

# Day 1 - Lecture 1

## Collective Dynamics of Biomolecules using

### Elastic Network Models

Ivet Bahar

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

June 28-July 1, 2021

# MMBioS Resources

**Anisotropic Network Model Web Server 2.0 (2014)** 

What's new in this version? Having Java problems?

Enter the PDB id of your protein  
  pdb coordinates  biological unit

or

Submit your own protein  
 No file chosen

**iGNM 2.0 – Gaussian Network Model Database**

[Home](#) | [Tutorial](#) | [Theory](#) | [References](#) | [oGNM 2.0](#) | [ANM 2.0](#) | [Computational & Systems Biology](#) | [NTHU site](#)

What is the GNM DB? Which questions can be answered?

Several studies in the last decade have drawn attention to the significance of intrinsic dynamics as a major determinant of the mechanism of action of proteins and their complexes (1–5). Intrinsic dynamics refers to conformational changes intrinsically favored by 3D structure, which often underlie the adaptation of biomolecules to functional interactions (6). As a consequence, an important question is to assess which structural elements (e.g. residues, secondary structures, domains, or entire subunits) undergo large fluctuations away from their mean positions (i.e. those enjoying high *mobility*), or which ones provide adequate *flexibility* to enable conformational changes (e.g. hinge-bending sites) that may be relevant to function. Furthermore, it is often of interest to determine which structural elements are subject to strongly correlated (or anticorrelated) motions, toward gaining insights into allosterically coupled regions. The GNM (7,8) addresses these questions. It further allows to dissect these properties into the contributions of individual modes, thus elucidating the cooperative (*global*) couplings (cross-correlations) underlined by low frequency modes. For more information see [Theory](#) and [Tutorial](#).

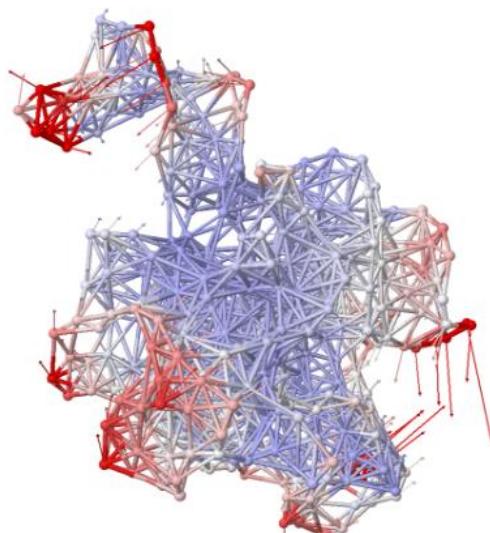
Note: Query the GNM DB (iGNM 2.0) with a single PDB code (e.g., 101M and 4NIH, etc.); or, search the database with customized condition(s) using the "Advanced search".

PDB ID:  Go iGNM

Biological assembly:  Yes  No

Molecular viewer:  JsMol  Jmol (fast response for big structures)

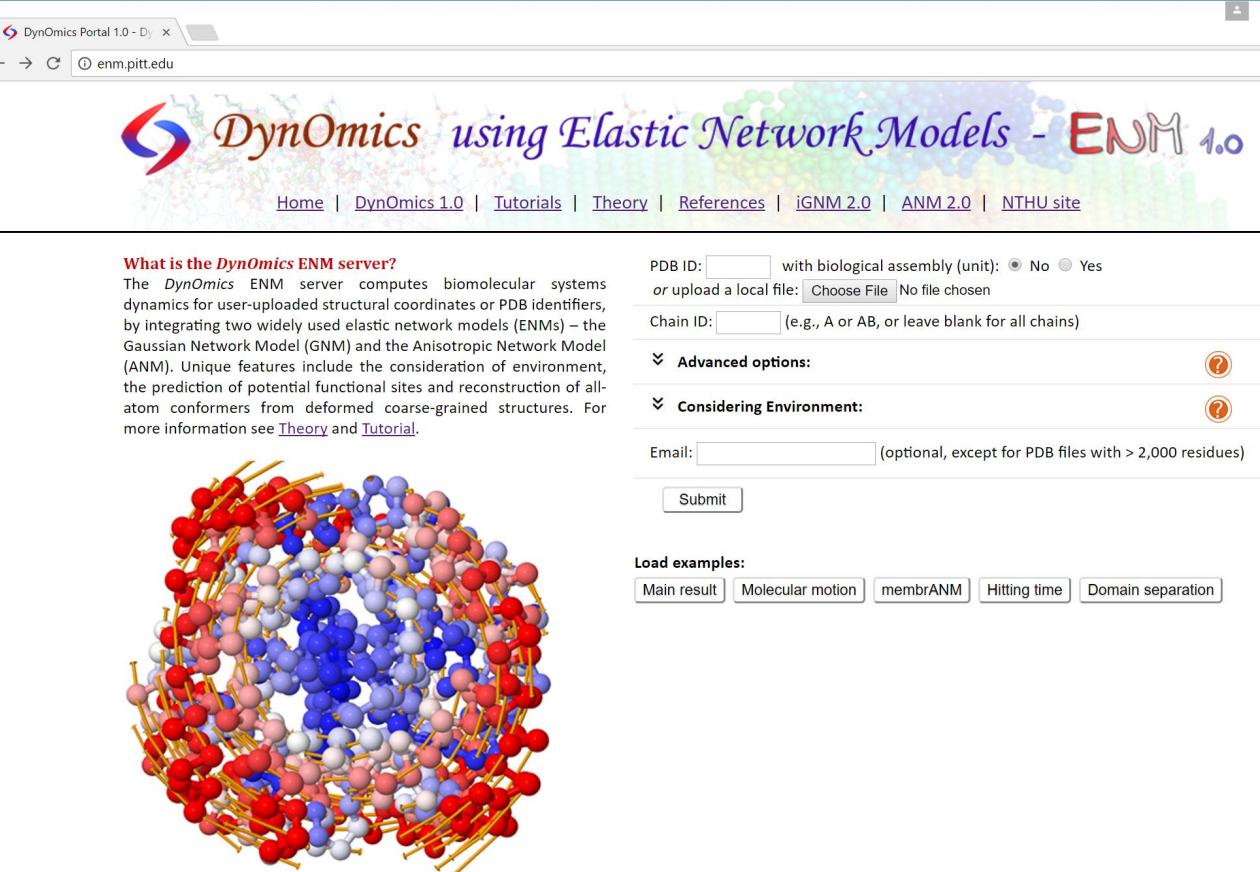
Advanced search:



Eyal et al., *Bioinformatics* 2015

Li et al., *Nucleic Acids Res* 2016

# MMBioS Resources



The screenshot displays two web-based resources side-by-side. On the left is the ProDy software homepage, featuring a large logo and sections for the ProDy Project, Structure analysis, and Dynamics analysis. On the right is the DynOmics using Elastic Network Models - ENM 1.0 portal, which includes a main landing page with navigation links and a detailed input form for protein dynamics analysis.

**ProDy Software** (Left):

- ProDy Project:** Describes ProDy as a free and open-source Python package for protein structural dynamics analysis, designed for flexible and responsive API usage.
- Structure analysis:** Details fast and flexible PDB and DCD file parsers, atom selection for contact identification, and structure comparisons.
- Dynamics analysis:** Lists principal component analysis for heterogenous X-ray structures, normal mode analysis using ANM or GNM, and ANM/GNM with distance and property dependent force constants.

**DynOmics using Elastic Network Models - ENM 1.0** (Right):

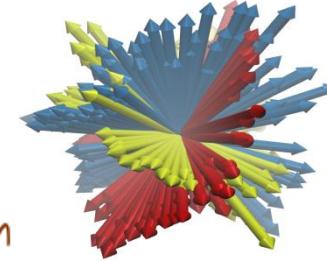
- What is the DynOmics ENM server?**: The DynOmics ENM server computes biomolecular systems dynamics for user-uploaded structural coordinates or PDB identifiers, integrating Gaussian Network Model (GNM) and Anisotropic Network Model (ANM).
- PDB ID:** Input field for PDB ID with biological assembly unit, with options for "No" or "Yes".
- Chain ID:** Input field for chain identifier (e.g., A or AB).
- Advanced options:** Collapsible section for additional parameters.
- Considering Environment:** Collapsible section for environmental factors.
- Email:** Input field for email address (optional for PDB files with > 2,000 residues).
- Submit**: Button to submit the form.
- Load examples:** Buttons for Main result, Molecular motion, membrANM, Hitting time, and Domain separation.

**People**, **Community**, **Source Code**, and **Problems?** sections are also visible at the bottom of the DynOmics portal.

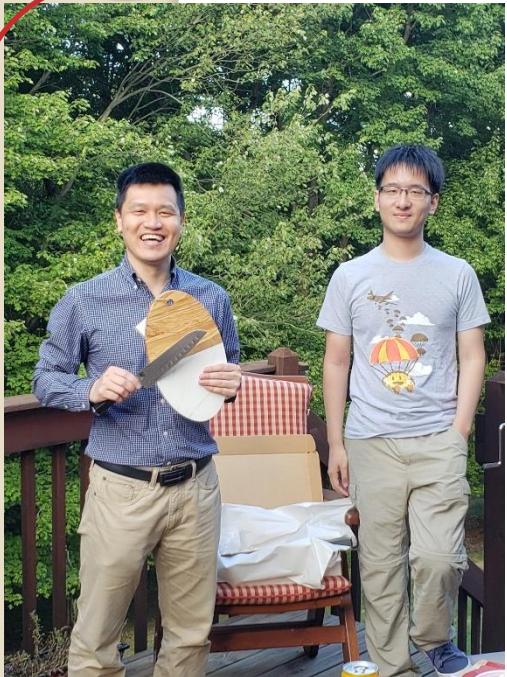


# ProDy

Protein Dynamics Analysis in Python



**Dr. Timothy Lezon**  
Assistant Prof, DCSB, Pitt



**Dr. Hongchun Li**  
Assoc Professor,  
Shenzhen Institute

**Dr. She (John) Zhang**  
Postdoc at OpenEye



**Dr. James Krieger**  
Postdoc, U of Madrid



**Dr. Anindita Dutta**  
Principal Deep Learning/AI  
Engineer at Illumina

**Dr. Ahmet Bakan**  
Senior Software Engineer,  
Google Inc.



**Dr. Ying Liu**  
Software Engineer,  
Google Inc.



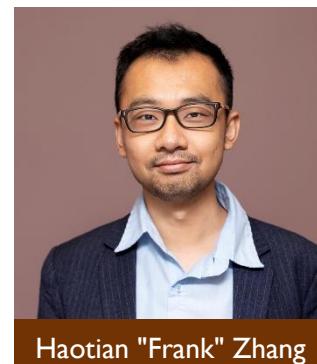
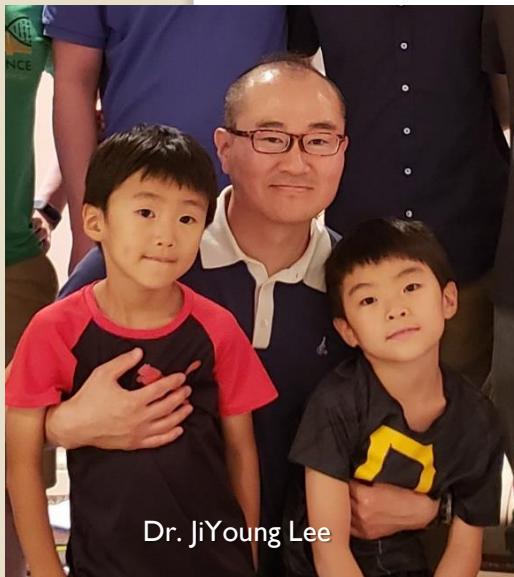
**Dr. Chakra Chennubhotla**  
Assoc Prof, DCSB, Pitt

## Reference:

Bakan et al (2011) ProDy: Protein dynamics inferred from theory and experiments *Bioinformatics* **27**:1575-7  
 Bakan, Dutta et al, (2014) *Bioinformatics* **30**: 2681-2683; Zhang, Krieger et al., (2021) *Bioinformatics*, *in press*.



## The team



# ProDy References

Bakan A,\* Dutta A,\* Mao W, Liu Y, Chennubhotla C, Lezon TR, Bahar I (2014) [Evol and ProDy for Bridging Protein Sequence Evolution and Structural Dynamics](#) *Bioinformatics* **30**: 2681-3

Bakan A, Meireles LM, Bahar I (2011) [ProDy: Protein dynamics inferred from theory and experiments](#) *Bioinformatics* **27**: 1575-1577.

Zhang S, Krieger JM, Zhang Y, Kaya C, Kaynak B, Mikulska-Ruminska K, Doruker P, Li H, Bahar I. (2021) [ProDy 2.0: Increased Scale and Scope after 10 Years of Protein Dynamics Modelling with Python](#). *Bioinformatics* Apr 5:btab187.

# ProDy: Usage and dissemination statistics

Date	Releases	Downloads <sup>1</sup>	Visits <sup>2</sup>	Unique <sup>3</sup>	Pageviews <sup>2</sup>	Countries <sup>5</sup>
Nov'10 - Oct'11	19	8,530	8,678	2,946	32,412	45
Nov'11 - Oct'12	6+9*	35,108	16,472	6,414	71,414	59
Nov'12 - Oct'13	8*	87,909	19,888	8,145	86,204	66
Nov'13 - Oct'14	5*	140,101	24,134	11,170	112,393	69
Nov'14 - May'15	1*	68,230	15,941	8,479	66,641	50
June '15- June'16	5*	124,613	32,491	15,402	140,818	132
June'16- June 17			31,374	16,201	129,900	136
<b>Total (6/17)</b>	<b>53+</b>	<b>464,491+</b>	<b>148,978</b>	<b>68,757</b>	<b>639,782</b>	<b>136</b>
<b>Total (5/18)</b>		<b>979,356</b>	<b>182,415</b>	<b>86,063</b>	<b>784,430</b>	
<b>Total (5/19)</b>		<b>1,670,461</b>	<b>218,811</b>	<b>106,130</b>		
<b>Total (10/20)</b>		<b>2,161,939</b>	<b>280,862</b>	<b>140,905</b>		

<sup>1</sup> Download statistics retrieved from PyPI (<https://pypi.python.org/pypi/vanity>).

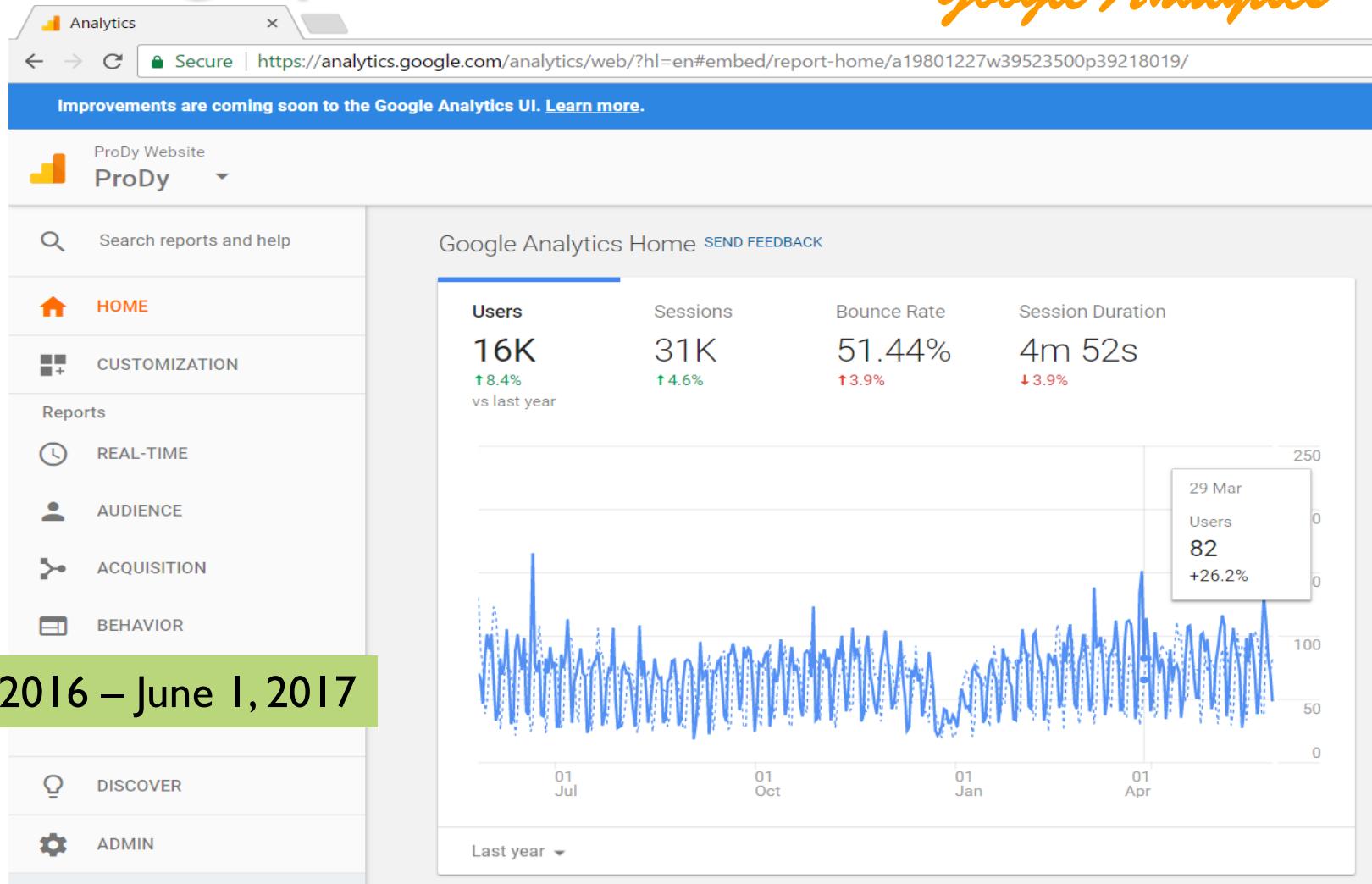
<sup>2</sup> Google Analytics ([www.google.com/analytics](http://www.google.com/analytics)) was used to track:

<sup>3</sup> Unique indicates number of unique visitors;

**55,263 lines of code**

# Usage pattern

Google Analytics

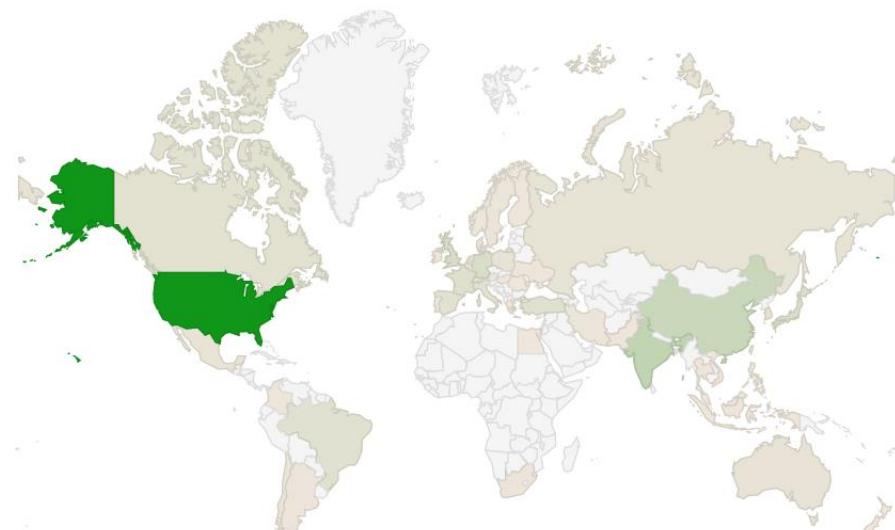


## Statistics

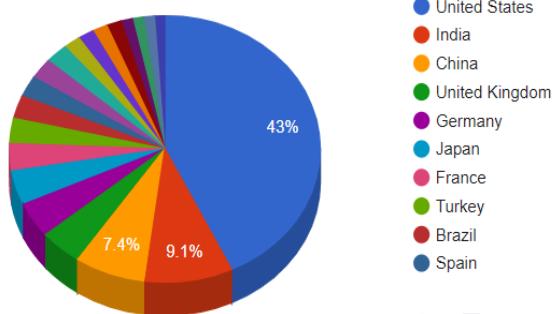
Prody has been downloaded 2,161,939 times as of yesterday since October 2011.

