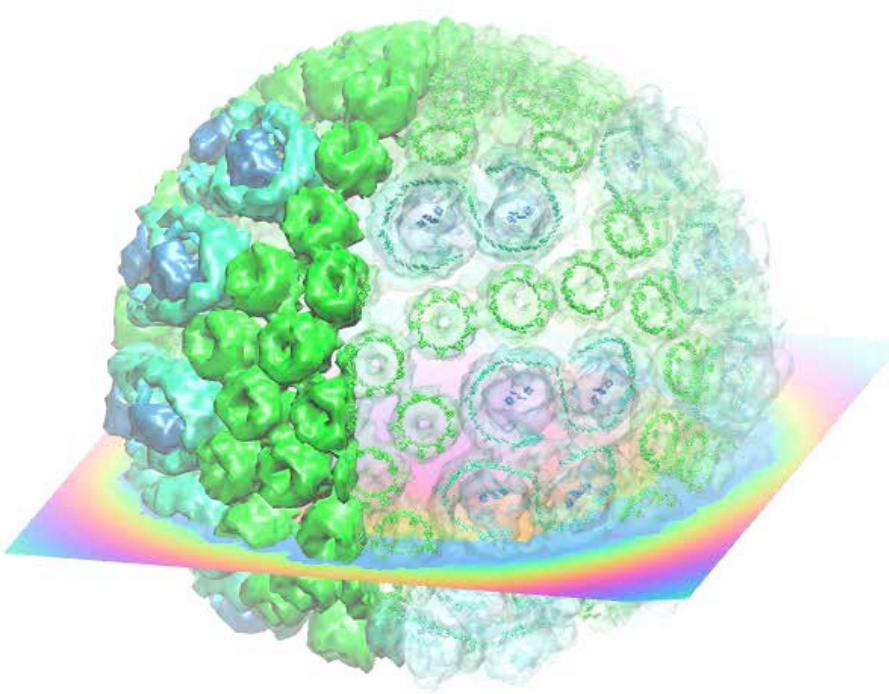
Stone et al. (2007) *J Comp Chem* 28:2618-2640

# Photobiology of Vision and Photosynthesis

## Investigations of the chromatophore, a photosynthetic organelle




Partial model:
~10M atoms



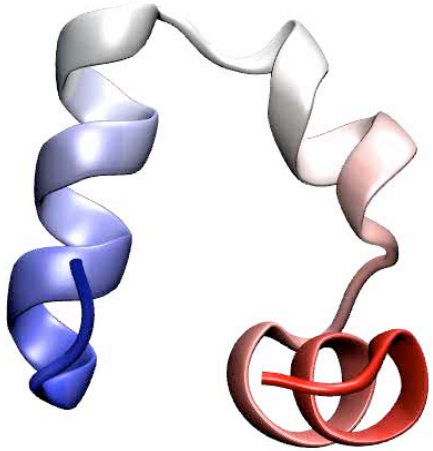
Electrostatics needed to build full structural model, place ions, study macroscopic properties

Electrostatic field of chromatophore model from multilevel summation method: computed with 3 GPUs in ~90 seconds, 46x faster than single CPU core

**Full chromatophore model will permit structural, chemical and kinetic investigations at a structural systems biology level**

# Timeline Tool: identify events in an MD trajectory

- We have MD trajectories:

