Petascale Computing in the Biosciences
Three Projects Starting Today
National Science Foundation- 2006
Klaus Schulten

1 Million Atoms
Virus

3 Million Atoms
Ribosome

10 Million Atoms
Chromatophore

Theoretical and Computational Biophysics Group, Beckman Institute, Univ. Illinois
NAMD permits large scale simulations

<table>
<thead>
<tr>
<th>System</th>
<th>Simulation</th>
<th>Size</th>
<th>Platform (# proc.)</th>
<th>Year, Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>bilayer of 200 lipids</td>
<td>NVE, NC</td>
<td>27K</td>
<td>60 node transputer</td>
<td>1993, [69]</td>
</tr>
<tr>
<td>membrane-water interface</td>
<td>PMA</td>
<td>32K</td>
<td>SGI Crimson</td>
<td>1995, [150]</td>
</tr>
<tr>
<td>binding of estrogen receptor to DNA</td>
<td>FMA, NVT</td>
<td>36K</td>
<td>HP cluster (8)</td>
<td>1997, [73]</td>
</tr>
<tr>
<td>apolipoprotein A-I</td>
<td>NVE</td>
<td>46K</td>
<td>HP cluster (8)</td>
<td>1997, [107]</td>
</tr>
<tr>
<td>calmodulin</td>
<td>NVE</td>
<td>33K</td>
<td>T3E (64)</td>
<td>1998, [143]</td>
</tr>
<tr>
<td>Rieske subunit motion in cytochrome δc1 complex</td>
<td>SMD</td>
<td>91K</td>
<td>T3E (64)</td>
<td>2000, [60]</td>
</tr>
<tr>
<td>bacteriorhodopsin and purple membrane</td>
<td>PME, NVE, NPT</td>
<td>24K</td>
<td>Alpha cluster (8)</td>
<td>2001, [7]</td>
</tr>
<tr>
<td>MscL</td>
<td>PME, NPT</td>
<td>55K</td>
<td>T3E (64)</td>
<td>2001, [47]</td>
</tr>
<tr>
<td>aquaporin-1</td>
<td>PME, NPT</td>
<td>60K</td>
<td>T3E (64)</td>
<td>2001, [152]</td>
</tr>
<tr>
<td>BamHI endonuclease binding to DNA</td>
<td>PME, NPT</td>
<td>65K</td>
<td>T3E (64)</td>
<td>2001, [88]</td>
</tr>
<tr>
<td>photosynthetic light harvesting system</td>
<td>PME, NPT</td>
<td>87K</td>
<td>SGI Origin 2000 (4)</td>
<td>2001, [23]</td>
</tr>
<tr>
<td>fibronectin type III16</td>
<td>PME, SMD</td>
<td>120K</td>
<td>Linux cluster (32)</td>
<td>2002, [39], [2003, here]</td>
</tr>
<tr>
<td>aquaglyceroporin</td>
<td>PME, NPT</td>
<td>106K</td>
<td>TCS (128)</td>
<td>2002, [130], 2003, here</td>
</tr>
<tr>
<td>rhodopsin</td>
<td>PME, NPT</td>
<td>40K</td>
<td>T3E (128)</td>
<td>2002, [120]</td>
</tr>
<tr>
<td>CD2-CD58 complex</td>
<td>PME, SMD</td>
<td>91K, 104K</td>
<td>Linux Cluster (32)</td>
<td>2003, [8]</td>
</tr>
<tr>
<td>F6-ATPase</td>
<td>PME, NPT, SMD</td>
<td>112K</td>
<td>T3E (128), TCS (256), Linux cluster (32)</td>
<td>2003, here</td>
</tr>
<tr>
<td>F1-ATPase</td>
<td>PME, NPT, SMD</td>
<td>327K</td>
<td>Platinum (448), TCS (512)</td>
<td>2003, here</td>
</tr>
</tbody>
</table>

Ankyrin, 340,000 atoms (2005)
Estrogen Receptor 36K atoms, 100 ps (1996)
Hemolysin, 360,000 atoms (2005)
LacI-DNA complex 310,000 atoms, 100 ns (2005)
Satellite tobacco mosaic virus, 1,000,000 atoms, 50 ns (2006)
Our Tools: VMD, NAMD, BioCoRE

They support typical simulation workflow

Build simulation using VMD:
• visually analyze crystal structure
• generate molecular structure files
• add membrane, solvent, and ions

Minimize, equilibrate, and test using NAMD on local on-demand clusters.

