Bimolecular Structure Determination with NAMD: Computational Cryo-EM on Titan

May 16th, 2018

Abhishek Singharoy
School of Molecular Sciences
Arizona State University

OLCF Users Meeting
Oak Ridge National Laboratory

INCITE PI to “All-atom Simulations of Photosynthetic and Respiratory Energy Conversion”
Biomolecules in Action: Using Titan as a “Computational Microscope” with NAMD

Chemistry

\[ U(\vec{R}) = \sum \text{ bonds} \left( k_{\text{bond}} (r_i - r_0)^2 + \sum \text{ angles} \left( \theta_i - \theta_0 \right)^2 + \sum k_{\text{dihedral}} \left[ 1 + \cos(n_i \phi_i + \delta_i) \right] + \sum \sum_{i,j \neq i} \left[ \sum_{i, j \neq i} \left( \frac{r_{ij}^3}{r_{ij}^3} - 1 \right) - \left( \frac{r_{ij}^2}{r_{ij}^2} - 1 \right) \right] + \sum \sum_{i, j \neq i} \delta_{ij} \right) \]

Physics

\[ m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R}) \]

Math

\[ \vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \vec{F}_i(t) \]

(repeat one billion times = microsecond)

Molecular Dynamics (MD) simulations

NAMD Software

Virus

Protein Folding

ns/day

cores

128 256 512 1024 2048 4096 8192 16384 32768

and

Supercomputers

David Hardy

Jim Phillips
Why Does One Need a Supercomputer?

Structural transitions

Ensemble of trajectories

Multiple replica required!!

Why Does One Need a Supercomputer?

![Graph showing the number of atoms in various biological structures over time.](Image)
Parallel Performance of NAMD on Summit

- Time to solution (sec/step)
- No. of nodes

- 128 M atoms (GPUs)
- 128 M atoms (CPUs)
- 128 replica of 1M atoms (GPUs)
- 128 replica of 1M atoms (CPUs)
Visualization and Analysis: VMD
In-situ Visualization of Billion Atoms: SIGHT

Noah Trebesch

Ben Hernandez (OLCF)
Scientific Accomplishment # 1: Energy Conversion in Bacterial Photosynthesis

**Figure 1B** Energy conversion processes starting after initial light absorption are divided into three stages: (1) quinol production at RC as a result of excitation transfer; (2) diffusion between RC and cyt c complexes as well as the lipid phase (olive; see also Figure 2A). (3) Utilization of proton gradient for ATP synthesis. 

\[ k_{ATP}(I) = \frac{1}{2} Iq \left( 1 + \frac{1}{2} Iq \tau_{RC}(I) \frac{1}{n_{RC}} \right) \]

\[ \tau_{RC}(I) = 1 + (\tau_H - 1) \left( 1 - e^{\frac{Iq}{2B}} \right) \]

\[ \tau_H = \frac{n_{RC}}{n_B} \tau_B; B = \frac{2n_B}{\tau_B} \]

---

*eLife* 2016, 5, e09541;
*Parallel Comput.* 2016, 55, 17

low-light adaptation
Can we make a model for growth from first principles??

**Summit Goals 1: From First-Principles to Phenotypic Behaviors**

**Figure 3.** Effect of vesicle composition on steady-state ATP production at different light intensities. Vesicle composition is given in terms of the number of cytbc dimers ($n_B$) and of RC-LH1-PufX dimers ($n_L$) for vesicles featuring identical surface area; LH2 composition of the vesicle is determined by $k_{ATP}$ (ATP/s).

Sener et al. eLife 2016;5:e09541. DOI: 10.7554/eLife.09541

*J.Phys.Chem.B 2017, 121, 3787–3797*

Hunter (DOE’s PARC center WUSTL)
Summit Goals 2: Conformational Transition in Molecular Motors

Design principles of 100% energy conversion efficiency ??

Scientific Accomplishment # 2

J. Am. Chem. Soc. 2017, 139, 293
Scientific Accomplishment # 3: Synthesis of Artificial Motors

Inchworm motion

**Tetrazine BPTz**

**Monomer Pathway**

**Dimer Pathway**

**RDS**

**Anion**

**cyanostar**

**cyanodimer**

**cyanosolo**

**bis-macrocycle**


(Accepted for special issue commemorating the 2016 Nobel Prize in Chemistry)
Scientific Challenge # 3: High-throughput design of Artificial Motors

<table>
<thead>
<tr>
<th>TTF$^{2+}$</th>
<th>CBPQT$^{4+}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Target free energy cost function

<table>
<thead>
<tr>
<th>Linear motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>State 0</td>
</tr>
<tr>
<td>Stimulus A</td>
</tr>
<tr>
<td>Stimulus B</td>
</tr>
<tr>
<td>State 1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rotatory motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxidation</td>
</tr>
<tr>
<td>Reduction</td>
</tr>
</tbody>
</table>

Space of components

Machine Learning application
Scientific Accomplishment # 4: Data-guided Structure Discovery

X-ray diffraction

Cryo-EM

Nat. Struct. Mol. Biol. 21, 244 (2014)

eLife 3, e03035 (2014)
PNAS 113, 10310 (2016)
eLife 5, e16105 (2016)
Methods 100, 50-60 (2016)
Structure 24, 2102 (2017)
Molecular Dynamics Flexible Fitting - creation of data-driven force fields

Two terms are added to the MD potential

$$U_{total} = U_{MD} + U_{EM} + U_{SS}$$

An external potential derived from the EM map is defined on a grid as

$$U_{EM}(\mathbf{R}) = \sum_j w_j V_{EM}(\mathbf{r}_j)$$

$$V_{EM}(\mathbf{r}) = \begin{cases} 
\xi \left(1 - \frac{\Phi(\mathbf{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right) & \text{if } \Phi(\mathbf{r}) \geq \Phi_{thr}, \\
\xi & \text{if } \Phi(\mathbf{r}) < \Phi_{thr}. 
\end{cases}$$

A mass-weighted force is then applied to each atom

$$\mathbf{f}^E_M = -\nabla U_{EM}(\mathbf{R}) = -w_i \partial V_{EM}(\mathbf{r}_i)/\partial r_i$$


*2017 Nobel Prize in Chemistry*
Data-acquisition Pipelines at NSF BioXFEL Center

Serial (femto-second) X-ray Crystallography

Raw Data -> 10-100TB

Hit finding -> 1-10 TB

Find Indexable patterns -> < 1 TB

Index -> ~GB

merge -> ~MB

model

Quality assessment & refinement

Deposit Data (CXIDB)

Deposit Structure (PDB)

Publish

Fromme (ASU)
Scientific Challenge # 4: Structure Discovery with Supercomputers
Combination of “Enhanced-sampling” with Flexible Fitting

Multiple stereoisomers in the same 1.8 Angstrom data
Results of the Cryo-EM Structure Challenge (EM193)

emb_5778

emb_5995

Journal of Structural Biology (Under review at for special section commemorating 2017 Nobel Prize)
Quality of fitting

TRPV1

β-galactosidase

Sequence Match
Vision: Structure Discovery with Summit

Currently on BlueWaters

Tajkhoshid
Dill (Stony-Brooks U.)
Richardson (Duke)