Collective Dynamics of Biomolecules using ProDy & Elastic Network Models

Ivet Bahar

Department of \(^1\)Computational and Systems Biology
School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260
MMBioS Resources

Eyal et al., Bioinformatics 2015

Li et al., Nucleic Acids Res 2016
MMBioS Resources

What is the DynOmetics ENM server?
The DynOmetics ENM server computes biomolecular systems dynamics for user-uploaded structural coordinates or PDB identifiers, by integrating two widely used elastic network models (ENMs) – the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM). Unique features include the consideration of environment, the prediction of potential functional sites and reconstruction of all-atom conformers from deformed coarse-grained structures. For more information see Theory and Tutorial.

ProDy Project
ProDy is a free and open-source Python package for protein structural dynamics analysis. It is designed as a flexible and responsive API suitable for interactive usage and application development.

Structure analysis
ProDy has fast and flexible PDB and DCD file parsers, and powerful and customizable atom selections for contact identification, structure comparisons, and rapid implementation of new methods.

Dynamics analysis
- Principal component analysis can be performed for
  - heterogeneous X-ray structures (missing residues, mutations)
  - mixed structural datasets from Blast search
  - WiR models and MD snapshots (essential dynamics analysis)
- Normal mode analysis can be performed using
  - Anisotropic network model (ANM)
  - Gaussian network model (GNM)
  - ANM/GNM with distance and property dependent force constants

Dynamics from experimental datasets, theoretical models and simulations can be visualized.

Reference
Bakan A, Martinez LH, Bahar I. ProDy: Protein Dynamics inferred from Theory and Experiments. 2011

Funding
Continued development of ProDy is supported by NIH through R01 GM109978 award.

DynOmetics using Elastic Network Models - ENM 1.0

Bakan et al., Bioinformatics 2011; 2014

Li et al, Nucleic Acids Res 2016
ProDy
Protein Dynamics Analysis in Python


Reference:


Reference:

### ProDy: Usage and dissemination statistics

<table>
<thead>
<tr>
<th>Date</th>
<th>Releases</th>
<th>Downloads(^1)</th>
<th>Visits(^2)</th>
<th>Unique(^3)</th>
<th>Pageviews(^2)</th>
<th>Countries(^5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov’10 - Oct’11</td>
<td>19</td>
<td>8,530</td>
<td>8,678</td>
<td>2,946</td>
<td>32,412</td>
<td>45</td>
</tr>
<tr>
<td>Nov’11 - Oct’12</td>
<td>6+9*</td>
<td>35,108</td>
<td>16,472</td>
<td>6,414</td>
<td>71,414</td>
<td>59</td>
</tr>
<tr>
<td>Nov’12 - Oct’13</td>
<td>8*</td>
<td>87,909</td>
<td>19,888</td>
<td>8,145</td>
<td>86,204</td>
<td>66</td>
</tr>
<tr>
<td>Nov’13 - Oct’14</td>
<td>5*</td>
<td>140,101</td>
<td>24,134</td>
<td>11,170</td>
<td>112,393</td>
<td>69</td>
</tr>
<tr>
<td>Nov’14 - May’15</td>
<td>1*</td>
<td>68,230</td>
<td>15,941</td>
<td>8,479</td>
<td>66,641</td>
<td>50</td>
</tr>
<tr>
<td>June ’15- June’16</td>
<td>5*</td>
<td>124,613</td>
<td>32,491</td>
<td>15,402</td>
<td>140,818</td>
<td>132</td>
</tr>
<tr>
<td>June’16- June 17</td>
<td></td>
<td>31,374</td>
<td>16,201</td>
<td>129,900</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>53+</strong></td>
<td><strong>464,491+</strong></td>
<td><strong>148,978</strong></td>
<td><strong>68,757</strong></td>
<td><strong>639,782</strong></td>
<td><strong>136</strong></td>
</tr>
</tbody>
</table>

* Indicates software release made during the grant period.

1 Download statistics retrieved from PyPI ([https://pypi.python.org/pypi/ProDy](https://pypi.python.org/pypi/ProDy)) using ([https://pypi.python.org/pypi/vanity](https://pypi.python.org/pypi/vanity)).

2 Google Analytics ([www.google.com/analytics](http://www.google.com/analytics)) was used to track:

3 Unique indicates number of unique visitors;

5 Country of origin for visits.
Usage in the last year

June 1, 2016 – June 1, 2017

Google Analytics
Who? Where?