Long NAMD runs at NSF centers:
• constant temperature and pressure ensembles
• alchemical and conformational free energy perturbation
• variety of steering forces, plus Tcl scripting for new ideas

Analyze trajectories locally:
• VMD trajectory animation
• VMD Tcl scripts and plugins
• NAMD interaction analysis
• Communicate over distance

$50K, 48 CPUs runs 100K atoms 5 ns/week

up to 5 ns/day
VMD Molecular Graphics

- > 60,000 registered users
- Platforms:
  - Unix (16 builds)
  - Windows
  - MacOS X
- Display of large biomolecules and simulation trajectories
- Multiple sequence - structure analysis

VMD view of F1-ATPase

Electrostatic potential for an ATPase obtained with VMD’s PME plugin
VMD: PDB Search and Analysis

• Load entire PDB (30K molecules, 32GB RAM)
• Perform atom selection-based searches and analyses on entire PDB database
• Refine search results with successive queries
• Query structural properties, e.g. “How much does the length of arginine vary?”
• PDB scan measured 591,338 arginine residues
• Dominant factor in one-time scan is disk I/O
• Doing scan in memory runs 100x to 1000x faster
• Multiple instances of VMD can scan in parallel
NAMD Performance on Small Linux Clusters

*New Xeon clusters are only 12% faster for typical production runs, but much more compact, versatile, and manageable.*

Older (2003):
- 24 Dual Athlon 2133 MHz
- Desktop tower
- No hard drives
- Clustermatic 3 on RedHat 8.0
- Floppy boot
- $1200 per node

Newer (2004):
- 24 Dual Xeon 3.06 GHz
- Rackmount
- Fully loaded
- Clustermatic 4 on RedHat 9.0
- Network boot
- $2000 per node

NAMD Registrants
12,095 Registrants (16% NIH)
2,122 Repeat Users (18% NIH)
Because of its scalability and portability, NAMD has been selected for the next SPEC HPC benchmark suite.
NAMD Scalable on Current Machines

Simulation of large biomolecular systems on parallel computers.

File compatible with original community codes CHARMM & AMBER.

Ten-year collaboration with UIUC Parallel Programming Lab.

2002 Gordon Bell Award for parallel scalability.

Runs at NSF centers, on clusters, and on desktop.

Available for FREE as precompiled binaries, includes source code.

ApoA1
92K atoms
with PME

2.3 s/step

PSC XT3

Linear scaling

8 ms

73% efficient on 256 CPUs

NAMD Scalable Molecular Dynamics

NAMD

Scalable Molecular Dynamics

Simulation of large biomolecular systems on parallel computers.

File compatible with original community codes CHARMM & AMBER.

Ten-year collaboration with UIUC Parallel Programming Lab.

2002 Gordon Bell Award for parallel scalability.

Runs at NSF centers, on clusters, and on desktop.

Available for FREE as precompiled binaries, includes source code.
Cray XT3 Performance Results: ApoA1
NAMD Scalable on Current Machines

NAMD on BlueGene/L

<table>
<thead>
<tr>
<th>Procs</th>
<th>Time (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9000</td>
</tr>
<tr>
<td>32</td>
<td>347</td>
</tr>
<tr>
<td>128</td>
<td>97.2</td>
</tr>
<tr>
<td>512</td>
<td>23.7</td>
</tr>
<tr>
<td>1024</td>
<td>13.8</td>
</tr>
<tr>
<td>2048</td>
<td>8.6</td>
</tr>
<tr>
<td>4096</td>
<td>6.2</td>
</tr>
<tr>
<td>8192</td>
<td>5.2</td>
</tr>
</tbody>
</table>

Using 512 nodes (1024 processors)

Oct’05

CP (Mar 06)
VN (Mar 06)

Step Time (ms)

Processors
On BlueGene/L (recent tuning), 1024 procs
Charm++’s “Projections” Analysis tool
Time intervals on x axis, activity added across processors on Y axis
Red: integration, Blue/Purple: electrostatics
Orange: PME, turquoise: angle/dihedral
green: communication
Shallow valleys, high peaks, nicely overlapped
PME
Solvate the virus in a 220Åx220Åx220Å water box, add Mg$^{2+}$ ions to neutralize RNA and Cl$^{-}$ ions to neutralize the protein.