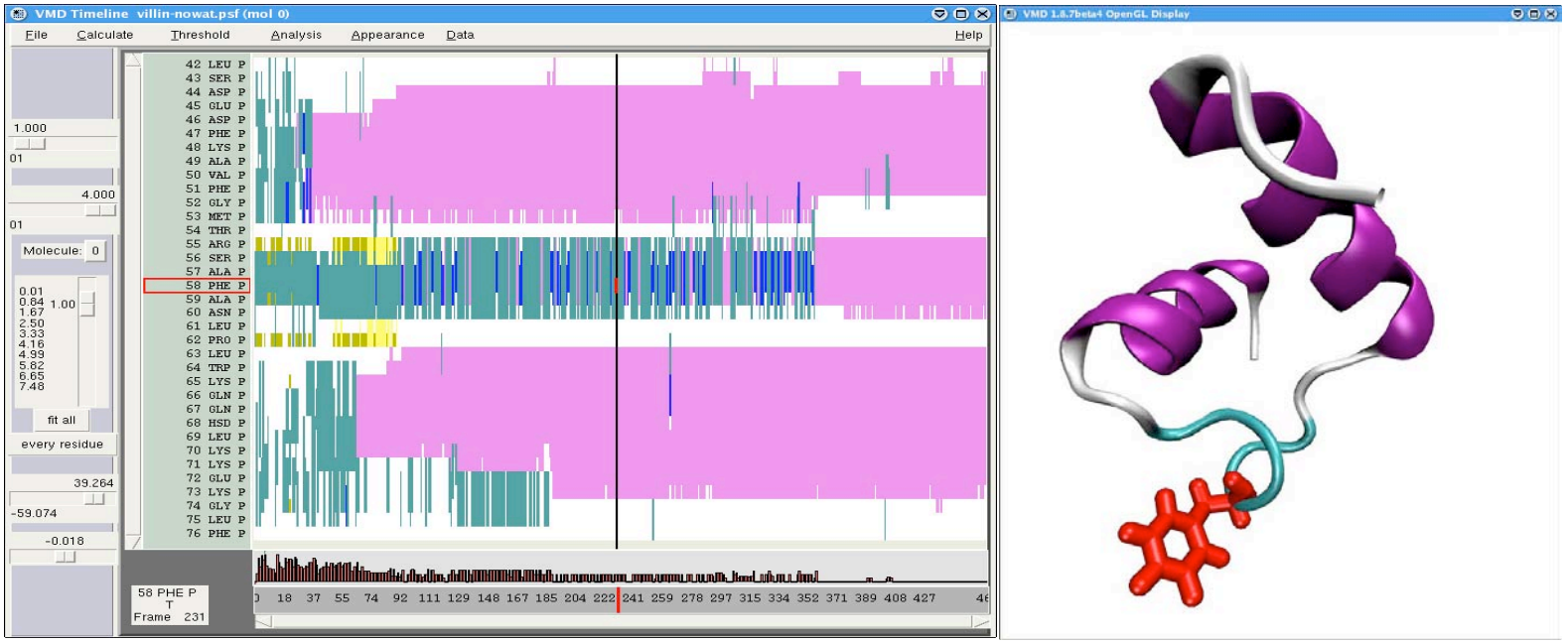


- We want to identify events in a trajectory:
  - 7.1 μs, 600 GB of trajectory data
  - events: *0.5 μs, helix 3 forms; 3.0 μs, helix 1 forms; etc.*
  - How long would it take an expert user to visually inspect this trajectory to find motional changes of events?

    *2 days!* (*plus:* tiring task; one is liable to miss much)

# Timeline: a graphing and analysis tool to identify events in an MD trajectory

Events during 7 μs villin headpiece folding




Alpha helix  Extended beta  Isolated bridge  3-10 helix  Beta turn  None (coil)

