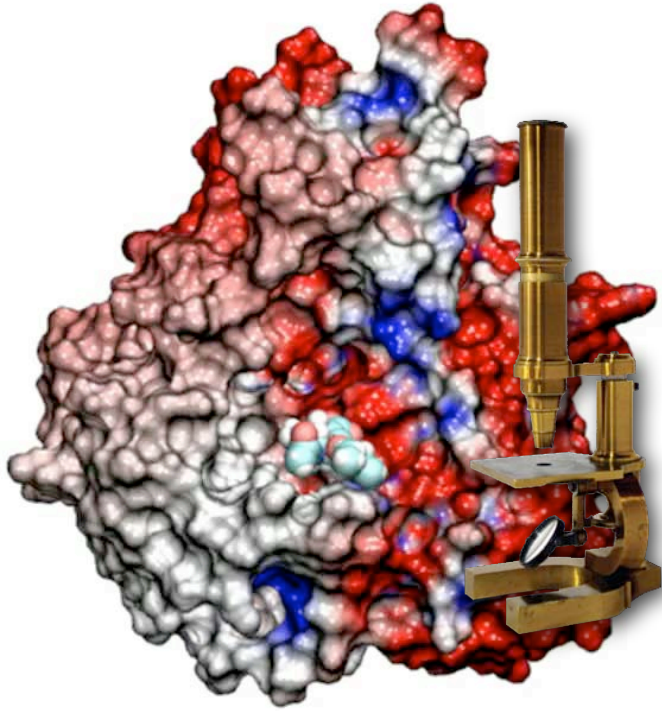# Discoveries Through the Computational Microscope

**Accuracy • Speed-up • Unprecedented Scale**



Investigation of drug (Tamiflu) resistance of the "swine" flu virus demanded **fast response!**
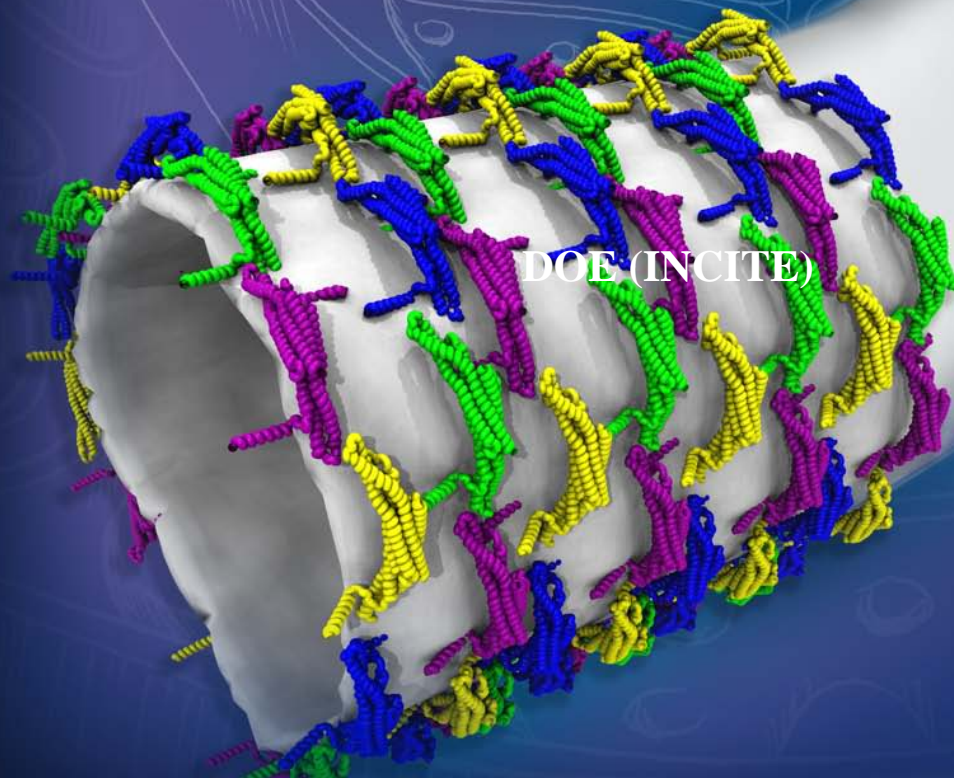
Klaus Schulten

Department of Physics and

Theoretical and Computational Biophysics Group

University of Illinois at Urbana-Champaign

The Computational Microscope

100 - 1,000,000 processors

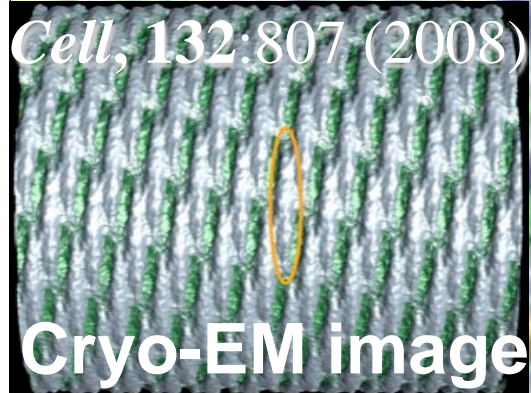Viewing the Morphogenesis of a Cellular Membrane from Flat to Tubular in 200 µs

DOE (INCITE)

# Viewing the Morphogenesis of a Cellular Membrane from Flat to Tubular in 200 µs

0 ns

CPC-D-10-00292



Cryo-EM image

Cell, 132:807 (2008)

A. Arkhipov, Y. Yin, and K. Schulten. **Four-scale description of membrane sculpting by BAR domains.** *Biophysical J.*, 95: 2806-2821 2008.
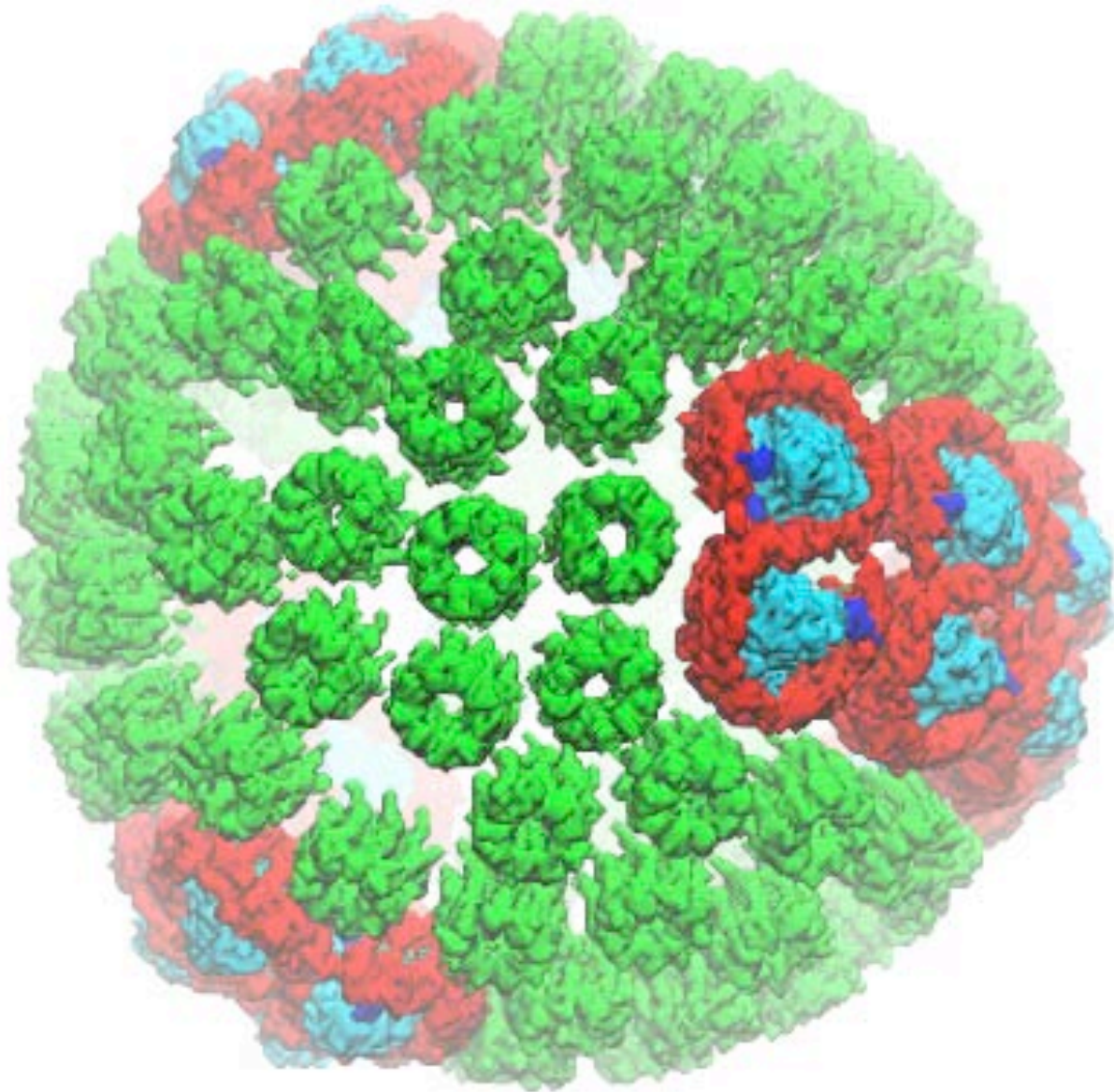
Ying Yin, Anton Arkhipov, and Klaus Schulten. **Simulations of membrane tubulation by lattices of amphiphysin N-BAR domains.** *Structure* **17,** *882-892,* 2009.
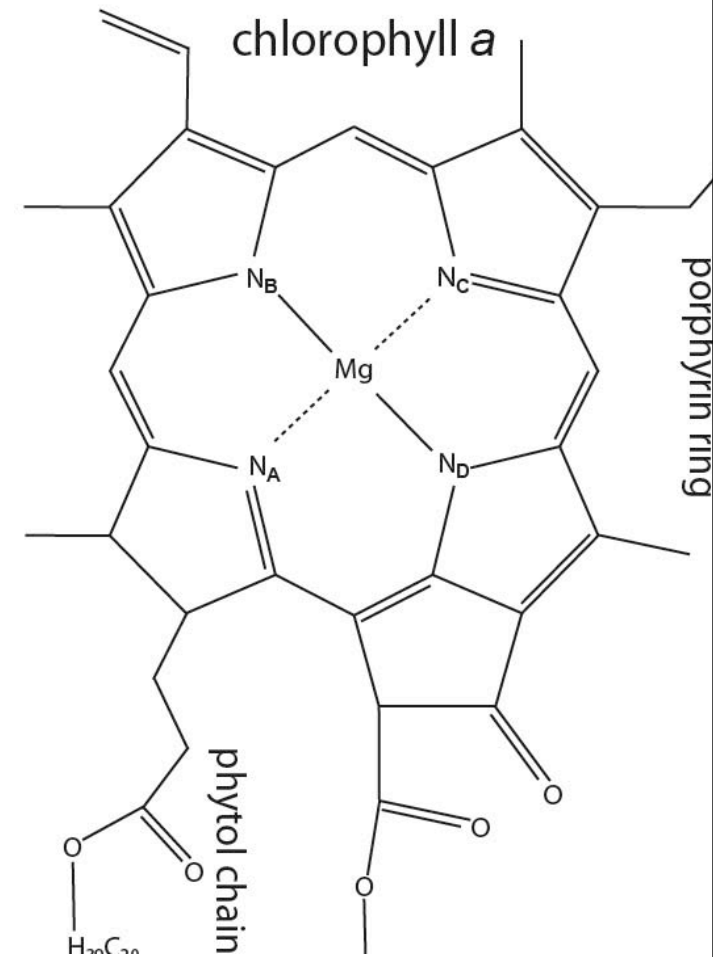
# Viewing How Nature Harvests Sun Light

