June 10-14, 2013

Hands-on Workshop on Computational Biophysics

by

The Theoretical and Computational Biophysics Group (TCBG)

and

The National Center for Multiscale Modeling of Biological Systems (MMBioS)

Workshop Program

Mon, June 10: Introduction to Protein Structure and Dynamics - *Klaus Schulten*

Tue, June 11: Statistical Mechanics of Proteins; Force Field Parameterization- *Klaus Schulten and Emad Tajkhorshid*

Wed, June 12: Simulating Membrane Proteins - *Emad Tajkhorshid*

Thu, June 13: Collective Dynamics of Proteins Using Elastic Network Models - *Ivet Bahar, Tim Lezon and Ahmet Bakan*

Fri, June 14: Druggability Simulations, and Analyzing Sequence Patterns and Structural Dynamics - *Ivet Bahar and Ahmet Bakan*

Workshop Program

Thu, June 13: Collective Dynamics of Proteins Using Elastic Network Models - *Ivet Bahar, Tim Lezon and Ahmet Bakan*

Fri, June 14: Druggability Simulations, and Analyzing Sequence Patterns and Structural Dynamics - *Ivet Bahar and Ahmet Bakan*

Workshop Program

Thu, June 13: Collective Dynamics of Proteins Using Elastic Network Models -

Ivet Bahar, Tim Lezon and Ahmet Bakan

Fr

St

9:00 – 9:45 am: Elastic Network Model, Collective (Normal) Modes: Definitions, Assumptions (1) *Ivet Bahar*

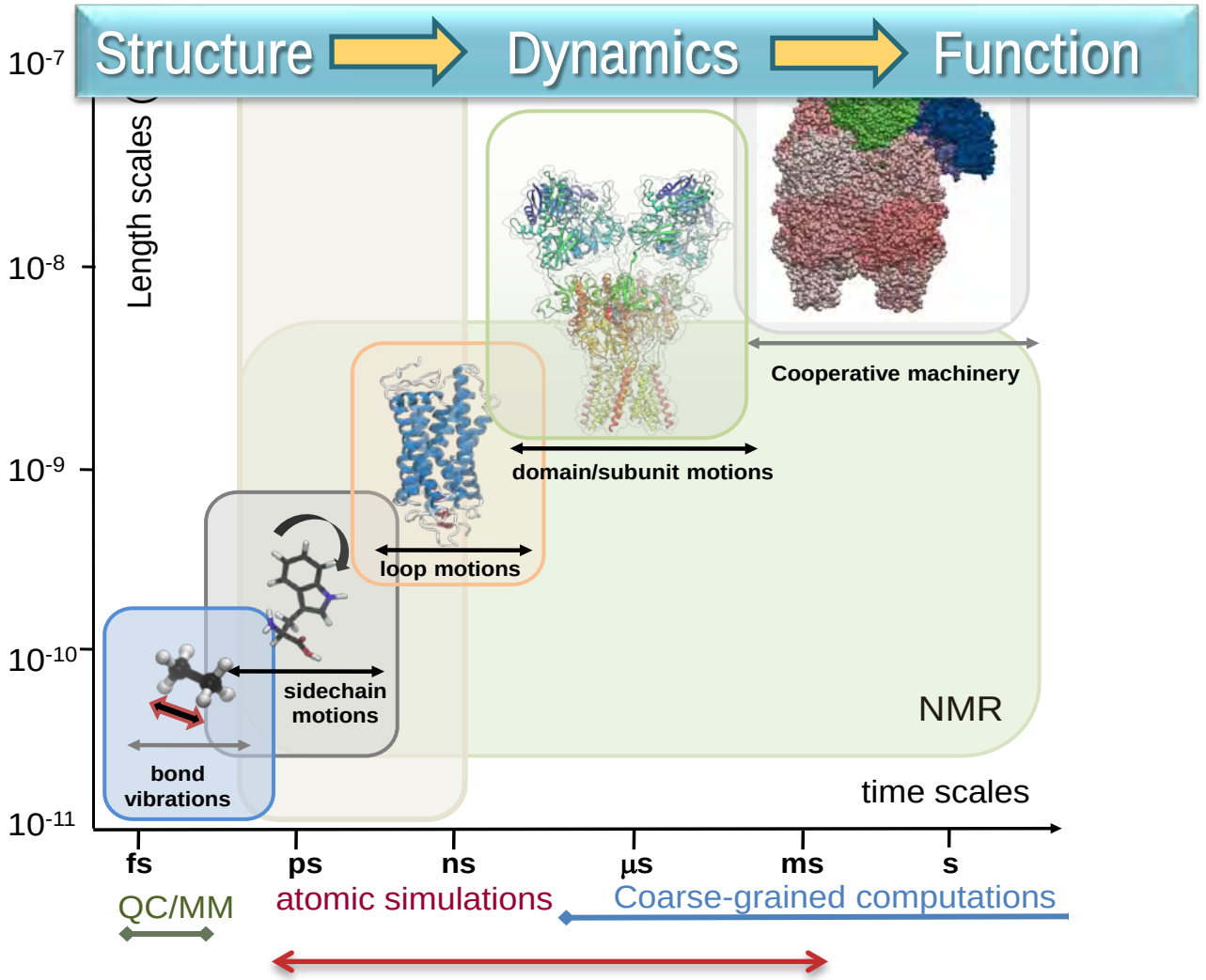
9:45 – 10:15 am: Applications and Comparison with Ensembles of Experimental Structures (2) *Ivet Bahar*

Coffee Break

10:30 –12:00pm ProDy Overview and Applications (3) *Tim Lezon, Ahmet Bakan*

1. **Bahar, Lezon, Yang & Eyal (2010)** *Annu Rev Biophys* **39**: 23-42;
2. **Bakan & Bahar** (2009) *Proc Natl Acad Sci* **106**, 14349-54; 3. **Bakan** et al. (2011) *Bioinformatics* **27**:1575-77.

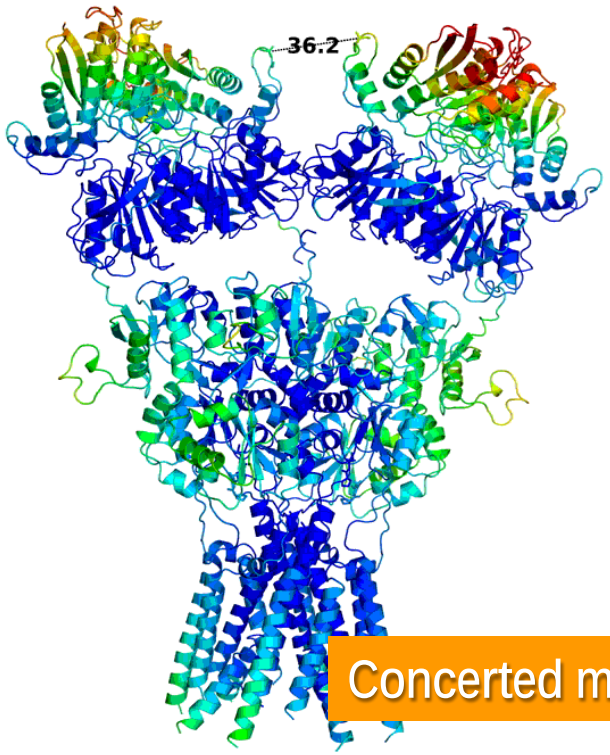
Each structure encodes a unique dynamics