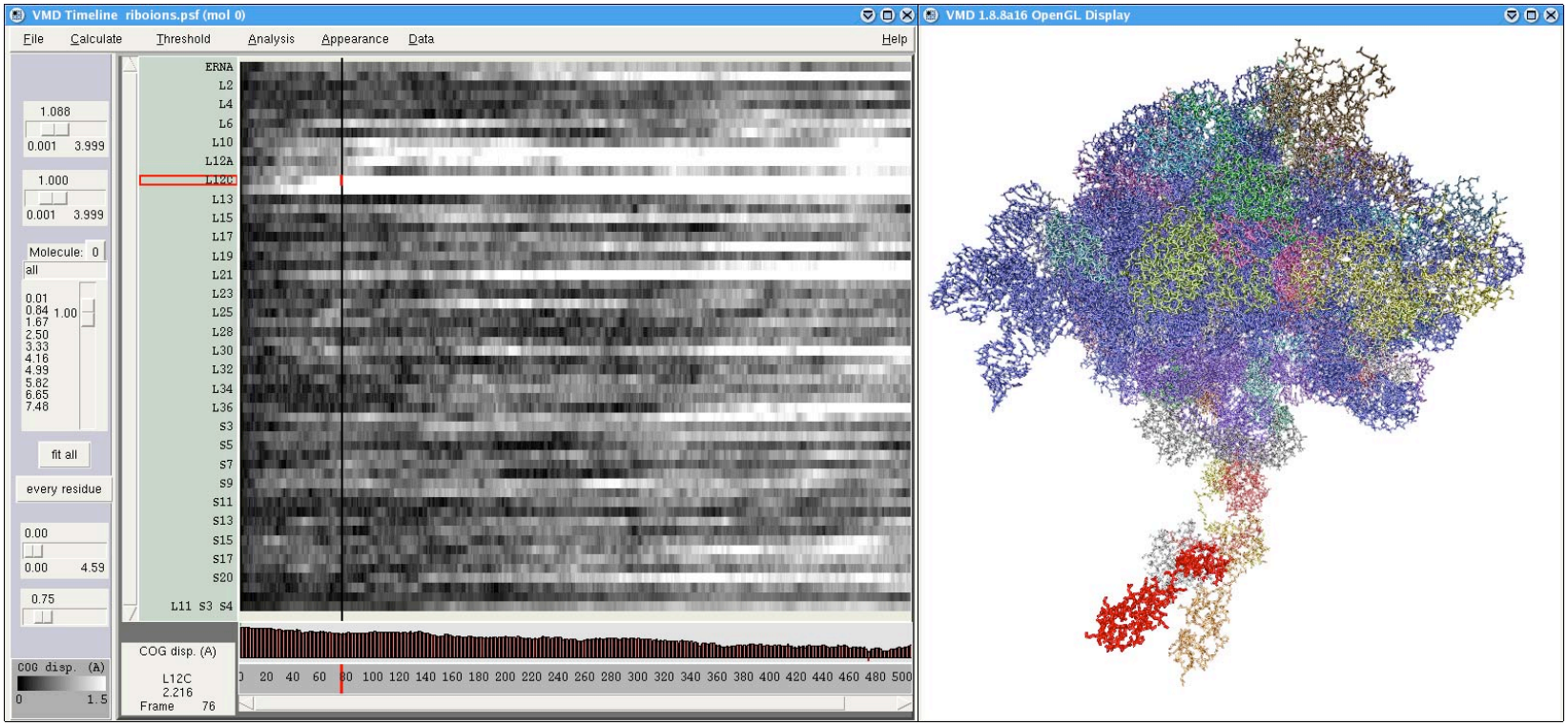
Per-residue secondary structure: villin headpiece folding from a fully denatured state.

7μs simulation; 654 atoms; over 1 million frames to examine

VMD **Timeline plug-in**: live 2D plot linked to 3D structure

- a single picture shows changing properties across entire structure, entire trajectory.
- explore time vs. attribute (per-residue or per-selection) linked to molecular structure
- many analysis methods available; user-extendable

