Collective Dynamics of Biomolecules using ProDy & ENMs - II

Ivet Bahar

Department of Computational and Systems Biology
School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260
Leveraging the PDB since 2010

- High-throughput analysis of structural data
- Application Programming Interface (API) for development of tools
- Suitable for interactive usage

User inputs a sequence

Usage example

>1A9U:A | PDBID | CHAIN
GSSHHHHHHHSGLVPRGSHMSQ
ERPTFYRQELNKTIWEVPERYQ
NLSPVGSQAYGSVCAAFDTK TG
......

ProDy identifies, retrieves, aligns, and analyzes (PCA) structures matching input sequence

User can
- Compare experimental and theoretical models
- Sample conformations along normal modes

Experimentally observed structural changes are usually reconfigurations along soft modes

- Correlation cosine of $0.75 \pm 0.15$ between one of the softest modes and the experimentally observed change in structure

- Significant decrease in RMSD between the endpoints upon moving along a single soft mode (out of $3N-6$ modes)

An Interactive Tool

Languages
- Python: 80%
- C: 13%
- Other: 7%

More at Ohloh

IPython: Interactive Computing

Matplotlib NumPy
Suite of tools

- **ProDy**: Protein Dynamics & Sequence Analysis
- **Evol**: Bridging Evolution & Dynamics
- **DruGUI**: Druggability Suite for VMD
- **NMWiz**: Normal Mode Wizard for VMD

**Elastic Network Model** (ANM/GNM) Analysis
Principal component analysis of experimentally resolved structures

**Multiple Sequence Alignment**
Sequence conservation
Correlated Mutation

**Computational Drug Discovery**
Binding Site Prediction
Affinity Estimation

**A VMD plugin**
Visualization of collective motions
Animations/movies
Suite of tools

- membrANM: Membrane Anisotropic Network Model
  - Modeling coupled protein-lipid dynamics
  - Useful for membrane proteins

- MechStiff: Mechanical Stiffness Calculations
  - Response to external forces
  - Identification of mechanical stiffness

- coMD: Collective Molecular Dynamics
  - ENM guided MD simulations
  - Efficient sampling of energy landscape
Tutorials: ProDy & Structure Analysis

- Retrieving PDB Files
- BLAST-Searching the PDB
- Constructing Biomolecular Assemblies
- Determining functional motions
- Aligning and Comparing Structures
- Identifying Intermolecular Contacts
Tutorial: Elastic Network Models

- Gaussian Network Model (GNM)
- Anisotropic Network Model (ANM)
Tutorial: Trajectory Analysis

- Fast processing of long trajectories
- Enables comparison of MD trajectories and ENM predictions
Tutorial: Trajectory Analysis

- NMR Models
- Homologous Proteins
- Multiple X-ray Structures
- Multimeric Proteins
ANM for membrane proteins: membrANM

- Evaluating membrane proteins’ dynamics in the presence of lipid bilayer
- Comparing global motions in the presence and absence of membrane
- Understanding mechanisms of protein-membrane remodeling

Implemented in ProDy and DynOmics
MembrANM for γ-secretase

The effect of the membrane environment can be incorporated in the ANM analysis by adopting envANM-membrANM or Substructure-membrANM.

The MembrANM results for the γ-secretase (PDB ID: 5FN2) can be viewed on the ENM server:

The movies in PyMol session format for ANM motions can be downloaded from the DynOomics ENM server by generating “Full Atomic Structures for ANM-Driven Conformers”.

http://enm.pitt.edu/oGNM_ANM.php?gnm_id=5FN2&slice_mem=1&sub_all=1
MembrANM for γ-secretase

The effect of the membrane environment can be incorporated in the ANM analysis by adopting envANM-membrANM or Substructure-membrANM.

The MembrANM results for the γ-secretase (PDB ID: 5FN2) can be viewed on the ENM server:

The movies in PyMol session format for ANM motions can be downloaded from the DynOomics ENM server by generating “Full Atomic Structures for ANM-Driven Conformers”.

http://enm.pitt.edu/oGNM_ANM.php?gnm_id=5FN2&slice_mem=1&sub_all=1
Collective dynamics of AMPA receptors

Experimentally verified by cross-linking experiments. Substitution of cysteines in the presence of an oxidizing agent promotes cross-linking, which in turn inhibits AMPAR function.