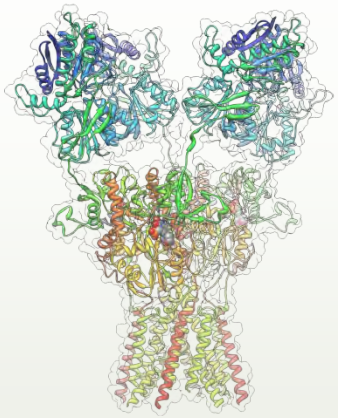


Each structure encodes a unique dynamics

Structure → Dynamics → Function

Signaling dynamics of AMPARs and NMDARs



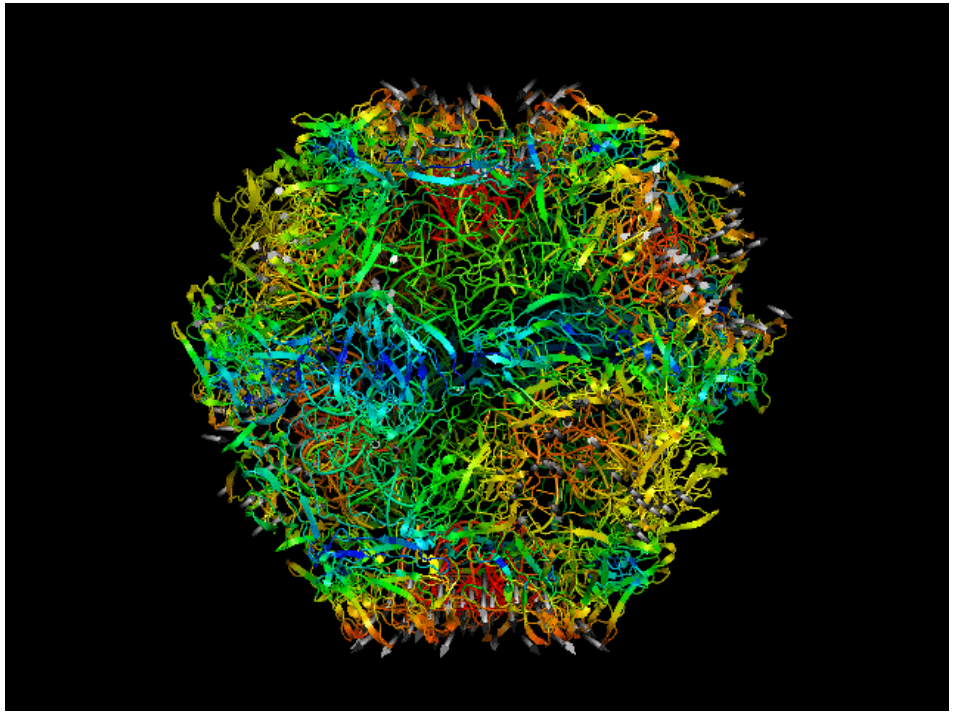


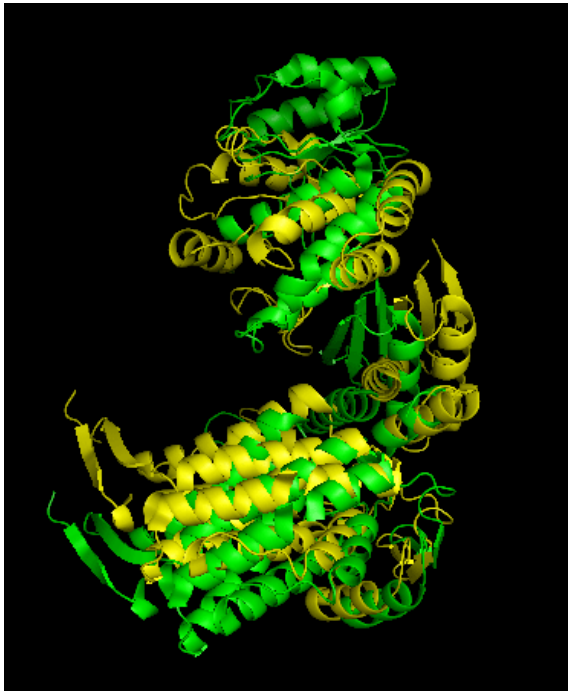
domain/subunit motions

Concerted movements of signaling molecules

Many proteins are molecular machines

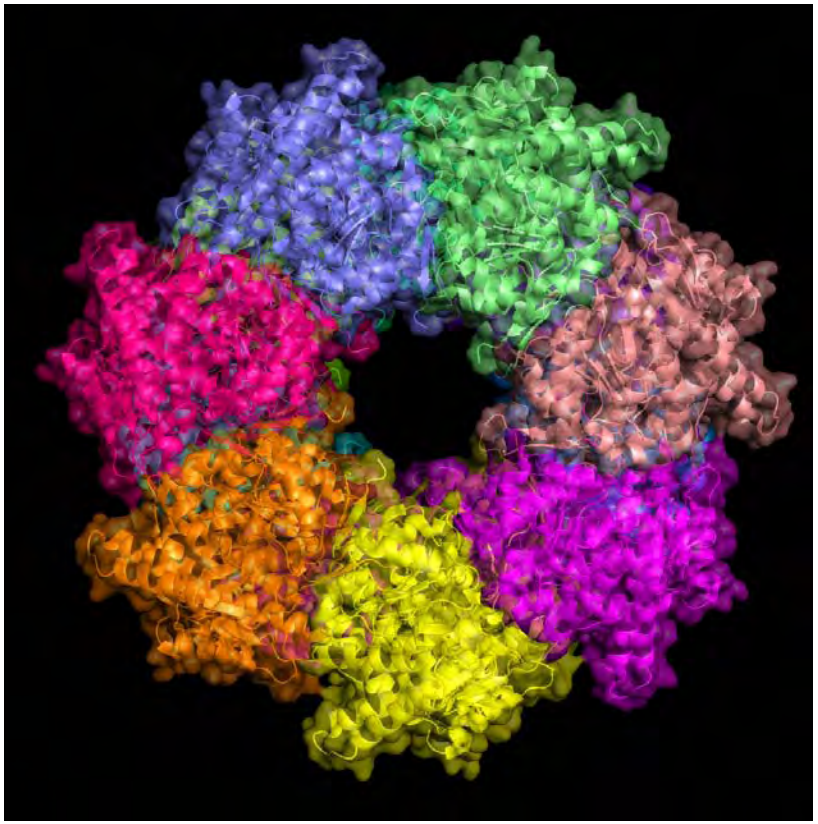
And mechanical properties become more important in complexes/assemblies





STMV dynamics (Zheng Yang)

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 of polymer physics, spectral graph methods, etc.

Physics-based approach

- Statistical Mechanics of Polymers
- Theory of Rubber Elasticity

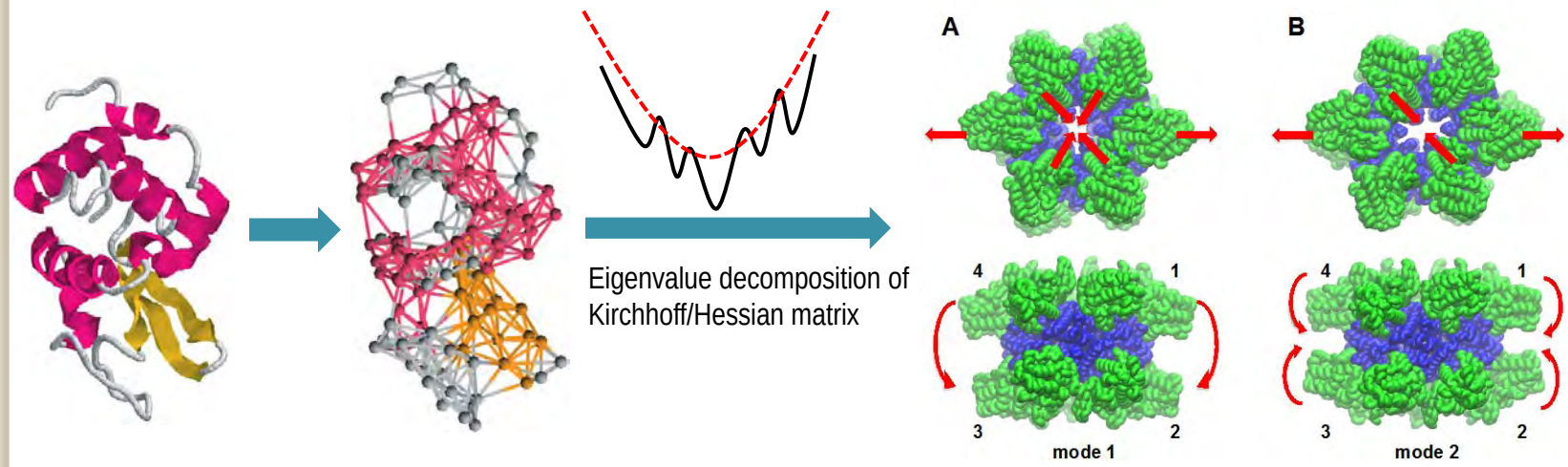
Paul J. Flory (1910-1985)
Nobel Prize in Chemistry 1974