The table and map below displays the data from [Google Analytics](#) on the total number of visitors to ProDy API website since Jun 2011. More detailed statistics from Google Analytics are given below

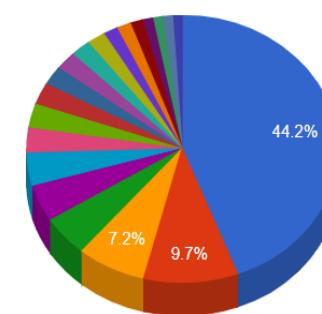
	Country	Sessions
1	United States	104,451
2	India	22,092
3	China	18,052
4	United Kingdom	10,768
5	Germany	10,376
6	Japan	10,103
7	France	8,091
8	Turkey	7,391
9	Brazil	6,751
10	Spain	6,171
11	Canada	6,042
12	Italy	5,875
13	Russia	4,286
14	Poland	4,019



### Visitor distribution across the world (top 20 countries)



#### Unique visitors across the world (top 20 countries)

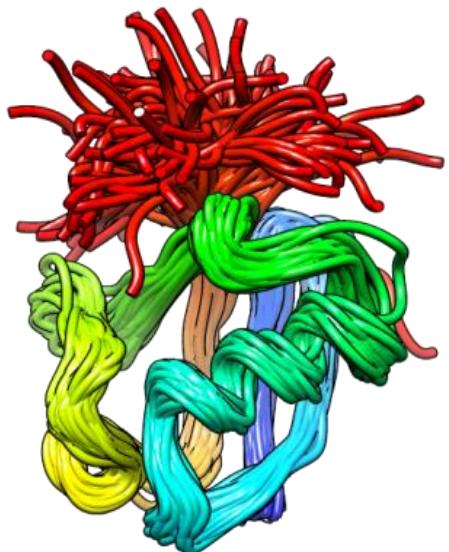


- United States
- India
- China
- United Kingdom
- Germany
- Japan
- France
- Canada
- Brazil
- Turkey

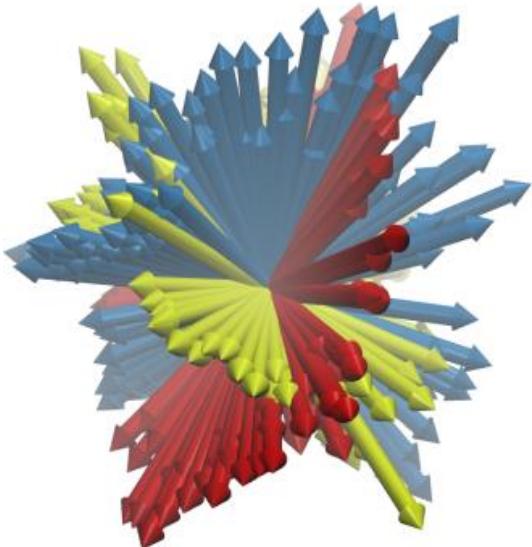
Oct 2020

# Tutorials

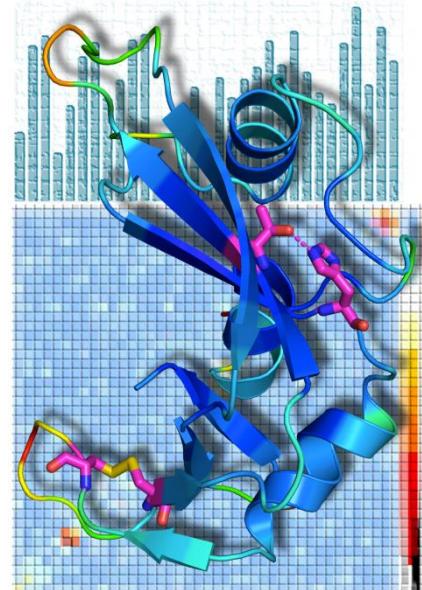
<http://prody.csb.pitt.edu/tutorials/>



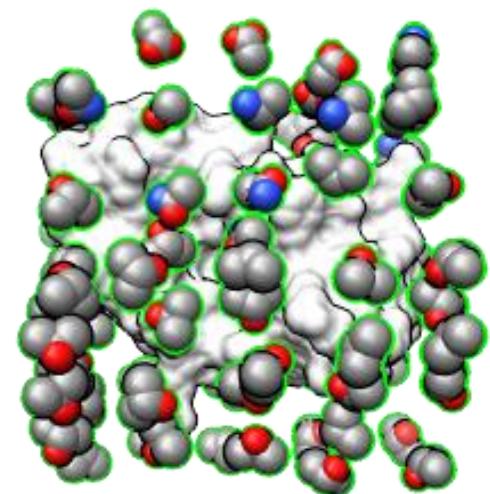
ProDy



NMWiz



Evol



Druggability

# Workshop files on ProDy website

The screenshot shows the ProDy website homepage. The main header features the ProDy logo with a stylized protein structure and the text "ProDy Protein Dynamics & Sequence Analysis". Below the header is a navigation bar with links to various tools and resources: ProDy, Evol, NMWiz, SignDy, membrANM, MechStiff, PRS, DruGUI, coMD, Downloads, Tutorials, and Statistics. A search bar for "Manual and Tutorials" is also present.

**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
  - NMR 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 using [NMWiz](#).

**Reference**

Bakan A, Meireles LM, Bahar I [ProDy: Protein Dynamics Inferred from Theory and Experiments](#) 2011 *Bioinformatics* 27(11):1575-1577

Bakan A, Dutta A, Mao W, Liu Y, Chennubhotla C, Lezon TR, Bahar I [Evol and ProDy for Bridging Protein Sequence Evolution and Structural Dynamics](#) 2014 *Bioinformatics* 30(18):2681-2683

**Funding**

Continued development of ProDy is supported by NIH through the R01 GM099738 award.

**Workshops**

The ProDy development team hosts annual workshops together with the NAMD/VMD development team as part of our joined center MMBioS funded by NIH through the P41

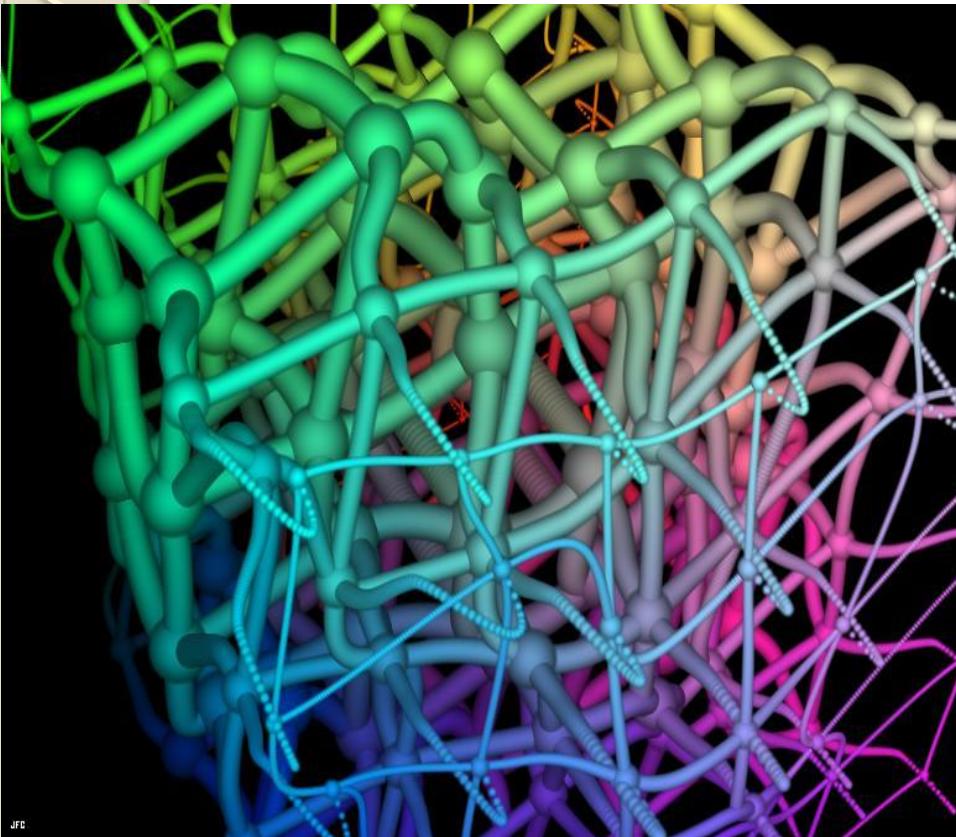
**Compare Dynamics from Experiments and Theory (2/4)**

Two PCA plots comparing dynamics from experiments and theory for RNase H. The top plot shows ANM2 (Å) vs PC1 (Å) with r = 99, and the bottom plot shows ANM3 (Å) vs PC1 (Å) with r = 95. Both plots include data points for Unbound (blue), inhibitor (red), +Prestash (green), and +GDP/ATP (purple). The plots are overlaid on a 3D ribbon model of the RNase H protein structure, showing the N-lobe, C-lobe, and fingers.

**new ANM server**

**new iGNM database**

# Representation of structure as a network

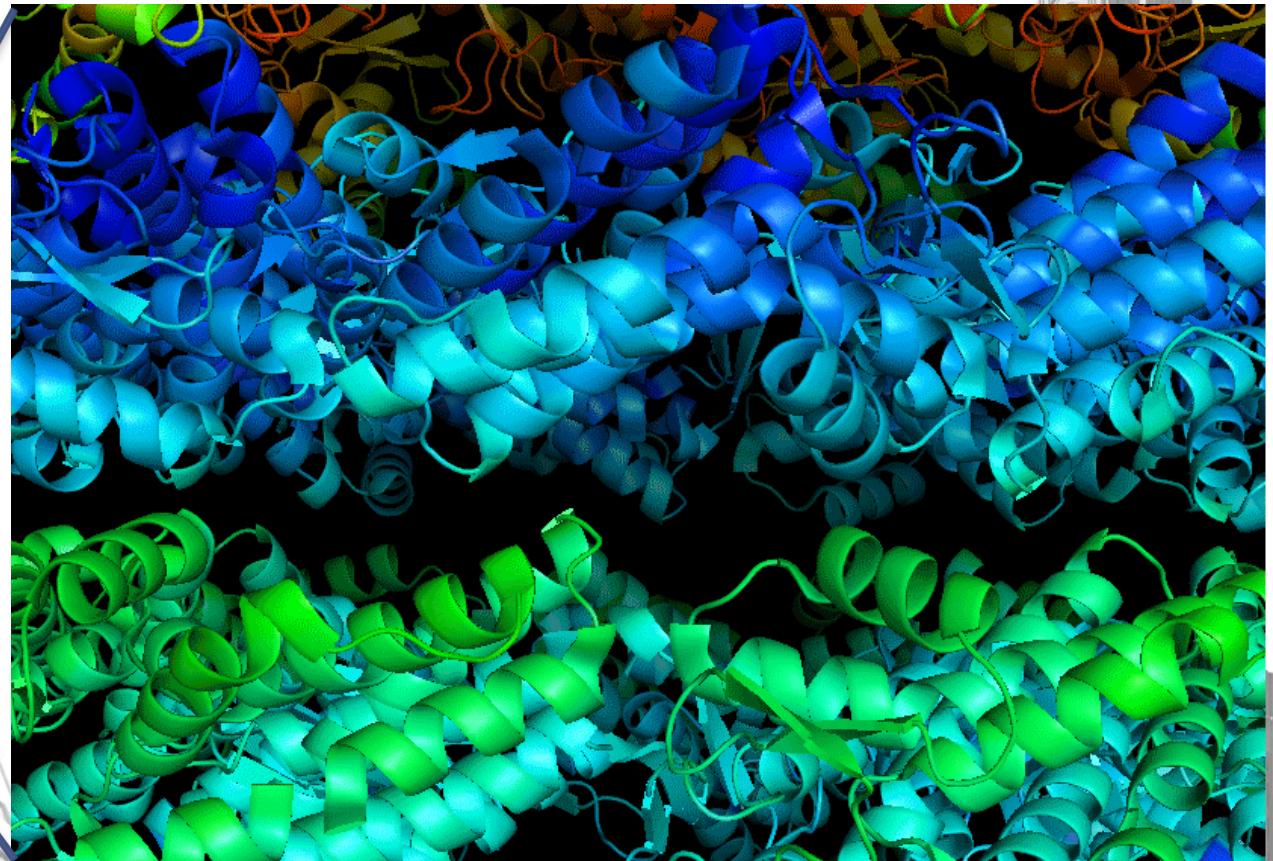


<http://www.lactamme.polytechnique.fr/>

## Why network models?

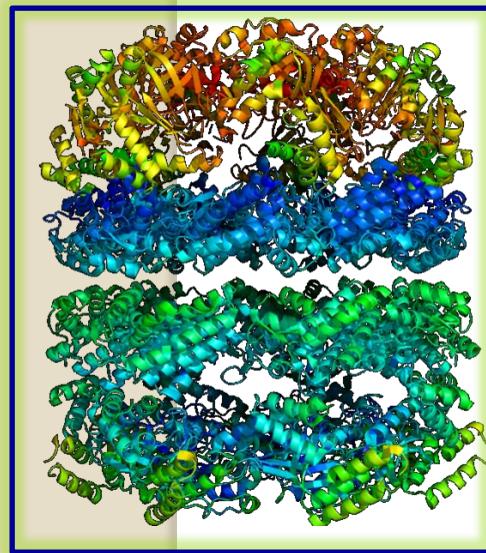
- for large systems' collective motions & long time processes beyond the capability of full atomic simulations
- to incorporate structural data in the models – at multiple levels of resolution
- to take advantage of theories developed in other disciplines: polymer physics, graph theory, spectral graph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

Proteins are not static:  
They move, breath, work, dance, interact with each other

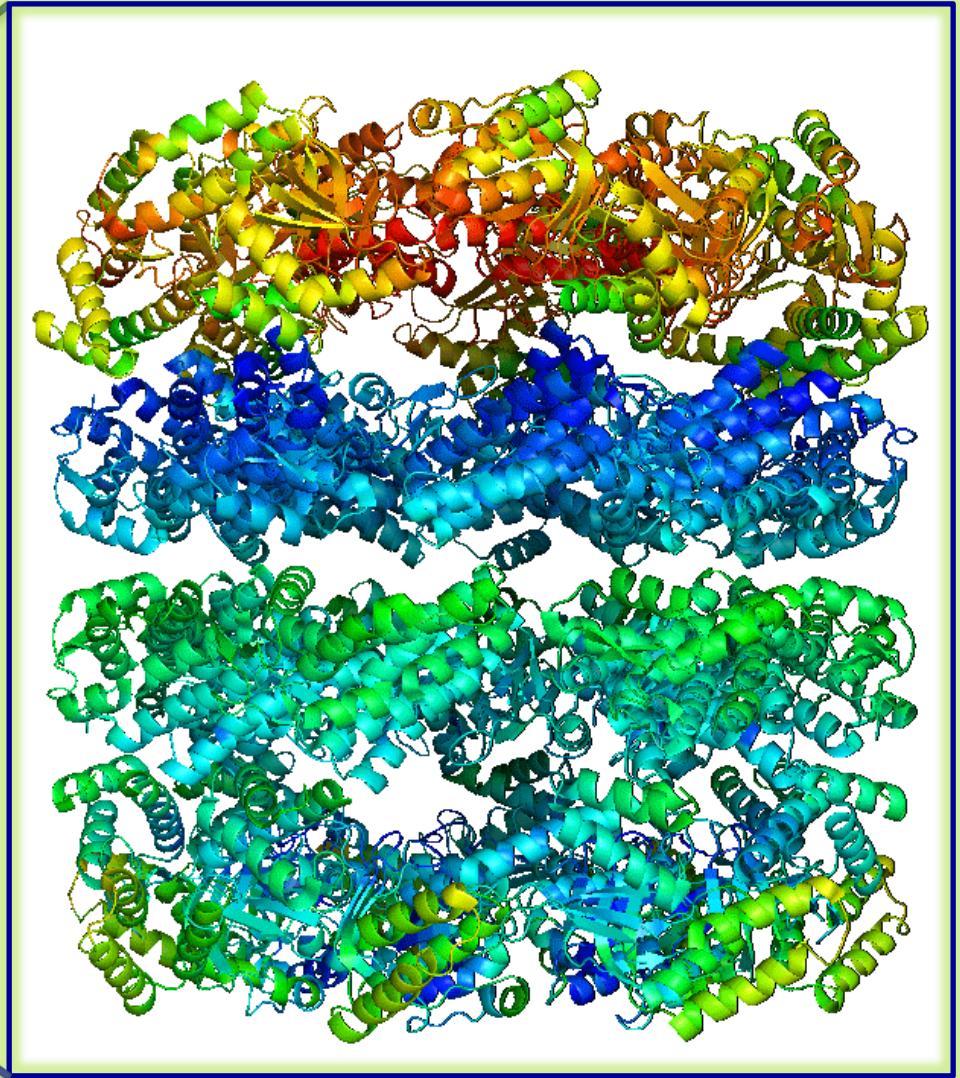


Local motions

Proteins are not static:  
They move, breath, work, dance, interact with each other