Glutamate transporter - $\text{Glt}_{\text{ph}}$ structure
C-terminal core - functional region in each subunit

N-terminal Cylinder (gray; TM1-TM6)

C-terminal Core (HP1-TM7-HP2-TM8)
Opening of HP2 loop confirmed by X-ray structure

TBOA: DL-threeo-beta-Benzyloxyaspartate, potent blocker for EAAT

Reference:
Alternating-access model

Reference

Global motions viewed from the EC side
ANM global modes enable neurotransmission

Jiang et al., PNAS 2011

Glutamate uptake blocked by disulfide cross-linking between ‘distant’ pairs

Susan Amara
Global transitions

Glutamate transporters – trimeric membrane proteins, each subunit composed of two domains: **scaffolding** + **transport** domains

Outward Facing (OF)  
Inward Facing (IF)

En bloc movement of transport core

Global transitions

Transport domain undergoes elevator-like motions, while the trimerization scaffolding domains is rigidly affixed to the membrane.

Single subunit showing the transport domain moving across the membrane.
Substates may be identified along soft modes.

Is this transition along a soft mode?
Correlation between soft modes & observed structural change

1. Evaluate of the deformation vector $d_{3N}$

2. Calculate the correlation cosine $\cos(u_i, d)$ between each mode directional vector, $u_i$, and $d$

3. Cumulative overlap $= \left[ \sum_i \cos^2(u_i, d) \right]^{1/2}$

Reference:
Correlation between soft modes & observed structural change

Correlation cosine between $u_k$ and $d$

$d = [\Delta x_1 \ \Delta y_1 \ \Delta z_1 \ \ldots \ \Delta z_N]^T$
Correlation between soft modes & structural changes

1. Evaluate of the deformation vector $\mathbf{d}_{3N}$

2. Calculate the correlation cosine $\cos(u_i, d)$ between each mode directional vector, $u_i$, and $d$

3. Cumulative overlap $= \left[ \sum i \cos^2(u_i, d) \right]^{1/2}$

Reference:
OF ↔ IF transition is enabled by a single ANM mode

Overlap of > 0.6 achieved with a single mode!

Softest nondegenerate mode out of > 3,000 modes

EnM global motions

Membrane facilitates alternating access to EC/IC regions

Overlap of > 0.6 achieved with a single mode!

Softest nondegenerate mode out of > 3,000 modes

Reference:
Lezon TR, Bahar I. (2012) Constraints imposed by the membrane selectively guide the alternating access dynamics of the glutamate transporter Glt

Membrane facilitates alternating access to EC/IC regions

Overlap of > 0.6 achieved with a single mode!

Softest nondegenerate mode out of > 3,000 modes

ENM global motions

If the predicted modes were ‘random’, each mode would contribute by $1/3N$ to the cumulative overlap, i.e.

$$\cos(u_k \cdot d_{\text{exp}}) = 1/3N^{1/2} = 0.0167$$

Reference:
As the environment fluctuates randomly, the effective motion of the system is given by

\[ V_{\text{eff}}(s) = \frac{1}{2} \Delta s^T \left( H^{ss'} \right) \Delta s \]

\[ H^{ss'} = H^{ss} - H^{SE} \left( H^{EE} \right)^{-1} H^{ES} \]
Exploring structural transitions: Glutamate transporter

ANM predicts large radial motions of the trimer. Can we design a better model?

\[
H_{ij} = -\frac{\gamma}{(R_{ij}^0)^2} \begin{bmatrix}
(x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & x_{ij}^0 z_{ij}^0 \\
x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & y_{ij}^0 z_{ij}^0 \\
x_{ij}^0 z_{ij}^0 & y_{ij}^0 z_{ij}^0 & (z_{ij}^0)^2
\end{bmatrix}
\]