June 1, 2016 – June 1, 2017

1. United States 11,833 (37.72%)
2. India 2,415 (7.70%)
3. China 1,662 (5.30%)
4. United Kingdom 1,283 (4.09%)
5. Germany 1,157 (3.69%)
6. Brazil 1,043 (3.32%)
7. France 996 (3.17%)
8. Japan 851 (2.71%)
9. Canada 799 (2.55%)
10. Italy 755 (2.41%)
Tutorials

Day 1
http://prody.csb.pitt.edu/tutorials/

Day 2

ProDy

NMWiz

Evol

Druggability
Workshop files on ProDy website
Representation of structure as a network

Why network models?

- for large systems’ collective motions & long time processes beyond the capability of full atomic simulations
- to incorporate structural data in the models – at multiple levels of resolution
- to take advantage of theories developed in other disciplines: polymer physics, graph theory, spectral graph methods, etc.

http://www.lactamme.polytechnique.fr/
Proteins are not static: They move, breathe, work, dance, interact with each other.
Proteins are not static:
They move, breath, work, dance, interact with each other.
Many proteins are **molecular machines**

And mechanical properties become more important in complexes/assemblies

STMV dynamics (Zheng Yang)
Each structure encodes a unique dynamics

Signaling dynamics of AMPARs and NMDARs

Concerted movements of signaling molecules

GOAL: TO GENERATE DATA FOR MESOSCOPIC SCALE

Developing integrated methodology to enable information transfer across scales

Microphysiological simulations

to subcellar events

from molecules

from 6 x 6 x 5 μm³ sample of adult rat hippocampal stratum radiatum neuropil
Goal: to generate data for mesoscopic scale

**Developing integrated methodology for complex systems dynamics, to enable information transfer across scales**
Each structure encodes a unique dynamics

Structure ➔ Dynamics ➔ Function

Time scales:
- fs: femtoseconds (10^-15 s)
- ps: picoseconds (10^-12 s)
- ns: nanoseconds (10^-9 s)
- µs: microseconds (10^-6 s)
- ms: milliseconds (10^-3 s)
- s: seconds (s)

Length scales:
- 10^-11 m
- 10^-10 m
- 10^-9 m
- 10^-8 m
- 10^-7 m

Methods:
- QC/MM: Quantum Chemical/Molecular Mechanics
- Atomic simulations
- Coarse-grained computations

Cooperative machinery
- Domain/subunit motions
- Loop motions
- Sidechain motions

NMR
Summary

1. **Theory**
   a. Gaussian Network Model (GNM)
   b. Anisotropic Network Model (ANM)
   c. Resources/Servers/Databases (ProDy, DynOmics)

2. **Allosteric Changes in Structure**

3. **Ensemble analysis. Experiments vs Predictions**
   Adaptability/evolution

4. **Recent Extensions and Applications**
   a. Membrane Proteins
   b. AMPA Receptor
   c. Chromatin
Two elastic network models:

**Gaussian Network Model (GNM)**


**Anisotropic Network Model (ANM)**


Physics-based approach

- Statistical Mechanics of Polymers
- Theory of Rubber Elasticity

Elastic Network Model for Proteins

Paul J. Flory (1910-1985)
Nobel Prize in Chemistry 1974

And Pearson (1976), Eichinger (1980), Klockzkowski, Erman & Mark (1989)…
Collective motions using elastic network models (ENM)

Eigenvalue decomposition of Kirchhoff/Hessian matrix

Based on theory of elasticity for polymer networks by Flory, 1976


Gaussian Network Model (GNM)

- Each node represents a residue
- Residue positions, $R_i$, identified by $\alpha$-carbons’ coordinates
- Springs connect residues located within a cutoff distance (e.g., 10 Å)
- Nodes are subject to Gaussian fluctuations $\Delta R_i$
- Inter-residue distances $R_{ij}$ also undergo Gaussian fluctuations
  $\Delta R_{ij} = \Delta R_j - \Delta R_i$

Fluctuations in residue positions

Bahar, Atilgan & Erman, *Fold & Des* 1997
Gaussian Network Model (GNM)

Fluctuation vector:

\[ \Delta \mathbf{R} = \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \Delta R_3 \\ \vdots \\ \Delta R_{N} \end{bmatrix} \]