Elastic Network Model for Proteins

And Pearson (1976), Eichinger (1980), Klockzkowski, Erman & Mark (1989)…

Collective motions

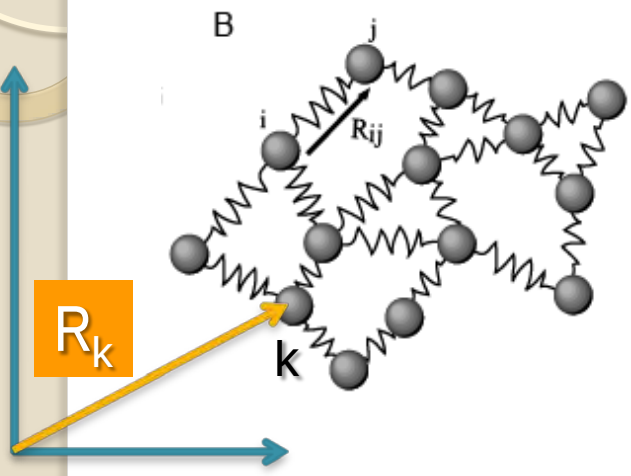
using elastic network models (ENM)

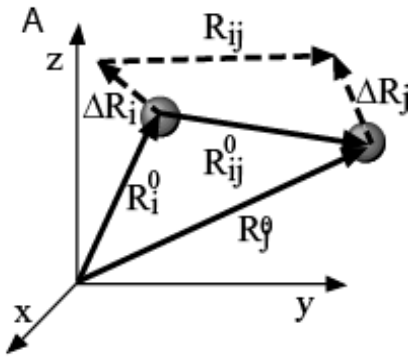


GNM: Bahar et al *Fold & Des* 1996; Haliloglu et al. *Phys Rev Lett*1997
ANM: Doruker et al. *Proteins* 2000; Atilgan et al, *Biophys J* 2001

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 their α-carbons' coordinates
- Springs connect residues located within a cutoff distance (e.g., 10 Å)

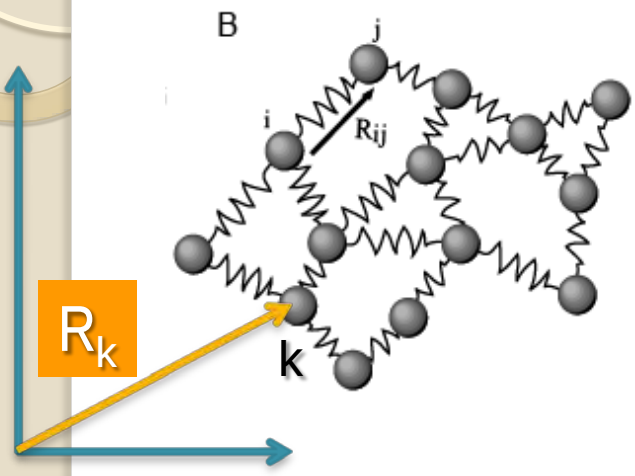
→ Nodes are subject to **Gaussian fluctuations** ΔR_i
→ Inter-residue distances R_{ij} also undergo Gaussian fluctuations

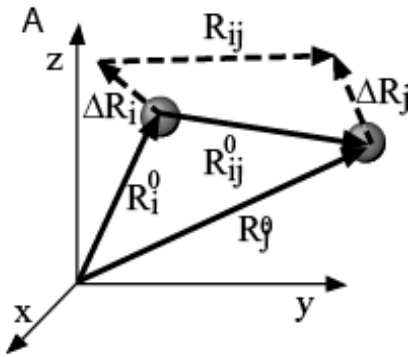
→ $\Delta \mathbf{R}_{ij} = \Delta \mathbf{R}_j - \Delta \mathbf{R}_i$

Fluctuations in residue positions

Bahar, Atilgan & Erman, *Fold & Des* 1997; Haliloglu et al. *Phys Rev Lett* 1997

Gaussian network model (GNM)





Fluctuation vector:

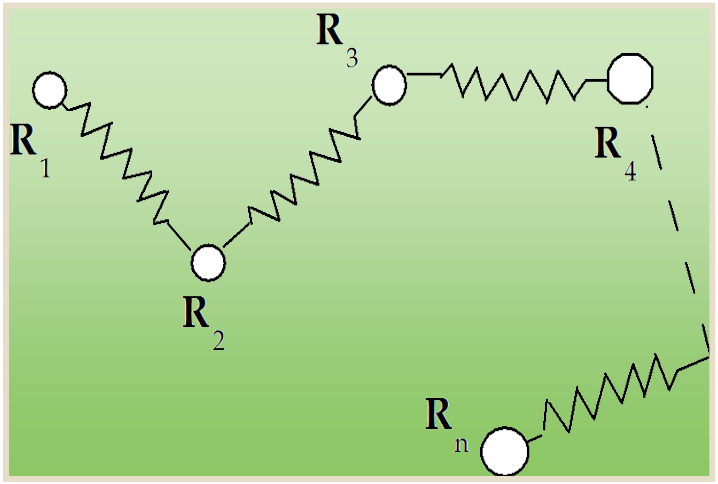
$$\rightarrow \Delta \mathbf{R} = \begin{bmatrix} \Delta \mathbf{R}_1 \\ \Delta \mathbf{R}_2 \\ \Delta \mathbf{R}_3 \\ \Delta \mathbf{R}_4 \\ .. \\ .. \\ .. \\ .. \\ \Delta \mathbf{R}_N \end{bmatrix}$$

Fluctuations in residue positions

Bahar, Atilgan & Erman, *Fold & Des* 1997; Haliloglu et al. *Phys Rev Lett* 1997

Rouse model for polymers

Classical bead-and-spring model



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

Kirchhoff matrix

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

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

$$= (\gamma/2)\ [\ (\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \ldots\ldots$$

Rouse model for polymers

Kirchhoff matrix

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

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

$$= (\gamma/2)\ [\ (\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \ldots\ldots$$

Rouse model for polymers

Fluctuation vector

Kirchhoff matrix

$$(\gamma/2)\,[\Delta R_1 \;\; \Delta R_2 \;\; \Delta R_3 \;\ldots\; \Delta R_N] \begin{bmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & -1 & & & \\ & & & \ddots & \ddots & & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix} \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \Delta R_3 \\ \vdots \\ \Delta R_N \end{bmatrix} =$$

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

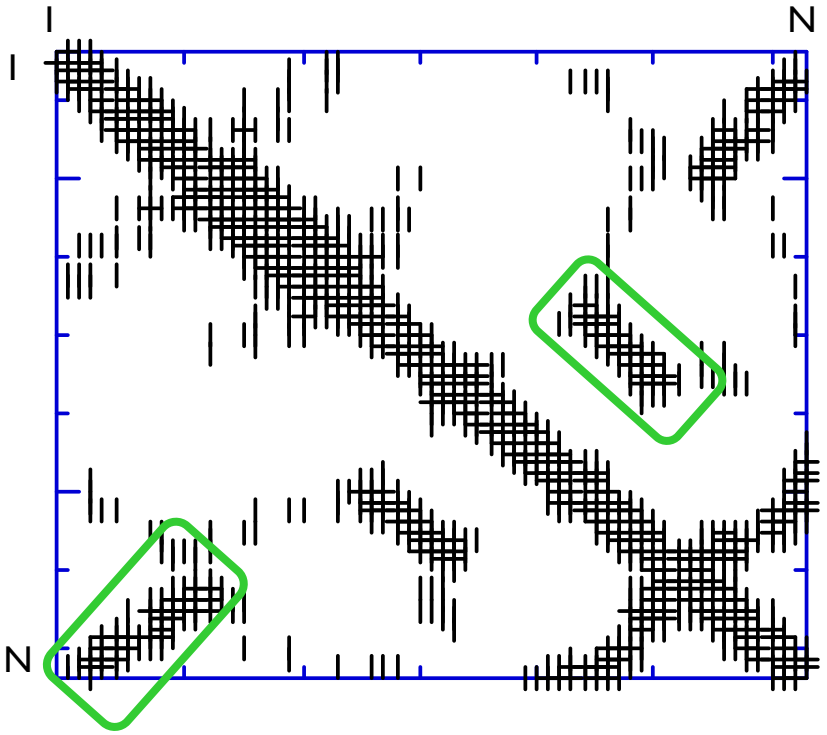
$$V_{tot} = (\gamma/2)\,[\,(\Delta R_{12})^2 + (\Delta R_{23})^2 + \ldots\ldots (\Delta R_{N-1,N})^2\,]$$

$$= (\gamma/2)\,[\,(\Delta R_2 - \Delta R_1)^2 + (\Delta R_3 - \Delta R_2)^2 + \ldots\ldots$$

Kirchhoff matrix for inter-residue contacts

For a protein of N residues

$\Gamma =$



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

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

Γ provides a complete description of contact topology!