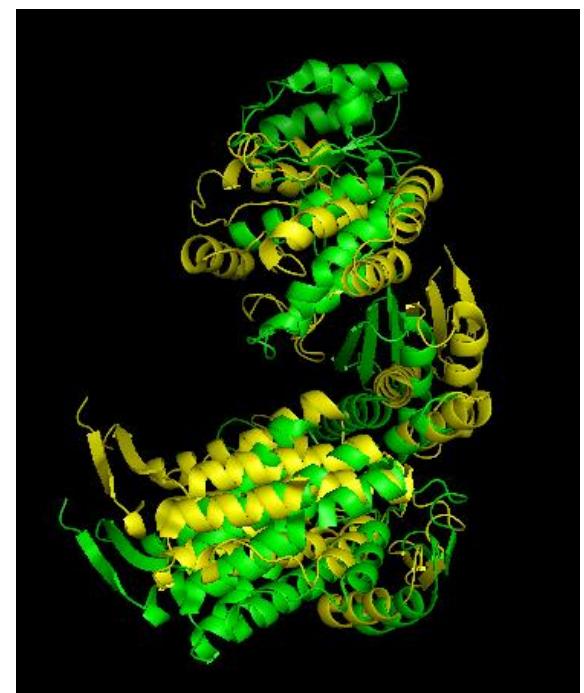
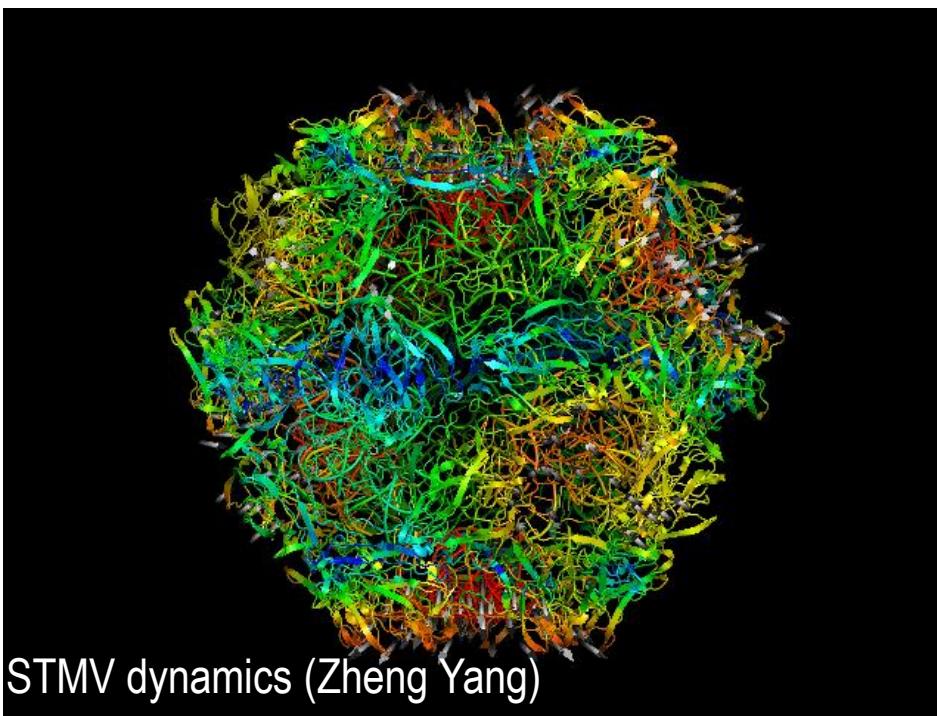


Global motions



# Many proteins are molecular machines

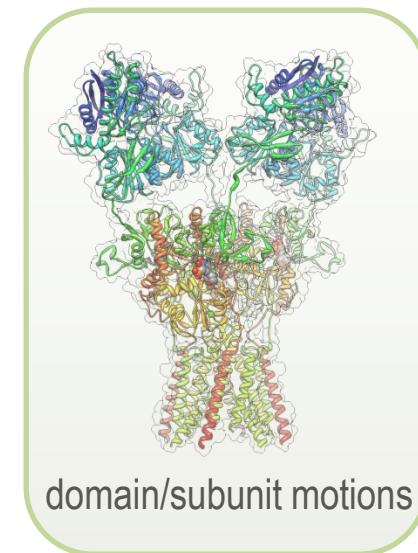
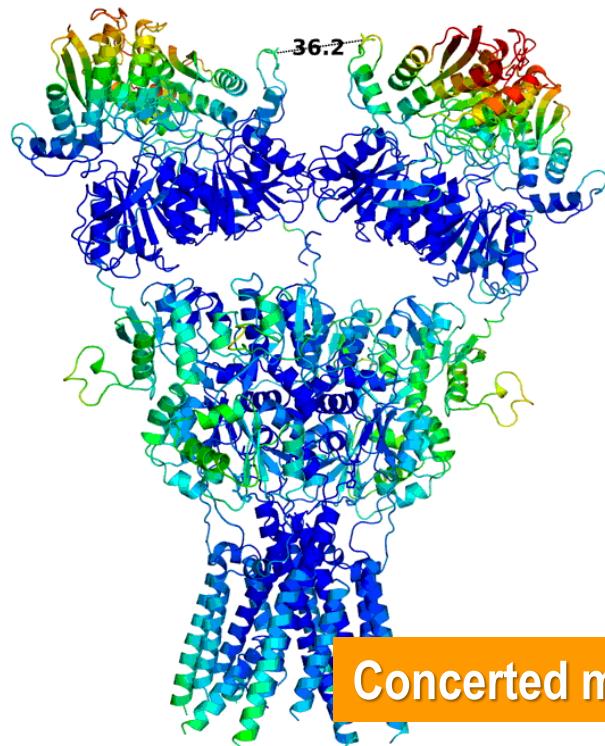
And mechanical properties become more important in complexes/assemblies



# Each structure encodes a unique dynamics

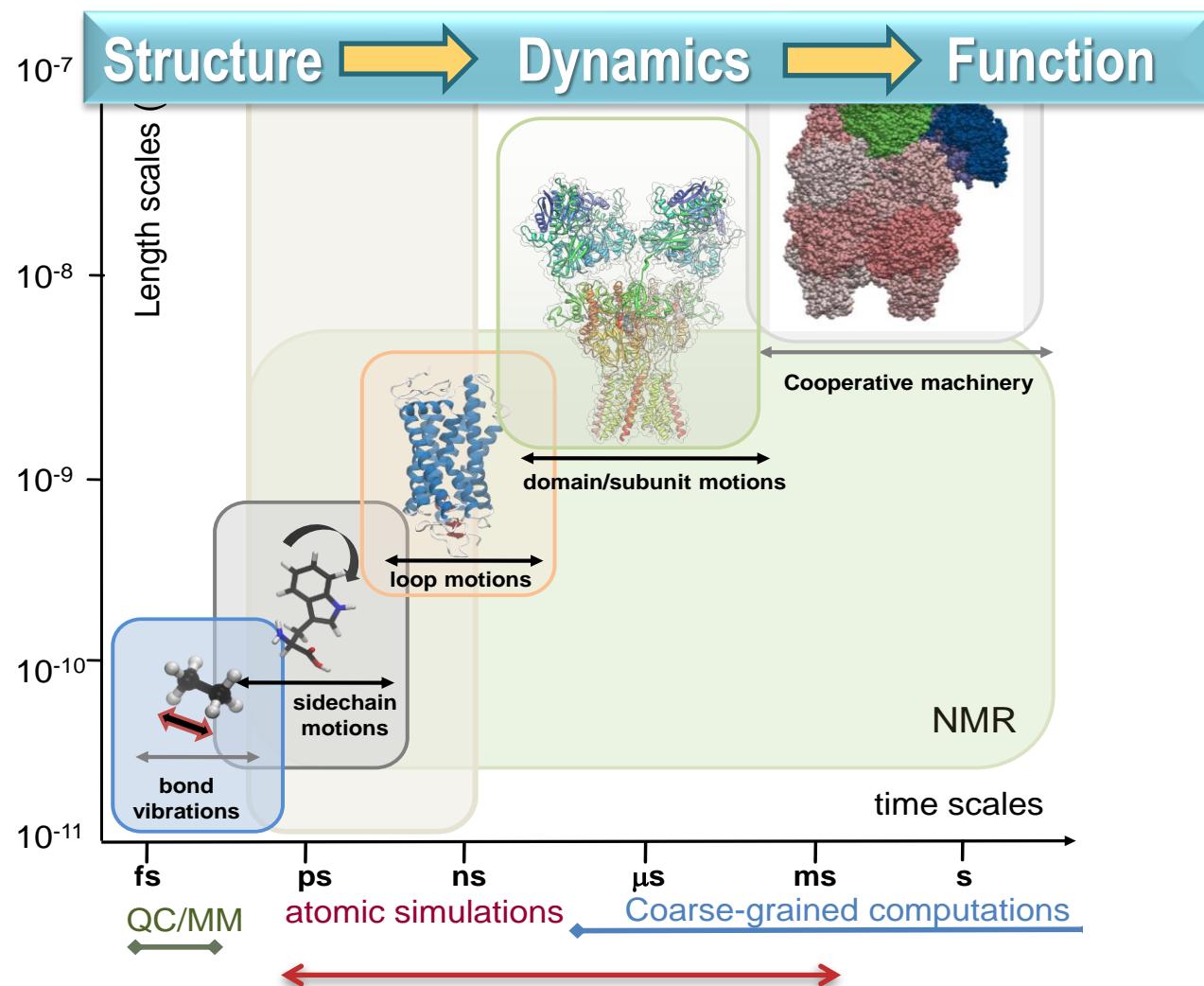
Structure → Dynamics → Function

Signaling dynamics of AMPARs and NMDARs



Concerted movements of signaling molecules

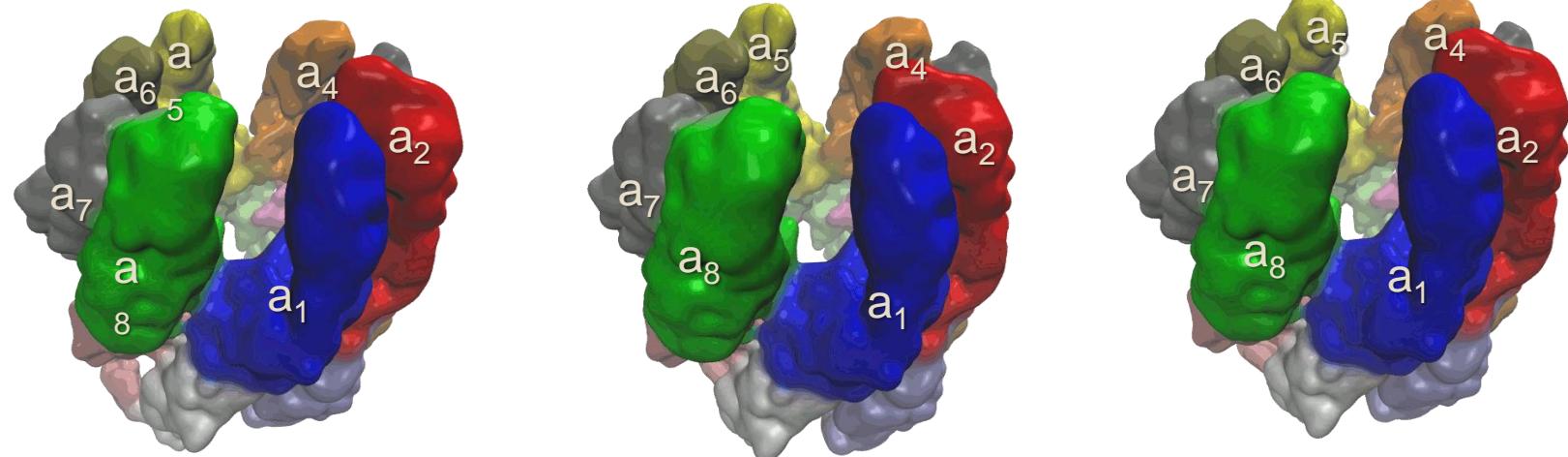
# Each structure encodes a unique dynamics



# Modeling the machinery of cryo-EM structures



Yan Zhang



Collective modes of the mammalian chaperonin TRiC/CCT reveals a state-dependent sequence of asymmetric movements

Zhang Y, Krieger JM, Mikulska-Ruminska K, Kaynak B, Sorzano COS, Carazo JM, Xing J, Bahar I (2020) [State-dependent sequential allostery exhibited by chaperonin TRiC/CCT revealed by network analysis of Cryo-EM maps](#). *Prog Biophys Mol Biol* S0079-6107(20)30082-1

# 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

# Two elastic network models:



## Gaussian Network Model (GNM)

- Li H, Chang YY, Yang LW, Bahar I (2016) [iGNM 2.0: the Gaussian network model database for bimolecular structural dynamics](#) Nucleic Acids Res **44**: D415-422
- Bahar I, Atilgan AR, Erman B (1997) [Direct evaluation of thermal fluctuations in protein](#) *Folding & Design* **2**: 173-181.



## Anisotropic Network Model (ANM)

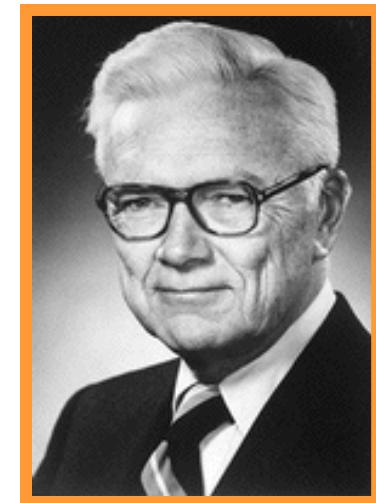
- Eyal E, Lum G, Bahar I (2015) [The Anisotropic Network Model web server at 2015 \(ANM 2.0\)](#) *Bioinformatics* **31**: 1487-9
- Atilgan AR, Durrell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I (2001) [Anisotropy of fluctuation dynamics of proteins with an elastic network model](#) *Biophys J* **80**: 505-515.