# Timeline and large structures: events during ribosome equilibration



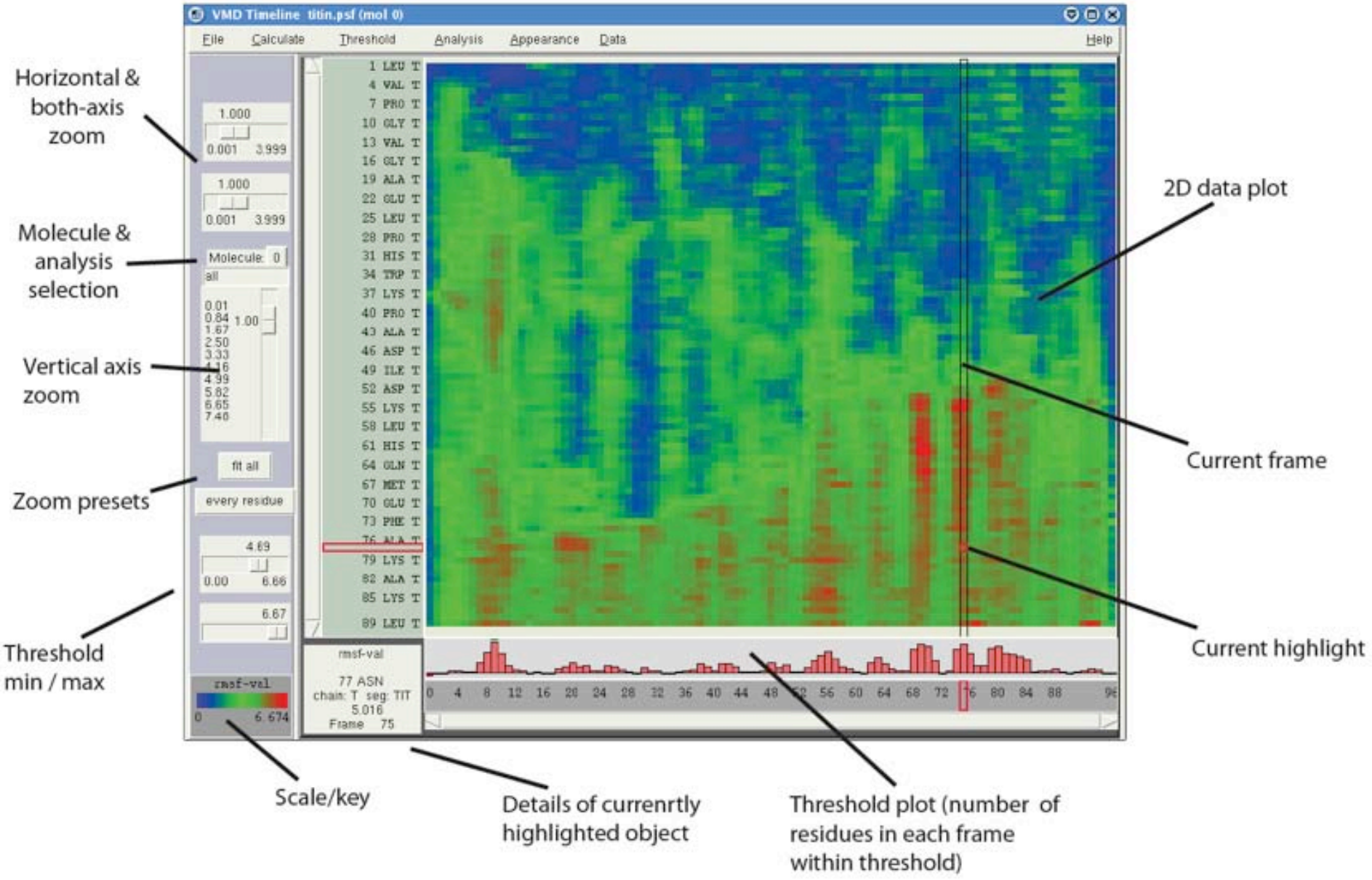

Ribosome equilibration, 17,000+ protein/nucleic residues + ions

*Example analysis*: displacement (Å) of center-of-geometry of each component protein

(calculation here is per-component-protein, not per-indiviudal-residue)