Statistical mechanical averages

For a protein of N residues

$$< \Delta \boldsymbol{R}_i \bullet \Delta \boldsymbol{R}_j > \;\; = (1/Z_N) \int (\Delta \boldsymbol{R}_i \bullet \Delta \boldsymbol{R}_j)\, e^{-V/k_B T}\, d\{\Delta \boldsymbol{R}\}$$

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

Γ provides a complete description of contact topology!

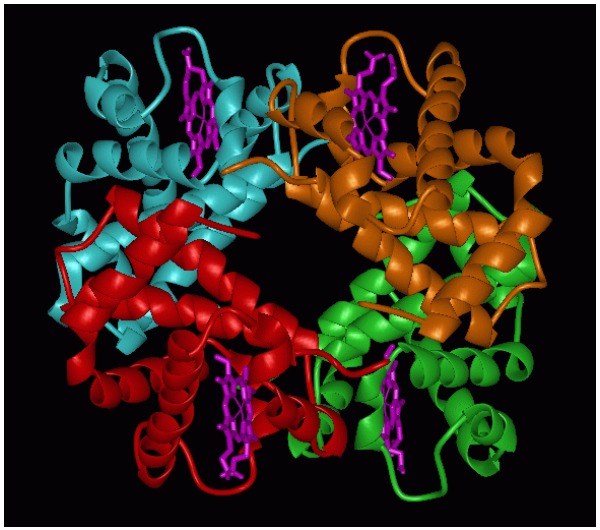
Kirchhoff matrix determines the mean-square fluctuations

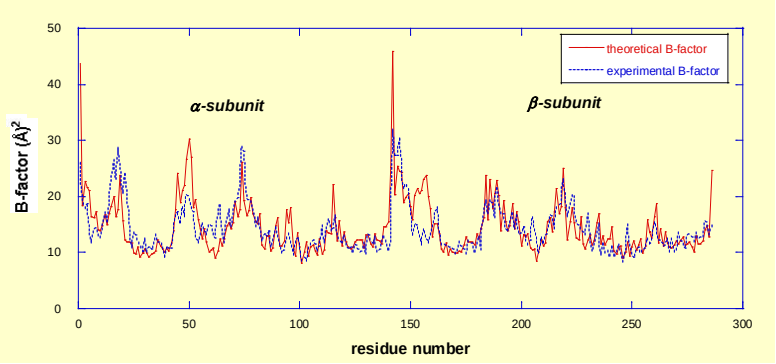
$$[\Gamma^{-1}]_{ii} \sim \langle(\Delta \mathbf{R}_i)^2\rangle$$

And cross-correlations between residue motions

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

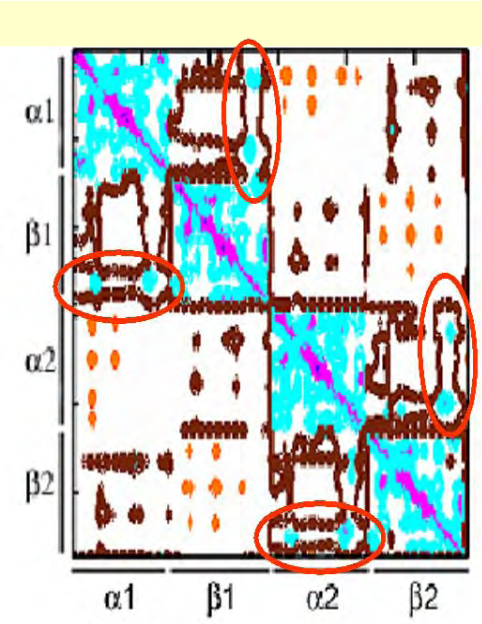
I. Application to hemoglobin





B- factors – Comparison with experiments

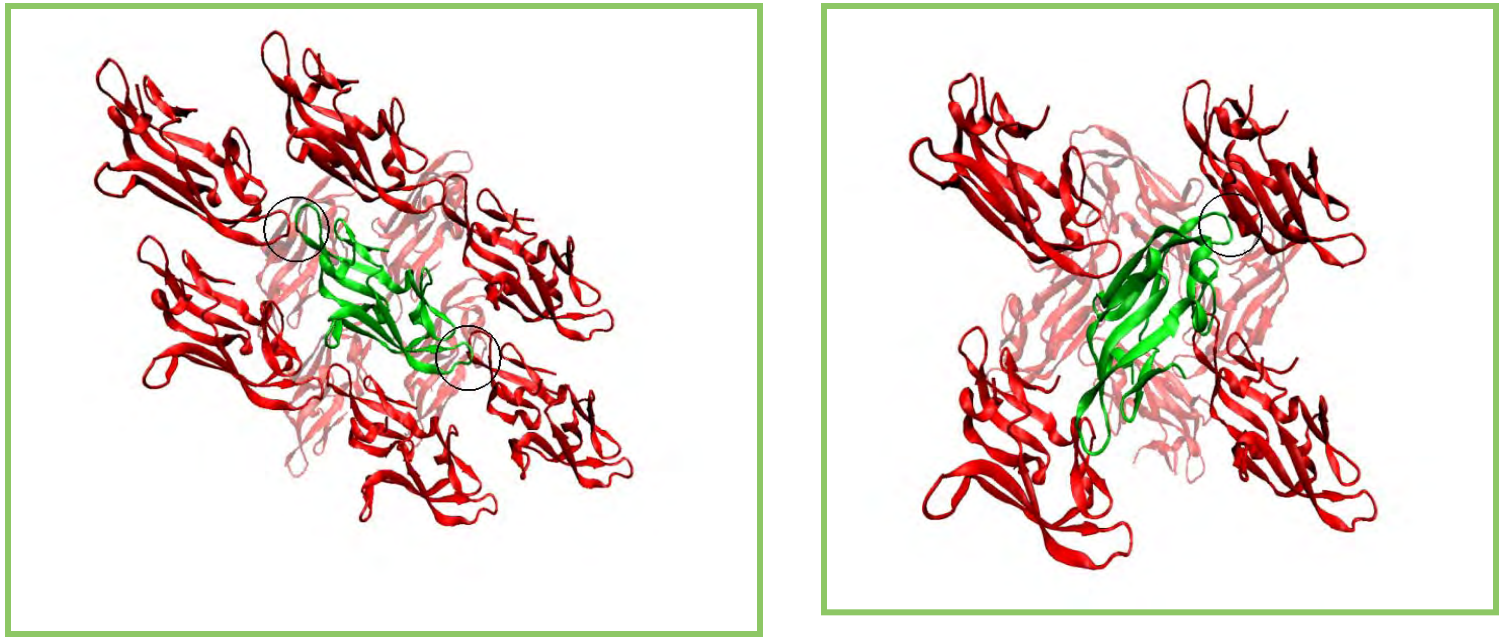
$$B_i = 8\pi^2/3 \langle(\Delta R_i)^2\rangle$$