Altered radial force constants:

\[
H_{ij} = -\left(R_{ij}^0\right)^2 \begin{bmatrix}
(x_{ij}^0 \sqrt{\gamma_x})^2 & x_{ij}^0 y_{ij}^0 \sqrt{\gamma_x \gamma_y} & x_{ij}^0 z_{ij}^0 \sqrt{\gamma_x \gamma_z} \\
x_{ij}^0 y_{ij}^0 \sqrt{\gamma_x \gamma_y} & (y_{ij}^0 \sqrt{\gamma_y})^2 & y_{ij}^0 z_{ij}^0 \sqrt{\gamma_y \gamma_z} \\
x_{ij}^0 z_{ij}^0 \sqrt{\gamma_x \gamma_z} & y_{ij}^0 z_{ij}^0 \sqrt{\gamma_y \gamma_z} & (z_{ij}^0 \sqrt{\gamma_z})^2
\end{bmatrix}
\]

\[
H_{ij} = -\frac{\gamma}{(R_{ij}^0)^2} \begin{bmatrix}
(x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & cx_{ij}^0 z_{ij}^0 \\
x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & cy_{ij}^0 z_{ij}^0 \\
ix_{ij}^0 z_{ij}^0 & cy_{ij}^0 z_{ij}^0 & (cz_{ij}^0)^2
\end{bmatrix}
\]
Exploring global transitions: Glutamate transporter

ANM: Large radial motions

imANM

Lezon & Bahar, Biophys J 2012
TM3 and TM6 form an integral part of the transport core (TM7, TM8, HP1 and HP2).

MembrANM - DynOmics
Coupled motions of AMPAR and lipid bilayer

Motions without membrane

- Mode4
- mode6
- mode8
- New mode

Motions with membrane

- mode5
- mode9
- mode10
Comparison of mode shapes

with membrane (ANM-mem)

without membrane (ANM-iso)
Other ENM applications

Pulling

Transitions

Steered MD

Ensemble Analysis

Environmental Effects

Model Optimization
The mechanical response of proteins

- Probed by AFM experiments
- Usually involves partial or total unfolding
- Depends on
  - the application points of tension
  - or the direction of deformation

- The structural change is accommodated by the collective motions of the protein
- If soft modes can accommodate the change, then the ‘effective’ resistance to stress, or the effective mechanical stiffness of the protein is smaller.
Constructing a mechanical resistance map for the entire protein

ANM permits us to calculate an effective stiffness/resistance against deformation under uniaxial tension applied to residues $i$ and $j$.

The idea is simple: $i$ and $j$ already undergo long distance fluctuations by virtue of intrinsically accessible slow modes, the molecule is more ‘yielding’.

$< \kappa_{ij} >$ is the ‘average force constant’

$< \kappa_{ij} > = \sum_k d_{ij}(k) \lambda_k / \sum_k d_{ij}(k)$

where $d_{ij}(k)$ is the contribution of mode $k$ to the change $\Delta R_{ij}$ given by the projection of $\Delta R_{ij}^{(k)}$ onto $R_{ij}^0$, i.e.,

$d_{ij}(k) = \Delta R_{ij}^{(k)} \cos(R_{ij}^0, \Delta R_{ij}^{(k)})$
Bridging Sequence Evolution and Structural Dynamics

- Sequence encodes structure
- Each structure has a unique dynamics (Tobi & Bahar, PNAS 2005)
- Dynamics is functional (Bakan & Bahar, PNAS 2009)

Did sequences evolve to enable functional dynamics?
Bridging Sequence Evolution and Structural Dynamics

1. Obtain MSA
2. Obtain structure
3. Find the corresponding sequence in MSA
4. MSA refinement
5. Entropy/MI calculation
6. GNM calculation
7. Comparison

- Dynamics is functional (Bakan & Bahar, PNAS 2009)

Did sequences evolve to enable functional dynamics?