# Physics-based approach

- Statistical Mechanics of Polymers
- Theory of Rubber Elasticity

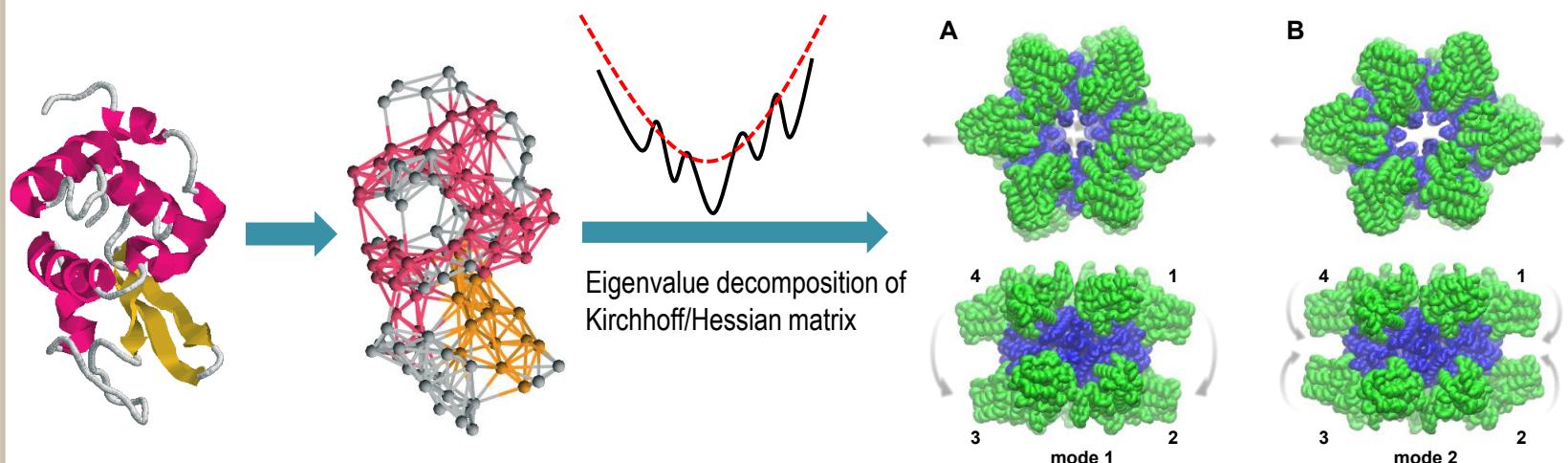


**Elastic Network Model for Proteins**



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

# Collective motions using elastic network models (ENM)

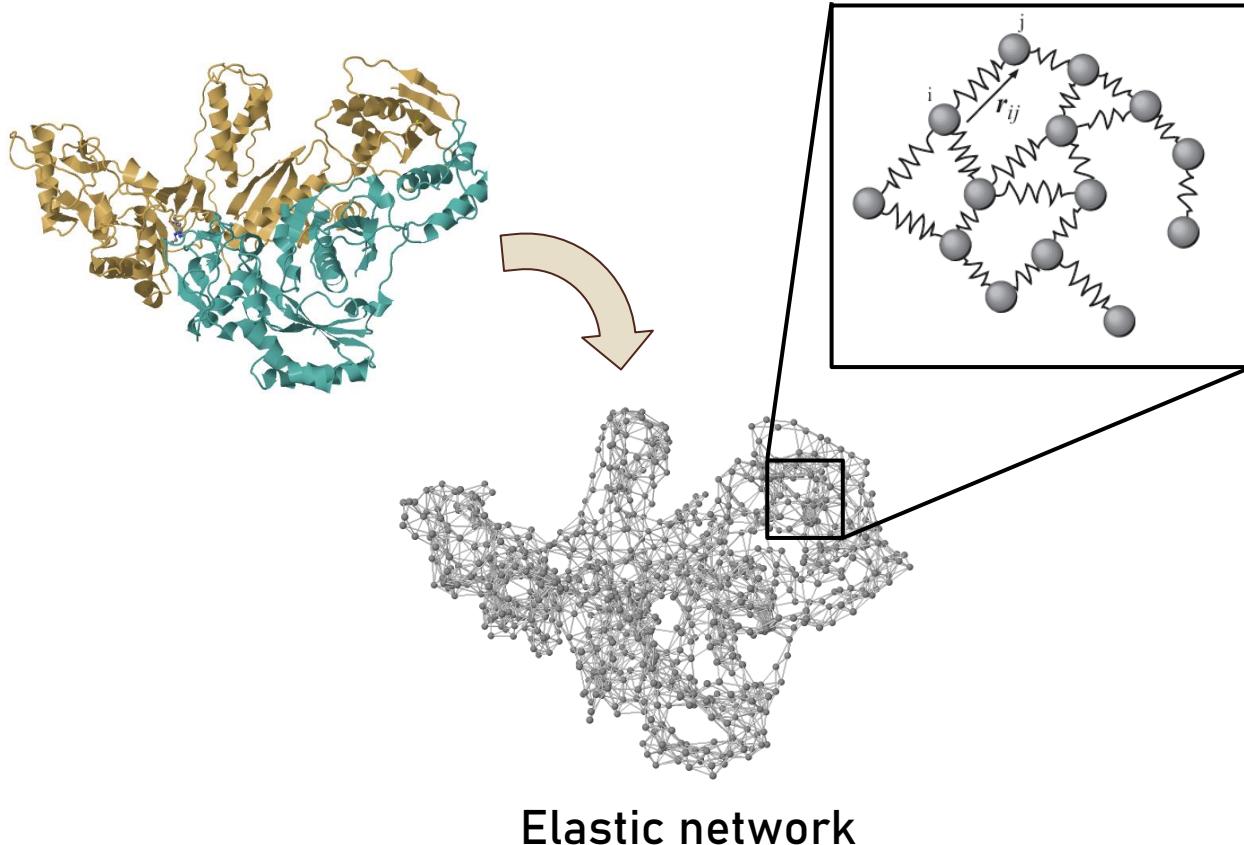


**GNM:** Bahar et al *Fold & Des* 1996; Haliloglu et al. *Phys Rev Lett* 1997

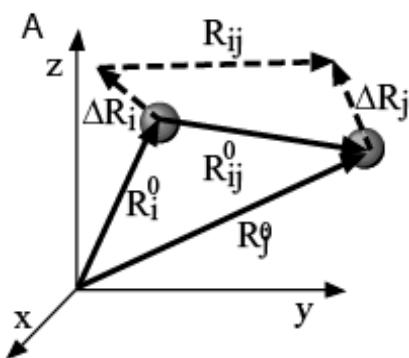
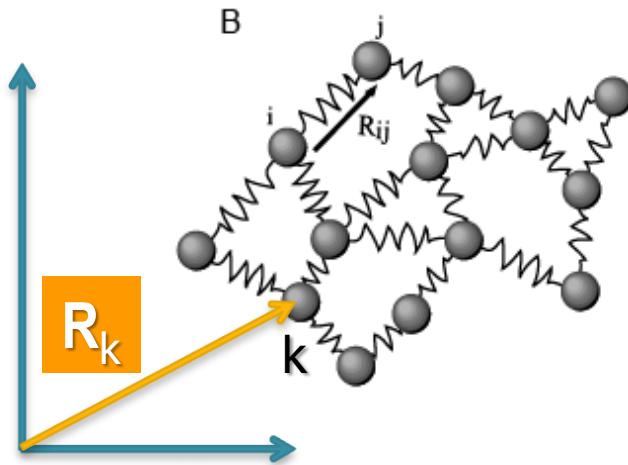
**ANM:** Doruker et al. *Proteins* 2000; Atilgan et al, *Biophys J* 2001

**Basic approach:**

**Mapping the structure to a network, the beads of which are the residues, and springs connect nearest spatial neighbors**



# Gaussian Network Model (GNM)



- Each node represents a residue
- Residue positions,  $\mathbf{R}_i$ , identified by  $\alpha$ -carbons' coordinates
- Springs connect residues located within a cutoff distance (e.g., 10 Å)

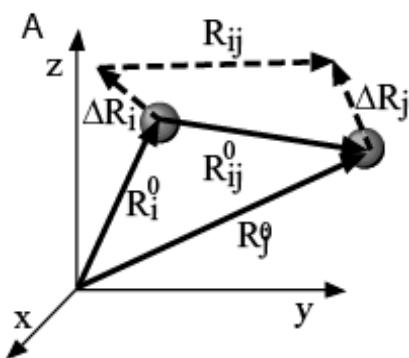
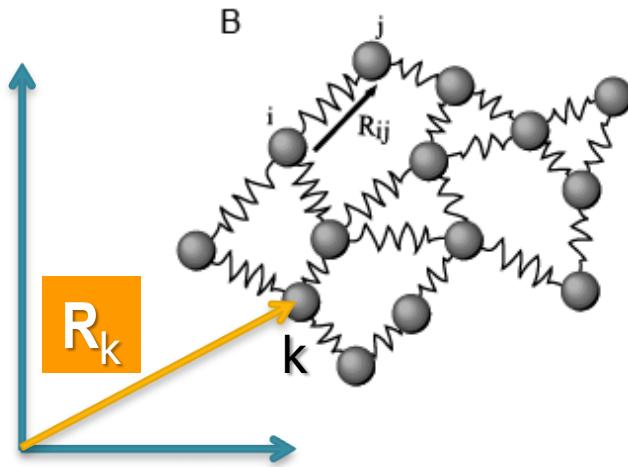
→ Nodes are subject to **Gaussian fluctuations**  $\Delta R_i$

→ Inter-residue distances  $R_{ij}$  also undergo Gaussian fluctuations

$$\rightarrow \Delta \mathbf{R}_{ij} = \Delta \mathbf{R}_j - \Delta \mathbf{R}_i$$

**Fluctuations in residue positions**

# Gaussian Network Model (GNM)



Fluctuation vector:

$$\rightarrow \Delta \mathbf{R} =$$

$$\begin{bmatrix} \Delta \mathbf{R}_1 \\ \Delta \mathbf{R}_2 \\ \Delta \mathbf{R}_3 \\ \Delta \mathbf{R}_4 \\ \dots \\ \dots \\ \dots \\ \Delta \mathbf{R}_N \end{bmatrix}$$

Fluctuations in residue positions

# Fluctuation with respect to starting structure $R(0)$

Instantaneous deviation for atom i

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

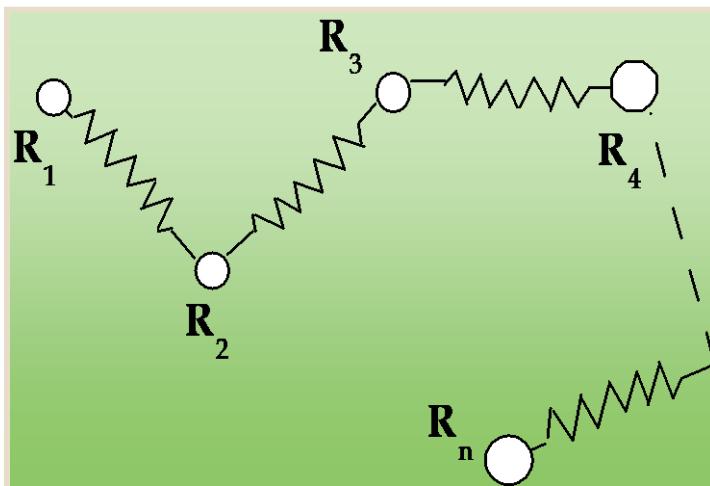
Under equilibrium conditions:

Average displacement from equilibrium:  $\langle \Delta R_i(t_k) \rangle = 0$

But the mean-square fluctuation (MSF),  $\langle (\Delta R_i(t_k))^2 \rangle \neq 0$

# Rouse model for polymers

Classical bead-and-spring model



$$\Delta R_{12} = \mathbf{R}_{12} - \mathbf{R}_{12}^0$$

$$\begin{aligned} V_{\text{tot}} &= (\gamma/2) [(\Delta R_{12})^2 + (\Delta R_{23})^2 + \dots + (\Delta R_{N-1,N})^2] \\ &= (\gamma/2) [(\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \dots] \end{aligned}$$

Kirchhoff matrix

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

# Rouse model for polymers

Kirchhoff matrix

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

Force constant

$$\begin{aligned} V_{\text{tot}} &= (\gamma/2) [(\Delta R_{12})^2 + (\Delta R_{23})^2 + \dots + (\Delta R_{N-1,N})^2] \\ &= (\gamma/2) [(\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \dots] \end{aligned}$$

# Rouse model for polymers

Fluctuation vector

$$(\gamma/2) [\Delta R_1 \ \Delta R_2 \ \Delta R_3 \dots \ \Delta R_N]$$

Kirchhoff matrix

$$\begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & \ddots & \\ & & & \ddots & -1 & 2 & -1 \\ & & & & 1 & -1 & \end{bmatrix} = \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \Delta R_3 \\ \vdots \\ \vdots \\ -1 \\ 2 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

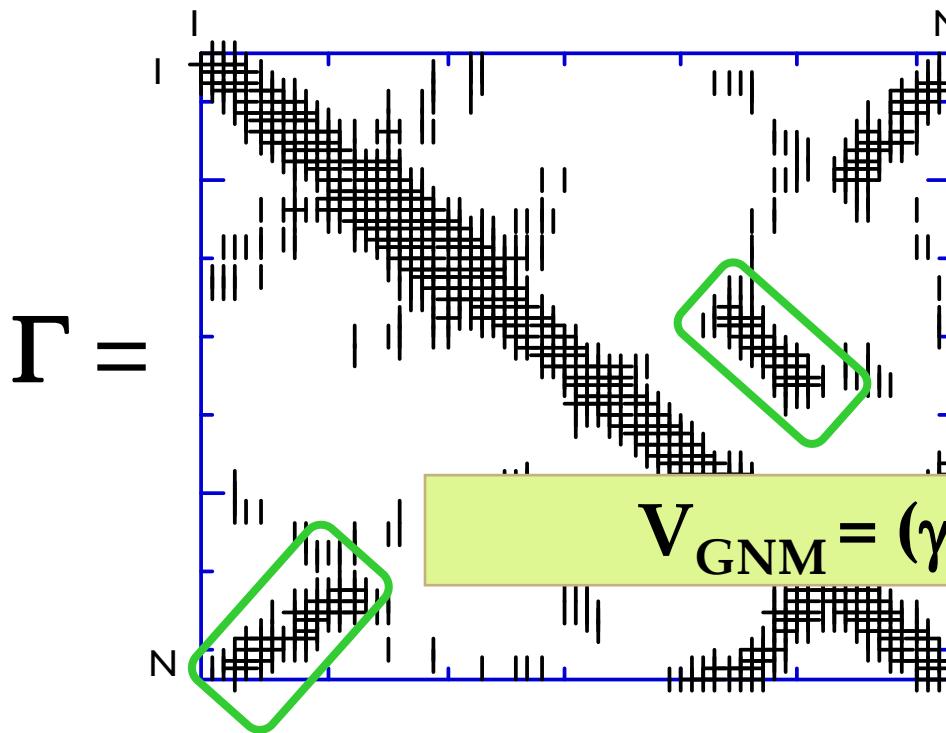
$$V_{\text{tot}} = (\gamma/2) \Delta R^T \Gamma \Delta R$$

Force constant

$$\begin{aligned} V_{\text{tot}} &= (\gamma/2) [(\Delta R_{12})^2 + (\Delta R_{23})^2 + \dots + (\Delta R_{N-1,N})^2] \\ &= (\gamma/2) [(\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \dots] \end{aligned}$$

# Kirchhoff matrix for inter-residue contacts

For a protein of  $N$  residues



$$\Gamma_{ik} = \begin{cases} -1 & \text{if } r_{ik} < r_{\text{cut}} \\ 0 & \text{if } r_{ik} > r_{\text{cut}} \end{cases}$$
$$\Gamma_{ii} = - \sum_k \Gamma_{ik}$$

$$V_{GNM} = (\gamma/2) \Delta R^T \Gamma \Delta R$$

$\Gamma$  provides a complete description of contact topology!

# An alternative definition of spring constant: distance dependent $\gamma$

$$U_{\text{elastic}} = \frac{1}{2} \sum_{i < j} k(R_{ij}) (r_{ij} - R_{ij})^2,$$

HCA model

Hinsen et al Harmonicity in slow protein dynamics. Chem Phys. 2000; 261:25–37.

$$k(R) = \begin{cases} 205.5 \cdot R - 571.2 & \text{if } r \leq 4.0 \text{ \AA} \\ 305.9 \times 10^3 \cdot R^{-6} & \text{if } r > 4.0 \text{ \AA}, \end{cases}$$

where the unit for  $k(R)$  is kcal mol<sup>-1</sup> Å<sup>-2</sup>

See also: Riccardi D, Cui Q, Phillips G-N. Application of elastic network models to proteins in the crystalline state. Biophys J. 2009;96:464–475.

# Statistical mechanical averages

$$\langle f(x) \rangle = \int f(x) p(x) dx = \frac{\int f(x) w(x) dx}{\int w(x) dx} = \frac{\int f(x) w(x) dx}{Z}$$

Suppose  $f$  = cross-correlation between residue fluctuations =  $\langle (\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j) \rangle$   
And  $x$  represents the conformational changes (multiple modes of motion)

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = (1/Z_N) \int (\Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j) e^{-V/k_B T} d\{\Delta \mathbf{R}\}$$

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

$\Gamma$  provides a complete description of equilibrium fluctuations!

# Kirchhoff/connectivity matrix fully defines

the **cross-correlations** between residue motions

$$[\Gamma^{-1}]_{ij} \sim \langle (\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j) \rangle$$

and the **mean-square fluctuations of residues**

$$[\Gamma^{-1}]_{ii} \sim \langle (\Delta\mathbf{R}_i)^2 \rangle$$

# 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  $i$ ,

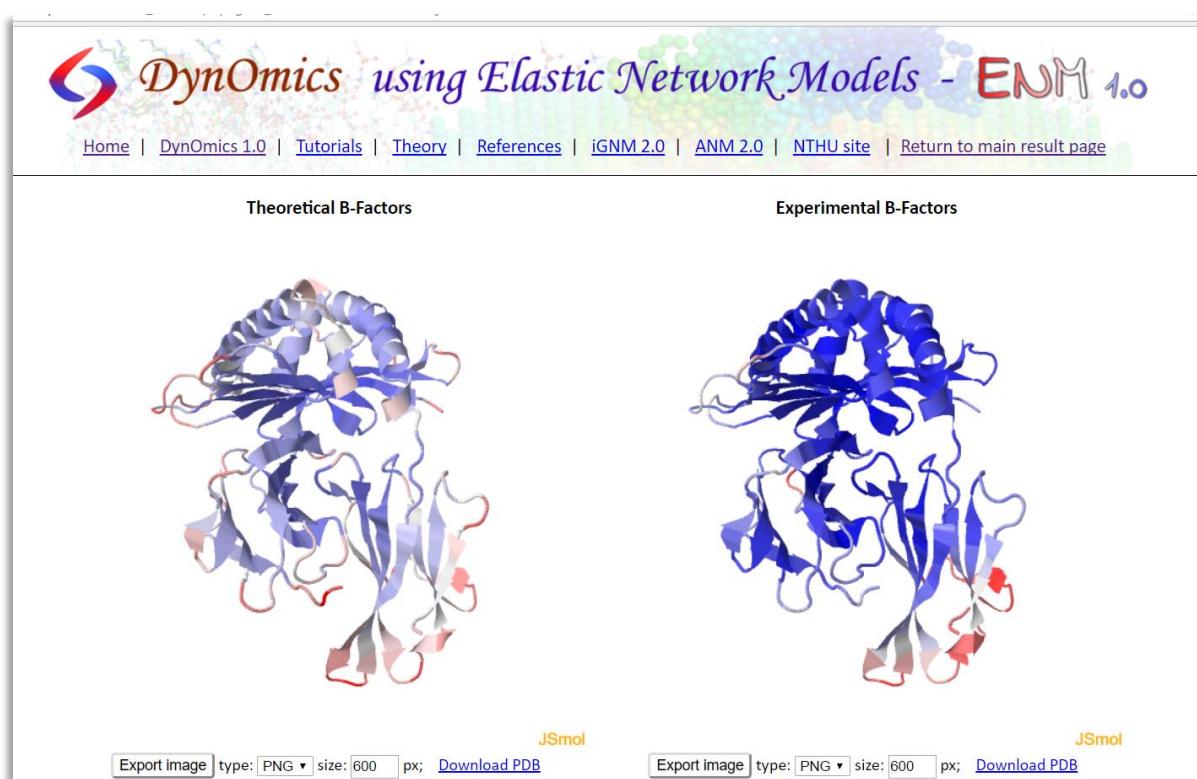
$$B_i = [8\pi^2/3] \langle (\Delta R_i)^2 \rangle$$

How do residue MSFs compare with the B-factors?