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

C. Xu, D. Tobi and I. Bahar (2003) *J. Mol. Biol.* 2003, 153-168

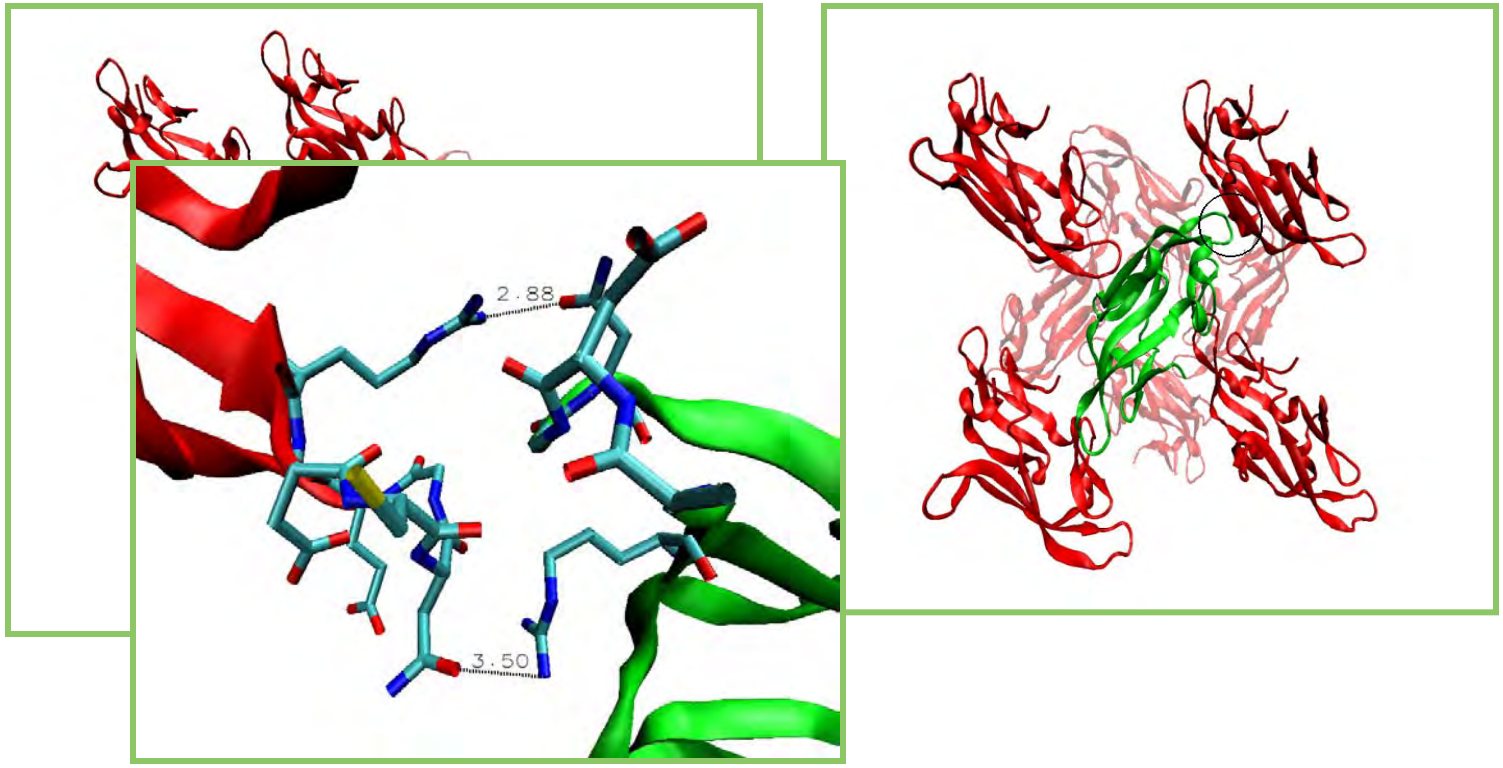
B-factors are affected by crystal contacts



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

1

B-factors are affected by crystal contacts



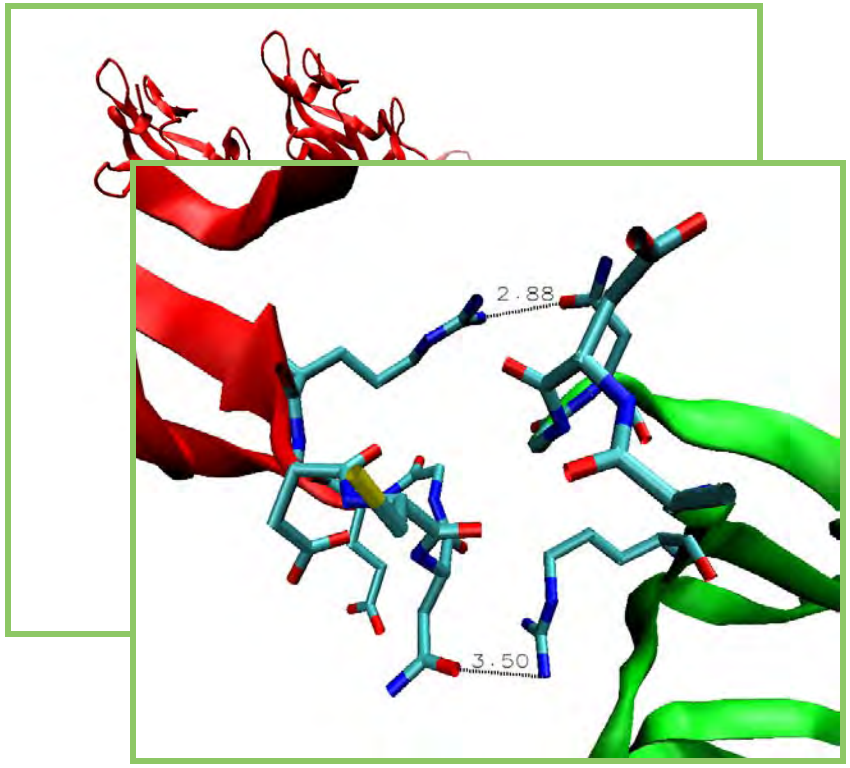
Particular loop motions are curtailed by intermolecular contacts in the crystal environment causing a discrepancy between theory and experiments

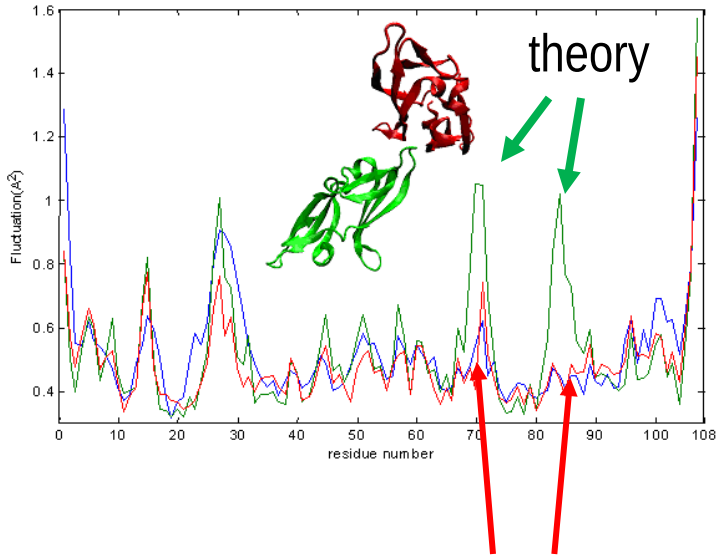
FOR MORE INFO...

Liu, Koharudin, Gronenborn & Bahar (2009) *Proteins 77, 927-939.*

1

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...

Liu, Koharudin, Gronenborn & Bahar (2009) *Proteins 77, 927-939.*

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

Several modes contribute to dynamics

$$< \Delta \mathbf{R}_i \bullet \Delta \mathbf{R}_j > = \sum_k [\Delta \mathbf{R}_i \bullet \Delta \mathbf{R}_j]_k$$

Contribution of mode k

$$< \Delta \mathbf{R}_i \bullet \Delta \mathbf{R}_j > = (3k_B T / \gamma) \left[\boldsymbol{\Gamma}^{-1} \right]_{ij}$$

Contribution of mode k

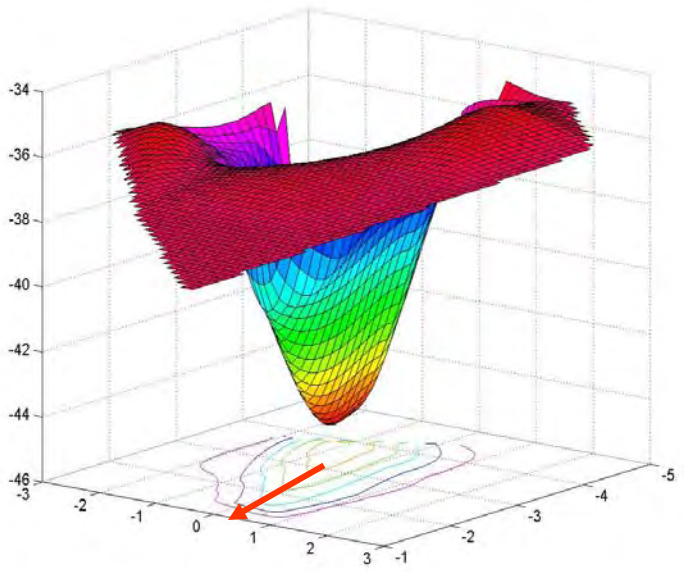
$$[\Delta \mathbf{R}_i \bullet \Delta \mathbf{R}_j]_k = (3k_B T / \gamma) \left[\lambda_k{}^{-1} \mathbf{u}_k \mathbf{u}_k{}^T \right]_{ij}$$