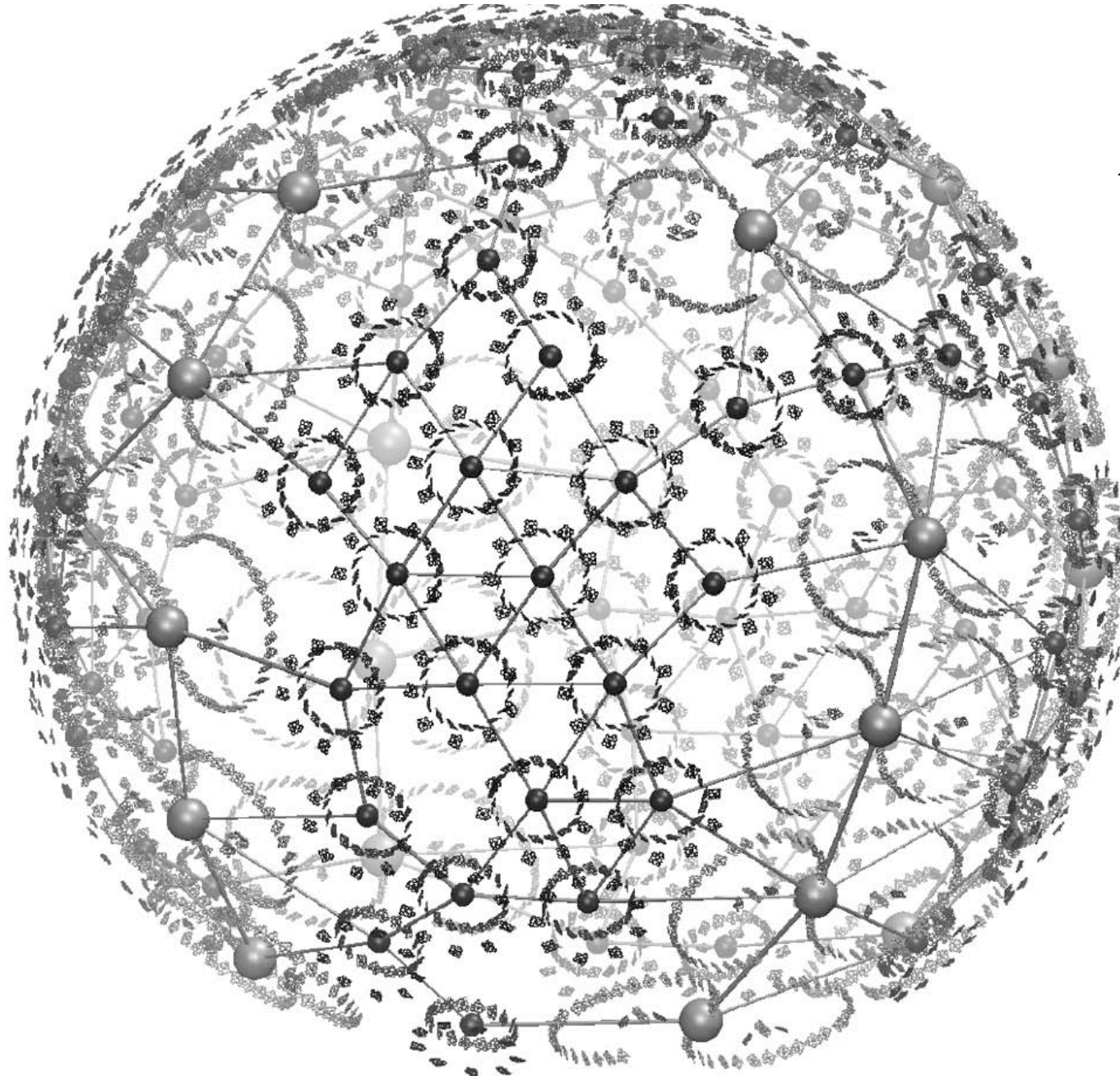## 95% of the energy in the biosphere comes from this energy source



LIGHT

20

200

ADP

ATP

Purple Photosynthetic Bacterium

Chromatophore (700 Å)

6

1
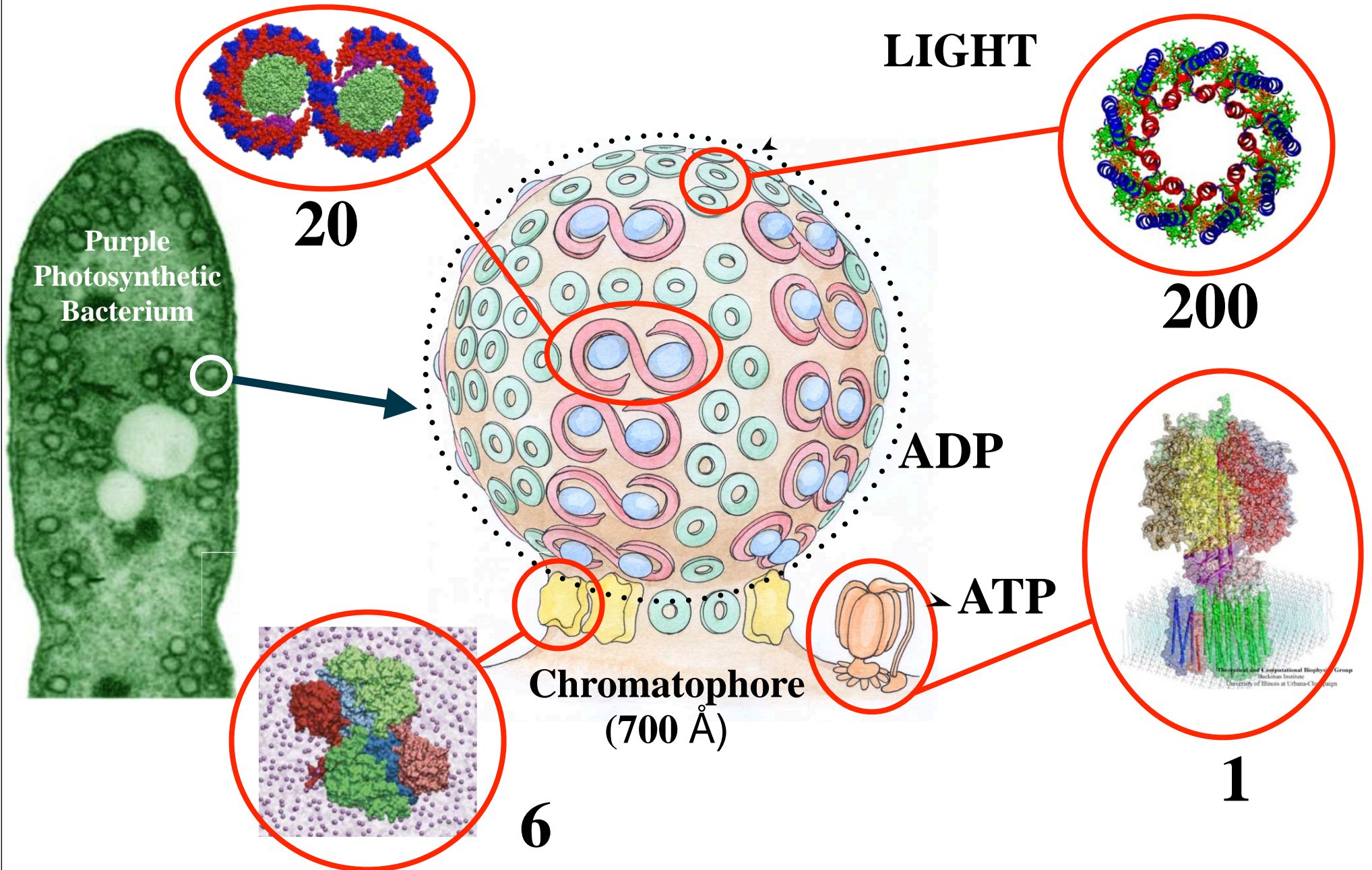
# Photosynthetic Chromatophore of Purple Bacteria
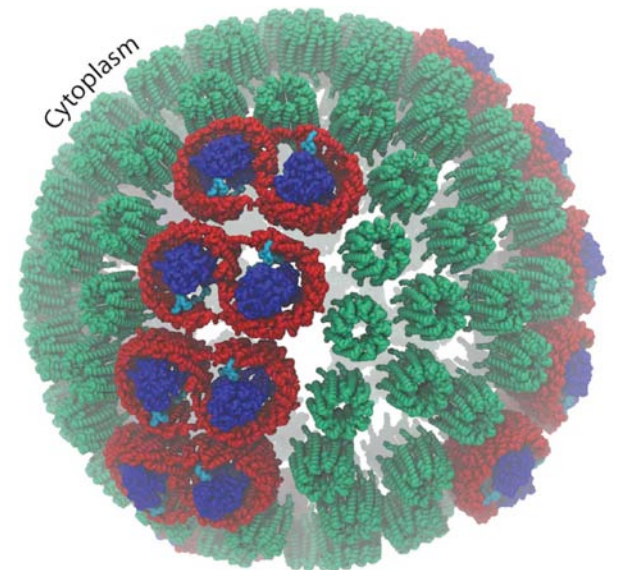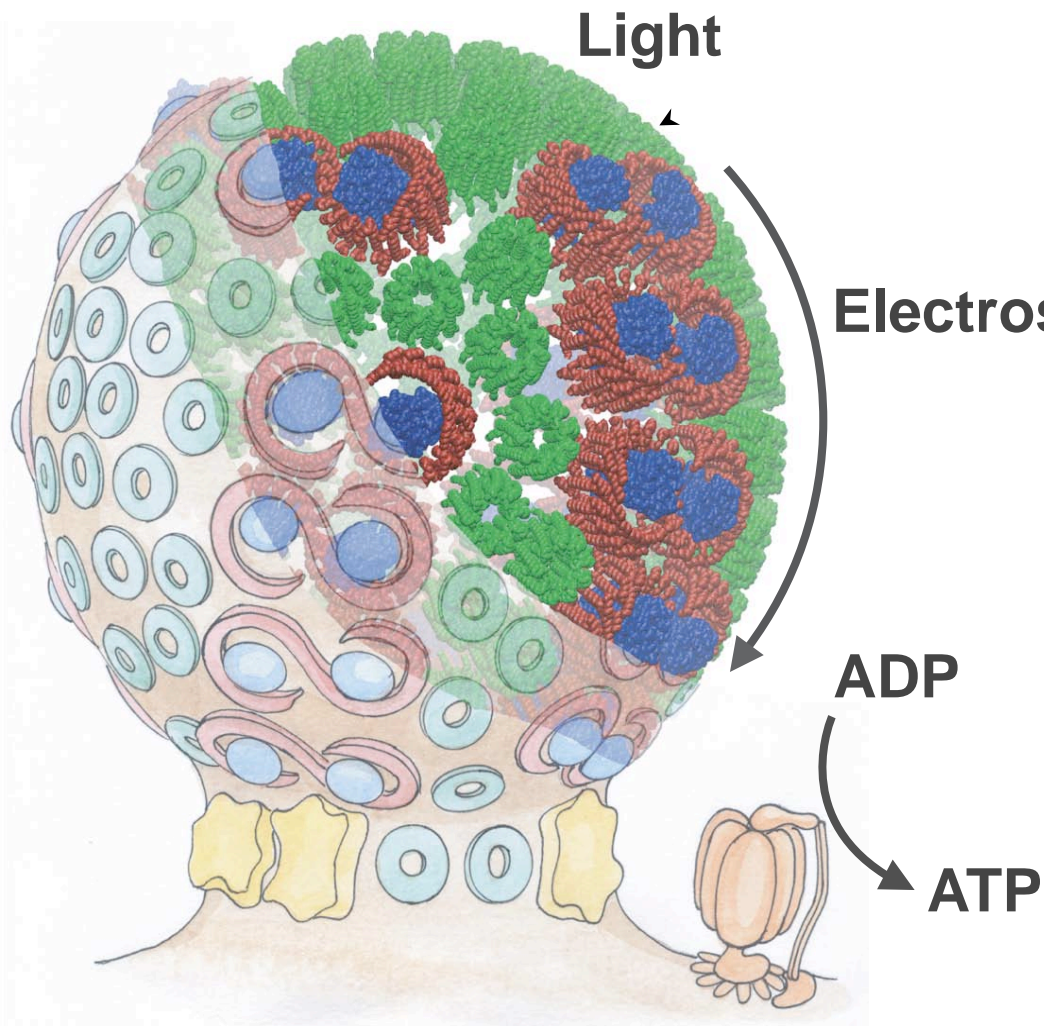