Fluctuations in residue positions

Bahar, Atilgan & Erman, *Fold & Des* 1997
\[ \Delta R_{ij} = \Delta R_j - \Delta R_i \]
Fluctuation
with respect to starting structure $R(0)$

**Instantaneous deviation for atom $i$**

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

**Under equilibrium conditions:**

Average displacement from equilibrium: $< \Delta R_i(t_k)> = 0$

But the mean-square fluctuation (MSF), $< (\Delta R_i(t_k))^2> \neq 0$
Rouse model for polymers

Classical bead-and-spring model

Kirchhoff matrix

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

Force constant

\[ \Delta R_{12} = R_{12} - R_{12}^0 \]

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

\[ = (\gamma / 2) \left[ (\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \ldots \right] \]
Rouse model for polymers

Kirchhoff matrix

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

Force constant

\[ V_{\text{tot}} = \left( \frac{\gamma}{2} \right) \left[ (\Delta R_{12})^2 + (\Delta R_{23})^2 + \ldots \ldots (\Delta R_{N-1,N})^2 \right] \]

\[ = \left( \frac{\gamma}{2} \right) \left[ (\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \ldots \ldots \right] \]
Rouse model for polymers

Fluctuation vector

\[
\frac{\gamma}{2} \begin{bmatrix} \Delta R_1 & \Delta R_2 & \Delta R_3 & \ldots & \Delta R_N \end{bmatrix}
\]

Kirchhoff matrix

\[
\begin{bmatrix}
1 & -1 & & & \\
-1 & 2 & -1 & & \\
& -1 & 2 & -1 & \\
& & & \ddots & \\
& & & & -1 & 2 & -1
\end{bmatrix}
\]

Force constant

\[
V_{\text{tot}} = \frac{\gamma}{2} \begin{bmatrix} \Delta R_1 \\
\Delta R_2 \\
\Delta R_3 \\
\vdots \\
\Delta R_N \end{bmatrix}
\]

\[
V_{\text{tot}} = \frac{\gamma}{2} \left[ (\Delta R_{12})^2 + (\Delta R_{23})^2 + \ldots \ldots + (\Delta R_{N-1,N})^2 \right]
\]

\[
= \frac{\gamma}{2} \left[ (\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \ldots \ldots \right]
\]
Kirchhoff matrix for inter-residue contacts

For a protein of $N$ residues

\[
\Gamma = \begin{cases} 
-1 & \text{if } r_{ik} < r_{\text{cut}} \\
0 & \text{if } r_{ik} > r_{\text{cut}} 
\end{cases}
\]

$\Gamma_{ii} = -\sum_k \Gamma_{ik}$

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

$\Gamma$ provides a complete description of contact topology!
Statistical mechanical averages

For a protein of $N$ residues

\[
\langle \Delta R_i \cdot \Delta R_j \rangle = \frac{1}{Z_N} \int (\Delta R_i \cdot \Delta R_j) e^{-V/k_BT} d\{\Delta R\}
\]

\[
= (3 k_B T / \gamma) [\Gamma^{-1}]_{ij}
\]

$\Gamma$ provides a complete description of contact topology!
Kirchhoff matrix determines the mean-square fluctuations

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

And cross-correlations between residue motions

$$\left[ \Gamma^{-1} \right]_{ij} \sim \langle (\Delta R_i \cdot \Delta R_j) \rangle$$
Comparison with B factors

- X-ray crystallographic structures deposited in the PDB also report the B-factors (Debye-Waller factors) for each atom, in addition to atomic coordinates

- B-factors scale with mean-square fluctuations (MSFs), i.e. for atom i,

\[ B_i = \left[ \frac{8\pi^2}{3} \right] \langle (\Delta R_i)^2 \rangle \]

How do residue MSFs compare with the B-factors?
Output from DynOmics

Example: 1vaa

PDB title: CRYSTAL STRUCTURES OF TWO VIRAL PEPTIDES IN COMPLEX WITH MURINE MHC CLASS I H-2KB
Output from DynOmics

Correlation: 0.72

Theoretical and Experimental B-Factors

The effective force constant of the GNM springs is $9.4652e-01 \ k_B T \AA^{-2}$, and corresponding rescaling prefactor is 83.4180.
B-factors are affected by crystal contacts

Two X-ray structures for a designed sugar-binding protein LKAMG
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.
Agreement between theory and experiments upon inclusion of crystal lattice effects into the GNM

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

B- factors – Comparison with experiments

Intradimer cooperativity – Symmetry rule (Yuan et al. JMB 2002; Ackers et al. PNAS 2002.)

