Understanding function by

- Recognition of functional sequences or motifs
- Analysis of different structures in different forms
- Computational analyses and simulations

Molecular dynamics (MD) simulations

- A deterministic method based on the solution of Newton's equation of motion

$$ F_i = m \ddot{x}_i $$

for the $i$th particle. The acceleration at each step is calculated from the negative gradient of the overall potential, using

$$ \ddot{x}_i = -\nabla V_i $$

$$ V_i = \sum_k \text{(energies of interactions between } i \text{ and all other residues } k \text{ located within a cutoff distance of } R_c \text{ from } i) $$

$$ \nabla V_i = \text{Gradient of potential?} $$

- Derivative of $V$ with respect to the position vector $r_i = (x_i, y_i, z_i)$ at each step

  $$ a_x = -\nabla V_x $$
  $$ a_y = -\nabla V_y $$
  $$ a_z = -\nabla V_z $$

Interaction potentials include:

- **Non-Bonded Interaction Potentials**
  - Electrostatic interactions of the form $E_i^{es} = q_i q_j / r_{ij}$
  - Van der Waals interactions $E_i^{vdW} = -a_{vdW} r_{ij}^{6} + b_{vdW} r_{ij}^{12}$

- **Bonded Interaction Potentials**
  - Bond stretching $E_i^{bs} = (k_i/2) (|r_i| - |r_{i0}|)^2$
  - Bond angle distortion $E_i^{bad} = (k_{θ_i}/2) (θ_i - θ_{i0})^2$
  - Bond torsional rotation $E_i^{tor} = (k_{ϕ_i}/2) f(\cos ϕ_i)$
Example 1: gradient of vdw interaction with respect to $\mathbf{r}_i$

- $E_{vdW}(\mathbf{r}) = -a_i |\mathbf{r}_i|^6 + b_i |\mathbf{r}_i|^{12}$
- $\mathbf{r}_i = \mathbf{r}_i - \mathbf{r}_j$
- $x_{ij} = x_i - x_j$
- $y_{ij} = y_i - y_j$
- $z_{ij} = z_i - z_j$
- $r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$

$$\nabla E_{vdW}/\nabla x_{ii} = \hat{\mathbf{e}} \cdot \left[ -a_i/r_{ij}^6 + b_i/r_{ij}^{12} \right] / \partial x_i$$

where $r_{ij}^6 = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$

Example 2: gradient of bond stretching potential with respect to $\mathbf{r}_i$

- $E_{bs}(\mathbf{r}) = \frac{1}{2} k_{bs} (l_{bi} - l_{li})^2$
- $l_{ii} = r_{i+1} - r_i$
- $x_{ii} = x_{i+1} - x_i$
- $y_{ii} = y_{i+1} - y_i$
- $z_{ii} = z_{i+1} - z_i$
- $l_{ii} = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 + (z_{i+1} - z_i)^2}$

$$\nabla E_{bs}/\nabla x_{ii} = -m_i a_{bs} (l_{bi} - l_{li}) \frac{\partial}{\partial x_i}$$

Periodic boundary conditions

The Verlet algorithm

Perhaps the most widely used method of integrating the equations of motion is that initially adopted by Verlet [1967]. The method is based on positions $\mathbf{r}(t)$, accelerations $\mathbf{a}(t)$, and the positions $\mathbf{r}(t - \delta t)$ from the previous step.

The equation for advancing the positions reads as

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta^2 \mathbf{a}(t)$$

There are several points to note about this equation. It will be seen that the velocities do not appear at all. They have been eliminated by addition of the equations obtained by Taylor expansion about $\mathbf{r}(t)$:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta \mathbf{v}(t) + \frac{1}{2!} \delta^2 \mathbf{a}(t) + \ldots$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \delta \mathbf{v}(t) + \frac{1}{2!} \delta^2 \mathbf{a}(t)$$

The velocities are not needed to compute the trajectories, but they are useful for estimating the kinetic energy (and hence the total energy). They may be obtained from the formula

$$\mathbf{v}(t) = (\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t))/2\delta t$$
Initial velocities ($v_i$)

using the Boltzmann distribution at the given temperature

$$v_i = (m_i / 2\pi kT)^{1/2} \exp (- m_i v_i^2 / 2kT)$$

How to generate MD trajectories?

1. Known initial conformation, i.e. $r_i(0)$ for all atom $i$
2. Assign $v_i(0)$, based on Boltzmann distribution at given $T$
3. Calculate $r_i(t) = r_i(0) + v_i(0) \delta t$
4. Using new $r_i(t)$ evaluate the total potential $V_i$ on atom $I$
5. Calculate negative gradient of $V_i$ to find $a_i(t) = -\nabla V_i / m_i$
6. Start Verlet algorithm using $r_i(t), r_i(t+\delta t)$ and $a_i(t)$
7. Repeat for all atoms (including solvent, if any)
8. Repeat the last three steps for ~ $10^6$ successive times (MD steps)

Limitations of MD simulations

- Full atomic representation → noise
- Empirical force fields → limited by the accuracy of the potentials
- Time steps constrained by the fastest motion (bond stretching of the order of femptoseconds)
- Inefficient sampling of the complete space of conformations
- Limited to small proteins (100s of residues) and short times (subnanoseconds)

