Network Visualization of Conformational Sampling during Molecular Dynamics Simulation

Osamu Miyashita
Computational Structural Biology Unit

ADVANCED INSTITUTE FOR COMPUTATIONAL SCIENCE

MIAT, INRA of Toulouse, Dec 2014
Proteins and Nucleotides (DNA/RNA)

- Players of cell life activities

Hemoglobin  Myosin  Virus
Lines of defence

Antiviral treatments are a critical component of an effective healthcare response to influenza, but drug resistance to the treatment-of-choice has public health officials searching for other options.

Zachary Taylor, an infectious disease fellow at the Kaiser Permanente Fontana Medical Center in Sacramento, California. In part to safeguard against the possibility of such game-changing developments, drug developers are slowly filling the pipeline with alternative therapies (see "Drugs to treat influenza infection"). Each drug comes with side effects, which make them only worthwhile for those whom the flu could be potentially lethal — the elderly and the immunocompromised.

Given the wily history of the influenza virus, any sudden appearance of drug resistance is certain to concern public health officials. The first antiviral drugs to combat the disease — the adamantanes, which target the M2 channel protein to block virus entry into host cells — are now essentially useless. The US Centers for Disease Control and Prevention (CDC) found that 100% of seasonal H3N2 flu in the 2009–2010 season and 99.8% of 2009 pandemic H1N1 flu were resistant to adamantanes.

Oseltamivir belongs to a class of drugs called neuraminidase inhibitors. These agents block the active site of a viral protein called neuraminidase (N), thereby arresting the influenza virus' ability to leave the host cell after it proliferates. The most common way for the influenza virus to evade oseltamivir is via the H275Y mutation (also known as H274Y) of neuraminidase, which replaces a single histidine amino acid with a tyrosine. This alteration interferes with the drug's ability to bind to the protein — a problem acknowledged by the maker of oseltamivir.

"There remains a medical need and room for additional treatment options, especially for the management of severe infections and for improved pandemic preparedness," says Klaus Klumpp, Roche's top virologist. Klumpp says the Roche is supporting research into new therapies targeting viral replication as well as other mechanisms, but notes that these efforts are preclinical.

Fortunately, viruses with the H275Y mutation are still susceptible to a different neuraminidase inhibitor, the drug zanamivir.
NAs catalyze the hydrolysis of sialosides with net retention of stereochemistry at the site of substitution. A mechanism involving an ion-pair intermediate has long been suggested for the GHE34 enzymes. However, drug-resistant strains are now emerging, particularly against the more widely used and structurally divergent drug oseltamivir, highlighting an urgent need for new classes of NA inhibitors that differ minimally in structure from the parent sialic acid, given that the development of resistance to structurally conservative, mechanism-based inhibitors should be a much less probable event.

NAs catalyze the hydrolysis of sialosides with net retention of stereochemistry at the site of substitution. A mechanism involving an ion-pair intermediate has long been suggested for the GHE34 enzymes. However, drug-resistant strains are now emerging, particularly against the more widely used and structurally divergent drug oseltamivir, highlighting an urgent need for new classes of NA inhibitors that differ minimally in structure from the parent sialic acid, given that the development of resistance to structurally conservative, mechanism-based inhibitors should be a much less probable event.

Fig. 1. Structures of key influenza therapeutics, mechanism of action of DFSAs, and x-ray structure of inhibited enzyme. (A) Chemical structures of cell surface sialic acids, the neuraminidase transition state, zanamivir (Relenza), and oseltamivir (Tamiflu). (B) Mechanism of action of the DFSAs. (C) X-ray crystallographic structure of the active site of the enzyme trapped as its 3-fluoro(eq)-4-guanidino-sialyl-enzyme intermediate (elimination product is in pale cyan) overlaid with omit (22) electron density maps shown as a gray mesh contoured at 1σ within 1.6 Å of ligands. The electron density extends from the ligand molecule to Y406, suggesting a covalent link between the inhibitor’s C-2 atom and the OH of Y406. (D) Diagram of interactions (orange dashed lines; distances in Å) with the sialic acid in the covalently inhibited enzyme. The corresponding diagram of interactions for the elimination product is shown in fig. S4.
Protein Molecules

Globular protein structures are generally tightly packed, compact units.