Cross-correlations

- Provide information on the relative movements of pairs of residues

- Purely orientational correlations (correlation cosines) are obtained by normalizing cross-correlations as

\[
-1 \leq \frac{\langle (\Delta R_i . \Delta R_j) \rangle}{\sqrt{\langle (\Delta R_i)^2 \rangle \langle (\Delta R_j)^2 \rangle}} \leq 1
\]

- Fully anticorrelated
- Fully correlated
Output from iGNM

Li, Chang, Yang and Bahar (2016)
Nucleic Acids Res 44: D415-422
Output from DynOmics - ENM

Cross-Correlations may be organized in a Covariance Matrix \( C \)

\[ \Gamma^{-1} \sim C \]

Covariance scales with the inverse of the Kirchhoff matrix.

The proportionality constant is \( \frac{3kT}{\gamma} \)
### Covariance matrix (NxN)

\[
\mathbf{C} = \begin{pmatrix}
\langle \Delta \mathbf{R}_1 \cdot \Delta \mathbf{R}_1 \rangle & \langle \Delta \mathbf{R}_1 \cdot \Delta \mathbf{R}_2 \rangle & \ldots & \ldots & \langle \Delta \mathbf{R}_1 \cdot \Delta \mathbf{R}_N \rangle \\
\langle \Delta \mathbf{R}_2 \cdot \Delta \mathbf{R}_1 \rangle & \langle \Delta \mathbf{R}_2 \cdot \Delta \mathbf{R}_2 \rangle & & & \\
\vdots & \vdots & & & \\
\langle \Delta \mathbf{R}_N \cdot \Delta \mathbf{R}_1 \rangle & & & & \langle \Delta \mathbf{R}_N \cdot \Delta \mathbf{R}_N \rangle \\
\end{pmatrix}
\]

\[\mathbf{C} = \Delta \mathbf{R} \Delta \mathbf{R}^T\]

\[\Delta \mathbf{R} = \text{N-dim vector of instantaneous fluctuations } \Delta \mathbf{R}_i \text{ for all residues } (1 \leq i \leq N)\]

\[\langle \Delta \mathbf{R}_1 \cdot \Delta \mathbf{R}_1 \rangle = \text{ms fluctuation of site 1 averaged over all } m \text{ snapshots.}\]
Collective Motions Encoded by the Structure: Normal Modes
Several modes contribute to dynamics

\[ \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \sum_k \left[ \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \right]_k \]

\[ \langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = (3k_B T / \gamma) \left[ \Gamma^{-1} \right]_{ij} \]

Contribution of mode \( k \)

\[ \left[ \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \right]_k = (3k_B T / \gamma) \left[ \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T \right]_{ij} \]

expressed in terms of kth eigenvalue \( \lambda_k \) and kth eigenvector \( \mathbf{u}_k \) of \( \Gamma \)

Bahar et al. (1998) Phys Rev Lett. 80, 2733
Several modes contribute to dynamics

The first mode selects the ‘easiest’ collective motion
Output from DynOmics
Output from DynOmics

Mobility scale for slow modes (→ increase)

The highest energy residues (hotspots) for fast modes are colored red.

Mode shapes

Residue index

Hide/show: all chains ▼ slow modes 1-2 slow modes 1-3 slow modes 1-10 fast modes 1-10

Hide/show: slow modes ▼ all chains ▼ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Hide: All Chain A Chain B Chain P

Export: PNG JPEG SVG PDF CSV

Click a point on the 2D chart to show/hide the corresponding labels in both the 2D chart and the 3D windows above if the "Chart Control" is
Summary - Gaussian network model (GNM)

Several modes of motion contribute to dynamics

Contact: \( R_{ij} < 10\text{Å} \)

\[
\Gamma = \begin{bmatrix}
\Delta R_i \cdot \Delta R_i \end{bmatrix}_k = \left(3k_B T / \gamma \right) \lambda_k^{-1} u_k u_k^T
\]