100 Å

chlorophyll *a*

porphyrin ring

phytol chain

$$k_{D^*A \to DA^*} = \frac{2\pi}{\hbar} |\mathbf{d}_D|^2 |\mathbf{d}_A|^2 \frac{\left[ \left( \hat{\mathbf{d}}_D \cdot \hat{\mathbf{d}}_A \right) - 3 \left( \hat{\mathbf{d}}_D \cdot \hat{R} \right) \left( \hat{\mathbf{d}}_A \cdot \hat{R} \right) \right]^2}{R^6} \int_{-\infty}^{+\infty} dE \, S_D(E) \, S_A(E)$$
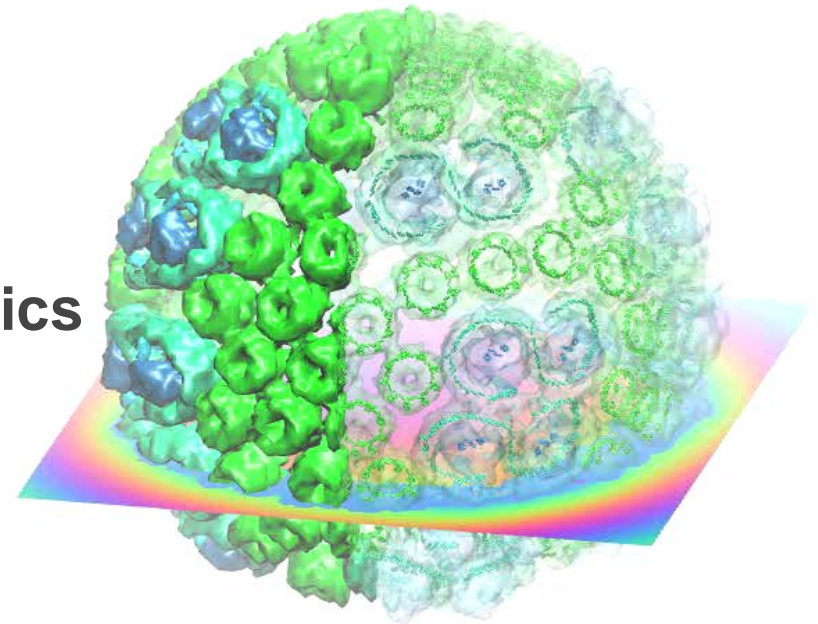
Excitation transfer through fluorescent resonant energy transfer (FRET) in photosynthetic light harvesting

# Viewing How Nature Harvests Sun Light



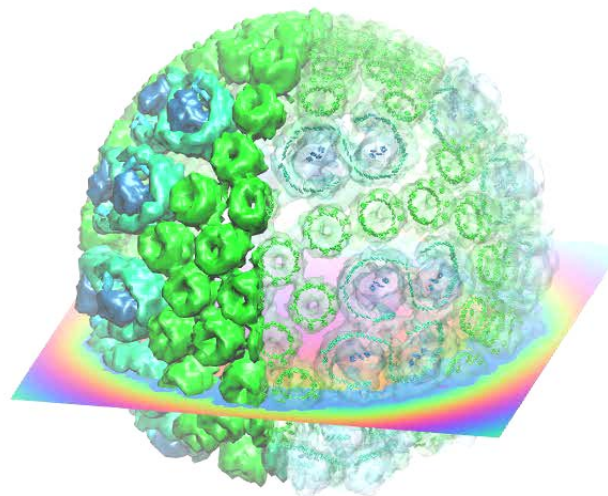Light

Electrostatics

ADP

ATP

~10M atoms
Electrostatic field calculated
with multilevel summation method

1 CPU core: 1 hr 10 min
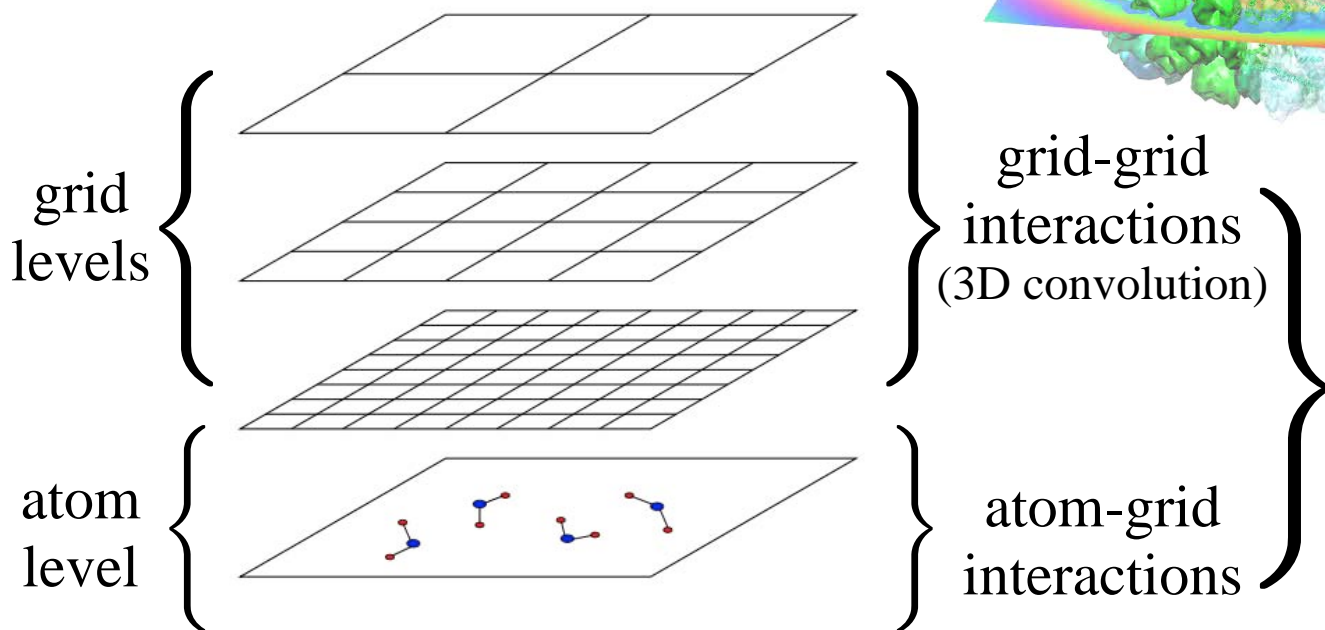3 GPUs (G80):  ~90 seconds

# GPU Solution: Multilevel Summation Method for Electrostatics on the GPU

Localized grid operations map well to GPUs

GPU speedups:

26x on 1 GPU,
46x on 3 GPUs

grid levels

grid-grid interactions
(3D convolution)