Topology-based models

- Near native fluctuations (springs acting on effective centroids, usually C atoms)
- Ben-Avraham (1993)
- Ciccotti et al. Proteins (2005)
- Go & Scheraga (Macromolecules 1978)
- Gallo & Finkelstein, PNAS (1999)
- Klimov & Thirumalai, PNAS (2000)
- Clementi et al. (Onuchic), JMB (2000)

- Folding/unfolding processes (free energy landscape)
- Micheletti et al., JMB (2000)
- Scheraga (Science 1970)
- Gallo & Finkelstein, PNAS (1999)
- Klimov & Thirumalai, PNAS (2000)
- Clementi et al. (Onuchic), JMB (2000)

"Native topology determines force-induced unfolding pathways"
Protein folding kinetics examined by a Go-like model


Topological and Energetic Factors: What determines the transition state ensemble, and folding intermediates?


Simulations with Go-like potential

"Topology plays a central role in determining folding mechanisms"
Can we use such simplified approaches for estimating amyloidogenic intermediates?

Can we predict fluctuations dynamics from native state topology only?

Gaussian Network Model

FOR MORE INFO...

"A single parameter potential is sufficient to reproduce the slow dynamics in good detail"
**Rouse chain**

Connectivity matrix

\[
\Gamma = \begin{bmatrix}
1 & -1 & 2 & -1 \\
-1 & 2 & -1 & 2 \\
-1 & 2 & -1 & 2 \\
.. & .. & .. & .. \\
-1 & 2 & -1 & 1 \\
\end{bmatrix}
\]

\[
V_{tot} = (\gamma/2) \Delta R^T \Gamma \Delta R = \]

**Total intramolecular potential**

\[
V_{tot} = (\gamma/2) \Delta R^T \Gamma \Delta R = \]

Kirchhoff matrix of contacts

\[
\Gamma = 1 \text{ if } r_{ik} < r_{cut} \\
0 \text{ if } r_{ik} > r_{cut} \\
\Gamma_{ii} = - \sum_k \Gamma_{ik}
\]

\[
V_{m} = (\gamma/2) \Delta R^2 \Gamma \Delta R
\]

Comparison with X-ray Temperature Factors

Debye-Waller factors:

\[
B_k = 8 \pi^2 <\Delta R_k \cdot \Delta R_k> \mathbb{I}
\]

Complex energy landscape near the folded state

Complexity can be reduced by observing the subspace near the native state and adopting a low resolution approach focusing on the 'dominant' modes/pathways.


Comparison with H/D Exchange - NMR data

\[ \Delta S = k \ln W(\Delta R) = -\gamma (\Delta R)^2 / (2T [\Gamma^-]_{ii}) \]


Understanding functional dynamics

1. Comparison of predicted deformed structures with the structures available in the PDB. Identification of domains and hinge sites.


Mode shapes provide information on functional motions

- Slowest (global) modes \( \rightarrow \) function (and unfolding)
- Fastest (local) modes \( \rightarrow \) stability (and folding)

FOR MORE INFO...

Knowledge of sequence or structure does not permit us to

- Understand the mechanism of function
- Device methods of controlling/inhibiting function
- Predict the behavior in different forms, different environments
- Answer the questions ‘how’ or ‘why’!

Biological function is a dynamic process

Structure → Dynamics → Function

FOR MORE INFO

Collective motions of GroEL-GroES complex


Keskin et al., Biochemistry, 2002

State-of-the-art in computational/mathematical biology

Molecular computations

- Limited to small systems (one macromolecule) or short times (~ ns)
- Dependent on force field
- Solvent effect – a problem

Subcellular/cellular computations

- Simple mass-action kinetics
- No spatial-structural realism
- Lack of data for model parameters
Progresses in molecular approaches:
Coarse-grained approaches for large complexes/assemblies

Example: elastic network models for modeling ribosomal machinery (Frank and coworkers, 2003)

Protein dynamics
- Folding/unfolding dynamics
- Passage over one or more energy barriers
- Transitions between infinitely many conformations
- Fluctuations near the folded state
- Local conformational changes
- Fluctuations near a global minimum

I. Dynamics of living systems
- Understanding the dynamics at the molecular level.
- Understanding the dynamics at the cellular level.
- Filling the gap between these two levels

Life’s complexity pyramid

Oltvai & Barabasi, Science 2002, 298, 763-764
The complexity pyramid is not specific to cells

Kirchhoff/connectivity matrix is analogous to the Transition Rate Matrix of Master Equation Formalism

Applications to macromolecular dynamics:

ATM signalling pathways

Computational models & methods for:
- Large assemblies, complexes of proteins
- Membranes and cytosolic fibrous systems
- Cellular pathways, signaling & regulation of cell cycle
**Challenge:** to understand the long-time dynamics of large systems

**Model:** Coarse-grained

**Method:** Analysis of principal modes of motion

(Frame transformation: Cartesian $\rightarrow$ collective coordinates)

- What is the optimal (realistic, but computationally efficient) model for a given scale (length and time) of representation?
- Which level of details is needed for representing global (collective) motions?
- How much specificity do we need for modeling large scale systems and/or motions?
- What should be the minimal ingredients of a simplified (reductionist) model?