expressed in terms of the kth eigenvalue λ_k and eigenvector $\mathbf{u}_k$

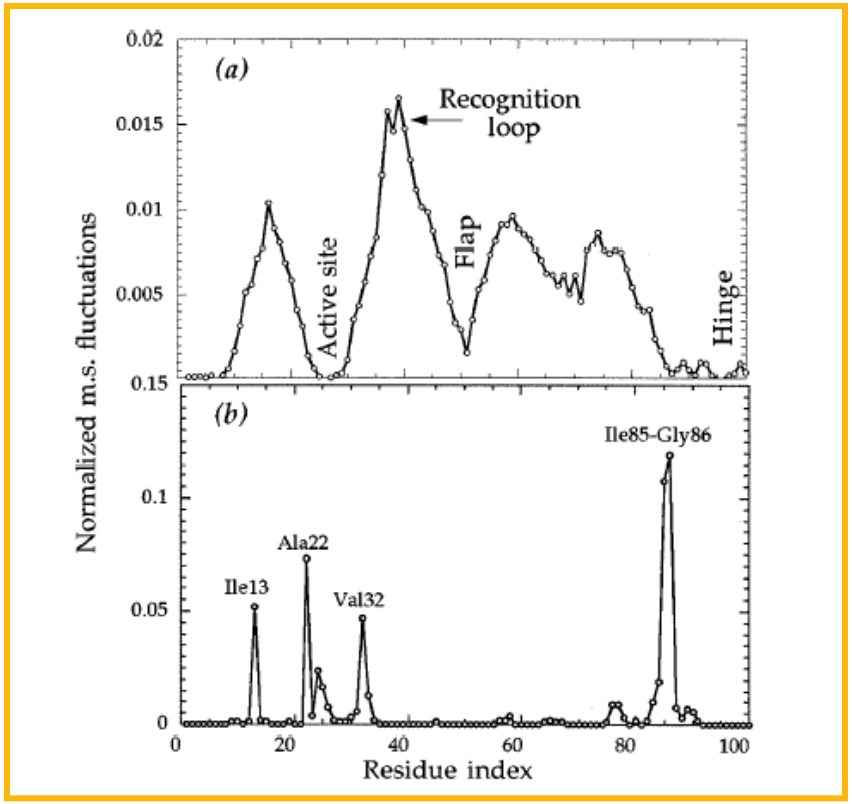
FOR MORE INFO...

Bahar et al. (1998) Phys Rev Lett. 80, 2733

Several modes contribute to dynamics



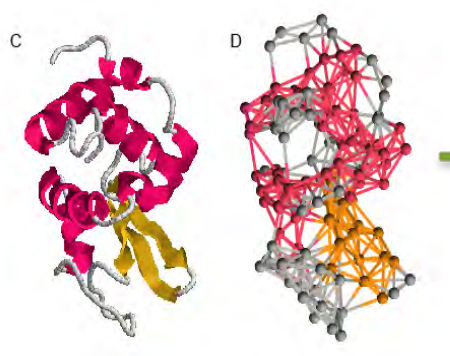
The first mode selects
the 'easiest' collective motion



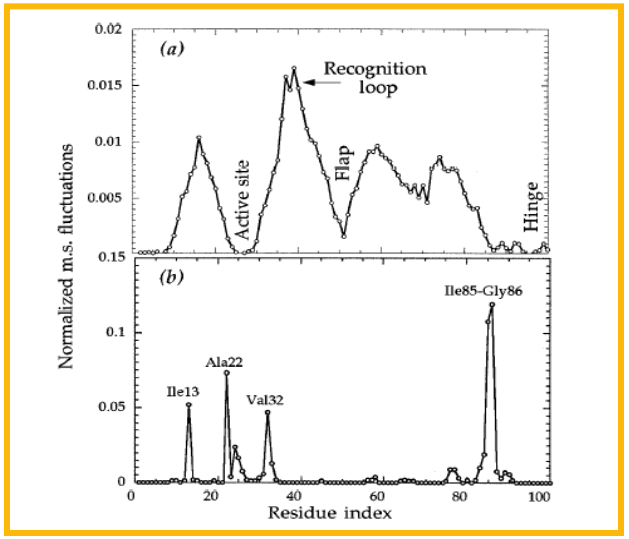
FOR MORE INFO...

Bahar et al. (1998) Phys Rev Lett. 80, 2733

Gaussian network model (GNM)

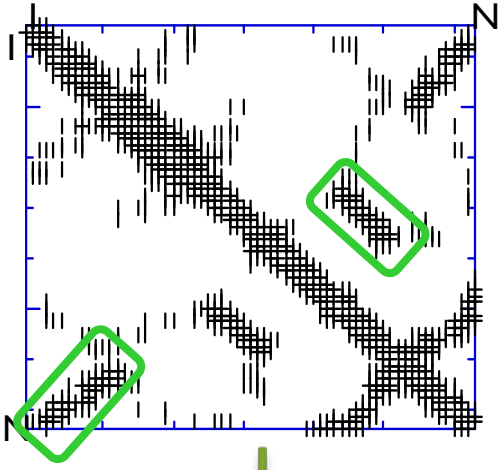


$$[\Delta \mathbf{R}_i \bullet \Delta \mathbf{R}_i]_k = (3k_B T/\gamma)\left[\lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^T\right]_{ii}$$



Kirchhoff matrix for inter-residue contacts

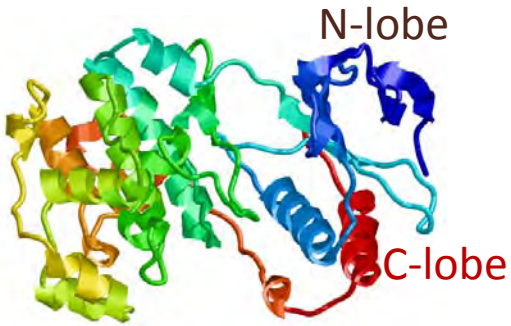
$\Gamma =$

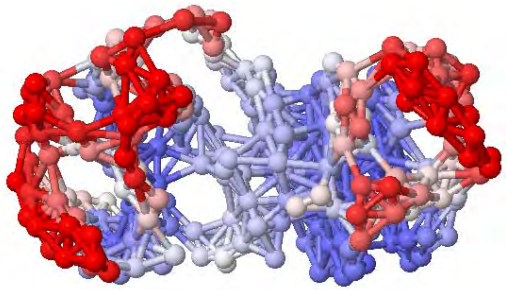


Several modes of motion contribute to dynamics

$$< \Delta \boldsymbol{R}_i \bullet \Delta \boldsymbol{R}_j >= (1/Z_N)\int (\Delta \boldsymbol{R}_i \bullet \Delta \boldsymbol{R}_j) e^{-V/k_B T} d\{\Delta \boldsymbol{R}\} = (3k_B T/\gamma)\left[\Gamma^{-1}\right]_{ij}$$

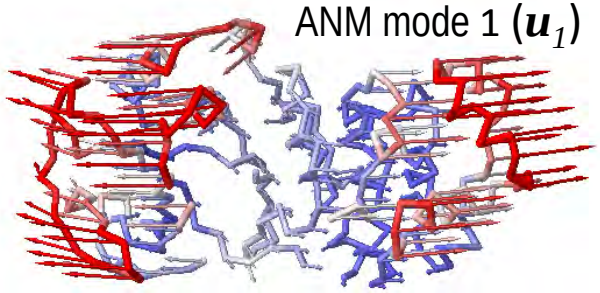
Anisotropic Network Model (ANM)