atom level
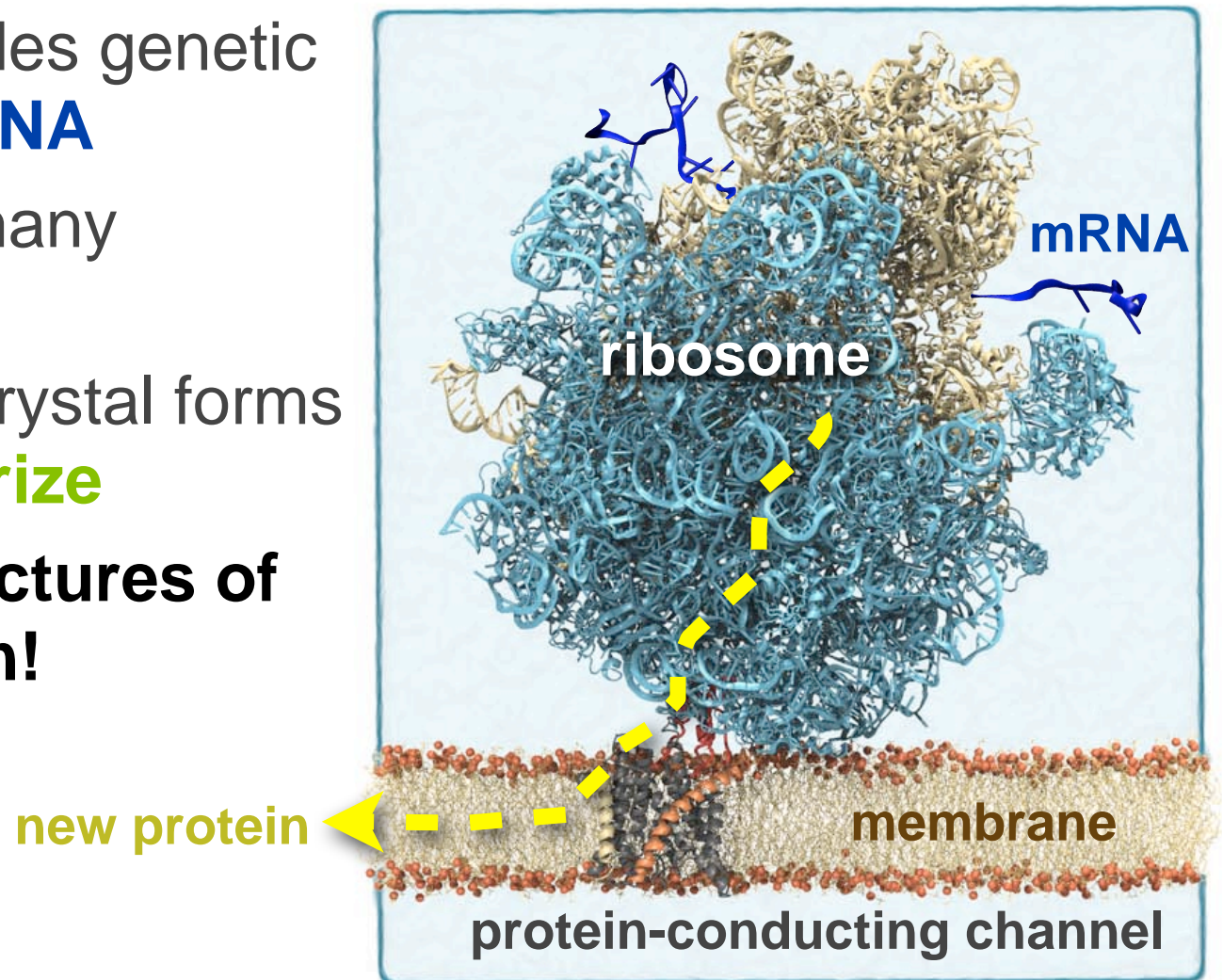
atom-grid interactions

**10 million atoms:**

**1 CPU core: 1 hr 10 min**
**3 GPUs (G80): 90 sec**

**Multilevel summation method** has linear time complexity well suited for GPUs; more flexible than other methods

# Viewing How Proteins are Made from Genetic Blueprint

- **Ribosome** — Decodes genetic information from **mRNA**

- Important target of many **antibiotics**

- Static structures of crystal forms led to 2009 **Nobel Prize**

- **But one needs structures of ribosomes in action!**



mRNA

ribosome

new protein

membrane
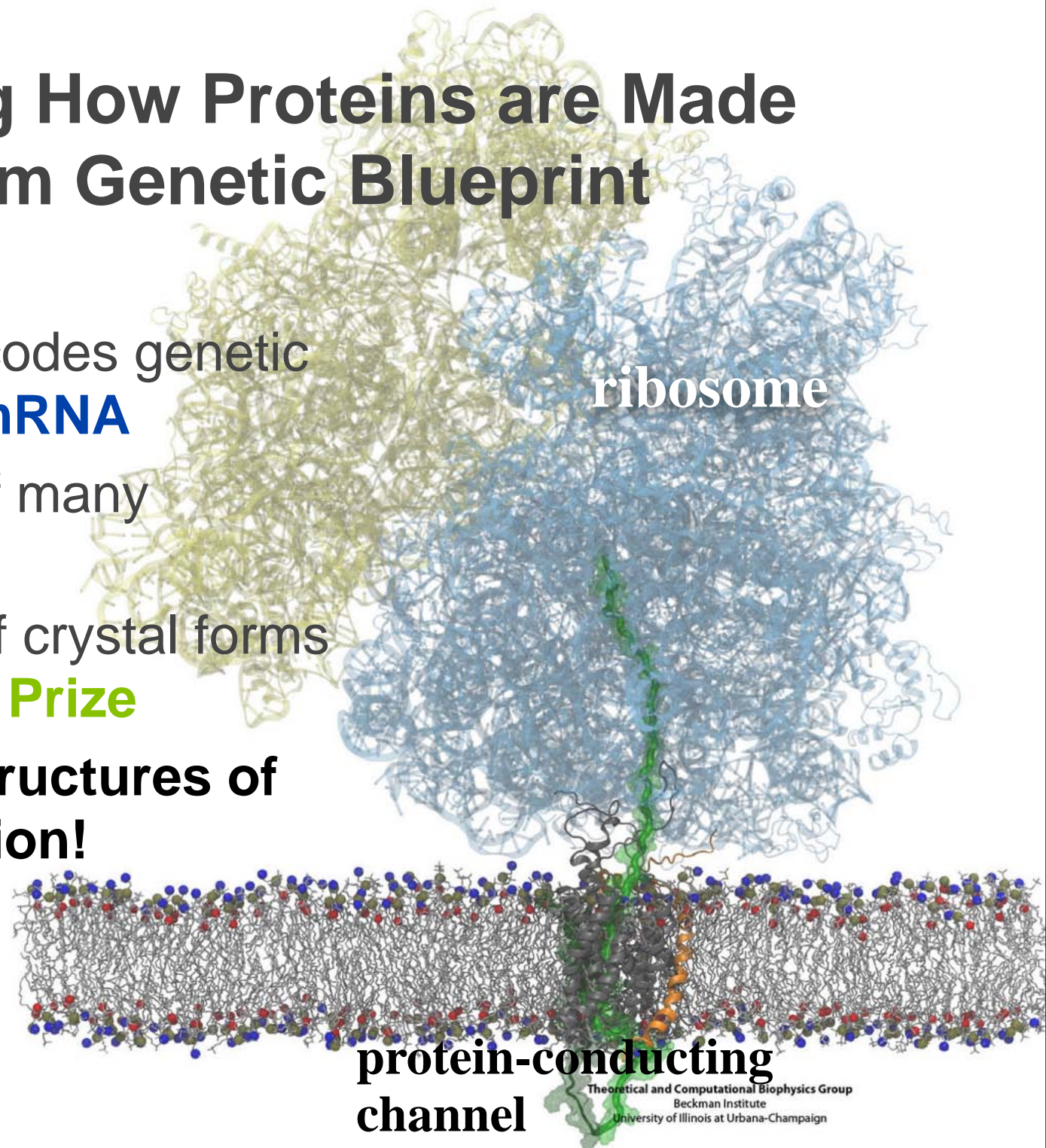
protein-conducting channel

# Viewing How Proteins are Made from Genetic Blueprint

- **Ribosome** — Decodes genetic information from **mRNA**

- Important target of many **antibiotics**

- Static structures of crystal forms led to 2009 **Nobel Prize**

- **But one needs structures of ribosomes in action!**

ribosome

protein-conducting channel

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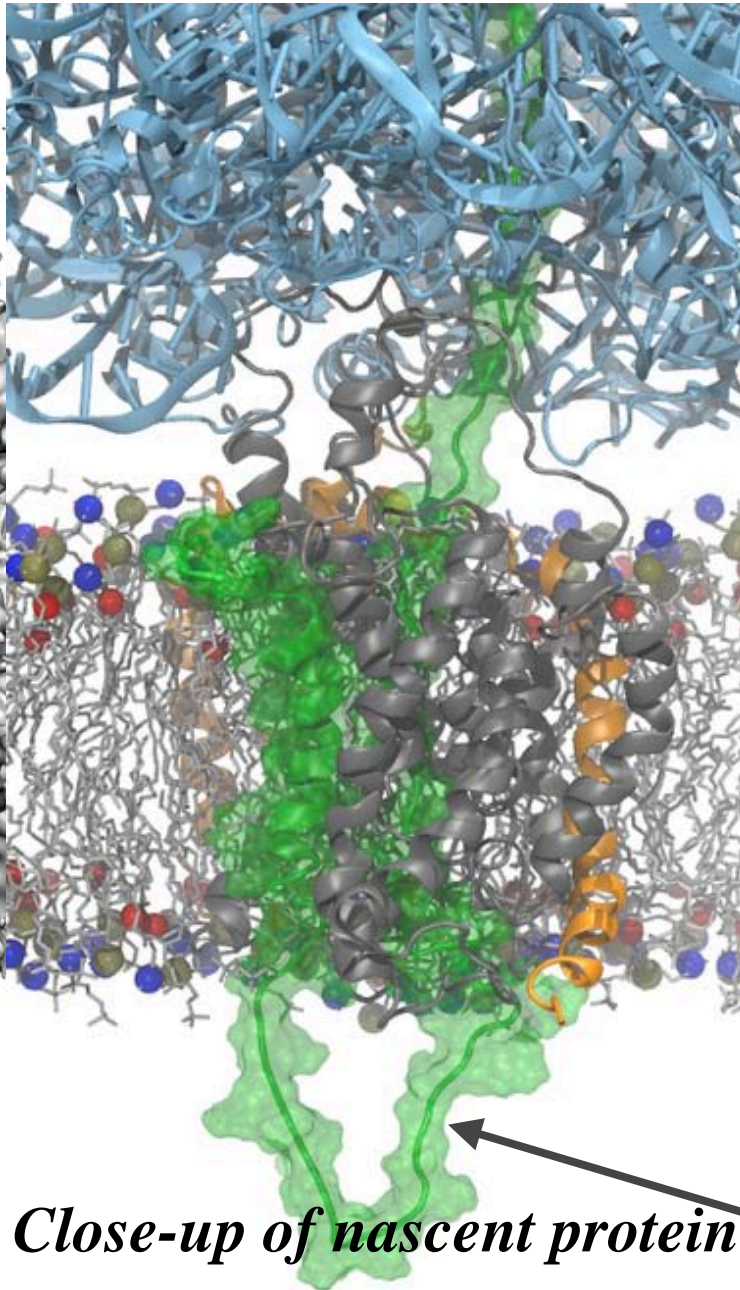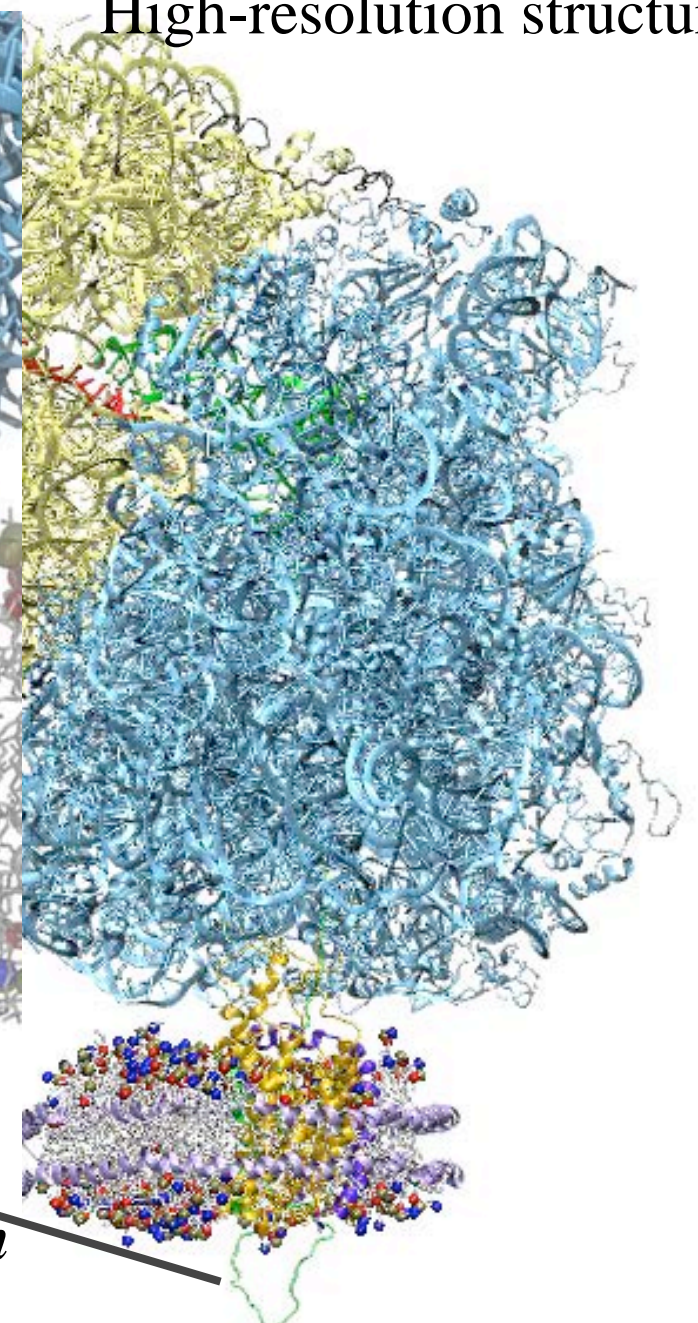University of Illinois at Urbana-Champaign