1 amino acid $\rightarrow$ 10 $\approx$ 30 atoms
Proteins:
Poly-peptide chain of 100 ≈ 500 amino acids
Biological Assemblies

Protein: amino acids linked together

amino acid (AA)

myoglobin ≈ 130 AA

Ribosome ≈ 11000 nucleotides/AA

Virus > 30000 AA

Myosin ≈ 4500 AA

supramolecular complexes of proteins / RNA
PDB – protein structure database

www.pdb.org

Ebola virus proteins
Computer Simulations
Molecular Dynamics Simulations

- Simulate the motions of atoms within biological molecules using physics theories
- Newton Equation

\[ F_n = -\nabla_n U(r_n) = m_n a_n = m_n \frac{d^2 r_n}{dt^2} \]

\[ F_n = \text{force on particle } n \]
\[ m_n = \text{mass} \]
\[ a_n = \text{acceleration} \]
\[ U(r_n) = \text{potential energy} \]
\[ r_n = \text{coordinate of particle } n \]

1. Numerical procedure to integrate the differential equation
2. Energy function for the biological molecules
Molecular Dynamics (MD) Simulation

starting coordinates = X-ray, NMR structure

Molecular Mechanics Potential ("Force Field")

\[
U(\mathbf{r}_n) = \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_0)^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4 \pi \varepsilon_0 r_{ij}} + 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]
\]

Newton’s second law

\[
\mathbf{F}_n = -\nabla U(\mathbf{r}_n) = m \mathbf{a}_n
\]

Use the accelerations to numerically integrate the equations of motion
Integration Algorithms

- Use a finite-difference approach: molecular coordinates and velocities at time $t = t + \Delta t$ are obtained from coordinates and velocities at time $t$

- A time-reversible integrator can be generated from Taylor expansions:

  \[
  r_n(t + \Delta t) = r_n(t) + v_n(t)\Delta t + \frac{1}{2} a_n(t)\Delta t^2 + O(\Delta t^3)...
  \]

  \[
  +/-
  \]

  \[
  r_n(t - \Delta t) = r_n(t) - v_n(t)\Delta t + \frac{1}{2} a_n(t)\Delta t^2 - O(\Delta t^3)...
  \]

  \[
  v_n = \frac{dr_n}{dt}
  \]

  \[
  a_n = \frac{dv_n}{dt} = \frac{d^2 r_n}{dt^2}
  \]

  \[
  r_n(t + \Delta t) = 2r_n(t) - r_n(t - \Delta t) + a_n(t)\Delta t^2 + O(\Delta t^4)...
  \]

  \[
  v_n(t) = \frac{r_n(t + \Delta t) - r_n(t - \Delta t)}{2\Delta t} + O(\Delta t^2)...
  \]
Verlet Integrator

\[ r_n(t + \Delta t) = 2r_n(t) - r_n(t - \Delta t) + a_n(t)\Delta t^2 + O(\Delta t^4) \ldots \]
Verlet Integrator

![Diagram showing the Verlet integrator process]

Compute forces at the current position using current positions

\[ a_n = \frac{F_n}{m_n} \]

\[ r_n(t + \Delta t) = 2r_n(t) - r_n(t - \Delta t) + a_n(t)\Delta t^2 + O(\Delta t^4) \ldots \]
Verlet Integrator

\[ r_n(t + \Delta t) = 2r_n(t) - r_n(t - \Delta t) + a_n(t)\Delta t^2 + O(\Delta t^4) \ldots \]

⇒ Advance to next step and repeat
Time Step in MD simulation

- If time step is too large, some energies (and forces) may suddenly get too large (e.g. because of hard sphere repulsion or bond over-stretching) and simulation “explodes”.

![Graph showing energy over time for different time steps](image)
Temperature Control

• Langevin Dynamics

\[ m_n \frac{d^2 r_n}{dt^2} = F_n - \gamma_n m_n \frac{dr_n}{dt} + R_n(t) \]

Friction and random force are added to the systematic forces.

Motivation: representation of a simple heat bath by accounting for molecular collisions

- \( \gamma \): collision parameter
- \( R \): random force vector

The strengths of random forces should follow

\[ <R(t)> = 0, \quad <R(t)R(t')> = 2\gamma k_B T_m \delta(t-t') \]
Solvent Box

- Protein in a box of water
- Used with *periodic boundary condition*
- No interface artifact
- But artifact of *mirror images*
Periodic Boundary Condition

- Protein and solvent are placed in the unit cell, surrounded by infinitely many copies in space.
- Atoms go out, then come back from opposite side.
- Cell has to be large enough to minimize artifacts:
  - At least a 10Å layer of water molecules.