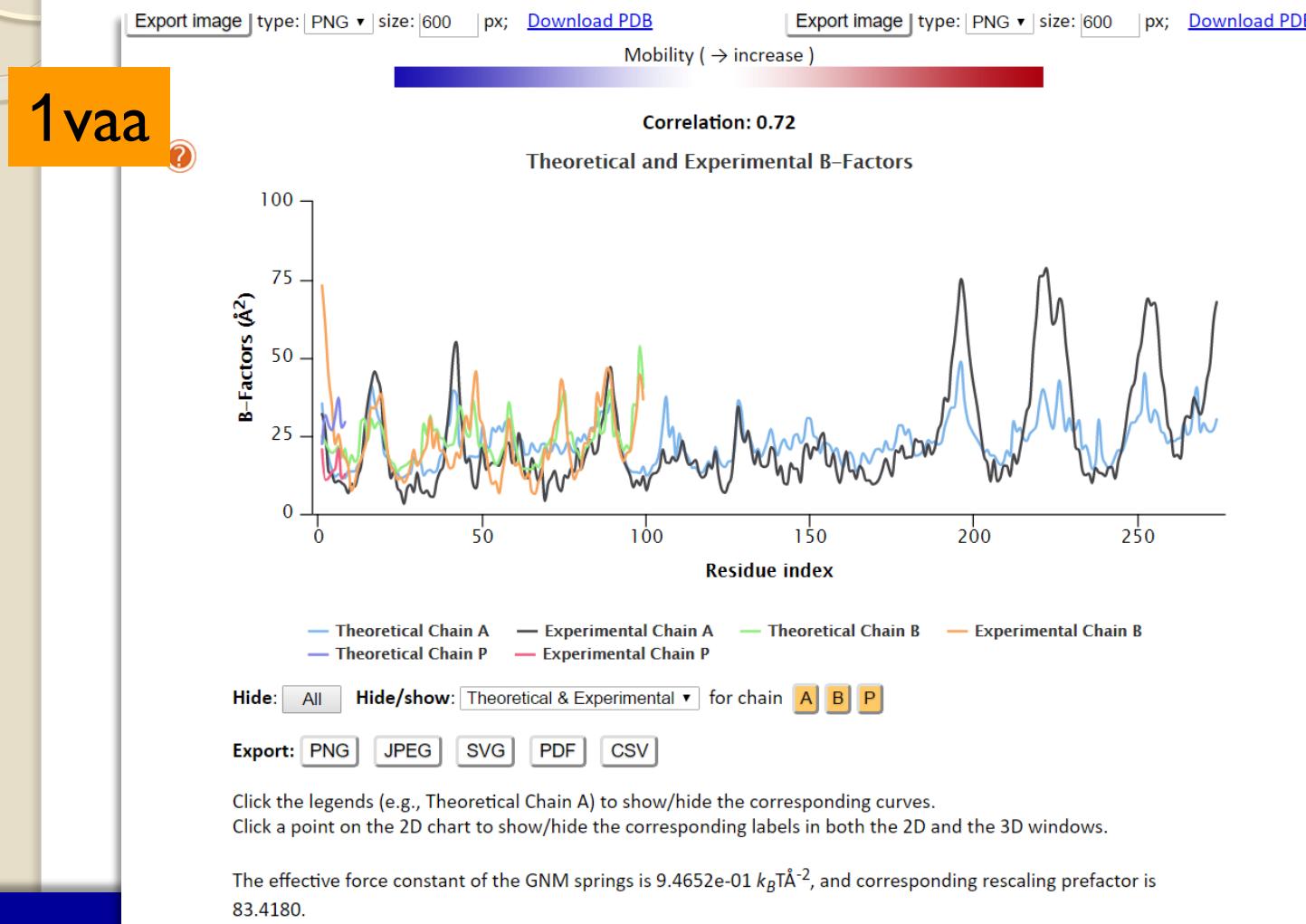
# Output from DynOmics

Example: 1vaa

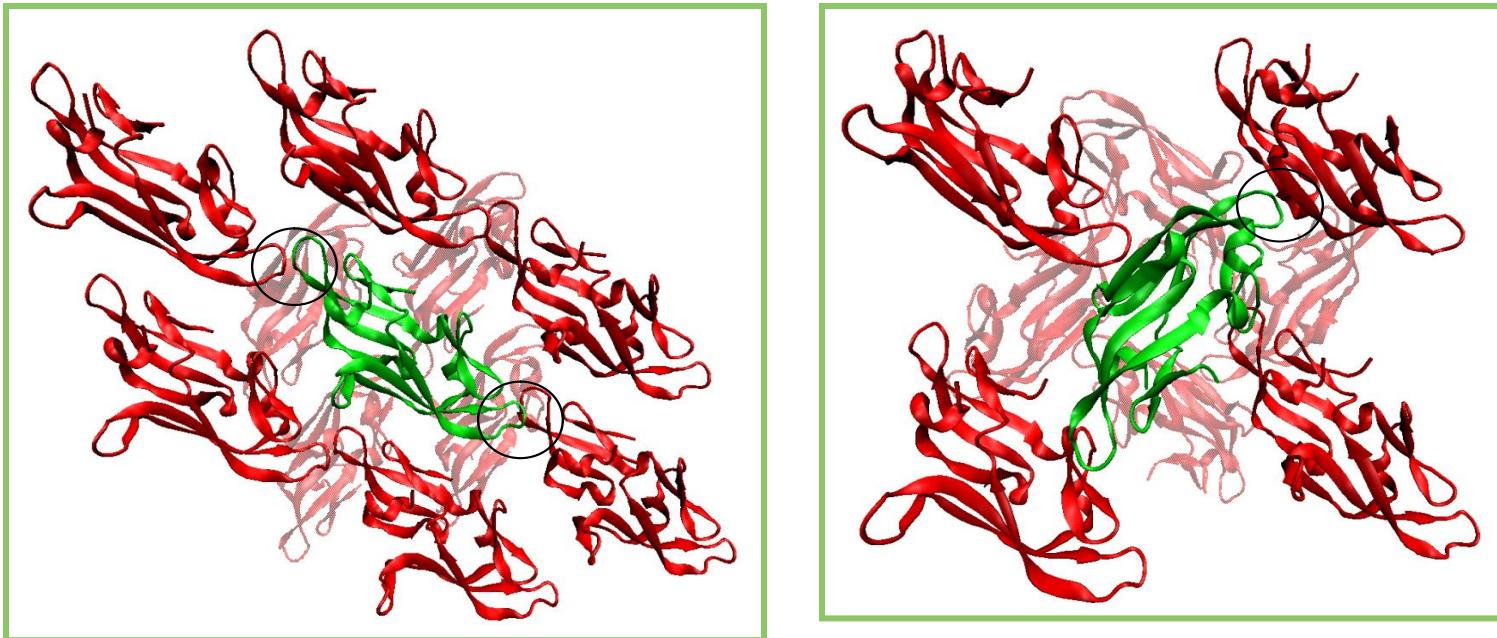
PDB title: CRYSTAL STRUCTURES OF  
TWO VIRAL PEPTIDES IN COMPLEX  
WITH MURINE MHC CLASS I H-2KB



# Output from DynOmics

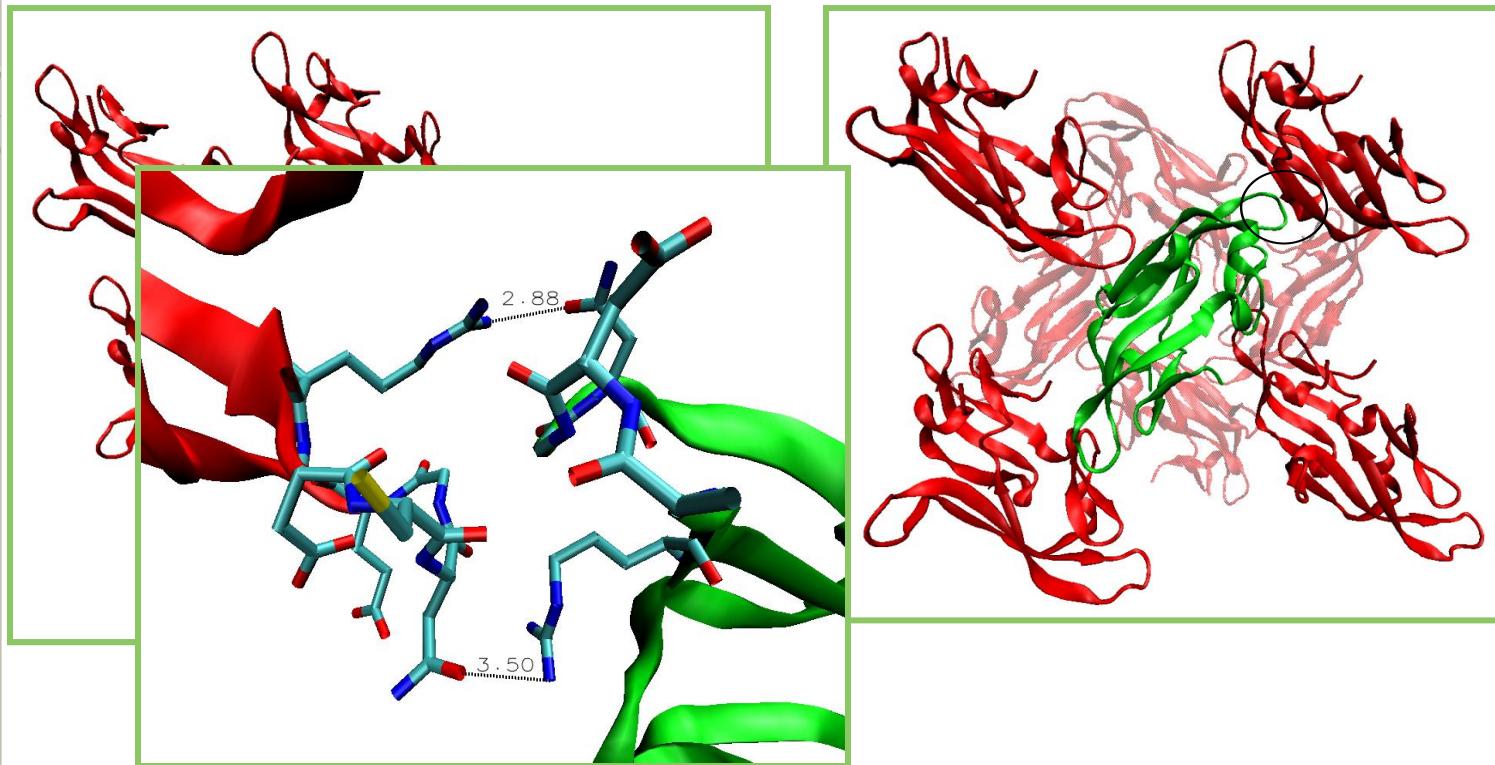


# B-factors are affected by crystal contacts



Two X-ray structures for a designed sugar-binding protein LKAMG

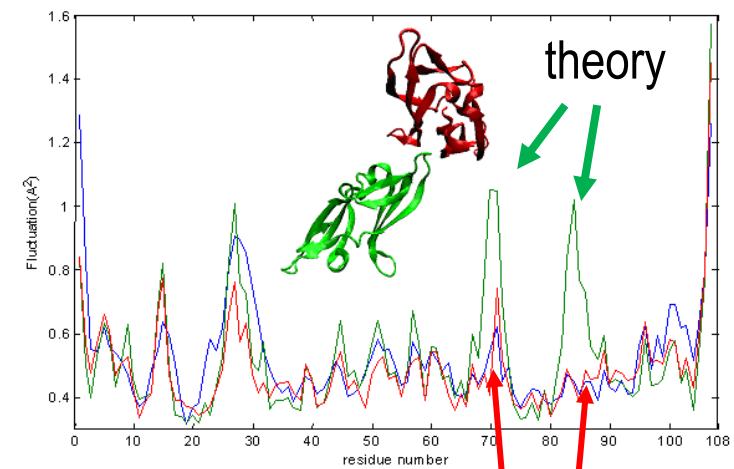
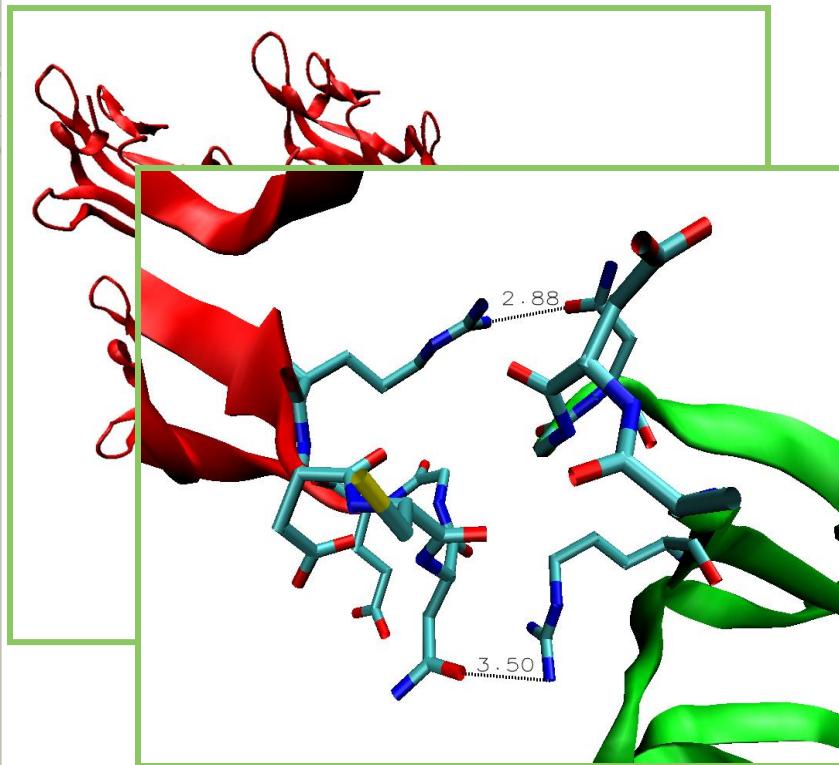
# B-factors are affected by crystal contacts



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

FOR MORE INFO...

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

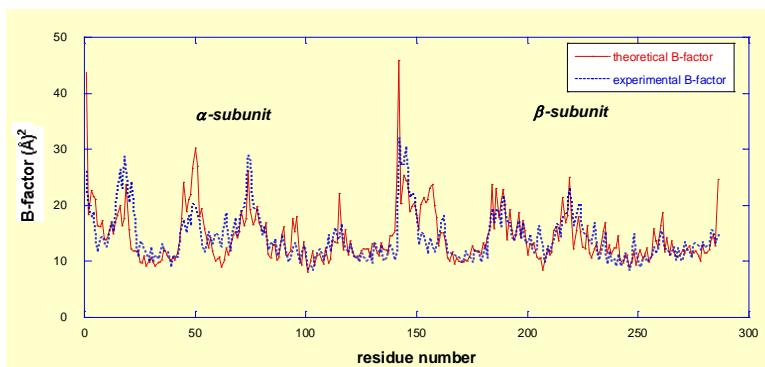
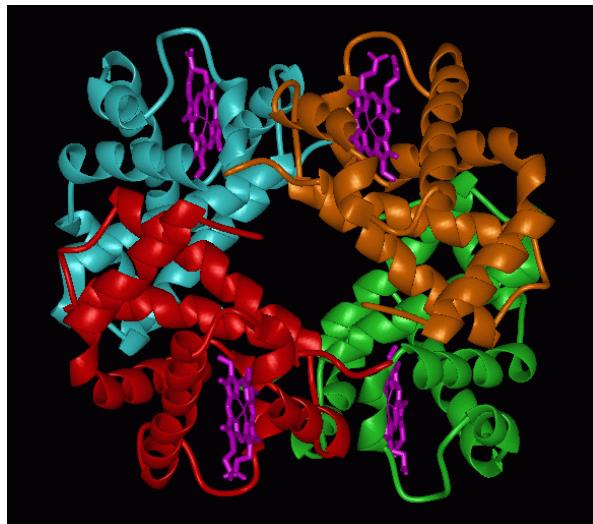


Crystal contacts

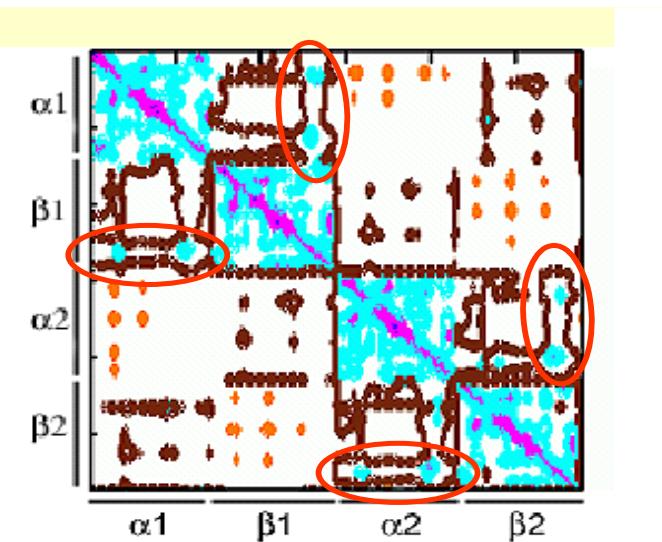
Particular loop motions are curtailed by intermolecular contacts in the crystal environment causing a discrepancy between theory and experiments

FOR MORE INFO...

# 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 ≤

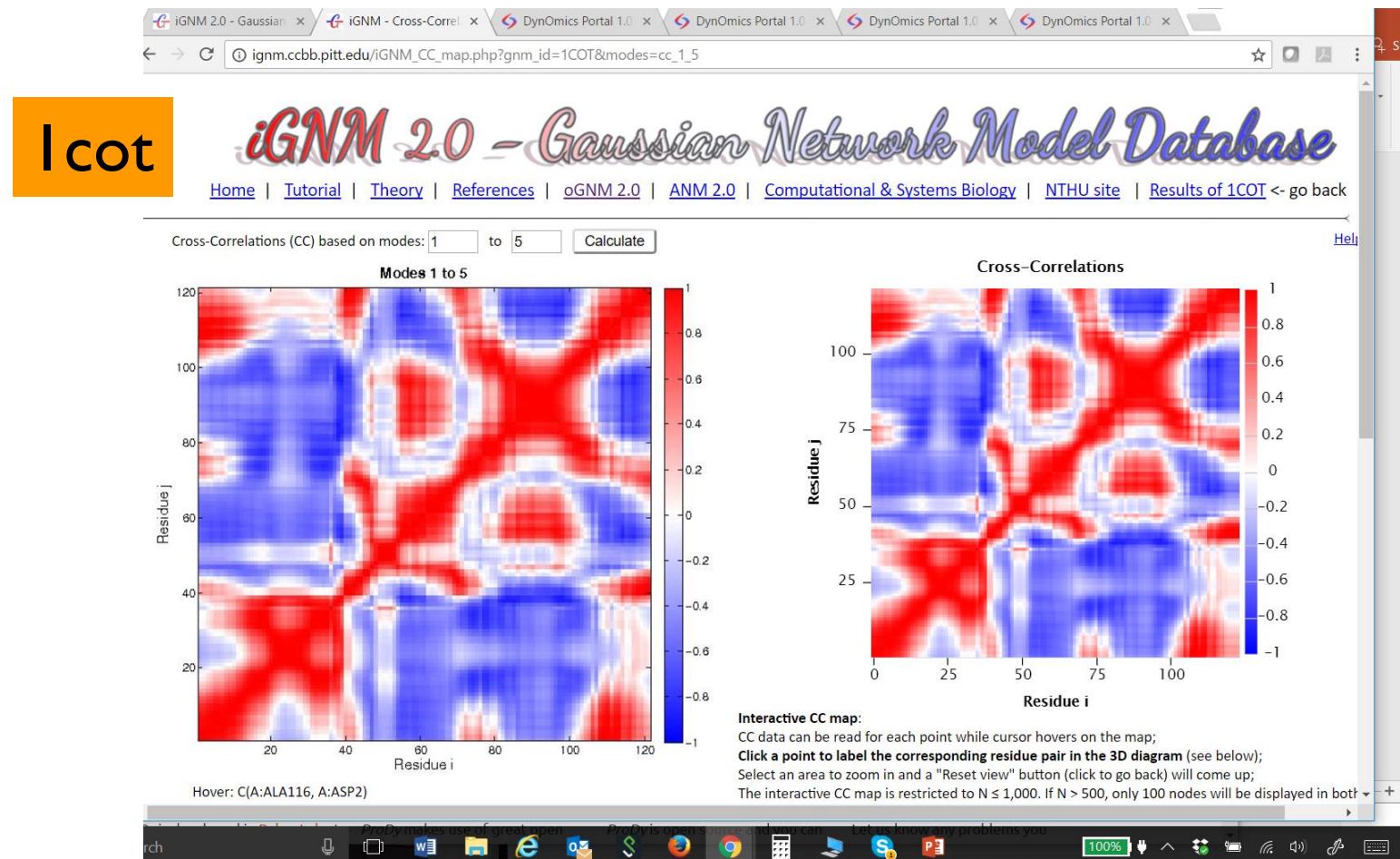
$$\frac{<(\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j)>}{[<(\Delta\mathbf{R}_i)^2> <(\Delta\mathbf{R}_j)^2>]^{1/2}}$$