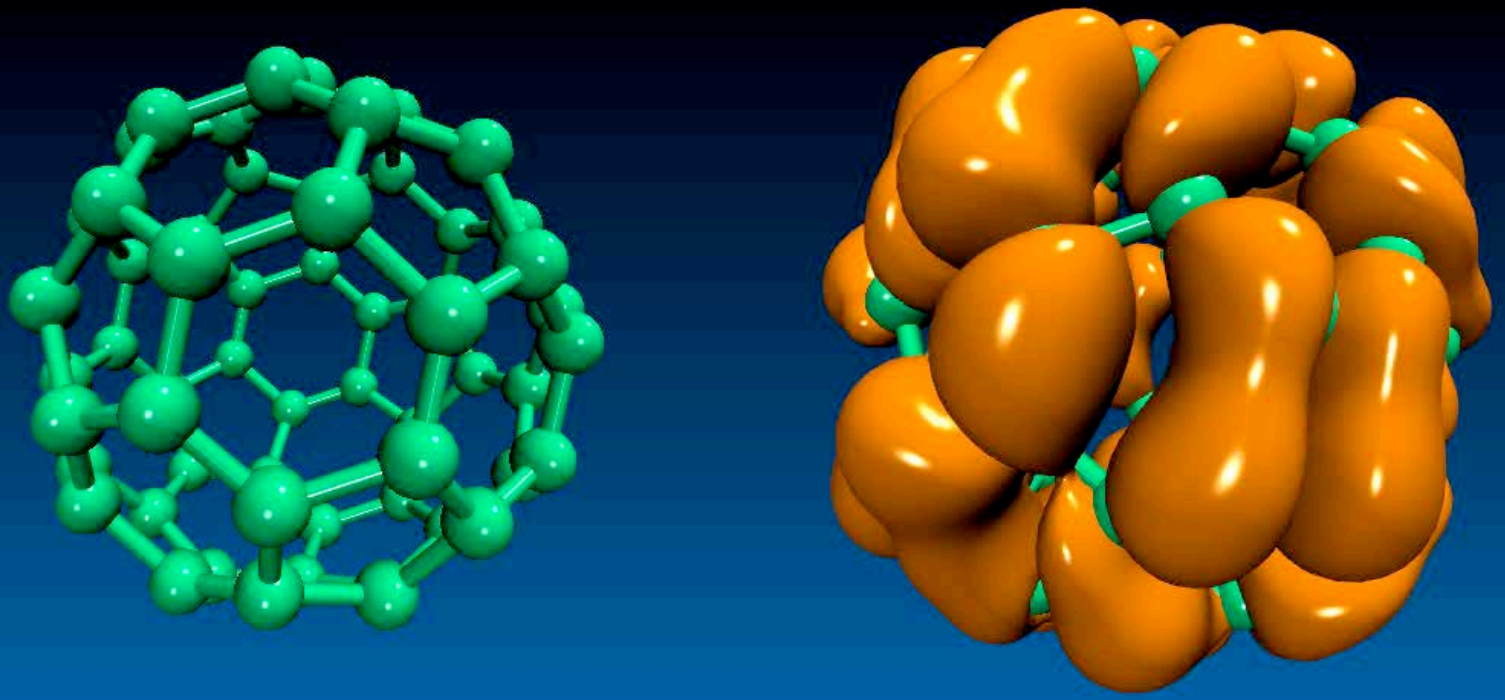
*Finding:* peripheral proteins show greater displacement than core proteins

# Main Interface Features of Timeline 2D Data Plot



# Computing Molecular Orbitals on GPUs

Quantum chemical calculation: Ivan Ufimtsev and Todd Martinez, Stanford U.



**The future is here: Classical and quantum mechanical simulations**

# Acknowledgements


## VMD team

*J. Stone (leader)*
*D. Hardy*
*B. Isralewitz*
*J. Saam*
*K. Vandivoort*

Center for the Physics of Living Cells

National Center for Research Resources

Funding: NIH, NSF