MSF of residue \( i \)

\[
< (\Delta R_i)^2 > = \left(3 k_B T / \gamma \right) \left[ \Gamma^{-1} \right]_{ii}
\]
Recipe (GNM)

- Obtain the coordinates of network nodes from the PDB
- Write the corresponding Kirchhoff matrix $\Gamma$
- Eigenvalue decomposition of $\Gamma$ yields the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{N-1}$ (and $\lambda_0 = 0$) and eigenvectors $u_1, u_2, u_3, \ldots, u_{N-1}$ (and $u_0$)

Properties

- the eigenvalues scale with the frequency squared ($\lambda_i \sim w_i^2$)
- eigenvector $u_k$ is an N-dim vector
- the $i^{th}$ element of $u_k$ represents the displacement of node $i$ in mode $k$
- the eigenvectors are normalized, i.e. $u_k \cdot u_k = 1$ for all $k$
- as such, the squared elements of $u_k$ represent the ‘mobility’ distribution
- dynamics results from the superposition of all modes
- $\lambda_k^{-1/2}$ serves as the weight of $u_k \Rightarrow$ low frequency modes have high weights
Database of GNM results

ignm.ccbb.pitt.edu

Li, Chang, Yang and Bahar (2016)
*Nucleic Acids Res* **44**: D415-422
Easy access to precomputed results for 95% of the PDB including
- the largest structures beyond the scope of MD
- protein-DNA/RNA complexes
- biological assemblies (intact, biologically functional structures)

Easy to understand, visualize, make functional inferences for any structure

13.9% of the structures in the iGNM 2.0 (14,899 out of 107,201) contain $>10^3$ nodes

The biological assembly of 39,505 PDB structures is different from the default structure reported in the PDBs (as asymmetric unit)
Collective motions are functional

Collectivity (2D) for a given mode $k$ is a measure of the degree of cooperativity (between residues) in that mode, defined as (*)

$$
Collectivity_k = \frac{1}{N} \sum_{i} \frac{u_{k,i}^2}{\ln u_{k,i}}
$$

where, $k$ is the mode number and $i$ is the residue index. A larger collectivity value refers to a more distributive mode and vice versa. Usually soft modes are highly collective.

Anisotropic Network Model (ANM)

\[ H = \sum_{i=1}^{3N-6} \lambda_i \mathbf{u}_i \mathbf{u}_i^T \]

\[ H^{(ij)} = \frac{y \Gamma_{ij}}{(R_{ij}^0)^2} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix} \]

ANM mode 1 \( (\mathbf{u}_1) \)

Mode 2

Mode 20

\( \lambda_1 < \lambda_2 < \lambda_3 < \ldots \)

Anisotropic Network Model

\[ V(\mathbf{r}) = \frac{\gamma}{2} \sum_{i=1}^{N} \sum_{j>i} (|\mathbf{r}_{ij}|^2 - |\mathbf{r}_{ij}^0|^2) \Theta(R_c - |\mathbf{r}_{ij}^0|) \]

Hessian

\( \begin{pmatrix} \frac{\partial^2 V}{\partial x_i \partial y_j} \end{pmatrix}_{\mathbf{r}_0} = -\frac{x_i^0 y_j^0}{|\mathbf{r}_{ij}^0|^2} \)  

Harmonic  Step function

Hessian is calculated directly from structure

\[ H_{ij} = -\frac{\gamma}{(R_j^0)^2} \begin{bmatrix} (x_i^0)^2 & x_i^0 y_i^0 & x_i^0 z_i^0 \\
 x_i^0 y_i^0 & (y_i^0)^2 & y_i^0 z_i^0 \\
x_i^0 z_i^0 & y_i^0 z_i^0 & (z_i^0)^2 \end{bmatrix} \]

3N x 3N Hessian of ANM replaces the NxN Kirchhoff matrix of GNM – to yield mode shapes in 3N-d space

ANM covariance matrix (3Nx3N)

\[
C_{3N} = \begin{bmatrix}
C_{11} & C_{21} & C_{13} & C_{1N} \\
C_{12} & C_{22} & & \\
C_{N1} & & C_{NN} & \\
\end{bmatrix}
\]

\[
\begin{array}{ccc}
\langle \Delta X_1 \Delta X_2 \rangle & \langle \Delta X_1 \Delta Y_2 \rangle & \langle \Delta X_1 \Delta Z_2 \rangle \\
\langle \Delta Y_1 \Delta X_2 \rangle & \langle \Delta Y_1 \Delta Y_2 \rangle & \langle \Delta Y_1 \Delta Z_2 \rangle \\
\langle \Delta Z_1 \Delta X_2 \rangle & \langle \Delta Z_1 \Delta Y_2 \rangle & \langle \Delta Z_1 \Delta Z_2 \rangle \\
\end{array}
\]