≤ 1

Fully  
anticorrelated

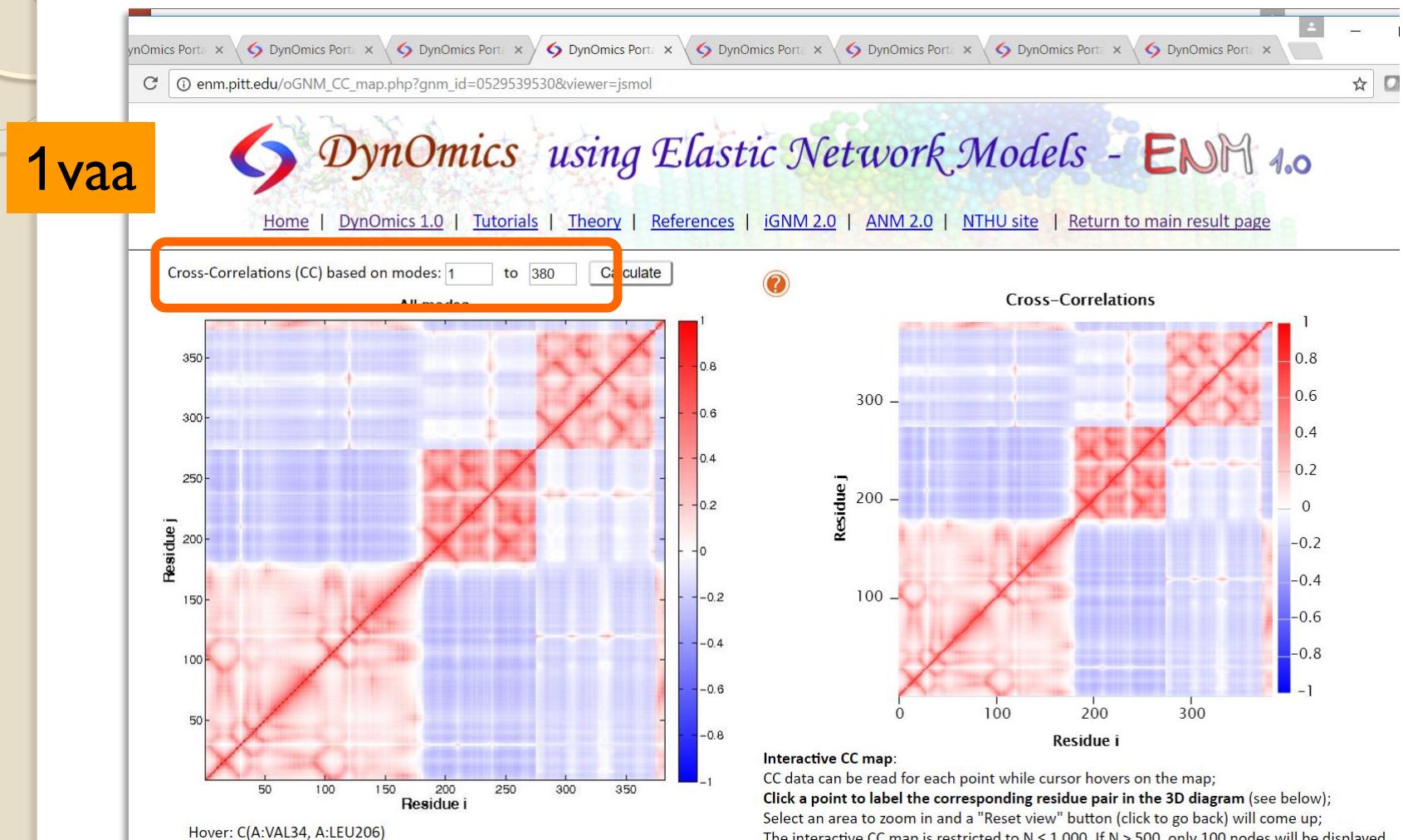
Fully  
correlated

# Output from iGNM



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

# Output from DynOmics - ENM



# Cross-Correlations

are elements of the

## Covariance Matrix $\mathbf{C}$

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

Covariance scales with the inverse of the Kirchhoff matrix.

The proportionality constant is  $3kT/\gamma$

# Covariance matrix ( $N \times N$ )

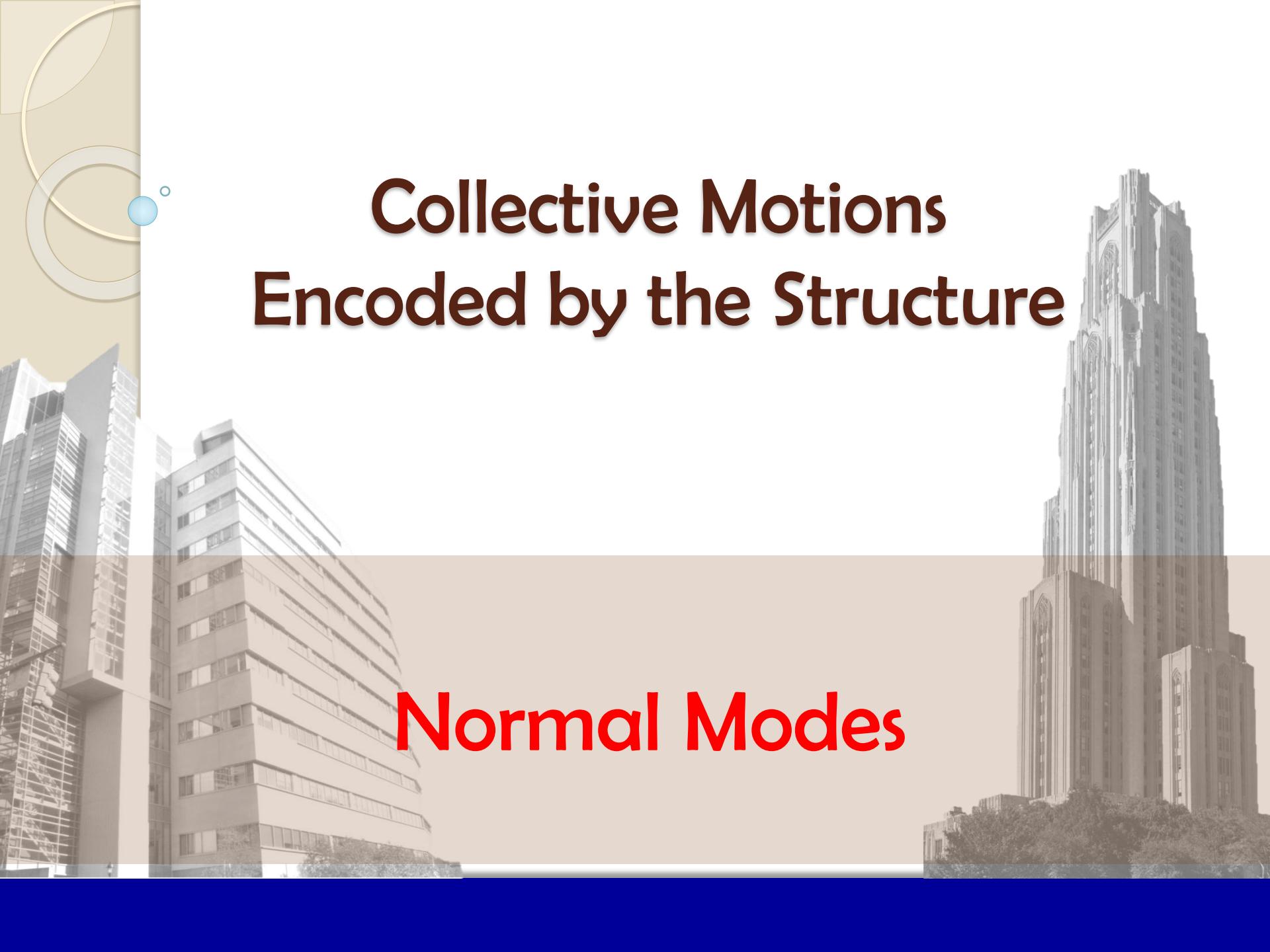
**C**

$\langle \Delta\mathbf{R}_1 \cdot \Delta\mathbf{R}_1 \rangle$	$\langle \Delta\mathbf{R}_1 \cdot \Delta\mathbf{R}_2 \rangle$	...	...	$\langle \Delta\mathbf{R}_1 \cdot \Delta\mathbf{R}_N \rangle$
$\langle \Delta\mathbf{R}_2 \cdot \Delta\mathbf{R}_1 \rangle$	$\langle \Delta\mathbf{R}_2 \cdot \Delta\mathbf{R}_2 \rangle$			
...				
...				
$\langle \Delta\mathbf{R}_N \cdot \Delta\mathbf{R}_1 \rangle$				$\langle \Delta\mathbf{R}_N \cdot \Delta\mathbf{R}_N \rangle$

$$= \Delta\mathbf{R} \Delta\mathbf{R}^T$$

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

$\langle \Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j \rangle$  = ms fluctuation of site  $i$  averaged over time (or all  $m$  snapshots).



# **Collective Motions Encoded by the Structure**

**Normal Modes**

# Eigenvalue decomposition of $\Gamma$

$$\Gamma = \mathbf{U} \Lambda \mathbf{U}^T$$

where  $\Lambda$  is the diagonal matrix of eigenvalues

$$\Lambda = \begin{matrix} & \lambda_0 & & & \\ & & \lambda_1 & & \\ & & & \lambda_2 & \\ & & & & \lambda_3 \\ & & & & \\ & & & & \lambda_{N-1} \end{matrix}$$

$$\lambda_0 = 0  
(\text{zero eigenvalue})$$

$$\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{N-1}$$

# Eigenvalue decomposition of $\Gamma$

$$\Gamma = \mathbf{U} \Lambda \mathbf{U}^T$$

and  $\mathbf{U}$  is the matrix of eigenvectors

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_{N-1} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_{11} & \mathbf{u}_{21} & \dots & \mathbf{u}_{N1} \\ \mathbf{u}_{12} & \mathbf{u}_{22} & \dots & \mathbf{u}_{N2} \\ \mathbf{u}_{13} & \mathbf{u}_{23} & \dots & \mathbf{u}_{N3} \\ \mathbf{u}_0 & \mathbf{u}_1 & \dots & \mathbf{u}_{N-1} \\ \mathbf{u}_{1N} & \mathbf{u}_{2N} & \dots & \mathbf{u}_{NN} \end{bmatrix}$$

$$\mathbf{U}^T = \begin{bmatrix} \mathbf{u}_0^T \\ \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_{N-1}^T \end{bmatrix}$$

# Eigenvalue decomposition of $\Gamma$

In component form

$$\Gamma_{ij} = \sum_k \mathbf{U}_{ik} \Lambda_k [\mathbf{U}^T]_{kj}$$

$$\Gamma = \sum_k \lambda_k \mathbf{u}_k \mathbf{u}_k^T$$

Note:

$$\mathbf{U}^T = \mathbf{U}^{-1}$$

Such that

$$\Gamma^{-1} = \mathbf{U} \Lambda^{-1} \mathbf{U}^T$$

Pseudoinverse

$$\Gamma^{-1} = \sum_{k=1}^{N-1} \lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T$$

# Several modes contribute to dynamics

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

Contribution of mode k

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

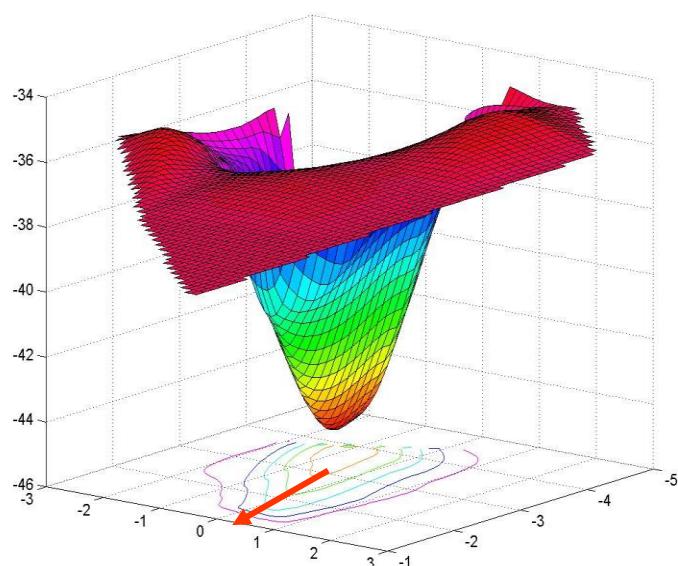
Contribution of mode k

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

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

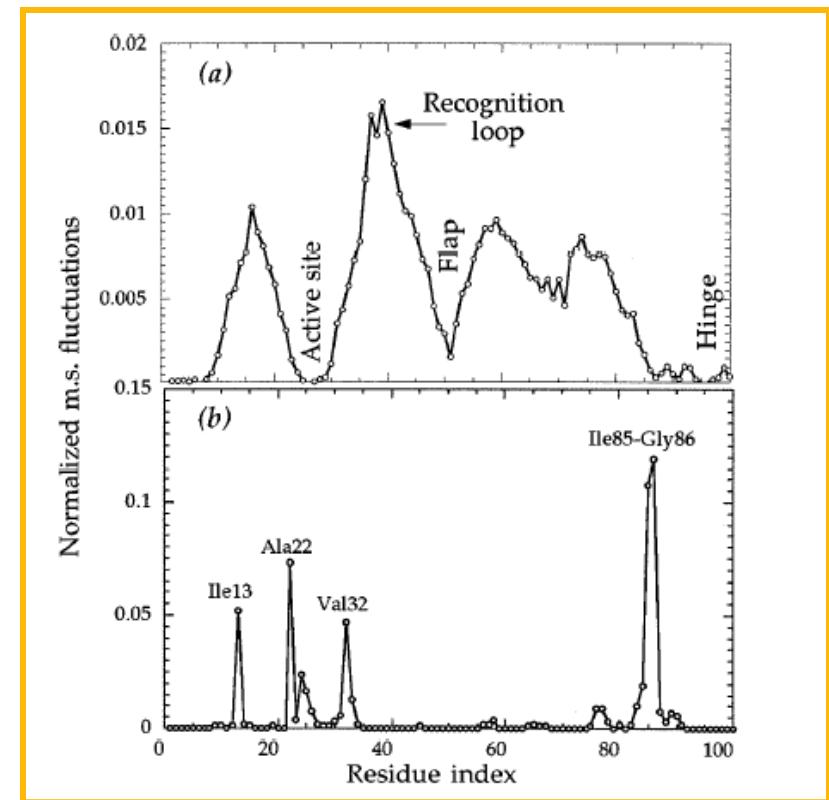
FOR MORE INFO...

