ProDy Workshop - Ivet Bahar

→ Protein Dynamics (ProDy)

Each structure encodes a unique dynamics

Structure ⇒ Dynamics ⇒ Function

motions { Global → entire structure; intrinsic dynamics ← overall structure favors
Local → Fluctuations

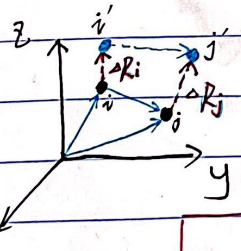
⇓

Biological Fuctions

Gaussian Network Model (GNM) ⟺ Anisotropic Network Model (ANM)

Statistical Mechanics of Polymers
Theory of Rubber Elasticity
⇒ Elastic Network Model for Proteins

GNM:
- Nodes are subject to Gaussian fluctuations ΔR_i
- Inter-residue distances R_{ij} also undergo Gaussian fluctuations ΔR_{ij}



$$\Delta R_{ij} = \Delta R_j - \Delta R_i$$

(Fluctuations in residue positions)

Fluctuation vector

$$\rightarrow \Delta R = \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \vdots \\ \Delta R_N \end{bmatrix} \quad (N \text{ residues})$$

(*) Under equilibrium conditions:

1/ Average displacement from equilibrium

vector ← $\langle \Delta R_i(t_k) \rangle = 0$ $\quad \overrightarrow{\Delta R_i(t_k)}$

2/ Mean-square fluctuation (MSF)

vector ← $\langle (\Delta R_i(t_k))^2 \rangle \neq 0$ $\quad \left| \overrightarrow{\Delta R_i(t_k)} \right|^2$

Instantaneous deviation for atom i:

$$\Delta R_i(t_k) = R_i(t_k) - R_i(0)$$

t_k: time k; $R_i(0)$: starting structure

Equilibrium conditions → structure moving while maintaining

* Proteins function under equilibrium conditions.

⇒ All the motions are accessible to the native structure (No folding)

⇒ (Equilibrium dynamics / Structural dynamics)

* A residue with small MSF	~ high MSF
⇒ tightly packed	⇒ high flexibity
⇒ conserved	⇒ easily accommodate mutations, bind substrates

⇒ Residues shown at peaks in MSFs usually locate on the surface.

⇒ likely to interact with substrates

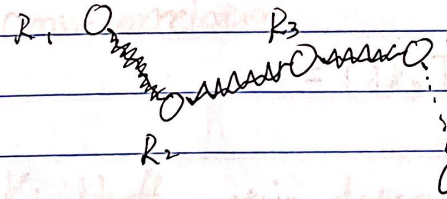
When binding sth., there's entropy loss

But energy of interaction is favourable which compensate entropy loss.

On the contrary, low MSFs ⇒ more constraint

Rouse model for polymers

Classical bead-and-spring model (only connect to first neighbour) $R_1 - R_2$ ✓ $R_2 - R_N$ ✗



$$\Delta R_{12} = R_{12} - R_{12}^{\circ}$$

total energy:

$$V_{tot} = (r/2)\left[(\Delta R_{12})^2 + (\Delta R_{23})^2 + \cdots + (\Delta R_{N-1,N})^2\right]$$

$$= (r/2)\left[(\Delta R_2 - \Delta R_1)^2 + \cdots + (\Delta R_N - \Delta R_{N-1})^2\right]$$

r: force constant

⇒ Kirchhoff matrix

$$\Gamma = \begin{array}{c} \\ R_1 \\ R_2 \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccccc} R_1 & R_2 & & & & & \end{array} \\ \begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & & & & \\ & & & \ddots & & & \\ & & & & -1 & 2 & 1 \\ & & & & & -1 & 1 \end{bmatrix} \end{array}$$

$R_i - R_j$ connected → -1

disconnected → 0

Diagonal → negative sum of off-diagonal terms

(represent Degree in graph theory)

GNM:

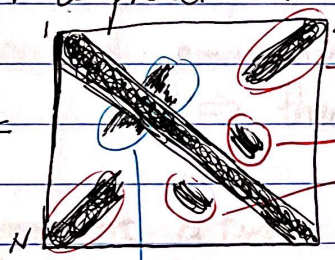
$$V_{tot} = (\gamma/2)\,\Delta R^T T \Delta R$$

$$= \underbrace{(\gamma/2)}_{\text{force constant}} \underbrace{[\Delta R_1\ \Delta R_2 \cdots \Delta R_N]}_{\text{fluctuation vector}} \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}_{\text{Kirchhoff matrix}} \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \vdots \\ \Delta R_N \end{bmatrix}$$

* Kirchhoff matrix for inter-residue contacts

T provides a complete description of contact topology!