3N x 3N
ANM server

http://anm.csb.pitt.edu/cgi-bin/anm2/anm2.cgi
Output from ANM server
Softest modes are functional

Experiments

open

closed

tense (T)

relaxed (R)


What is the DynOomics ENM server?
The DynOomics ENM server computes biomolecular systems dynamics for user-uploaded structural coordinates or PDB identifiers, by integrating two widely used elastic network models (ENMs) – the Gaussian Network Model (GNM) and the Anisotropic Network Model (ANM). Unique features include the consideration of environment, the prediction of potential functional sites and reconstruction of all-atom conformers from deformed coarse-grained structures. For more information see Theory and Tutorial.
New features

- sensors and effectors
- first passage times for signaling
- mechanically functional sites
- effect of oligomerization
- coupling to membrane

Dynamics of Structural Proteomics and Beyond

Plan

1. Theory
   a. Gaussian Network Model (GNM)
   b. Anisotropic Network Model (ANM)
   c. Resources/Servers/Databases (ProDy, DynOomics etc)

2. Allosteric Changes in Structure

3. Ensemble analysis. Experiments vs Predictions
   Adaptability/evolution

4. Recent Extensions and Applications
   a. Membrane Proteins
   b. AMPA Receptors
   c. Chromatin
Allosteric changes in conformation

Elastic Network Models are particularly useful for exploring the cooperative motions of large multimeric structures.

Comparison with experimental data shows that the functional movements are those predicted by the ANM to be intrinsically encoded by the structure.
Proteins exploit pre-existing soft modes for their interactions

Structural changes involved in protein binding correlate with intrinsic motions in the unbound state

Substates may be identified along soft modes

Hybrid ANM/MD methods include atomic details and specificity
Allosteric dynamics of GroEL

Passage between the R and T states

Z Yang, P Marek and I Bahar, PLoS Comp Biology 2009
What is the overlap between computations and experiments?

**Computations**

ANM yields a series of $3N$ dimensional deformation vectors

Mode 1 (slowest mode)
Mode 2
Mode 3
....
Mode $3N-6$ (fastest mode)

Given by eigenvectors $u_1, u_2, u_3, ....u_{3N-6}$, with respective frequencies of $\lambda_1, \lambda_2, \lambda_3, .... \lambda_{3N-6}$

$\mathbf{d} = [\Delta x_1 \; \Delta y_1 \; \Delta z_1 \; ... \; \Delta z_N]^T$
What is the overlap between computations and experiments?

Correlation cosine between $u_k$ and $d$

$$d = [\Delta x_1 \ \Delta y_1 \ \Delta z_1 \ \ldots \ \Delta z_N]^T$$

Z Yang, P Marek and I Bahar, PLoS Comp Biology 2009
The softest mode enables the passage $R \rightarrow T$
(with a correlation of 0.81)

$$d = [\Delta x_1, \Delta y_1, \Delta z_1, \ldots, \Delta z_N]^T$$
Mutations may stabilize conformers along soft modes – which may be impair function

E461 mutant is a deformed structure along mode 1

E461K mutation causes disruption of inter-ring transfer of ATP-induced signal (Sewell et al NSB 2004)
Plan

1. **Theory**
   a. Gaussian Network Model (GNM)
   b. Anisotropic Network Model (ANM)
   c. Resources/Servers/Databases (ProDy, DynOmics etc)

2. **Allosteric Changes in Structure**

3. **Ensemble analysis. Experiments vs Predictions**
   Adaptability/evolution

4. **Recent Extensions and Applications**
   a. Membrane Proteins
   b. AMPA Receptors
   c. Chromatin
A better comparison:

Consider more than 2 end points for a given structure, but all the known structures for a given protein, or the structurally resolved ensemble of structures.
Dynamics inferred from known structures

Comparison of static structures available in the PDB for the same protein in different form has been widely used is an indirect method of inferring dynamics.

