Bridging Structure and Function, Experiments and Computations

Day 2 – Lecture 1

Pemra Doruker

Department of Computational and Systems Biology
School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260
Summary

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

2. Bridging Sequence, Structure and Function
   a. Ensemble analysis and functional modes of motion
   b. Combining sequence and structure analyses – signature dynamics
   c. Modeling membrane proteins and lipid environment with ANM

3. Allostery and druggability
   a. Essential site scanning and allosteric pocket prediction
   b. Druggability simulations
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?
Two recent reviews:


**Sequence evolution**
**an information-theoretic approach**

Residue index (up to $N$)

<table>
<thead>
<tr>
<th>$i$</th>
<th>$i+5$</th>
<th>$i+7$</th>
<th>$i+9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>E</td>
<td>V</td>
<td>N</td>
</tr>
<tr>
<td>E</td>
<td>K</td>
<td>V</td>
<td>N</td>
</tr>
<tr>
<td>K</td>
<td>E</td>
<td>V</td>
<td>N</td>
</tr>
<tr>
<td>R</td>
<td>D</td>
<td>V</td>
<td>S</td>
</tr>
<tr>
<td>D</td>
<td>K</td>
<td>V</td>
<td>S</td>
</tr>
<tr>
<td>D</td>
<td>K</td>
<td>V</td>
<td>S</td>
</tr>
<tr>
<td>E</td>
<td>R</td>
<td>V</td>
<td>S</td>
</tr>
</tbody>
</table>

Conserved 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)**

K, R: + charge; E, D: - charge $\rightarrow$ salt bridge
Mutual Information without the influence of phylogeny

MI\textsubscript{p} - to eliminate random noise and phylogenetic components

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

Average product correction

\[ \text{APC}(i, j) = \left[ \langle I(i) \rangle \langle I(j) \rangle \right] / \langle I(i, j) \rangle \]

\[ \langle I(i) \rangle : \text{the mean mutual information of column } i \]

\[ \langle I(i, j) \rangle : \text{average over all MI values} \]

A systematic study of a set of enzymes

Correlation between sequence entropy & conformational mobility

Mobility increases with sequence entropy

Liu & Bahar  
Hinge sites are evolutionarily conserved
despite their moderate-to-high exposure to environment

Amino acids involved in intermolecular recognition exhibit high global mobility and co-evolution.

# 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>

SignDy: Signature dynamics of families

- How does functional differentiation take place while maintaining the fold?
- What are the shared/differentiated dynamics of family members?
- Can we categorize family members based on dynamics?
Sequence evolves faster than structure. Proteins with different sequences may share similar structures (functions).

What is special about these folds that lend themselves to different functionalities?

SignDy pipeline for evolution of dynamics

1. Sequence, PDB ID, UniProt ID
2. Selected sequences
3. Homologous structures
4. Mode-mode matches
5. Signature dynamics
6. Conservation/differentiation of subfamily dynamics
7. Dynamics-based dendrogram

Selected sequences:
PDB IDs:

Homologous structures:

PDB IDs:

CATH ID:

PDB:

DYNAMICS-based dendrogram

SignDy pipeline for evolution of dynamics

Case Study: LeuT Fold Superfamily

Neurotransmitter transporters

Bacterial Leucine Transporter (LeuT)

Case Study: LeuT Fold Superfamily

Minimal sequence identity

Same fold

Intrinsic Dynamics predicted by the ANM

SignDy results for LeuT family

Signature-dynamics of each family is robustly defined by the global motions that are unique to the fold.

Signature Dynamics: Shared fold-specific mobility profile

Central binding site:

LeuT  MhsT  DAT  Mhp1  BetP  AdiC

Signature Dynamics: Shared fold-specific mobility profile

Central binding site:

Signature Dynamics: Shared fold-specific mobility profile

Central binding site:

Signature modes match functions

How conserved are the global modes?

LeuT

Cosine similarity:

\[ CC = \mathbf{v}_i^{(\text{LeuT})} \cdot \mathbf{v}_j^{(\text{Mhp1})} \]

Mhp1

How conserved are the global modes?

LeuT

Mhp1

How conserved are the global modes?

![Graph showing residue index, mode index, LeuT, Mhp1, and vibrational frequency.]

**Cosine similarity:**

\[
SO_k(A, B) = 1 - \frac{2 \sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{\sigma_i(A) \sigma_j(B)}{\sigma_i(A) + \sigma_j(B)} \right)^2 \left( \psi_i^{(A)} \psi_j^{(B)} \right)^2}{\sum_{i=1}^{k} \sum_{j=1}^{k} \left( \frac{\sigma_i(A) \sigma_j(B)}{\sigma_i(A) + \sigma_j(B)} \right)^2}
\]

\( k \): a subset of modes

**Covariance/spectral overlap (Hess 2002):**

\[
\]
How conserved are the global modes?