For a protein of N residues



$T =$ → anti-parallel; → beta strand (parallel)

$$T_{ik} = \begin{cases} -1, & r_{ik} < r_{cut} \\ 0, & r_{ik} > r_{cut} \end{cases}$$

$$T_{ii} = -\sum_{k}^{k} T_{ik}$$

sth. close to diagonal ⇒ helix

Statistical mechanical averages

$$\underbrace{\langle \Delta R_i \cdot \Delta R_j \rangle}_{\text{cross-correlation}} = (1/Z_N) \int (\Delta R_i \cdot \Delta R_j) e^{-V/k_B T} \{\Delta R\} \quad \text{(weighted by probability)}$$

$$= (3k_B T/\gamma)[T^{-1}]_{ij} \rightarrow \text{inverse Kirchhoff matrix}$$

Kirchhoff matrix determines 1) MSFs

$$[T^{-1}]_{ii} \sim \langle (\Delta R_i)^2 \rangle$$

2) cross-correlations between residue motions

$$[T^{-1}]_{ij} \sim \langle (\Delta R_i \cdot \Delta R_j) \rangle$$

⇒ one residue moves, the other moves (even far away)
→ allosteric effect
intrinsic dynamics of the protein

Comparison with B factors

— B factor : deviation from the mean position (mean square fluctuation)

B factors scale with MSFs

for atom i

$$B_i = [8\pi^2/3]\langle(\Delta R_i)^2\rangle$$

↳ Inverse Kirchhoff matrix diagonal elements

⇒ check GNM

B-factors are affected by crystal contacts

Particular loop motions are curtailed by intermolecular contacts in the crystal environment causing a discrepancy between theory and experiments.

⇒ theory higher, experiment lower

* Agreement between theory and experiments upon inclusion of crystal lattice effects into the GNM.

Cross-correlations

- Provide information on the relative movements of pairs of residues
- Purely orientational correlations (correlation cosines) are obtained by normalizing cross-correlation as

$$-1 \leq \frac{\langle(\Delta R_i \cdot \Delta R_j)\rangle}{[\langle(\Delta R_i)^2\rangle\langle(\Delta R_j)^2\rangle]^{1/2}} \leq 1$$

Fully anticorrelated — Fully correlated

* White region between red and blue region ($cc = 0$)

⇒ hinge sites

Cross-correlations are elements of Covariance Matrix C

$$\Gamma^{-1} \sim C$$

Covariance scales with the inverse Kirchhoff matrix

The proportionality constant is $3kT/\gamma$

Higher connectivity ⇒ lower MSFs

Covariance matrix ($N \times N$)

$$C = \begin{bmatrix} \langle \Delta R_1 \cdot \Delta R_1 \rangle & \langle \Delta R_1 \cdot \Delta R_2 \rangle & \cdots & \langle \Delta R_1 \cdot \Delta R_N \rangle \\ \langle \Delta R_2 \cdot \Delta R_1 \rangle & & & \\ \vdots & & & \\ \langle \Delta R_N \cdot \Delta R_1 \rangle & & & \langle \Delta R_N \cdot \Delta R_N \rangle \end{bmatrix} = \Delta R \, \Delta R^T$$

ΔR = N-dim vector of instantaneous fluctuations ΔR_i for all residues ($1 \le i \le N$)

$\langle \Delta R_i \cdot \Delta R_i \rangle$ = ms (mean square) fluctuation of site i averaged over time (or all m snapshots)

Collective Motions Encoded by the Structure: **Normal Modes**

Several modes contribute to dynamics
- slowest (global) modes (most collective and softest) ⟶ function
- fastest (local) modes (at highest packing density regions) ⟶ stability

$$\begin{cases} \langle \Delta R_i \cdot \Delta R_j \rangle = \sum_k [\Delta R_i \cdot \Delta R_j]_k \\ \langle \Delta R_i \cdot \Delta R_j \rangle = (3k_B T/\gamma)[T^{-1}]_{ij} \end{cases}$$

⇒ $[\Delta R_i \cdot \Delta R_j]_k = (3k_B T/\gamma)[\lambda_k^{-1} U_k U_k^T]_{ij}$

Contribution of mode k: expresed in terms of kth eigenvalue λ_k and kth eigenvector U_k of T

* Eigenvalue decomposition of T

$T = U \Lambda U^T$ Λ: diagonal matrix of eigenvalues

U: matrix of eigenvectors

$$\Lambda = \begin{bmatrix} \lambda_0 & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_{N-1} \end{bmatrix}$$