132,000 atoms of protein, 30,000 atoms of RNA, ~1,000,000 atoms in total

Simulations are done using NAMD with CHARMM22/27 force field

Constant temperature and pressure: Langevin dynamics ($\gamma=5\text{ps}^{-1}$) and Nose-Hoover Langevin piston ($T=100\text{fs}$, $t=50\text{fs}$)

$\Delta t=1\text{fs}$, 12 Å cut-off

Each simulation is preceded by 1000 steps of conjugate gradient energy minimization and 150 ps of equilibration

Performance on 48 processors of Linux PC cluster Ariel: 1.3 ns per week

Performance on 256 processors at NCSA Altix: 1.1 ns per day

Simulation of the Complete Virion

RNA outside of the Capsid Keeps the Same Shape as inside the Capsid

Capsid without the RNA Core is Unstable

Capsid dynamics

Capsid without the RNA Core is Unstable

Capsid before dynamics

Capsid after dynamics

Petascale Project 1: Virus Capsid

Take next steps through coarse-graining (CG)
Poliovirus (CG)

- T number: pT3
- 60 subunits, each consisting of 4 different proteins
- 30 nm in diameter
- ~200 atoms per bead, 4500 beads total
- Simulated time: 10 µs

Polio virus - receptor complex - 3 million atoms
Simulated Virus Capsids (CG)

A

simSTMV 5 µs

simSTMVfull, 5 µs

B

simSPMV, 25 µs

C

simSTNV, 7 µs

D

simBMVCut 5 µs

simBMV, 5 µs

E

simPV 11 µs

F

simB174 3 µs

G

simRV, 1 µs
Petascale Project 2: Ribosome

- Very large system size:
  ~3,000,000 atoms
- Great biological and biomedical relevance
- Simulations with close collaboration with leading experimentalist
- All-atom, coarse graining, and multiscale simulations
- We collaborate with Joachim Frank (HHMI, U. Buffalo), a leading experimentalist in ribosome cryo-EM, and with Willy Wriggers (UT Houston), pioneer in the development of methods to fit atomic-structures into cryo-EM maps.

- The figure shows several examples of cryo-EM maps of the ribosome in different functional states obtained by our collaborator.

- Combining structural information from X-ray crystallography and cryo-EM will provide initial structures for simulations which will address long-standing questions in the molecular biology of protein synthesis.

Petascale Project 3: Chromatophore

Photosynthetic unit from purple bacteria (Bahatyrova et al., *Nature*, August 26 2004)

Photosynthetic unit from purple bacteria
Chromatophore
10 Million Atom Simulation of Cell Organelle

light energy converter
Chromotaphore - Six Proteins use sun light for ATP synthesis (LH2, LH1, RC, bc1, cyt c2, ATPase)

Schematic arrangement of all six proteins

photosynthetic unit

purple bacterium cell

energy converter
LH-II  LH-I  cytochrome  c2  ATPase  cytoplasm  periplasm  ADP  ATP  bc1  H+  e-  Q/QH2/Q  RC  hν  membrane  cytochrome  c2  Petascale Project 2: Chromatophore  Key* Organelle of Living Systems  * 95% of the energy consumed in Earth’s biosphere stems from it and similar modules
First Step: Physics of Chromotrophore

Links = dipole-dipole coupl.

\[ W_{jk} = C \left( \frac{\vec{d}_j \cdot \vec{d}_k}{r_{jk}^3} - \frac{3(r_{jk}^2 \cdot \vec{d}_j)(\vec{r}_{jk} \cdot \vec{d}_k)}{r_{jk}^5} \right) \]

M. Sener, N. Hunter, and K. Schulten (in preparation)

Bahatyrova et al., Nature, 2004
First Step: Physics of Chromatophore

Links = dipole-dipole coupl. \[ W_{jk} = C \left( \frac{\vec{d}_j \cdot \vec{d}_k}{r_{jk}^3} - \frac{3(\vec{r}_{jk} \cdot \vec{d}_j)(\vec{r}_{jk} \cdot \vec{d}_k)}{r_{jk}^5} \right) \]

M. Sener, N. Hunter, and K. Schulten (in preparation)
Petascale Project 2: Chromatophore
10-100 Million Atom Simulation of Cell Organelle
Acknowledgments

Theoretical Biophysics Group, Beckman Institute and Department of Physics UIUC

Center for Macromolecular Modeling and Bioinformatics