# Several modes contribute to dynamics

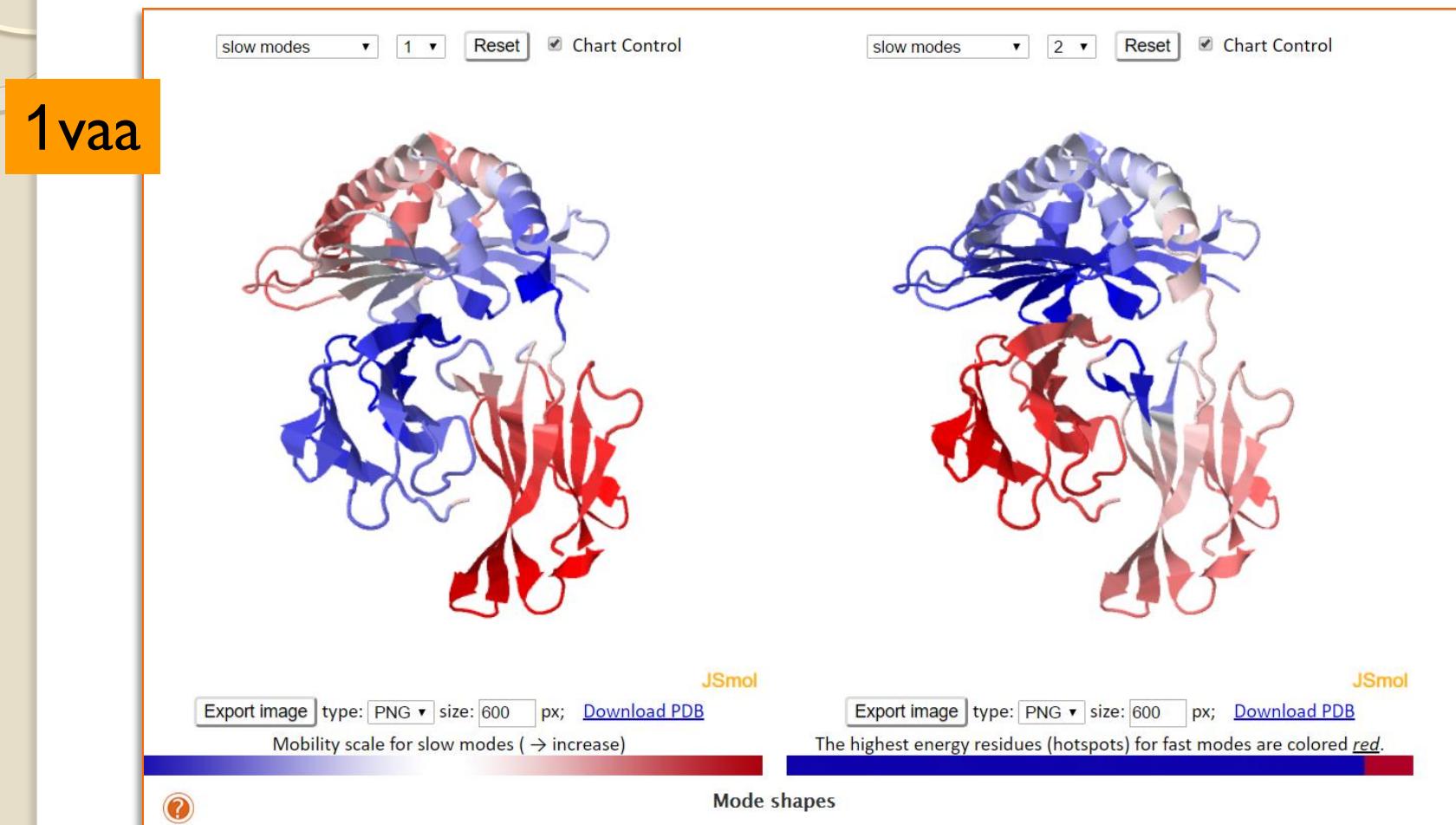


The first mode selects  
the 'easiest' collective motion

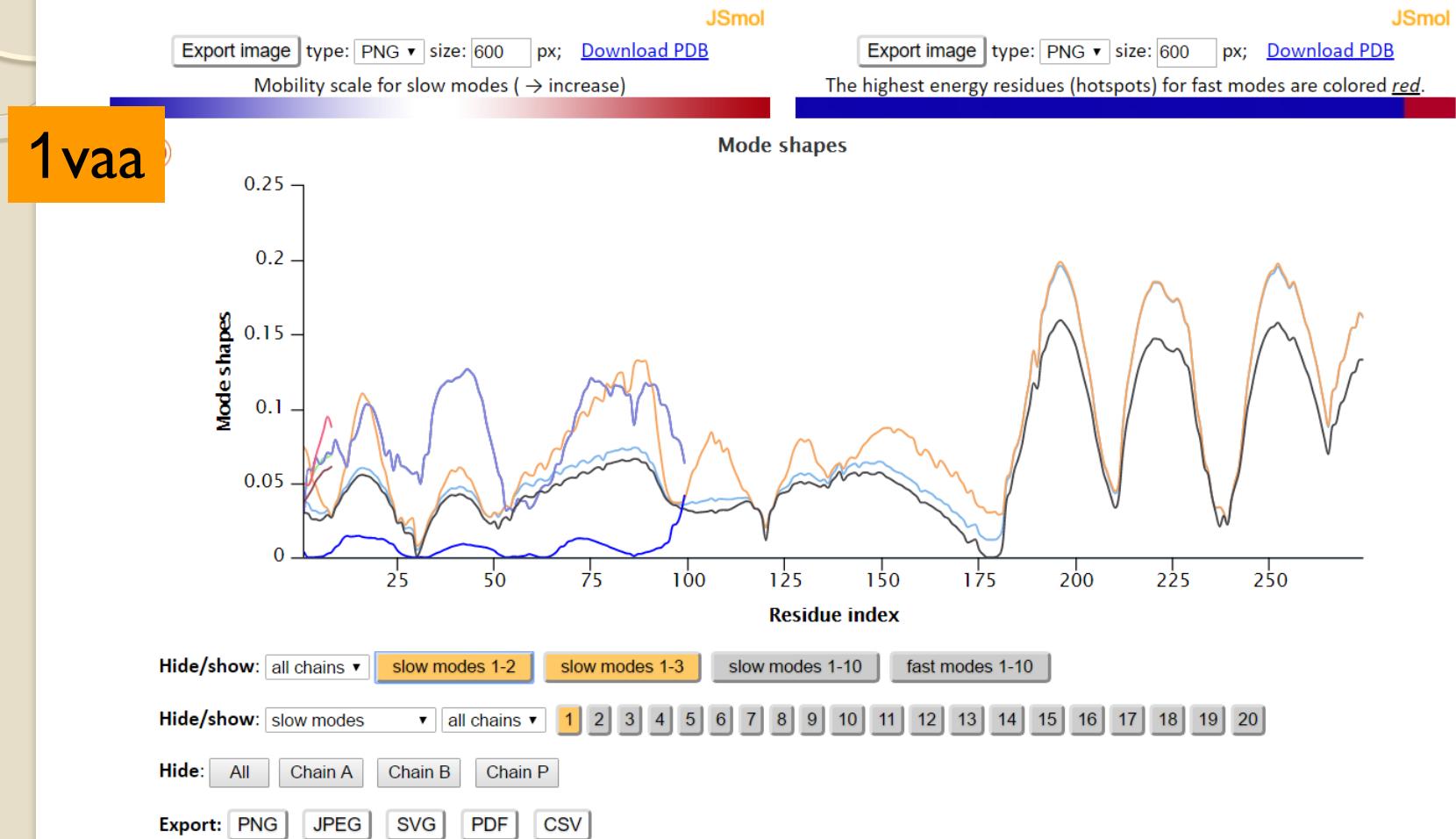
FOR MORE INFO...



# Output from DynOmics

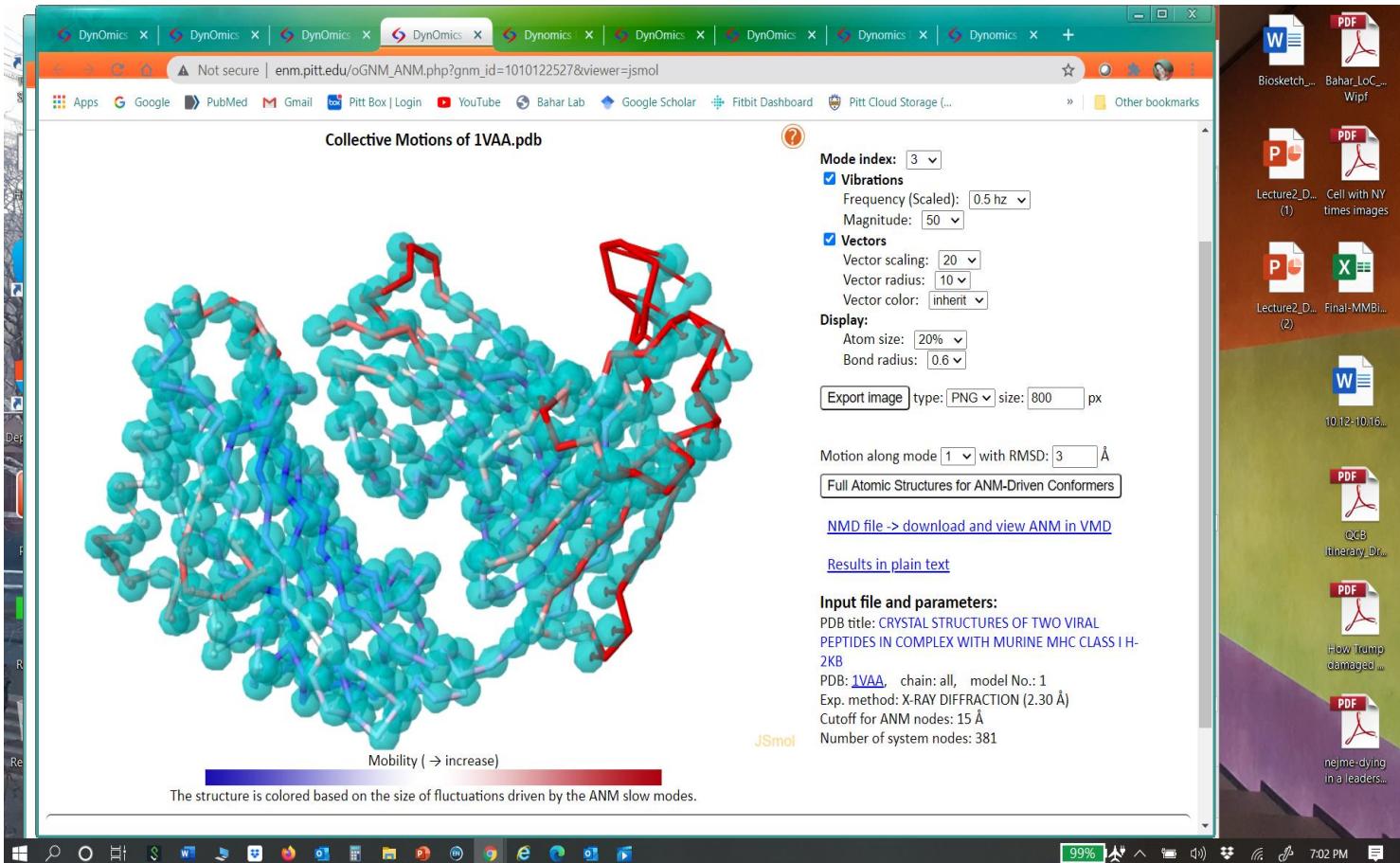


# Output from DynOmics

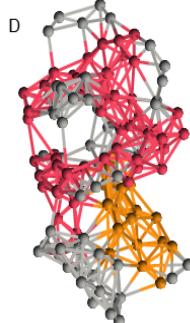


Click a point on the 2D chart to show/hide the corresponding labels in both the 2D chart and the 3D windows above if the "Chart Control" is

# Animations (different modes)



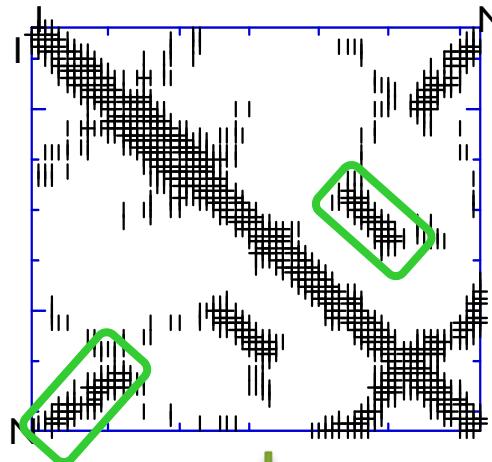
# Summary - Gaussian network model (GNM)



Kirchhoff matrix for inter-residue contacts

Contact:  $R_{ij} < 10\text{\AA}$

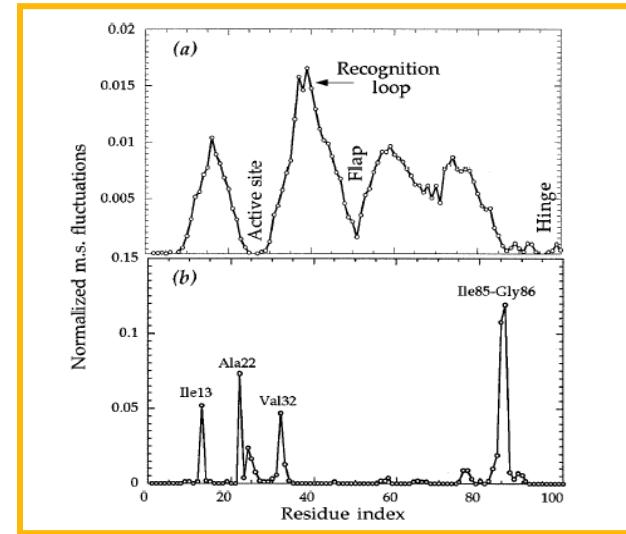
$$\Gamma =$$



MSF of residue  $i$   
 $= \langle (\Delta R_i)^2 \rangle$

$$\langle (\Delta R_i)^2 \rangle = (3 k_B T / \gamma) [\Gamma^{-1}]_{ii}$$

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



**Several modes of motion contribute to dynamics**

$$\langle (\Delta R_i)^2 \rangle = (3 k_B T / \gamma) [\Gamma^{-1}]_{ii}$$

# Recipe (GNM)

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



## Properties

- the eigenvalues scale with the frequency squared ( $\lambda_i \sim \omega_i^2$ )
- eigenvector  $u_k$  is an N-dim vector
- the  $i^{\text{th}}$  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, higher weights

# Database of GNM results

The screenshot shows a web browser window displaying the *iGNM 2.0 - Gaussian Network Model Database*. The URL [ignm.ccbb.pitt.edu](http://ignm.ccbb.pitt.edu) is highlighted with an orange box. The main content area features a 3D ribbon model of a protein-DNA complex, with the DNA shown in red. A green box highlights the URL [ignm.ccbb.pitt.edu](http://ignm.ccbb.pitt.edu) again. The page includes a detailed description of the GNM method, search options (PDB ID, Advanced search), and contact information for the University of Pittsburgh.

**Contact:**  
The server is maintained by Dr. Hongchun Li in the Bahar Lab at the Department of Computational & Systems Biology at the University of Pittsburgh, School of Medicine, and sponsored by the NIH awards #5R01GM099738-04 and #5P41GM103712-03 and the funding #104-2113-M-007-019 from MOST to the Yang lab at the National Tsing Hua University, Taiwan.

For questions and comments please contact [Hongchun Li](#).

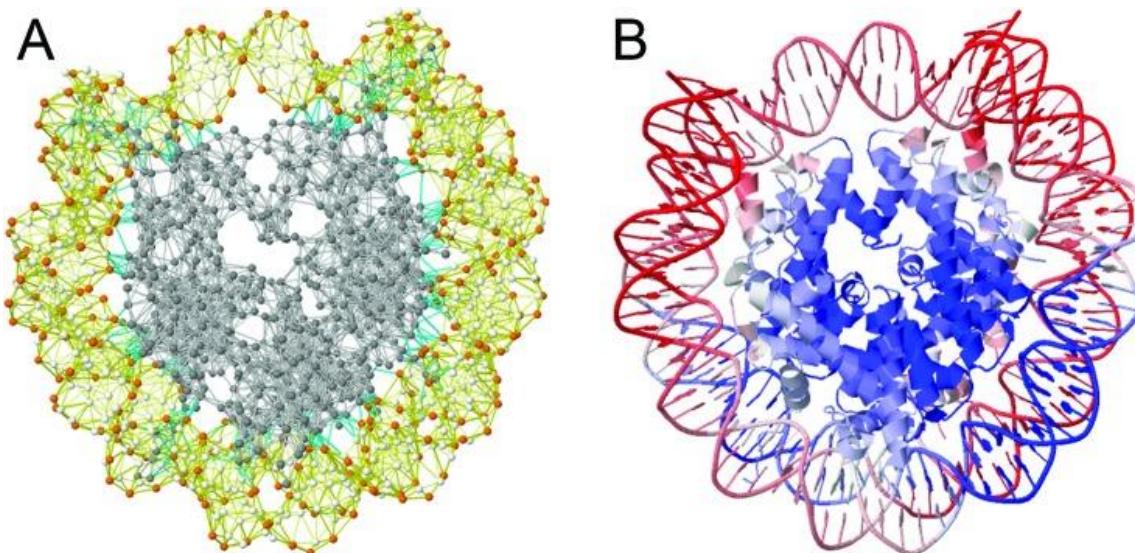
**Li, Chang, Yang and Bahar (2016)  
Nucleic Acids Res 44: D415-422**

# Why use iGNM2.0?

- Easy access to precomputed results for 95% of the PDB including
  - 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)





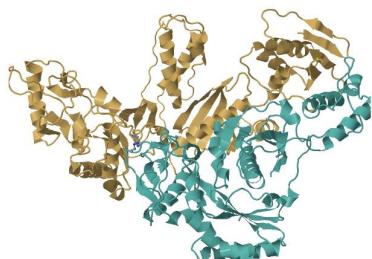
# Anisotropic Network Model (ANM)



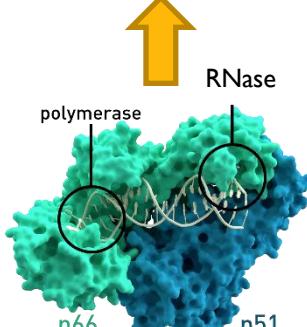
Motions in 3D

# Biological function entails both chemical and physical events

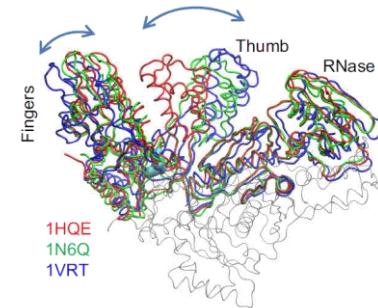
HIV-1 reverse transcriptase



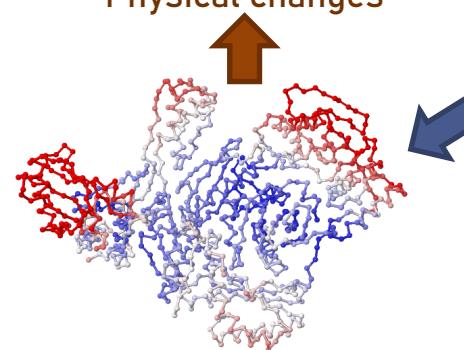
Molecular functions



Chemical changes



Physical changes



(Physical) dynamics

ANM

Bakan, A. and Bahar, I., 2009. *PNAS*, 106(34), pp.14349-14354.  
Tu, X., Das, K., Han, Q., et al., 2010. *Nature Nature Struc Mol Biol*, 17(10), p.1202.