$\lambda_0 = 0$

$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_{N-1}$

$$U = \begin{bmatrix} U_{11} & U_{21} & \cdots & U_{N1} \\ U_{12} & & & U_{N2} \\ \vdots & & & \vdots \\ U_{1N} & & & U_{NN} \end{bmatrix}$$

$U_{1N} \to U_0$, $U_{21} \to U_1$, $U_{NN} \to U_{N-1}$

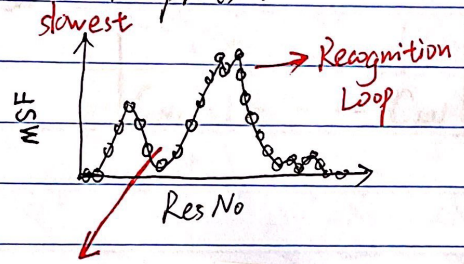
$$U^T = \begin{bmatrix} U_0^T \\ U_1^T \\ \vdots \\ U_{N-1}^T \end{bmatrix}$$

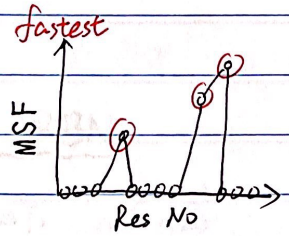
MSFs: sum of all modes

$$[\Delta R_i \cdot \Delta R_j]_k = (3k_B T/r)[\lambda_k^{-1} u_k u_k^T]_{ij}$$

$\lambda_k \downarrow \Rightarrow$ contribution $\uparrow$ (weight)
low frequency modes
soft modes

In MSF plots:





$\Rightarrow$ Individual single residues under high energy/frequency will not sustain substitutions or mutations normally
$\rightarrow$ evolutionally conserved

Active site
won't move at all in global motions
(catalytic residues position in precise way to function)

The eigenvalues scale with the frequency squared ($\lambda_i \sim \omega_i^2$)
Eigenvector u_k is an N-dim vector.
The ith element of u_k represents the displacement of node i in mode k.
The eigenvectors are normalized, i.e. $\vec{u}_k \cdot \vec{u}_k = 1$ for all k
The squared elements of u_k represent the mobility distribution
Dynamics result from the superposition of all modes.
$\lambda_k^{-\frac{1}{2}}$ serves as the weight of $u_k \rightarrow$ low frequency modes have high weights

Collective motions are functional

$$\text{Collectivity}_k = \frac{1}{N} e^{\left[-\sum_i^N u_{k,i}^2 \ln u_{k,i}^2\right]}$$

$\rightarrow$ Information entropy associated with residue fluctuations in mode k.

$k \rightarrow$ mode number $\quad i \rightarrow$ residue index

collectivity $\uparrow$ represents a more distribute mode and vice versa
Usually soft modes are highly collective.

Anisotropic Network Model

$3N \times 3N$ Hessian of ANM replaces the $N \times N$ Kirchhoff matrix of GNM
$\Rightarrow$ to yield mode shapes in $3N$-d shape
Hessian is calculated directly from structure

ANM covariance matrix ($3N \times 3N$)

$$C_{3N} = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{1N} \\ C_{12} & C_{22} & & \\ \vdots & \vdots & & \\ C_{N1} & | & --- & C_{NN} \end{bmatrix}$$

$3N \times 3N$

super element

$$\begin{matrix} \overrightarrow{\Delta X_1}\overrightarrow{\Delta X_2} & \overrightarrow{\Delta X_1}\overrightarrow{\Delta Y_2} & \overrightarrow{\Delta X_1}\overrightarrow{\Delta Z_2} \\ \overrightarrow{\Delta Y_1}\overrightarrow{\Delta X_2} & \overrightarrow{\Delta Y_1}\overrightarrow{\Delta Y_2} & \overrightarrow{\Delta Y_1}\overrightarrow{\Delta Z_2} \\ \overrightarrow{\Delta Z_1}\overrightarrow{\Delta X_2} & \overrightarrow{\Delta Z_1}\overrightarrow{\Delta Y_2} & \overrightarrow{\Delta Z_1}\overrightarrow{\Delta Z_2} \end{matrix}$$

- softest modes are functional
  The intrinsic mode tells what a protein can do. (potential)

Proteins exploit pre-exsiting soft modes for their interactions
→ Structure changes involved in protein binding correlate with intrinsic motions in the unbound state.

Allosteric changes in conformation
→ The functional movements are those predicted by the ANM to be intrinsically encoded by the structure.

Substates may be identified along soft modes.