# Viewing How Proteins Are Made from Genetic Blueprint
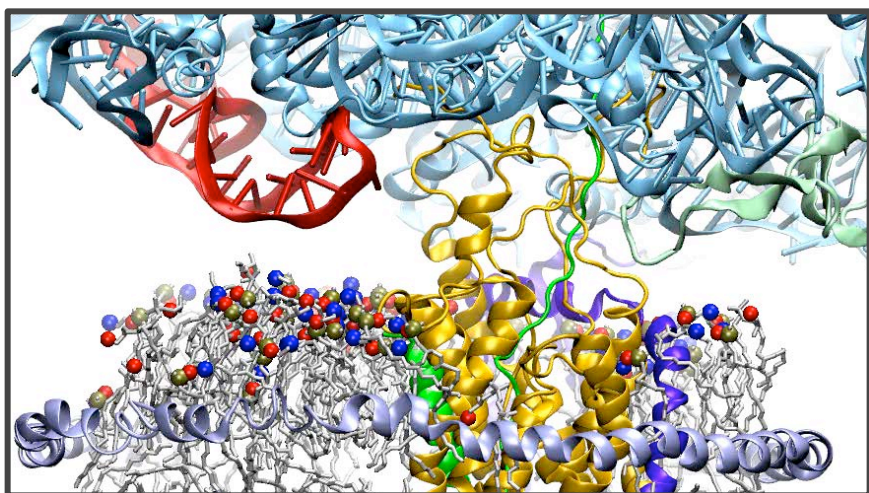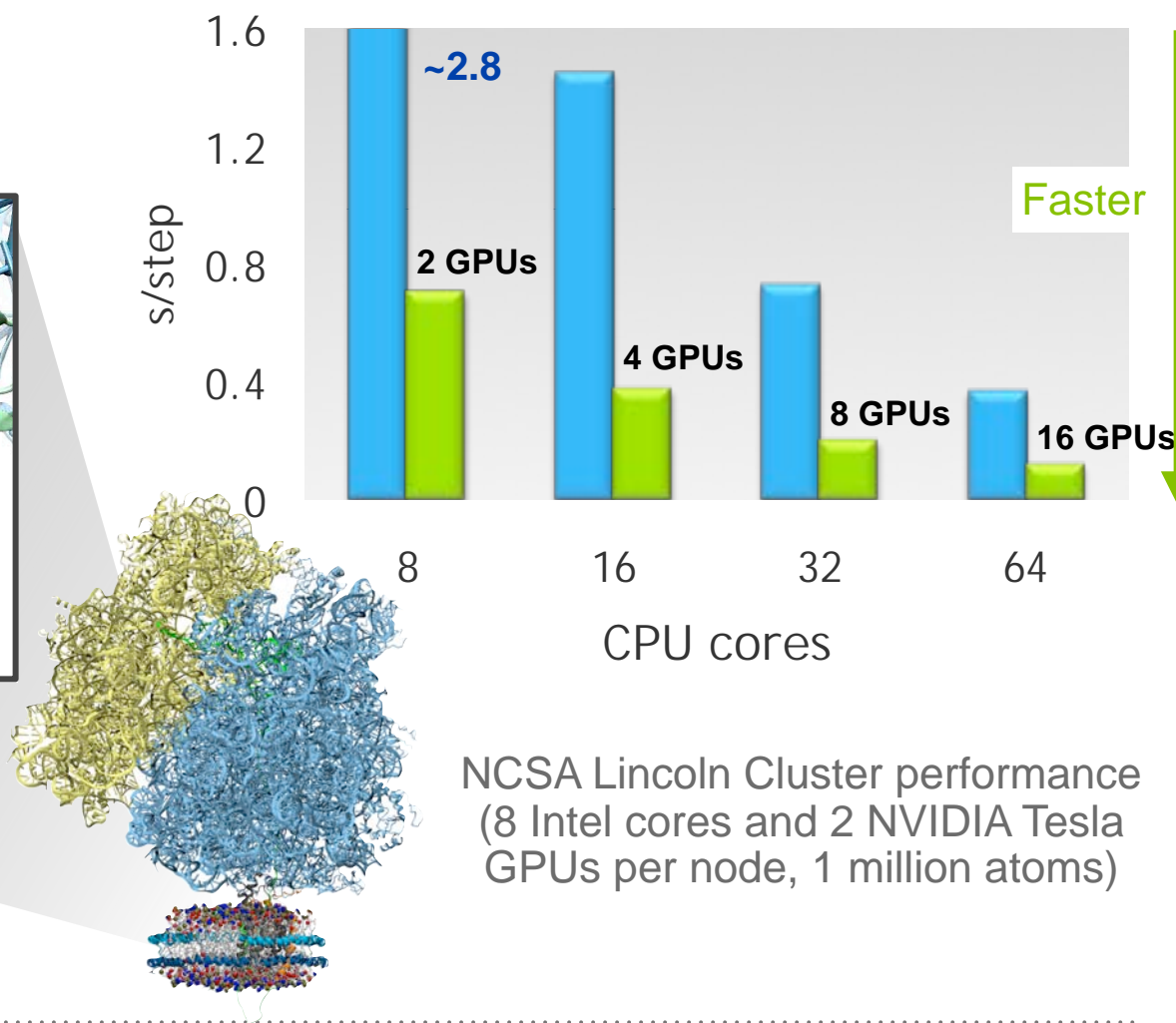
Low-resolution data

High-resolution structure



*Close-up of nascent protein*

# GPU Solution 3: Molecular Dynamics Simulations



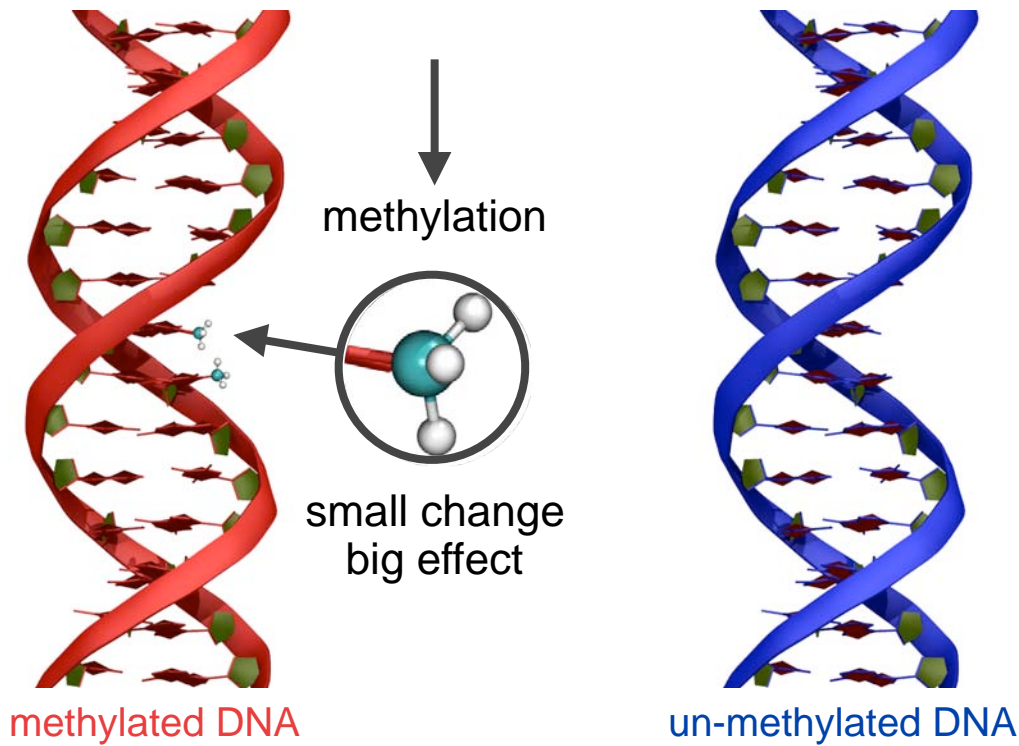Molecular dynamics simulation of protein insertion process

NCSA Lincoln Cluster performance (8 Intel cores and 2 NVIDIA Tesla GPUs per node, 1 million atoms)

GPUs reduced time for simulation from **two months to two weeks!**
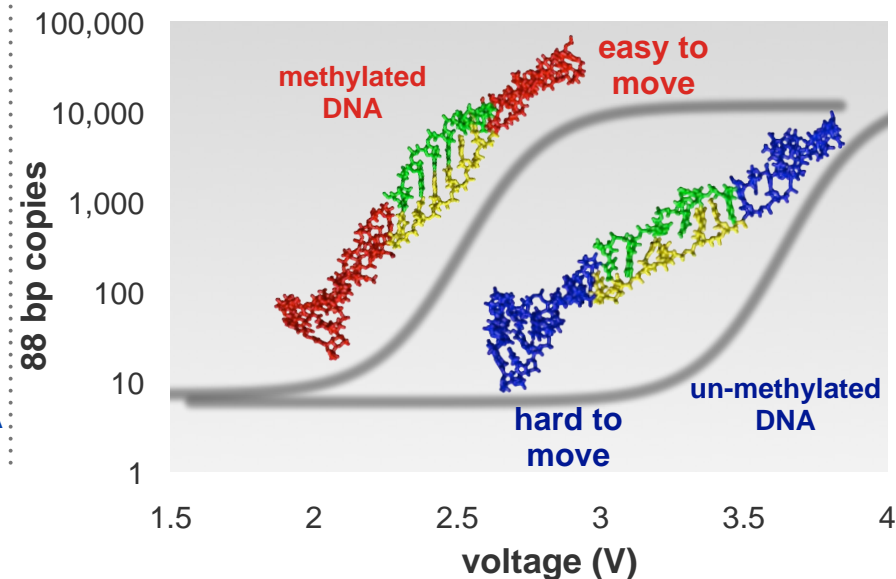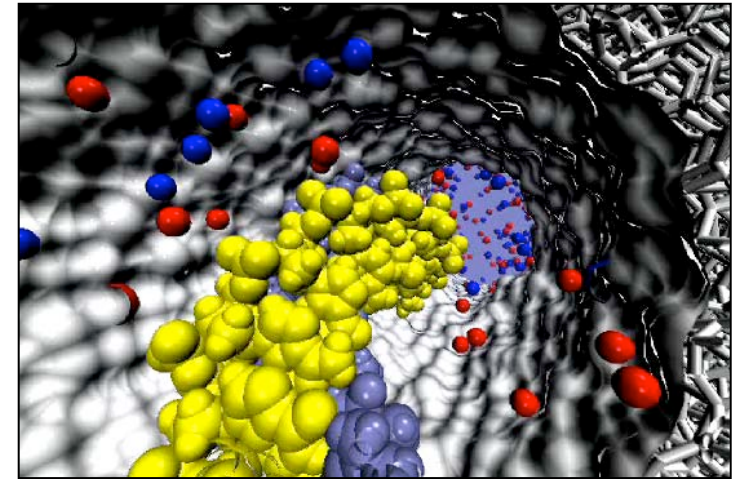
# Viewing Nanopore Sensors