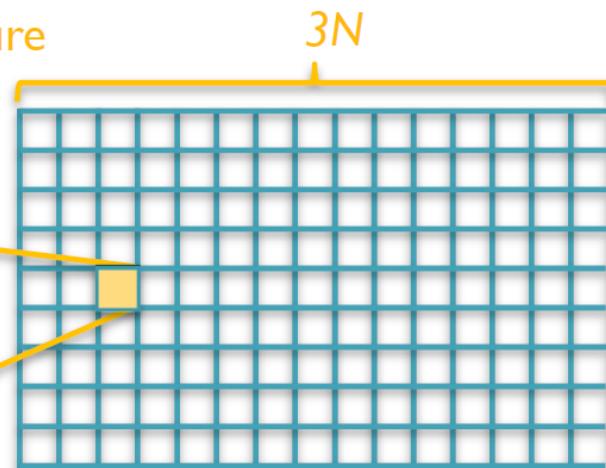
# Anisotropic Network Model

$$V(\mathbf{r}) = \frac{\gamma}{2} \sum_{i=1}^N \sum_{j>i} \underbrace{\left( |\mathbf{r}_{ij}| - |\mathbf{r}_{ij}^0| \right)^2}_{\text{Harmonic}} \underbrace{\Theta(R_c - |\mathbf{r}_{ij}^0|)}_{\text{Step function}}$$

$$\left( \frac{\partial^2 V}{\partial x_i \partial y_j} \right)_{\mathbf{r}^0} = - \frac{x_i^0 y_j^0}{|\mathbf{r}_{ij}^0|^2}$$

Hessian is calculated directly from structure

$$\mathbf{H}_{ij} = -\frac{\gamma}{(R_{ij}^0)^2} \begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & x_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & y_{ij}^0 z_{ij}^0 \\ x_{ij}^0 z_{ij}^0 & y_{ij}^0 z_{ij}^0 & (z_{ij}^0)^2 \end{bmatrix}$$



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



# Eigenvalue decomposition of H

$$H = \sum V K [V^T]$$

In component form

$$H = \sum_k \kappa_k v_k v_k^T$$

Note:

$$V^T = V^{-1}$$

Such that

$$H^{-1} = V \kappa^{-1} V^T$$

$$H^{-1} = \sum_{k=1}^{3N-6} \kappa_k^{-1} v_k v_k^T$$

ANM covariance matrix

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

$C_{3N} =$

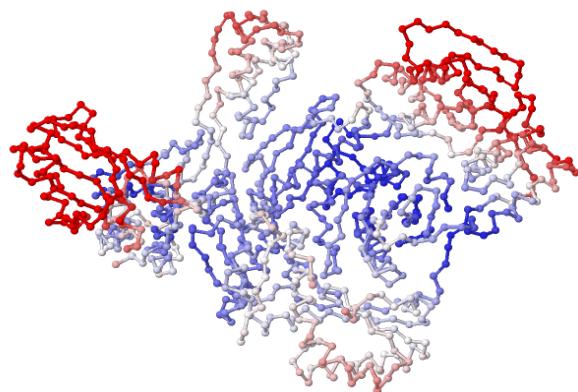
$C_{11}$	$C_{21}$	$C_{13}$		$C_{1N}$
$C_{12}$	$C_{22}$			
$C_{N1}$				$C_{NN}$

**$3N \times 3N$**

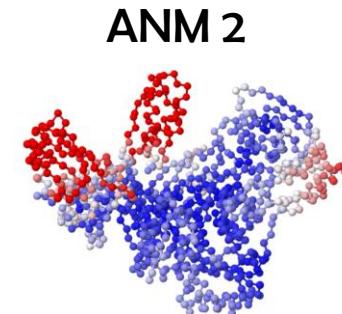
$\begin{matrix} <\Delta X_1 \Delta X_2> & <\Delta X_1 \Delta Y_2> & <\Delta X_1 \Delta Z_2> \\ <\Delta Y_1 \Delta X_2> & <\Delta Y_1 \Delta Y_2> & <\Delta Y_1 \Delta Z_2> \\ <\Delta Z_1 \Delta X_2> & <\Delta Z_1 \Delta Y_2> & <\Delta Z_1 \Delta Z_2> \end{matrix}$

# Collective motions (softest modes) intrinsically accessible to HIV-1 reverse transcriptase

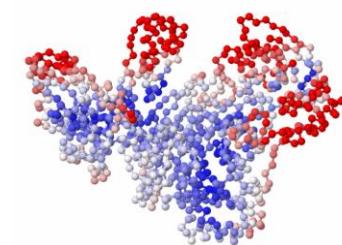
## Anisotropic Network Model (ANM)



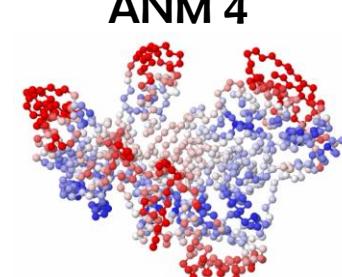
Collective motions (ANM 1)



ANM 2



ANM 3



ANM 4

Energetical favorability ↑

<http://dynamics.pitt.edu/>

Bahar, I., Lezon, T. R., Bakan, A., & Srivastava, I. H. (2010). *Chemical Reviews* 110, 1463-1497.

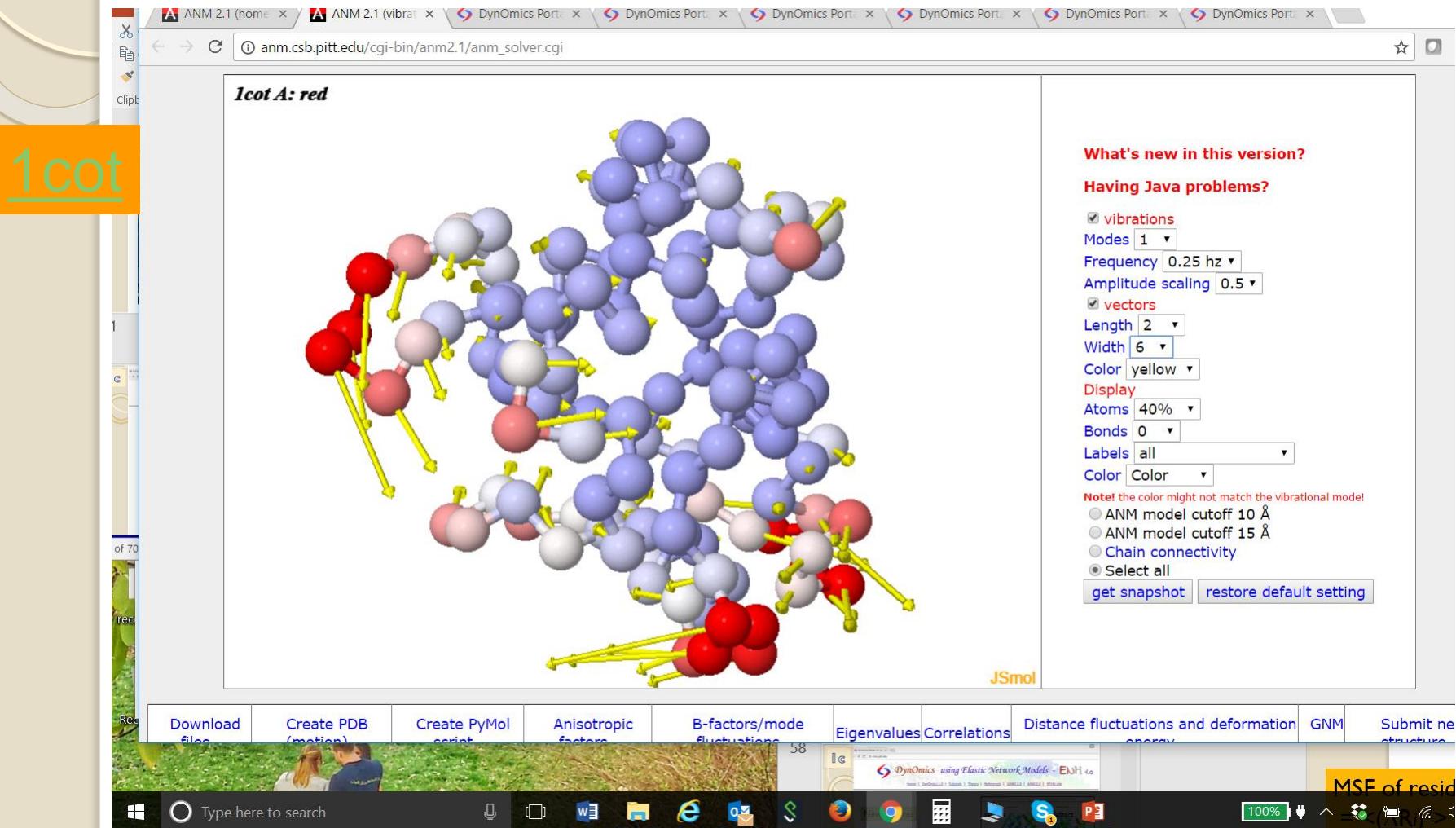
# ANM server

<http://anm.csb.pitt.edu/>

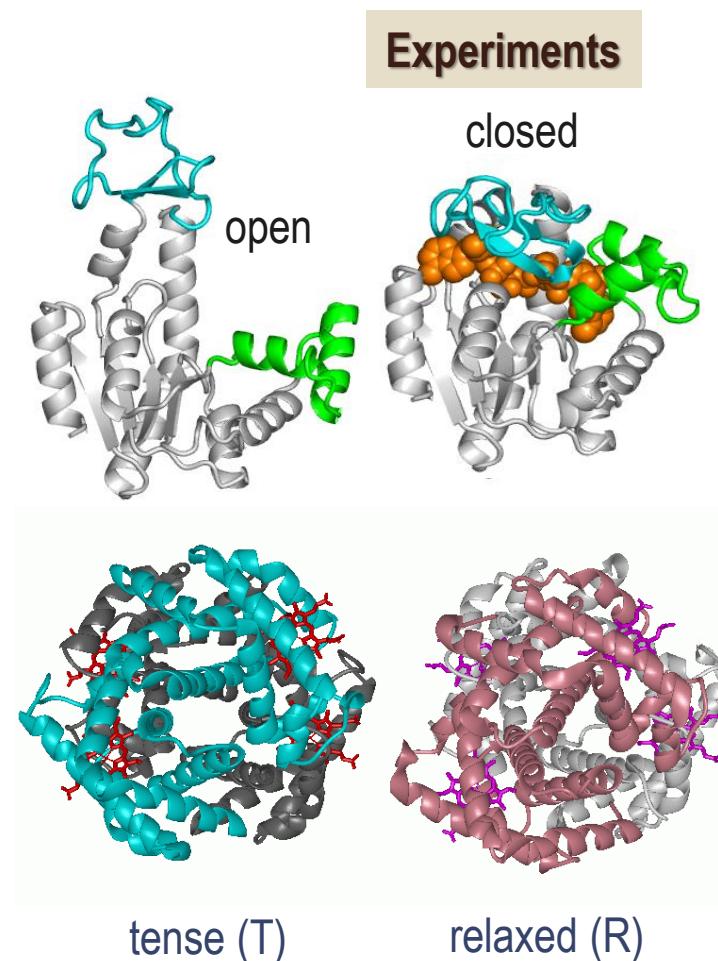
The screenshot shows the ANM server interface. At the top, the URL <http://anm.csb.pitt.edu/> is highlighted with an orange box. Below it, the browser address bar contains `anm.csb.pitt.edu/cgi-bin/anm2/anm2.cgi`. The main title is **Anisotropic Network Model Web Server 2.0 (2014)**. The page includes instructions for entering a PDB ID or uploading a protein file. To the right, a 3D molecular model is displayed with a complex network of red and blue lines representing interactions.

Eyal et al., *Bioinformatics* 2015

# Output from ANM server



# Softest modes are functional

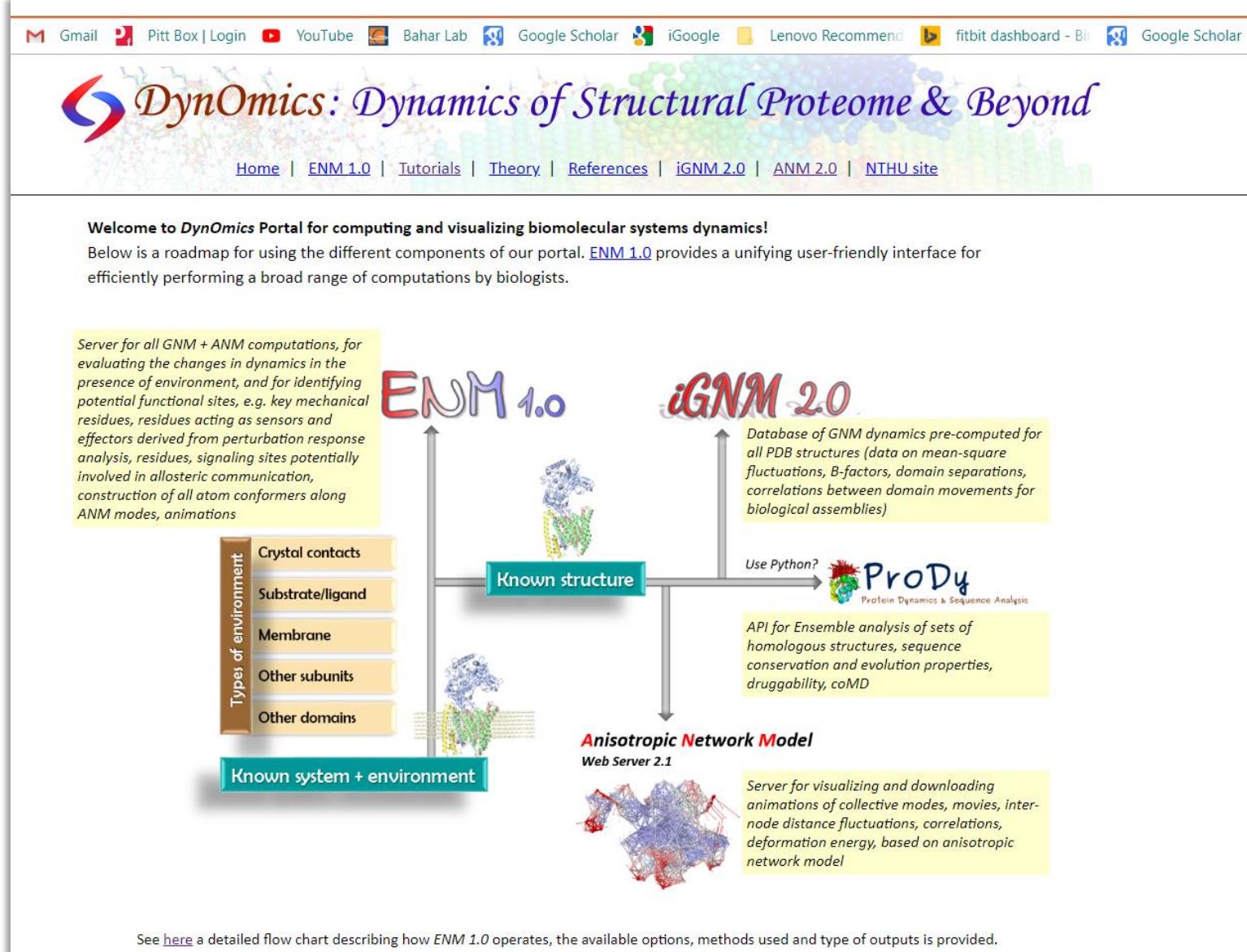


*E coli* adenylate kinase dynamics: comparison of elastic network model modes with  $^{15}\text{N}$ -NMR relaxation data Temiz NA, Meirovitch E, Bahar I. (2004) *Proteins* 57, 468.

T → R transition of Hb  
intrinsically favored by global  
dynamics Xu, Tobi & Bahar  
(2003) *J. Mol. Biol.* 333, 153;

# DynOmics Portal

<http://dynamics.pitt.edu/>



# ENM Server

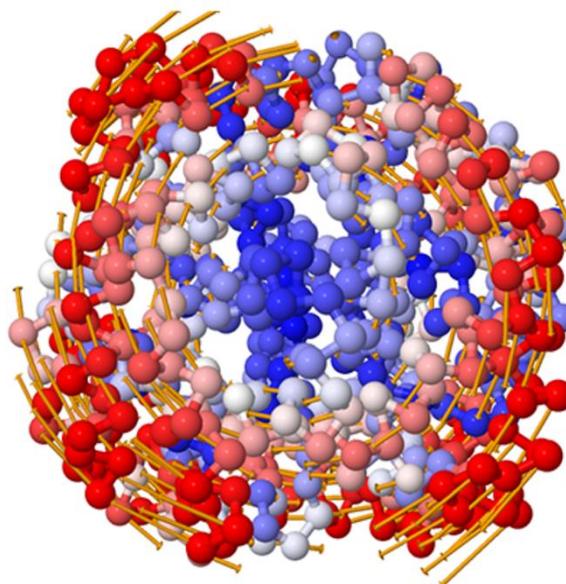
DynOmics Portal 1.0 - Dy X  
- → C ⓘ enm.pitt.edu



[Home](#) | [DynOmics 1.0](#) | [Tutorials](#) | [Theory](#) | [References](#) | [iGNM 2.0](#) | [ANM 2.0](#) | [NTHU site](#)

## What is the DynOmics ENM server?

The *DynOmics* 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](#).



PDB ID:  with biological assembly (unit):  No  Yes

or upload a local file:  No file chosen

Chain ID:  (e.g., A or AB, or leave blank for all chains)

▼ Advanced options:

▼ Considering Environment:

Email:  (optional, except for PDB files with > 2,000 residues)

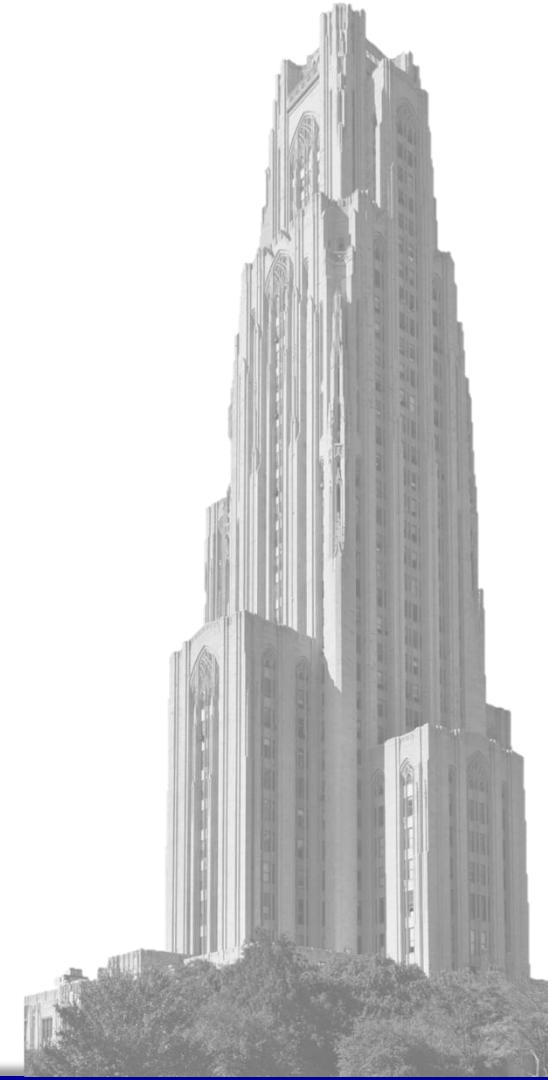
Load examples:

[Main result](#) [Molecular motion](#) [membrANM](#) [Hitting time](#) [Domain separation](#)

**enm.pitt.edu**



**Thank you!**



# Session I: Plotting $\langle(\Delta R_i)^2\rangle$ and contributions of selected modes

- from prody import \*
- from numpy import \*
- from matplotlib.pyplot import \*
- ion()
- anm, cot = calcANM('1cot', selstr='calpha')
- anm
- cot
- figure()
- showProtein(cot)
- figure()
- showSqFlucts(anm[:2], label= '2 modes')
- showSqFlucts(anm[:20], label= '20 modes')
- legend()

*Application to cytochrome c  
PDB: 1cot  
A protein of 121 residues*

cmd  
ipython

# 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

$\langle(\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j)\rangle$  between fluctuations

- `figure()`
- `showCrossCorr(anm[0])`
  
- `cross_corr = calcCrossCorr(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*)