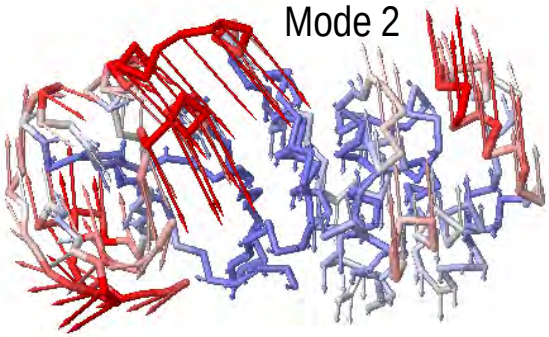




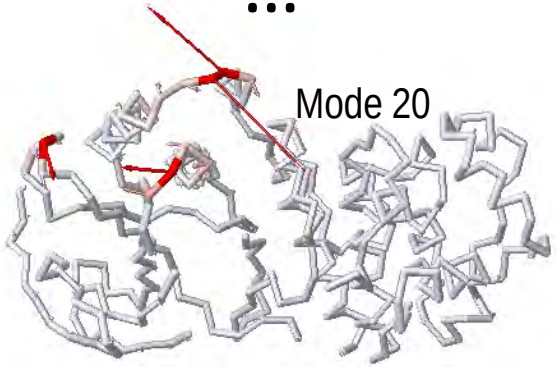
$$\mathbf{H} = \sum_{i=1}^{3N-6} \lambda_i \; \boldsymbol{u}_i \boldsymbol{u}_i^{\mathbf{T}}$$

$$\mathbf{H}^{(ij)} = \frac{\gamma \, \mathbf{\Gamma}_{ij}}{\left(R_{ij}^0\right)^2} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix}$$



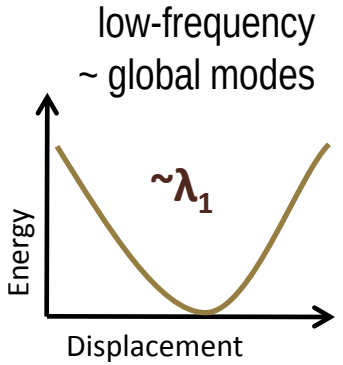


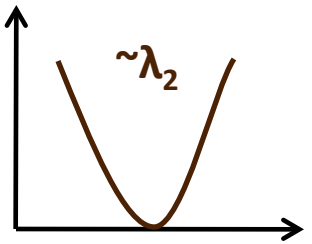
...



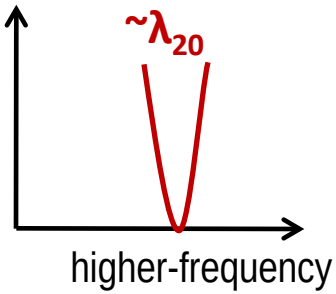
...

Total of 321x3 – 6 = **957 modes**





$\lambda_1 < \lambda_2 < \lambda_3 < \ldots$

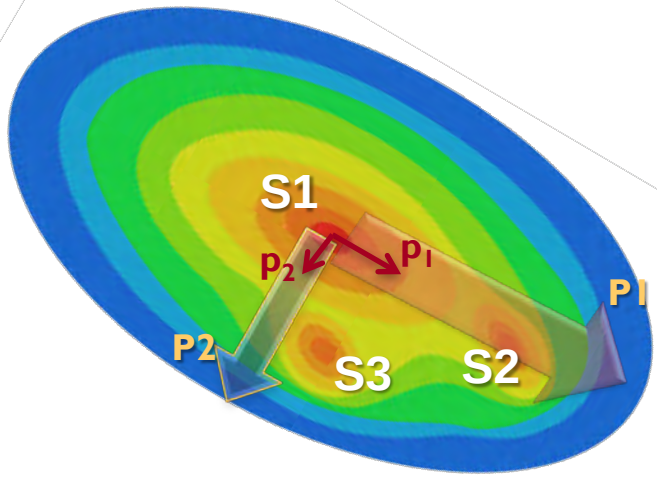


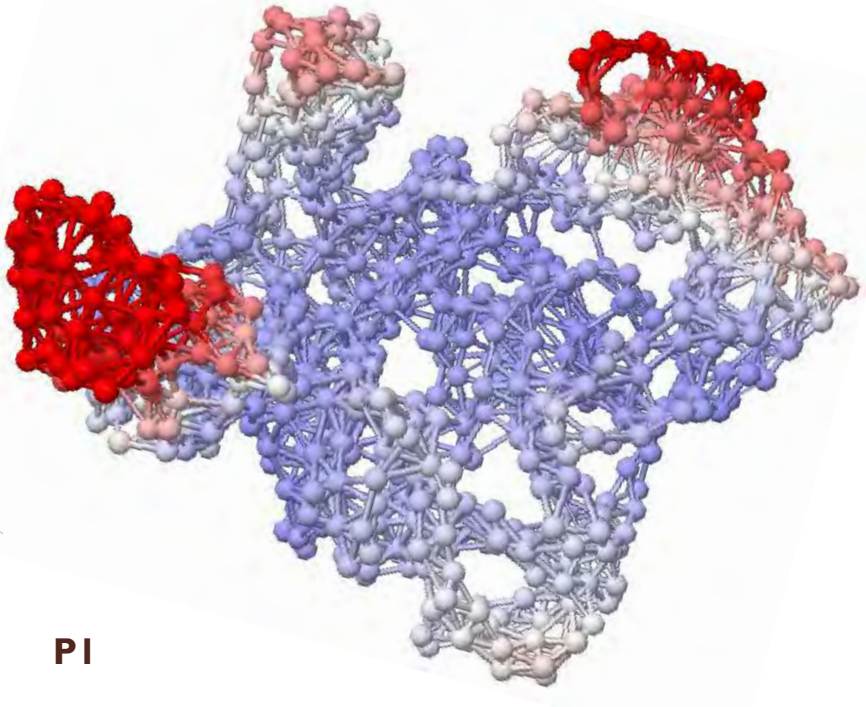
Doruker et al. (2000) *Proteins*; Atilgan AR et al. (2001) *Biophys J* 80; Eyal et al. (2006) *Bioinformatics* 22, 2619

Allosteric changes in conformation

ANM (anisotropic network model)

Elastic Network Models are particularly useful for exploring the allosteric dynamics of large multimeric structures





Comparison with experimental data shows that the functional movements are those predicted by the ENM to be intrinsically encoded by the structure

Atilgan et al (2001) *Biophys J*; Eyal et al (2006) *Bioinformatics* **22:** 2619-2627

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

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

figure()
showSqFlucts(anm)

figure()
 showSqFlucts(anm[0])
showSqFlucts(anm[:10])

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

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 $<(\Delta \mathbf{R}_i \,.\, \Delta \mathbf{R}_j)>$ between fluctuations

- cross_corr = calcCrossCorr?
- cross_corr = calcCrossCorr(anm[0])
- figure()
- showCrossCorr(anm[0])
- writeHeatmap('anm_cross1.hm', cross_corr)

Session 4:
Viewing cross-correlations using VMD

- *VMD – Load file*
- *Select cot_anm.nmd (from your local folder)*
- *Load HeatMap*
- *open anm_cross1.hm (from your local folder)*

Ensembles of structures

- Structural changes accompanying substrate (protein) binding

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

Ubiquitin
140 structures
1732 models

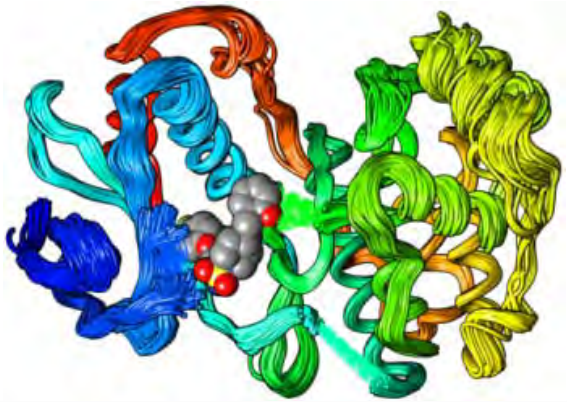
Ensembles of structures

- Structural changes accompanying substrate (protein) binding

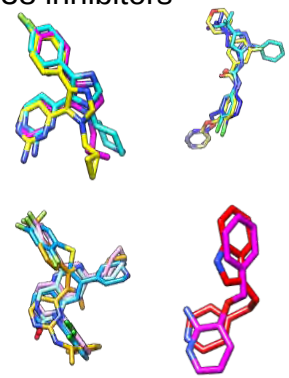
- Structural changes induced by, or stabilized upon, ligand binding

Ubiquitin
140 structures
1732 models

p38 MAP kinase
(**182 structures**)