**Genetics: Genes control our bodies and experiences!**
**Epigenetics: Our bodies and experiences control the genes!**

Epigenetics made possible through DNA methylation

methylation

small change
big effect

methylated DNA

un-methylated DNA

**Related pathologies: obesity, depression, cancer**

Detect methylation with **nanopores**

methylated DNA — easy to move

hard to move — un-methylated DNA

88 bp copies

100,000
10,000
1,000
100
10
1

voltage (V)
1.5    2    2.5    3    3.5    4
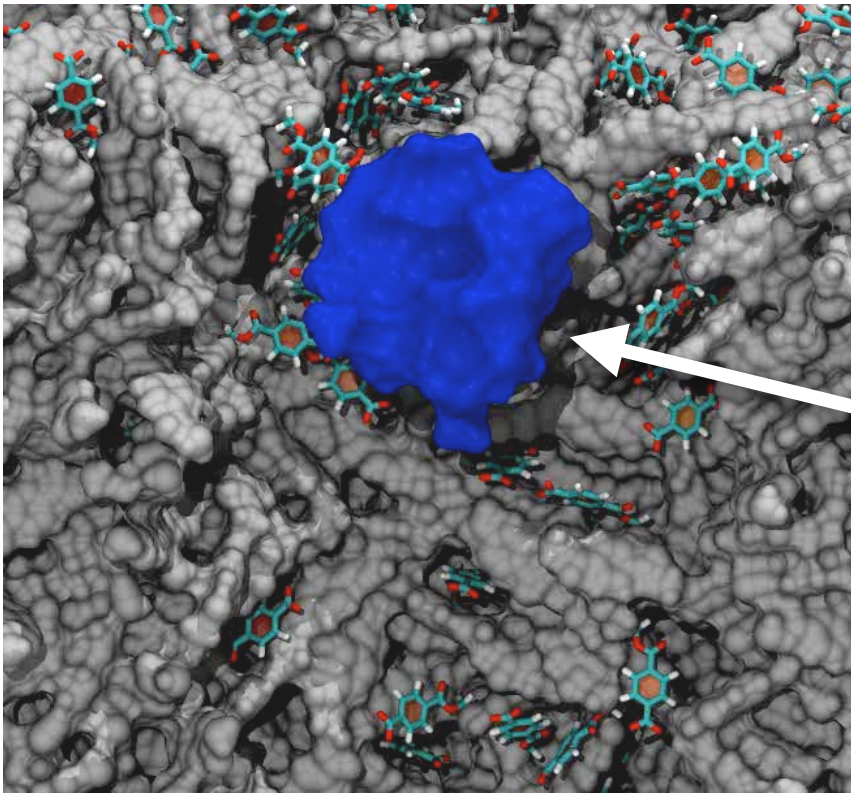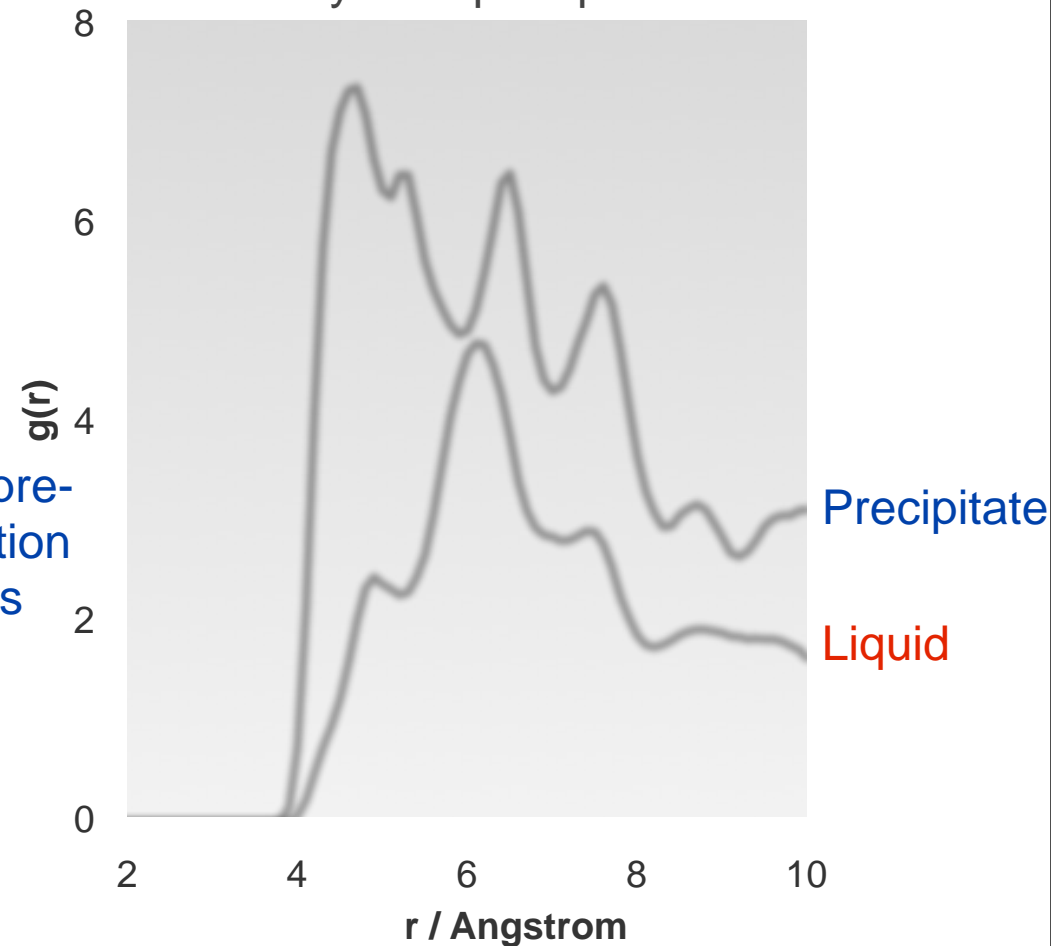
# Viewing Nanopore Sensors
## Create a **Better Nanopore** with Polymeric Materials

**New materials, new problems: Nanoprecipitation**



nanopre-cipitation of ions

**Radial distribution functions**
identify nanoprecipitation



Precipitate

Liquid

g(r)

r / Angstrom

# GPU Solution 4:
# Computing Radial Distribution Functions

- 4.7 million atoms

- 4-core Intel X5550 CPU: **15 hours**

- 4 NVIDIA C2050 GPUs: **10 minutes**

- Fermi GPUs ~3x faster than GT200 GPUs: larger on-chip shared memory
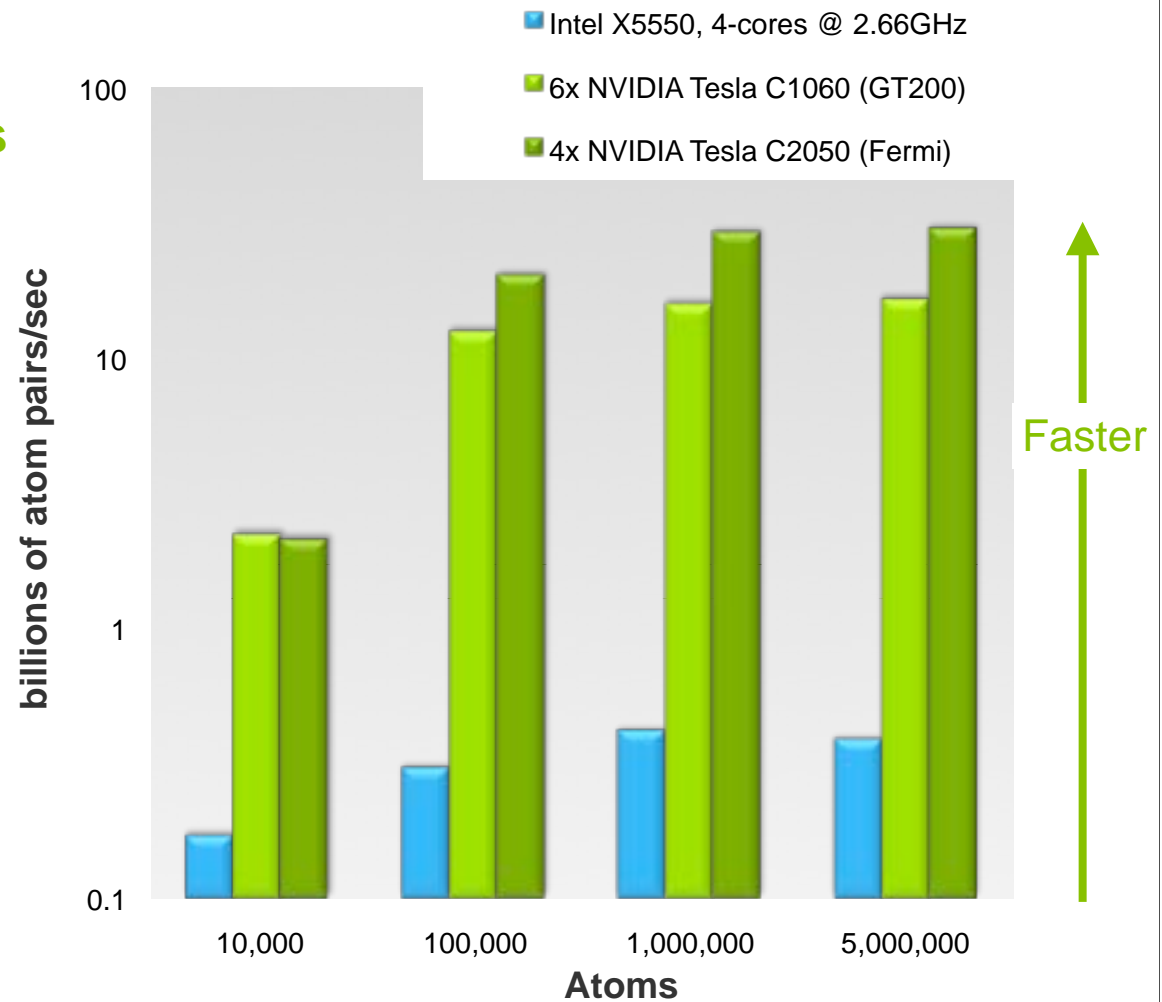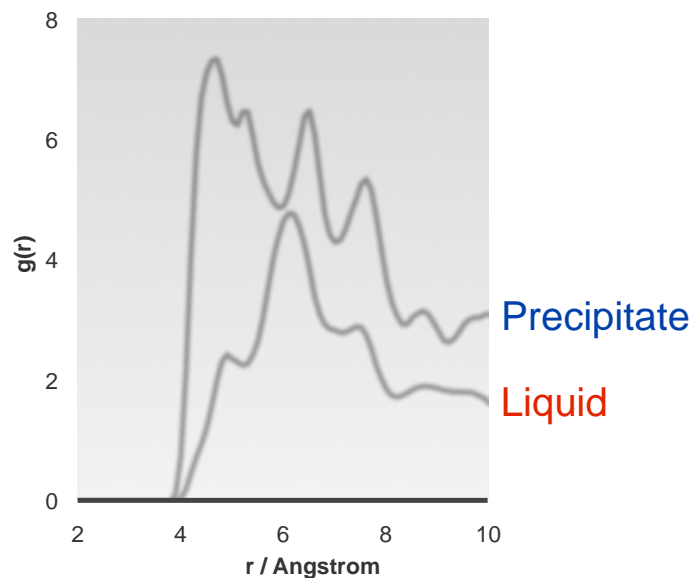
# Inspecting the mechanical Strength of a blood clot

Collaborator: Bernard C. Lim (Mayo Clinic College of Medicine)





A Blood Clot

Red blood cells within a network of fibrin fibers, composed of polymerized fibrinogen molecules.