Dr. Ying Liu
Comp & Systems Biol
Carnegie Mellon U – U Pitt

Correlation between sequence entropy & conformational mobility

Sequence evolution
an information-theoretic approach

Residue index

\[
\begin{array}{c|cccc}
  i   & i+5 & i+7 & i+9 \\
\hline
 R   & E   & V   & N   \\
 E   & K   & V   & N   \\
 K   & E   & V   & N   \\
 R   & D   & V   & S   \\
 D   & K   & V   & S   \\
 D   & K   & V   & S   \\
 E   & R   & V   & S   \\
\end{array}
\]

Conserved and correlated mutations

Information entropy (Shannon, 1951)

\[
S(i) = \sum_{x_i=1}^{20} P(x_i) \log \frac{1}{P(x_i)}
\]

Mutual information (MI)

\[
I(i, j) = \sum_{x_i=1}^{20} \sum_{y_j=1}^{20} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)}
\]

for correlated mutations analysis (CMA)
Mutual Information without the influence of phylogeny

\[ \text{MI}_p (i, j) = I(i, j) - \text{APC} \]

\[ \text{APC} = \text{Average product correction} = \left[ \frac{\text{I}(i, x) \text{I}(j, x)}{\langle \text{I}(i, j) \rangle} \right] \]

where \( \text{I}(i, x) \) is the mean mutual information of column \( i = \sum_j \text{I}(i, j) \)

HIV-1 protease correlated mutation analysis (CMA)

Shi and Malik (2000)

MI matrix $I_{ij} = I(i, j)$

residue index

reordered residue index

Dr. Ying Liu

Liu, Eyal & Bahar (2008) Bioinformatics
MDR mutations distinguished by CMA

MSA of HIV-1 protease
Stanford HIV Drug Resistance Database
http://hivdb.stanford.edu/

untreated

reordered residue index

Drug-resistant cluster

Phylogenetic cluster

treated by at least one drug

Drug-resistance cluster

mobility profile

high
low

47
Summary

- two groups of correlated mutation sites

<table>
<thead>
<tr>
<th>functional aspects</th>
<th>Structural location</th>
<th>structural dynamics</th>
</tr>
</thead>
<tbody>
<tr>
<td>phylogenetic</td>
<td>exposed</td>
<td>mobile</td>
</tr>
<tr>
<td>multi-drug resistant</td>
<td>dimerization interface</td>
<td>restrained</td>
</tr>
</tbody>
</table>

Questions:

- Are key mechanical sites (e.g. hinges) conserved?
- Is there any correlation between sequence variability and structural dynamics?
- How does the structure ensure substrate specificity and conformational adaptability?
A systematic study of a set of enzymes

Evol

http://www.csb.pitt.edu/prody/tutorials/evol_tutorial/index.html
Correlation between sequence entropy & conformational mobility

Mobility increases with sequence entropy

Hinge sites are evolutionarily conserved despite their moderate-to-high exposure to environment
Amino acids involved in intermolecular recognition are distinguished by their co-evolution propensities.

Amino acids involved in intermolecular recognition are distinguished by their high global mobility.

Summary

Four types of functional sites

<table>
<thead>
<tr>
<th>Functional site</th>
<th>Mobility in global modes</th>
<th>Sequence evolution</th>
<th>Dominant Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical (catalytic, ligand binding)</td>
<td>Minimal</td>
<td>Conserved</td>
<td>high fidelity, precision</td>
</tr>
<tr>
<td>Core</td>
<td>Minimal</td>
<td>Conserved</td>
<td>high stability</td>
</tr>
<tr>
<td>Hinge sites</td>
<td>Minimal</td>
<td>Conserved</td>
<td>rotational flexibility</td>
</tr>
<tr>
<td>Substrate recognition (specific)</td>
<td>High</td>
<td>High co-evolution propensity</td>
<td>adaptability</td>
</tr>
</tbody>
</table>

Adaptability requires easy access to substates

### Bridging Sequence Evolution and Structural Dynamics

<table>
<thead>
<tr>
<th>Structural dynamics</th>
<th>Properties</th>
<th>Functional feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>High mobility</td>
<td></td>
<td>adaptability</td>
</tr>
<tr>
<td>Low mobility</td>
<td></td>
<td>High fidelity, precision</td>
</tr>
</tbody>
</table>


Bakan et al. (2014) *Bioinformatics Evol and ProDy for Bridging Protein Sequence Evolution and Structural Dynamics*

CONCLUSION

- Proteins are designed to facilitate functional changes in their structure.
- The intrinsic dynamics endows enhanced mobility at substrate recognition sites.
- Same sites exhibit sequence variations to enable substrate specificity, but these variations exhibit co-evolutionary patterns.
- Chemically active sites are held almost fixed in space, and are highly conserved (as well as ligand binding sites).
There are several methods for evaluating sequence co-evolution.