LeuT

Mhp1

Cosine similarity:

A region where overlap is minimal → these modes enable functional differentiation
Functional differentiation of subfamilies is enabled by LTIF modes

Application to TIM barrel proteins

LTIF: low-to-intermediate frequency

Modes 1-3

Modes 4-20

Modes 21-60

Modes > 60

Largest spectral distance observed in modes 4-60.

Dynamics allows for classification
(complementing sequence- and structure-based classification)

SignDy reveals shared and divergent motions of domains/folds

Summary: SignDy analysis

- Characterization of the shared structural and dynamic features, or **signature dynamics**
- Identification of sites important for the specific function of subfamilies,
- Insights into **functional differentiation** based on the differentiation of dynamics
- Learning design principles: focus on LTIF regime for altering function
- Data generated for the most common 77 CATH superfamilies (containing 15,000 proteins)
- Generation of dendrograms based on evolution of structural dynamics and thereby function
Summary

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

2. **Bridging Sequence, Structure and Function**
   a. Ensemble analysis and functional modes of motion
   b. Combining sequence and structure analyses – signature dynamics
   c. Modeling membrane proteins and lipid environment with ANM

3. **Allostery and druggability**
   a. Essential site scanning and allosteric pocket prediction
   b. Druggability simulations
membrANM
Membrane Anisotropic Network Model
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 DynOomics
MembrANM for γ-secretase (DynOmics)

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

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

PyMol movies can be downloaded from DynOmics 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
The effect of the membrane environment can be incorporated in the ANM analysis by adopting:
- envANM-membrANM or
- substructure-membrANM.

MembrANM results for the γ-secretase (PDB ID: 5FN2) can be viewed on the ENM server: PyMol movies can be downloaded from DynOmics ENM server by generating “Full Atomic Structures for ANM-Driven Conformers”.

Collective dynamics of AMPA receptors

Experimentally verified by cross-linking experiments. Substitution of cysteines in the presence of an oxidizing agent promotes cross-linking.

Mode Comparison (DynOmics)

In the presence of membrane:

- Some intrinsic motions are preserved *(red)*
- Others are altered *(yellow)*
- New motions (mainly dominated by membrane fluctuations) emerge (blue rows)
- Diagonal shift due to new modes in the membrane
Alternating-access model. Transitions between outward-facing (OF) and inward facing (IF) states of membrane transporters.

Global transitions of Glutamate Transporters

Trimeric membrane proteins, each subunit composed of: scaffolding + transport domains

Outward Facing (OF)  Inward Facing (IF)

Domains of Aspartate Transporter

(also called EAAT – for excitatory amino acid transporter)

N-terminal (trimerization) domains in gray (TM1-TM6): **scaffolding domain**

C-terminal Core (HP1-TM7-HP2-TM8): **transport domain**
Global transitions

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

- Single subunit showing the transport domain moving across the membrane

- Translation of about 15 Å
Substates are sampled along soft modes

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

1. Evaluate the deformation $\mathbf{d}$ (3N unit vector)

2. Calculate the correlation cosine $\cos(\mathbf{v}_i, \mathbf{d})$ between each mode vector, $\mathbf{v}_i$, and $\mathbf{d}$

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

Reference:
Correlation between soft modes & observed structural change

1. Evaluate the deformation \( \mathbf{d} \) (3N unit vector)

2. Calculate the correlation cosine \( \cos(\mathbf{v}_i, \mathbf{d}) \) between each mode vector, \( \mathbf{v}_i \), and \( \mathbf{d} \)

3. Cumulative overlap = \( [ \sum_i \cos^2(\mathbf{v}_i, \mathbf{d}) ]^{1/2} \)

Reference:

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!

Reference:
Membrane facilitates alternating access

Overlap of > 0.6 achieved with a single mode!

Softest nondegenerate mode out of > 3,000 modes

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

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

Reference:
Two approaches for including the lipid bilayer:

- Explicit membrane (a network model for the membrane)
- Implicit membrane (change in Hessian force constants)
As the environment fluctuates randomly, the effective motion of the system is given by

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

\[
H^{ss'} = H^{ss} - H^{SE} \left( H^{EE} \right)^{-1} H^{ES}
\]
Implicit model for membrane effect

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

Constraining effect of the lipids in the radial direction (xy plane)

\[ \gamma_x = \gamma_y = c^2 \gamma_z \quad c > 1 \]

Altered radial force constants:

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

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

Rotations and Translation of Blocks (RTB)
Tama F, Gadea FJ, Marques O, Sanejouand YH. *Proteins* **2000** 41 1-7

Lipid bilayer favors elevator-like motions

ANM in the absence of membrane

ANM in the presence of membrane

Lezon & Bahar, Biophys J 2012
Thank you!