Different structures resolved for HIV-1 reverse transcriptase (RT)
Ensembles of structures

- Structural changes accompanying substrate (protein) binding

- Structural changes induced by, or stabilized upon, ligand binding

Ubiquitin
140 structures
1732 models
Ensembles of structures

- Structural changes accompanying substrate (protein) binding
- Structural changes induced by, or stabilized upon, ligand binding

Ubiquitin
140 structures
1732 models

p38 MAP kinase
(182 structures)

p38 inhibitors
Ensembles of structures

- Structural changes accompanying substrate (protein) binding
- Structural changes induced by, or stabilized upon, ligand binding
- Alternative conformations sampled during allosteric cycles

Yang et al. PLoS Comp Biol 2009
Ensembles of structures

- Structural changes accompanying substrate (protein) binding
- Structural changes induced by, or stabilized upon, ligand binding
- Alternative conformations sampled during allosteric cycles

What is the overlap between computations and experiments?

Correlation cosine between $u_k$ and $d$

$$d = [\Delta x_1 \Delta y_1 \Delta z_1 \ldots \Delta z_N]^T$$
The softest mode enables the passage \( R \rightarrow T \) (with a correlation of 0.81)

\[
d = [\Delta x_1, \Delta y_1, \Delta z_1, \ldots, \Delta z_N]^T
\]
Global motions inferred from theory and experiments

- PCA of the ensemble of resolved structures
- ANM analysis of a single structure from the ensemble
Global motions inferred from theory and experiments

The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding

Ahmet Bakan and Ivet Bahar

Department of Computational Biology, School of Medicine, University of Pittsburgh, 3064 BST3, 3501 Fifth Avenue, Pittsburgh, PA 15213

Reference:

Bakan & Bahar (2009) PNAS 106, 14349-54
What is Ensemble Analysis?

**Input:**
An ensemble of structures for a given protein
- NMR models (~40)
- X-ray structures resolved under different conditions (ligand-bound/unbound, different stages of molecular machinery or transport cycle
- MD snapshots/frames

**Output:**
Principal modes of conformational variations/differences between NMR models
- Rearrangements/changes under different functional states
- Dynamics/fluctuations observed in simulations

Principal component analysis
What is Ensemble Analysis?

**Method:**
- Superimpose of the structures
- Evaluate the covariance matrix (differences between individual coordinates and mean coordinates)
- Decompose it into a series of modes of covariance (3N-6 eigenvectors)

**Output:**
Principal modes of conformational variations/differences between NMR models
- rearrangements/changes under different functional states
- dynamics/fluctuations observed in simulations
Covariance matrix (NxN)

\[ C = \begin{bmatrix}
<\Delta R_1 \cdot \Delta R_1> & <\Delta R_1 \cdot \Delta R_2> & \ldots & \ldots & <\Delta R_1 \cdot \Delta R_N> \\
<\Delta R_2 \cdot \Delta R_1> & <\Delta R_2 \cdot \Delta R_2> & \ldots & \ldots & \ \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
<\Delta R_N \cdot \Delta R_1> & \ldots & \ldots & <\Delta R_N \cdot \Delta R_N> \\
\end{bmatrix} = \Delta \mathbf{R} \Delta \mathbf{R}^T \]

\( \Delta \mathbf{R} = \) N-dim vector of instantaneous fluctuations \( \Delta R_i \) for all residues \( 1 \leq i \leq N \)

\( <\Delta R_1 \cdot \Delta R_1> = \) ms fluctuation of site 1 averaged over all \( m \) snapshots.
Covariance matrix (3Nx3N)

\[
C_{3N} =
\begin{pmatrix}
C_{11} & C_{21} & C_{13} & C_{1N} \\
C_{12} & C_{22} & & \\
C_{N1} & & & C_{NN}
\end{pmatrix}
\]

\[
\begin{align*}
<\Delta X_1 \Delta X_2> & \quad <\Delta X_1 \Delta Y_2> & \quad <\Delta X_1 \Delta Z_2> \\
<\Delta Y_1 \Delta X_2> & \quad <\Delta Y_1 \Delta Y_2> & \quad <\Delta Y_1 \Delta Z_2> \\
<\Delta Z_1 \Delta X_2> & \quad <\Delta Z_1 \Delta Y_2> & \quad <\Delta Z_1 \Delta Z_2>
\end{align*}
\]

3N x 3N
Principal Component Analysis (PCA)

$$\mathbf{C}^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta x_j \rangle & \langle \Delta z_i \Delta y_j \rangle & \langle \Delta z_i \Delta z_j \rangle \end{bmatrix}$$

$$\mathbf{C} = \mathbf{P} \mathbf{S} \mathbf{P}^T = \sum_{i=1}^{3N} \sigma_i \mathbf{p}_i \mathbf{p}_i^T$$
Induced Dynamics or Intrinsic Dynamics?