Four possible outcomes:

- True positive (TP) – correctly predicted to be a hit
- False positive (FP); predicted but it is a miss
- True negative (TN) – correctly predicted to be a miss
- False negative (FN) – predicted as a miss, but is a hit
Two criteria for assessing the performance of different methods

- Minimizing false positives (signals between non-interacting proteins)
- Maximizing true positives (signals between contact making residues)
Screening of large databases

For testing 9 methods, including

- observed-minus-expected-squared ((OMES) (Kass and Horovitz, 2002)
- statistical coupling analysis (SCA) (Halabi et al., 2009; Lockless and Ranganathan, 1999).
- Direct Coupling Analysis (DCA or DI for Direct Information) (Morcos et al., 2011; Weigt et al., 2009),
- Protein Sparse Inverse COVariance (PSICOV) (Jones et al., 2012),
PSICOV and DI are the best

Average performance of the nine methods based on two criteria, absence of intermolecular FPs (a), and fraction of 3D contact making pairs (b) among different subsets of top-ranking signals. The signals are classified to 3 groups: strong coevolution signals (0.1-0.5%), intermediate signals (0.5-5%) and relatively weak signals (5-20%), which also refer to relatively small, intermediate, and high coverage of coevolving pairs. PSICOV and DI outperform other methods in the strong coevolution region. For the intermediate signal, OMES and Mlp exhibit performances similar to PSICOV and DI in panel a. Mlp(S) shows the best performance in the weak signal regime. SCA and MI (and its shuffled version) have lower performance compared to all others for both criteria over the whole range.
Proteins are designed to favor functional changes in their structure. Pre-existing soft modes facilitate substrate binding.

Collective mechanics/allosteric dynamics are mediated by conserved residues.

The intrinsic motions confer enhanced flexibility at substrate recognition sites.

Correlated mutations at recognition sites enable substrate specificity while conferring conformational adaptability.
**Mechanics vs chemistry?**

How does complexity scale with size of the system?

<table>
<thead>
<tr>
<th>Residues</th>
<th>Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10$</td>
<td>$-$</td>
</tr>
<tr>
<td>$10^2$</td>
<td>$1$</td>
</tr>
<tr>
<td>$10^3$</td>
<td>$2-10$</td>
</tr>
<tr>
<td>$10^4$</td>
<td>$10-100$</td>
</tr>
</tbody>
</table>
ProDy
Protein Dynamics Analysis in Python

Dr. Timothy R Lezon
Assistant Prof, DCSB, Pitt

Dr. Chakra Chennubhotla
Assist Prof, DCSB, Pitt

Drs. Ahmet Bakan and Anindita Dutta

Acknowledgment: NIH - 5 P41 GM10371202
Chromosomal dynamics predicted by Gaussian Network Model GNM explains genome-wide accessibility and long-range couplings

- Bahar Lab
  - She (John) Zhang

- Kingsford Lab
  - Natalie Sauerwald

Chromosome Conformation Capture (e.g. Hi-C)

Crosslink DNA

IN CELL NUCLEUS

Cut with restriction enzyme

Fill ends and mark with biotin

Ligate

Purify and shear DNA; pull down biotin

Sequence using paired-ends

Reference genome

Chromosome 17

Loci (50kb)

Contact map ($M$) determined by experiments ($n \times n$)

$M_{i,j} = \# \text{ reads mapped to locus } i \text{ and } j$
$\quad = \# \text{ interactions between locus } i \text{ and } j$

Chromosome Conformation Capture (e.g. Hi-C)