Figure 9.4. Periodic domain in 2D, showing the unit cell (center) and its images (surrounding replicas), and in 3D (rectangular), as used for a solvated protein (BPTI) system [609].
## MD Simulation History

\[ F_n = m_n a_n = -\nabla_n U(r_n) = m_n \frac{d^2 r_n}{dt^2} \]

<table>
<thead>
<tr>
<th>Year</th>
<th>Protein</th>
<th>Atoms</th>
<th>Simulation (ns)</th>
<th>CPU time / processors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1977</td>
<td>BPTI</td>
<td>885</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td>Myoglobin</td>
<td>1,423</td>
<td>0.3</td>
<td>50 days / 1</td>
</tr>
<tr>
<td>1992</td>
<td>HIV protease</td>
<td>25,000</td>
<td>0.1</td>
<td>4 days / 1</td>
</tr>
<tr>
<td>1998</td>
<td>Villin headpiece</td>
<td>12,000</td>
<td>1000</td>
<td>120 days / 256</td>
</tr>
<tr>
<td>2006</td>
<td>STM Virus</td>
<td>1,000,000</td>
<td>10</td>
<td>10 days / 256</td>
</tr>
<tr>
<td>2014</td>
<td>Protein*</td>
<td>24,000</td>
<td>2.5x10^6 ( = 2.5 ms)</td>
<td>30 days / 512</td>
</tr>
<tr>
<td>2020</td>
<td>Protein</td>
<td>40,000</td>
<td>10^9 (=1 sec)</td>
<td>????</td>
</tr>
</tbody>
</table>

1974: Rahman & Stillinger - simulation of liquid water

* Anton 2 – DE Shaw
What can Molecular Dynamics Simulation Do?

- Atomic detail information of dynamics
- Complete control
  - Output is created solely by inputs (parameters, structure, equations, algorithms, sampling etc...)
- But there are some limitations...
Molecular Dynamics Limitations

• Limited time scale
  – 1ns is easy, 1μs is difficult to reach
• Limited system size
  – Larger system needs more computation
• No chemical reactions
  – No bond breaking
• Statistical error due to insufficient sampling
  – Statistical accuracy needs to be checked
Aquaporin Water Channel

Information from Simulation

• Detail information inaccessible by experiments
  – Time course of biological events
  – Motion of each atom
  – Interaction energy
  – RMSD, energy, volume...
  – There are too many information...
  – *We need to analyze huge number of sampled structures*
    • *Clustering of the structures*
    • *Network visualization of conformational space*
Cyanovirin

- Sugar binding protein
- Anti HIV activity
1.3Å high resolution structure of Cyanovirin reveals the dynamics of Arg 76

**Fromme, Katiliene, Fromme, Ghirlanda, Protein Sci 2008**
Snapshots of protein dynamics from MD simulations

Some examples of clustering methods

- “Similar” conformations are grouped together
- Different algorithms use different criteria to join groups.
- “Similar” needs to be defined.

\[
\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta_i^2}
\]

Shao, J. Y., Tanner, S. W., Thompson, N. & Cheatham, T. E. (2007)
Network visualization of conformational ensemble

Conformational snapshots from MD simulation

Pair-wise RMSD (Å)

Network connectivity = Conformational similarity

Cutoff

Layout algorithm
Conformational Space of Arg76 sampled by MD simulation

• 1 point represents 1 conformation
• 1000 frames from 20 ns
• Edges between the frames with similar Arg76 conformations
Representative Structures

Vorontsov, Miyashita, Biophys J 2009
Cut-off for Network Connectivity

mean: 8.8 Å σ: 2.4 Å
Network v.s. Clustering

Baker, Ahlstrom, Ehrlich, Vorontsov, Patel, Campbell, Tama, Miyashita, 2013
Conformational Transitions

Network based on Structure Similarity

Edges showing Actual Transitions
Replica Exchange Molecular Dynamics Simulation of $\lambda$ Cro Protein

- 132 amino acids, 14,000 waters
- 15 ns equilibration, 15 ns production
Two Stable Conformation in Solution
2D Energy Landscape of Cro Conformational Space

Ahlstrom, L.S. and Miyashita, O. *Biophys J* 2011
Conformational Transitions Controlled by Salt-bridges

R4-E53′

closed-like

open-like

E53

NMR

Distance (Å)

19.7

3.1
Toward Larger and Larger Biological Systems

protein -> supramolecular complexes

amino acid ~ 100 aa
cytochrome ~ 5000 aa
ATP synthase ~ 5000 aa
Ribosome ~ 11000 nucleotides/aa
Virus > 30000 aa

Myosin ~ 4500 aa

http://www.pdb.org/
Dynamics in Biological Systems

- Short time scale: Vibration bond angle, Rotation: methyl group, Rotation: methylene group, change in hydration (water).
- Longer time scale: Ligand rotation, Protein rotation, Ligand binding, Libration main and lateral chains, Helix, Domain motions, Large conformational change, Intermediate lifetime (enzymatic reaction), Denaturation, Folding, Aggregation, Enzymatic reaction.
Computation

Molecular dynamics (MD) simulation: all atoms

Normal Mode Analysis (NMA)

Coarse grained models
Biased MD / Enhanced sampling

Amplitude (Å)

short time scale

ger longer time scale
Coarse Grained Models

POTENTIAL: $V_{HH} + V_{px} + V_{Nx} + V_{bond\ angle} + V_{dihedral}$

One point per residue

POTENTIAL: $V_{CA-CA}$
Coarse grained model: Go Model

Each atom is represented as a single bead of unit mass.