Experiments

- 1HQE
- 1N6Q
- 1VRT

Theory

References:


PC1 (Å)

ANM 2 (Å)

r = 0.99

http://www.youtube.com/watch?v=IOUzdzm68YY
Soft modes enable **functional** movements

Experiments

ANM 2 (Å) vs. PC1 (Å)

Thumb

RNase H

Fingers

1HQE
1N6Q
1VRT

References:


http://www.youtube.com/watch?v=1OUzdzm68YY
Experimental structures (for a given protein) are mainly variants along soft modes

Pre-existing paths

Meireles L, Gur M, Bakan A, Bahar I (2011) Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins *Protein Science* 20: 1645-58.
User inputs a protein sequence

ProDy identifies, retrieves, aligns, and analyzes (PCA) structures that match the input sequence

ProDy-ANM sampling of conformational space is more complete than that of MD

User can sample an ensemble of conformations along ANM modes for docking simulations

User can compare experimental and theoretical models

500,000+ downloads

Source: http://www.google.com/analytics/

Bakan & Bahar, PSB 2011, 181-192
Major advantages of ProDy:

- Simplicity
- Visualizing the global dynamics
- Applicability to large systems
- Assessing cooperative motions
- Efficiency – immediate results
- Relevance to observables, to functional mechanisms & allostery
Disadvantages

- Low resolution approach
- No specific interactions
- Lack of atomic details
- Linear theory — applicable near energy minimum
- Requires structural data — not a tool for structure prediction
Co-MD: Guiding MD simulations by ANM modes

coMD trajectories proceed along the minima of free energy landscape

ANM-guided transition pathways

Session I: Plotting \(<(\Delta R_i)^2>\) and contributions of selected modes

- from prody import *
- from pylab import *
- anm = calcANM('1cot', selstr='calpha')
- anm, cot = calcANM('1cot', selstr='calpha')
- anm
- cot
- figure()
- showProtein(cot)

- figure()
- showSqFlucts(anm)

- figure()
- showSqFlucts(anm[:10])

- figure()
- showSqFlucts(anm[:10], label='10 modes')

Application to cytochrome c PDB: 1cot
A protein of 121 residues
Session 2: Viewing color-coded animations of individual modes

- `writeNMD('cot_anm.nmd', anm, cot)`
- `Start VMD`
- `select Extensions → Analysis → Normal Mode Wizard`
- `Select ‘Load NMD File’`
Session 3: Cross-correlations $\langle (\Delta R_i \cdot \Delta R_j) \rangle$ between fluctuations

- `cross_corr = calcCrossCorr?`
- `cross_corr = calcCrossCorr(anm[0])`
- `figure()`
- `showCrossCorr(anm[0])`
Session 4:
Viewing cross-correlations using VMD

- writeHeatmap('anm_cross1.hm', cross_corr)
- VMD – Load file
- Select cot_anm.nmd (from your local folder)
- Load HeatMap
- open anm_cross1.hm (from your local folder)
Support from NIGMS, NLM, NIDDK & NIAID

Acknowledgment

Dr. Eran Eyal
Cancer Research Institute
Sheba Medical Center, Israel

Support from NIGMS, NLM, NIDDK & NIAID

Acknowledgment

Dr. Ahmet Bakan
Kabbage, Inc.

Dr. Ying Liu
Google, Inc.

Dr. Tim Lezon
Comp & Systems Biol,
U of Pittsburgh

Anindita Dutta
Agilent Technologies

Dr. Lee-Wei Yang
Assoc Prof,
Tsinghua U,
Taiwan

Dr. Karolina Mikulska-Dr. Hongchun Li
Ruminska

Dr. James Krieger

Cihan Kaya
Dr. JiYoung Lee
She (John) Zhang