Zhang and Wolynes (2015)

Umbarger et al. (2011)

Dixon et al. (2012)

Filippova et al. (2014)

Resolve 3D structures

Identifying topological domains

Chromosome 17

Loci (50kb)
Method Pipeline

Contact map determined by experiments \((n \times n)\)

Kirchhoff matrix (Laplacian, \(n \times n\))

Eigen decomposition

\[
\begin{bmatrix}
\lambda_1 & \lambda_2 \\
\vdots & \ddots \\
\lambda_{m-1} & \lambda_m \\
\end{bmatrix}
\]

Eigenvectors and eigenvalues

\[
cov = \sum_k \frac{u_k \cdot u_k^T}{\lambda_k}
\]

Covariance matrix \((n \times n)\)

Loci

Square fluctuations

\(0 \rightarrow 2.5 \times 10^{-5}\)
Square fluctuations vs chromatin accessibility

Open regions can be accessed by enzymes

Chromosome 17

Correlations with experiments

Chromosomes

Mobility/Accessibility

Locus Index (5kb)
Dynamical coupling consistent with ChIA-PET measurements

“Long range” (~100 kBp) interactions identified by ChIA-PET were found to show higher covariance in their movement calculated by GNM than the background pairs.
Collective GNM domains can be identified by combining hinge sites from different modes.

**Mode 1:**

1. **Rigid Parts**
2. **Hinge sites**

**Locus Index (5kb):**

- GNM Domains
- Cross-Correlated Distal Domains (CCDDs)
Compartments and Topologically associating domains (TADs)


GNM domains based on 5 modes agree with Lieberman-Aiden compartments. The loci-correlation matrix is computed by Lieberman-Aiden et al. Red and blue represents positive and negative correlations respectively. White lines denotes the GNM domains.

Conclusion

• The first successful application of the Gaussian Network Model (GNM) to the study of chromatin equilibrium dynamics

• Chromosomal fluctuation spectra as well as local and distal cross-correlations along the chromosomes at several levels of resolution, are in good agreement with DNase-seq, ATAC-seq and ChiA-PET experiments

• Significantly, strong correlations between distal regions along the chromosome are predicted

• Correlated pairs of genes exhibit enriched gene co-expression measurements.

• New avenues for bridging between 4D genome and gene regulation.
Support from NIGMS, NLM, NIDDK & NIAID

Acknowledgment

Group members
• JiYoung Lee
• Karolina Mikulska-Ruminski
• James Krieger
• Mary Cheng
• Cihan Kaya
• Luca Ponzoni
• Hongchun Li
• Bing Liu
• She Zhang
• Yan Zhang

Former members
• Ahmet Bakan
• Murat Can Cobanoglu
• Zerrin Bagci
• Chakra Chennubhotla
• Anindita Dutta
• Eran Eyal
• Ignacio General
• Mert Gur
• Basak Isin
• Tim Lezon
• Enrique Markus
• Lidio Meireles
• Ying Liu
• Lin Liu
• AJ Rader
• Dror Tobi
• Lee-Wei Yang
• Zheng Yang
• Elia Zomot
• Kaitlyn Fu
• Wenzhi Mao
• Philippo Pulara
• Indira Shrivastava
• Guang Hu

Susan Amara
Hulya Bayir
Tim Billiar
Kay Brummond
Takis Benos
Carlos Camacho
Rob Coalson
Billy Day
Kay Brummond
Mike Epperly
Bard Ermentrout
Jennifer Grandis
Angela Gronenborn
Joel Greenberger
Valerian Kagan
Hassan Karimi
Judith Klein-Seetharaman
John Lazo
Hagai Meirovitch
Penny Morel
Zoltan Oltvai
Stephen Pak
David Perlmutter
Alan Russell
Ed Prochownik
Nicholas Sluis-Kremer
Gary Silverman
Tom Smithgall
Joel Stiles
Lans Taylor
Michael Tsang
Yoram Vodovotz
Andreas Vogt
John Vries
Alan Wells
David Whitcomb
Mike Widom
Rick Wood
Peter Wipf
Jian Yu