**20ns SMD Simulation of fibrinogen, *1.06 million atoms,* 1.2 ns/day with pencil decomposition, 15 days on PSC XT3 Cray (1024 processors)**

B. Lim, E. Lee, M. Sotomayor, and K. Schulten. **Molecular basis of fibrin clot elasticity.** *Structure*, 16:449-459, 2008.

# Inspecting the mechanical Strength of a blood clot

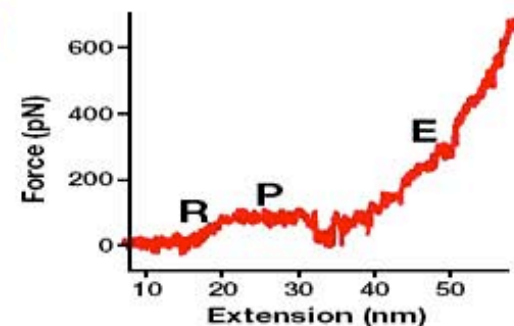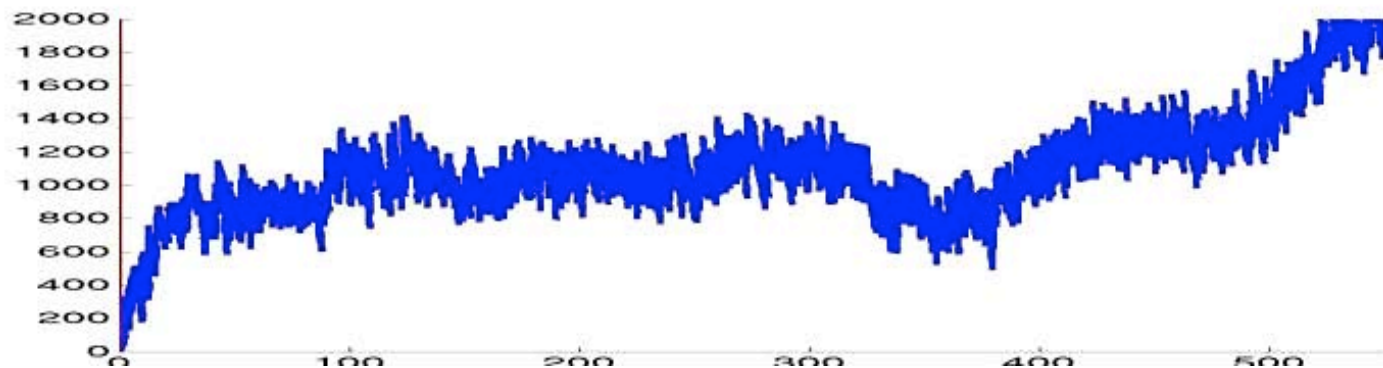Collaborator: Bernard C. Lim (Mayo Clinic College of Medicine)
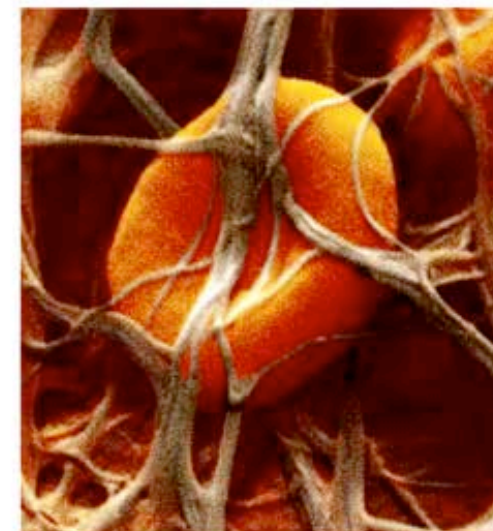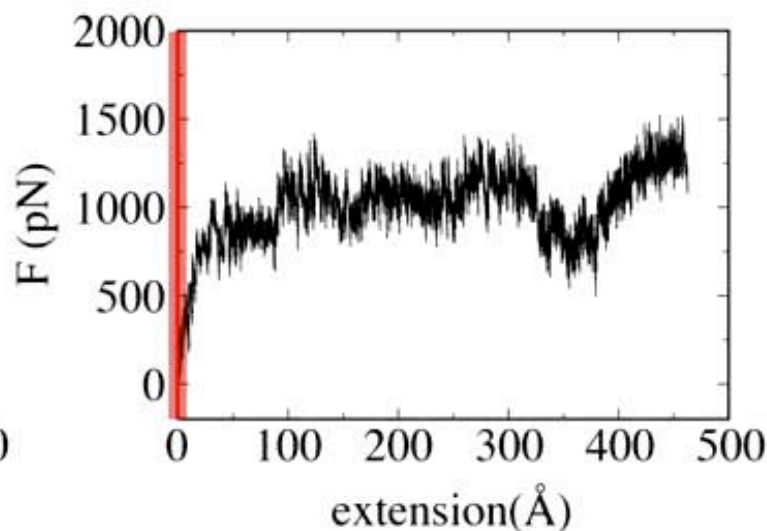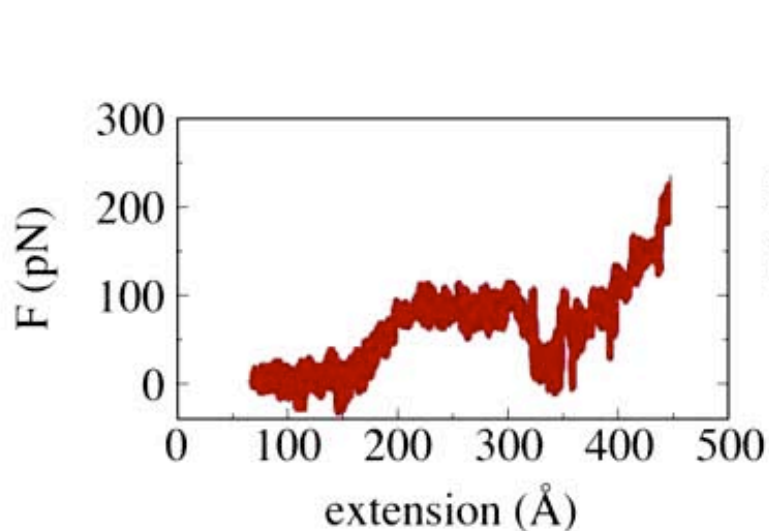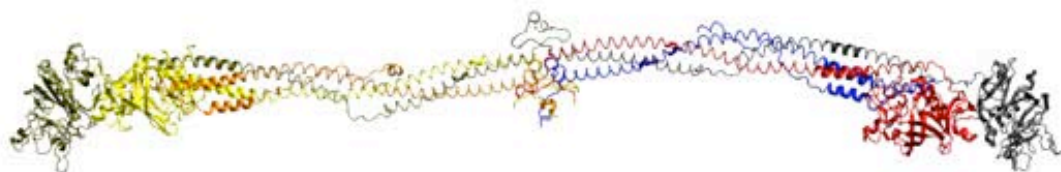


**20ns SMD Simulation of <u>fibrinogen,</u> *1.06 million atoms,* *1.2 ns/day with pencil decomposition, 15 days on PSC XT3 Cray (1024 processors)***

B. Lim, E. Lee, M. Sotomayor, and K. Schulten. **Molecular basis of fibrin clot elasticity.** *Structure*, 16:449-459, 2008.

**A Blood Clot**

Red blood cells within a network of fibrin fibers, composed of polymerized fibrinogen molecules.

# Petascale simulations will Permit Sampling
*For Example Carrying out a Second Simulation Required by a Referee*
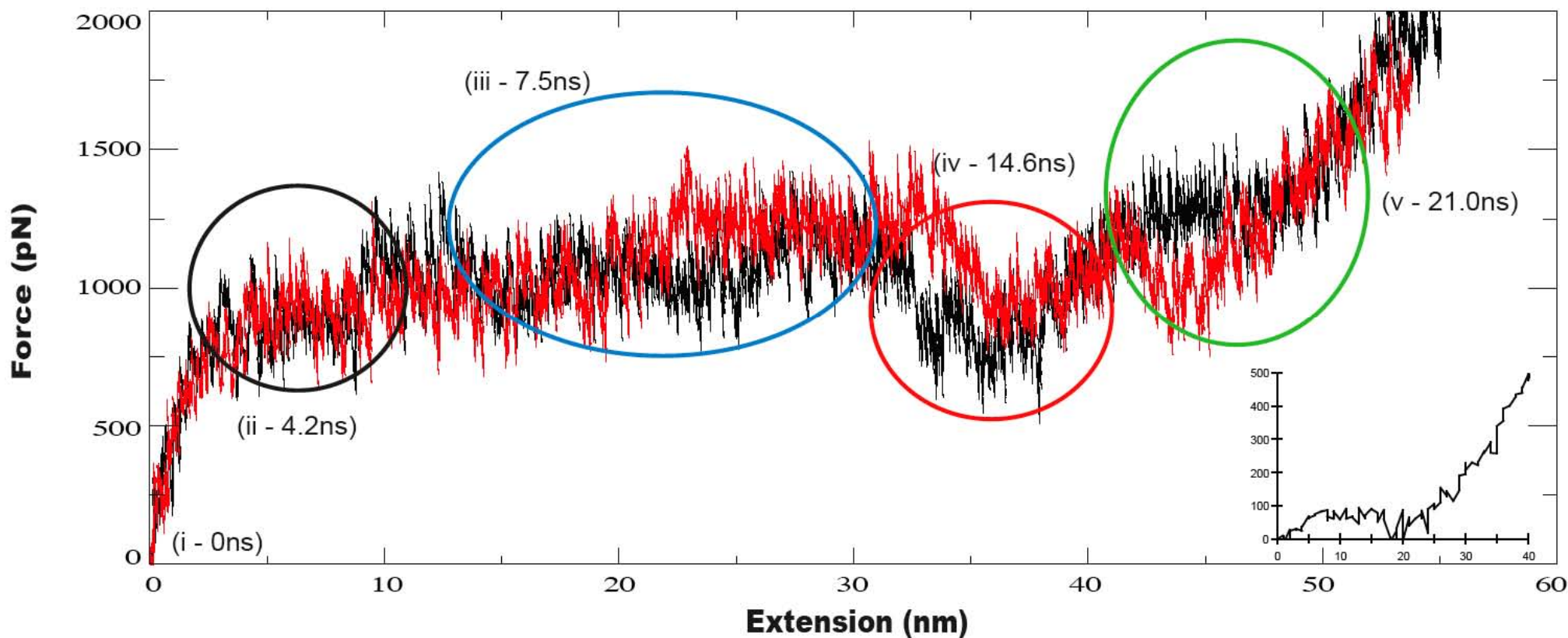
# Reaching for Overlapping Time Scales