Bond lengths, bond angles, improper dihedrals, and planar dihedrals are maintained by harmonic potentials.

Nonbonded atom pairs that are in contact in the native structure are given a Lennard-Jones contact potential.

all other nonlocal interactions are repulsive.

Protein Model: Whitford PC et al. Proteins, 2009
Normal mode analysis

Low frequencies  ↔  High frequencies

Related to conformational changes of biological systems

Tama & Sanejouand (2001)
Multi-scale modeling

Coarse-Grained Elastic network model:

\[
E(r_a, r_b) = \frac{C}{2} \left( |r_{a,b}| - |r_{a,b}^0| \right)^2
\]

\[
E_p = \sum_{r_{a,b}^0 < R_c} E(r_a, r_b)
\]

capture the shape of the ribosome

functional motions predicted

Normal Mode Analysis

Coarse Grained Elastic Network + RTB
Group Theory + Dihedral angle space / RTB

1 hour on a single processor

Coarse Grained Elastic Network + RTB

780 hours on a single processor

BPTI

DIMB

Elastic network

RTB

Years

1983
1996
2000
2004
Experimental Techniques for Structure Analysis

**High resolution**

- **X-ray crystallography/NMR**
  + atomic level detail
  - large systems are difficult to crystalize
  - NMR is only for small systems

- **X-ray Free Electron Laser (XFEL)**
  + Single molecule measurement
  + No crystallization
  + Higher resolution

- **Cryo-Electron Microscopy**
  - low resolution, only rough shape
  + applied to large important systems
  + can capture different conf states

**Other approaches:** SAXS, FRET etc
Cryo Electron Microscopy – Single Particle Analysis

Arrows => impact of the parallel electron beam upon the specimen

Many “copies” of the molecule are lying in random orientations.

Reconstruction of the 3D structure from thousands of 2D images / Image processing

- Low-resolution compared to X-ray (5 – 30 Å)
- Easier to catch molecules in different conformational states

shape of the molecule
no atomic positions
Structure Analysis using Cryo Electron Microscopy Data

Cryo-EM: low resolution

OPEN

X-ray: high resolution

CLOSED

Flexible fitting

MD simulation
Normal Mode

\( f_{\text{EM}} \)
Optimization function

Atomic model
Flexible fitting using NMA

Resolution = 10.8 Å

Medium resolution: works well
High resolution: limitations as more localized motions become evident
Some conformational changes are not well described

F. Tama et al.

Data from M. Valle et al.
Cell (2003)
Flexible fitting using targeted MD

\[ U = U_{\text{molecule}} + k(1 - f_{EM}) \]

\[ f_{EM} = \frac{\sum_{ijk} \rho^{\text{exp}}(i,j,k) \rho^{\text{sim}}(i,j,k)}{\sqrt{\sum_{ijk} \rho^{\text{exp}}(i,j,k)^2 \sum_{ijk} \rho^{\text{sim}}(i,j,k)^2}} \]

Lower resolution

Coarse-grained

+ Fast, Easy to set up
- Only Cα atoms (Go-model)

Higher resolution

All atoms

+ Full atomic description (amber, all atom Go model*)
- Slow, cumbersome

M Orzechowski and F. Tama, Biophys. J. 2008
I. Grubisic et al., JSB, 2010

*Protein Model: Whitford PC et al. Proteins, 2009
Flexible fitting using targeted MD
GroEL at 4.2 Å

Ribosome at 8.6 Å

Rhodopsin at 5.5 Å

Toward higher resolution data

Ribosome at 7.3 Å

TMV at 4.4 Å

SJ. Ludtke et al., Structure 2004

SJ. Ludtke et al., Structure 2008

M. Halic et al. Nature. 2006

A Krebs et al., JBC 2003

Schueler M, Nat Struct Biol 2006

X-ray Free Electron Laser (XFEL)

- High intensity X-ray laser, SACLA: June 2011
- Single molecule measurement, No crystallization
- Higher resolution (in theory) but techniques not established

Structure / Dynamics Study by XFEL

There is not enough data to construct 3D model. Needs Computational Modeling

Diffractions from the samples in different orientations and (probably) different conformations

Simulated diffractions

Hypothetical conformations in different orientations

Ab initio
Deformed low resolution models
Other experimental data
etc...
Acknowledgements

• Logan Ahlstrom
• Joe Baker
• Kent Ehrlich
• Ivan Vorontsov
• Zack Campbell
• Sunita Patel
• Florence Tama

• Aqeel Ahmed
• Marek Orzechowski
• Ivan Grubisic
• Max Shokhirev
• Christian Gorba
• Slavica Jonic (CNRS)

• FOCUS Establishing Supercomputing Center of Excellence
• University of Arizona High Performance Computing Center