## Microsecond simulations of muscle elasticity



- ● simulation
- ● experiment (Carrion-Vazquez et al. 1999)
- ● experiment (Rief et al., 1997)

*Stretched is I91, one of the 300 domains of titin*

model *Szabo and Hummer, 2003.*

Longest ever (1 us) SMD simulation closes gap between simulation and experiment!!!!!
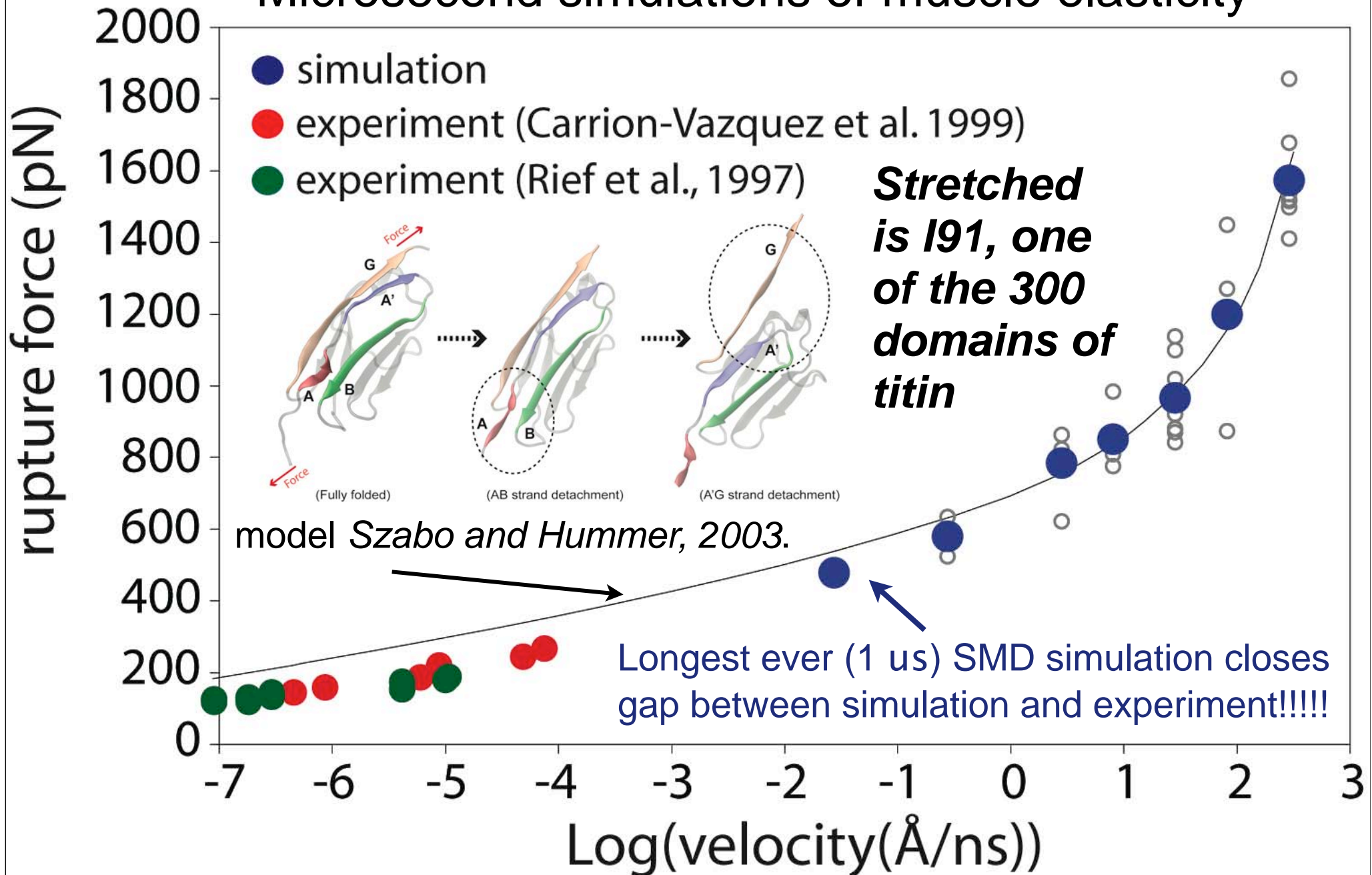
# Viewing Protein Folding

- Protein **misfolding** responsible for diseases:
  - Alzheimer's
  - Parkinson's
  - Huntington
  - Mad cow
  - Type II diabetes
  - ...
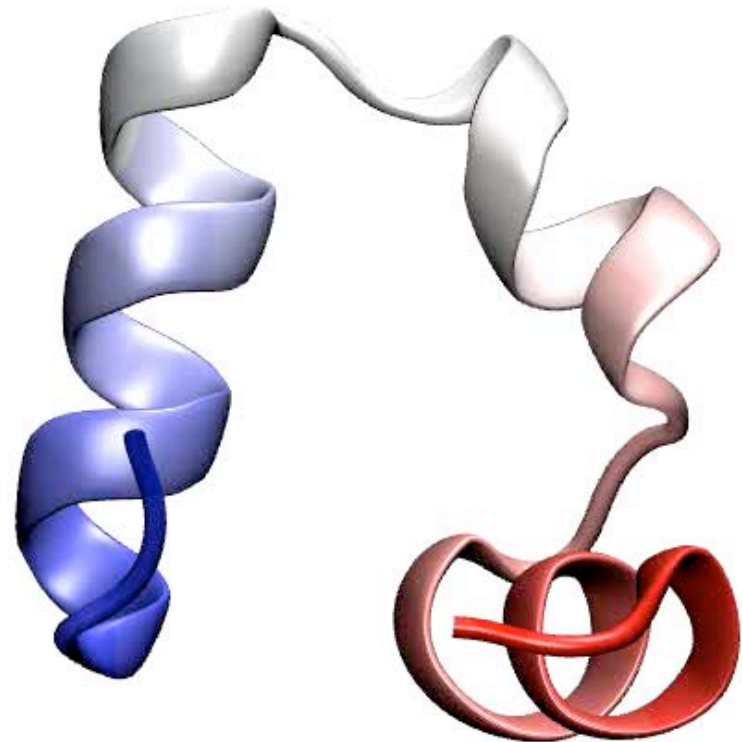


**villin headpiece**
**3 months on 329 CPUs**

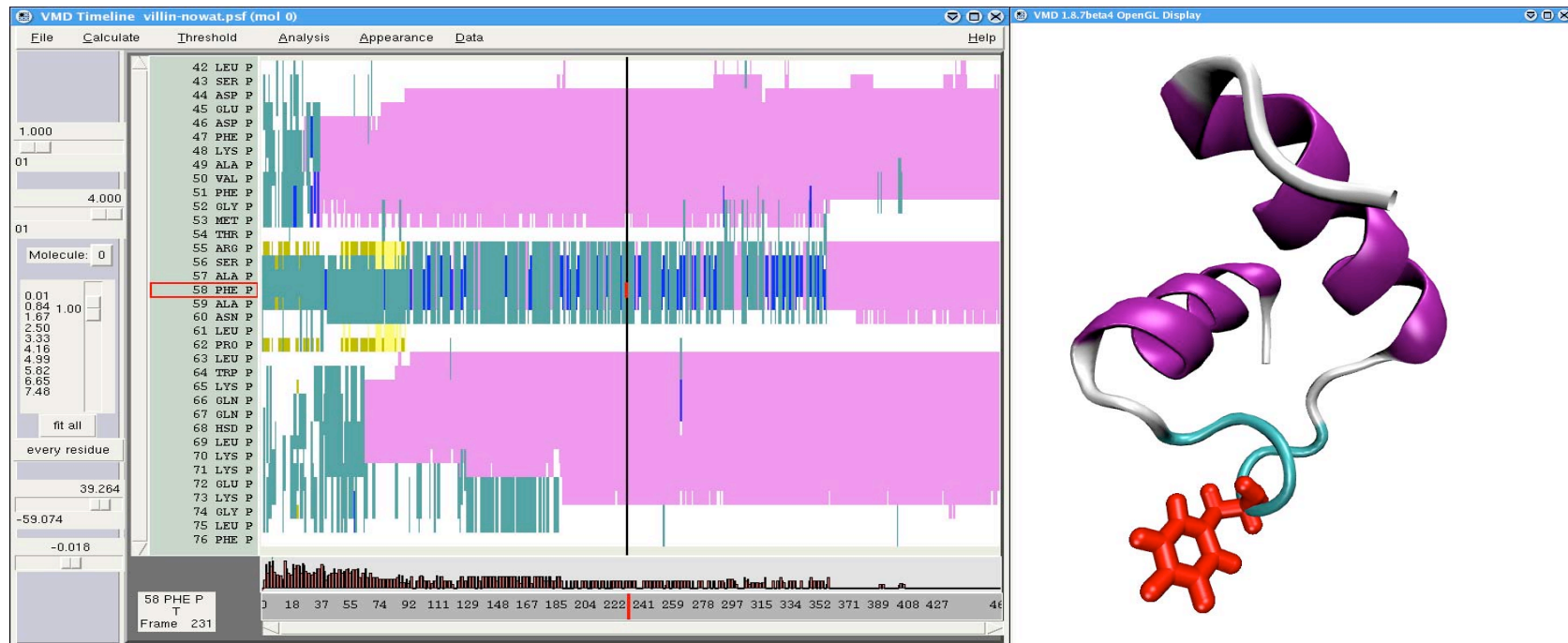Observe folding process in unprecedented detail

# 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

# Science 6: Protein Folding

- Atomic polarizability increases computation by 2x…

- …but, the additional computations are perfectly suited to the GPU!

- For now, NAMD calculates atomic polarizability on CPUs only...soon we will also use GPUs

Atomic polarizability of water, highly accurately simulated through additional particles (shown in green)



**NAMD CPU performance scaling**

**polarizable water**
**non-polarizable water**

Seconds per step

1

0.1

0.01

100          1000

**CPU cores**