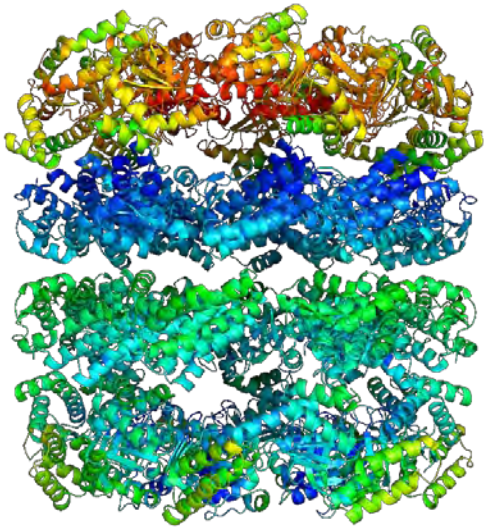


p38 inhibitors



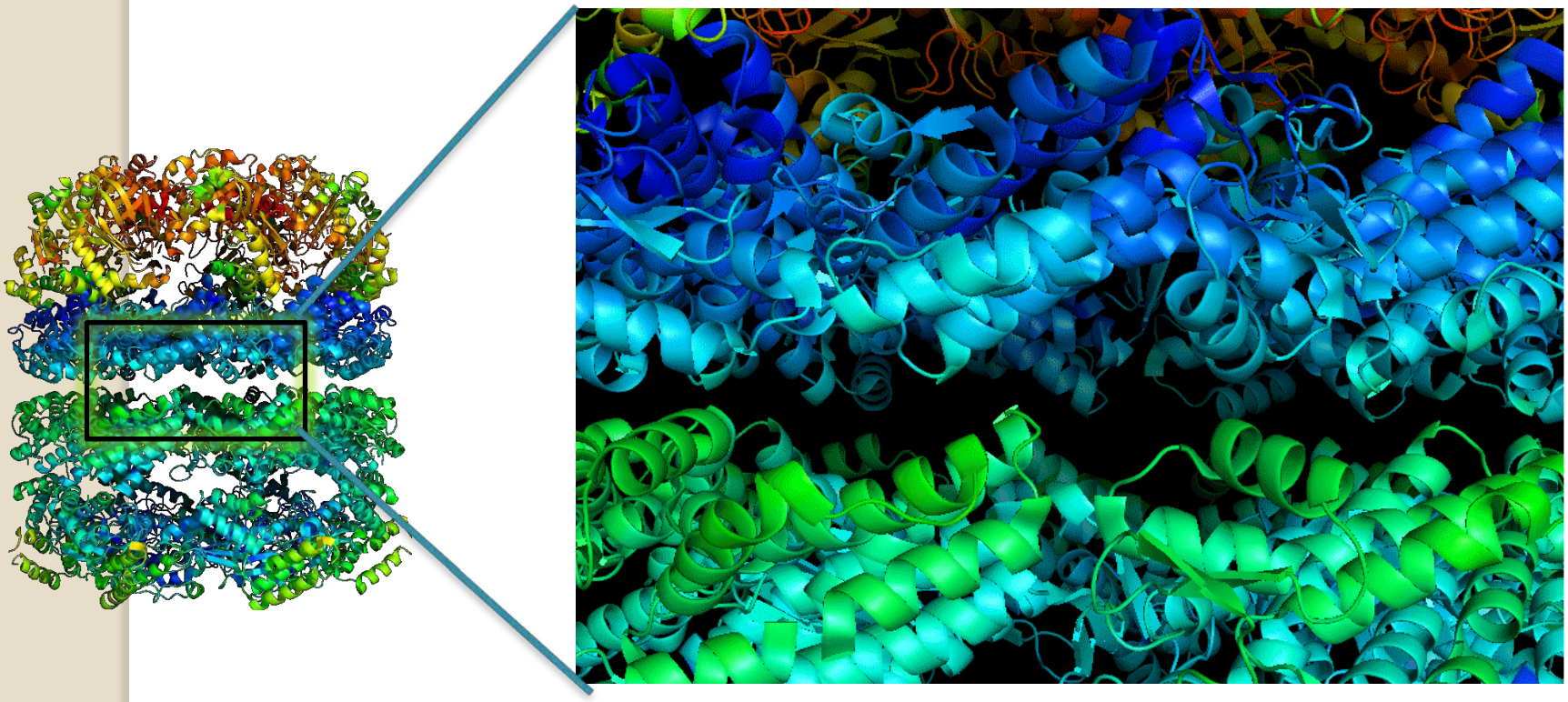
Ensembles of structures

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

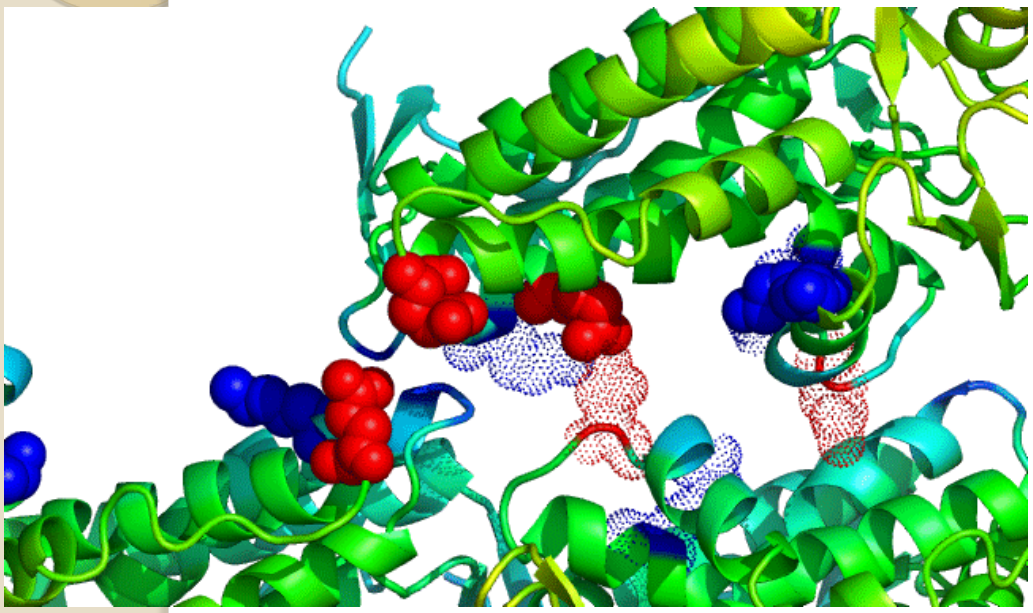


Yang et al. *PLoS Comp Biol* 2009

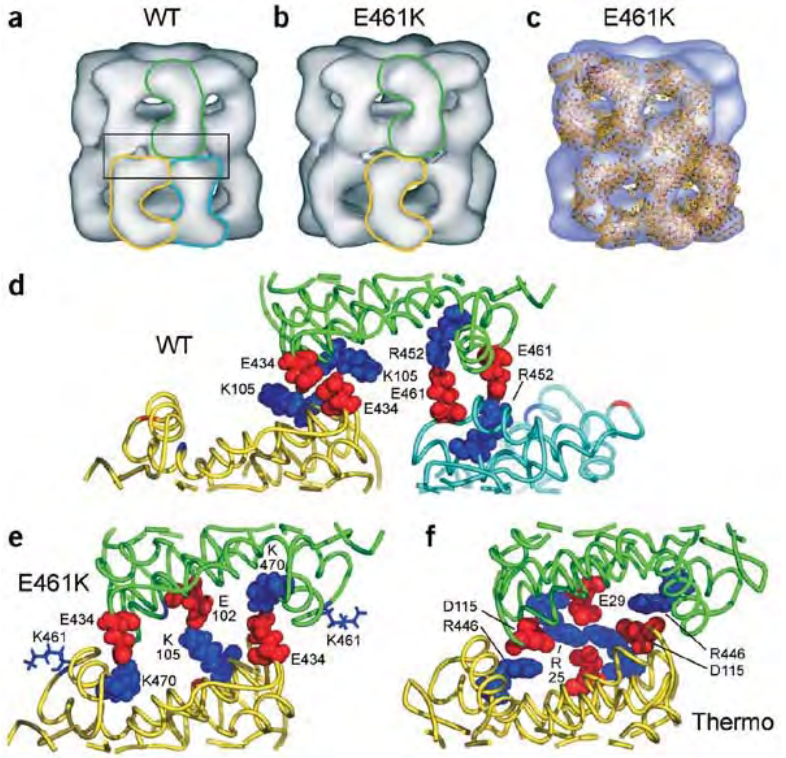
Redistribution of interactions at interfaces



Mutations may stabilizate conformers along soft modes – which may be dysfunctional



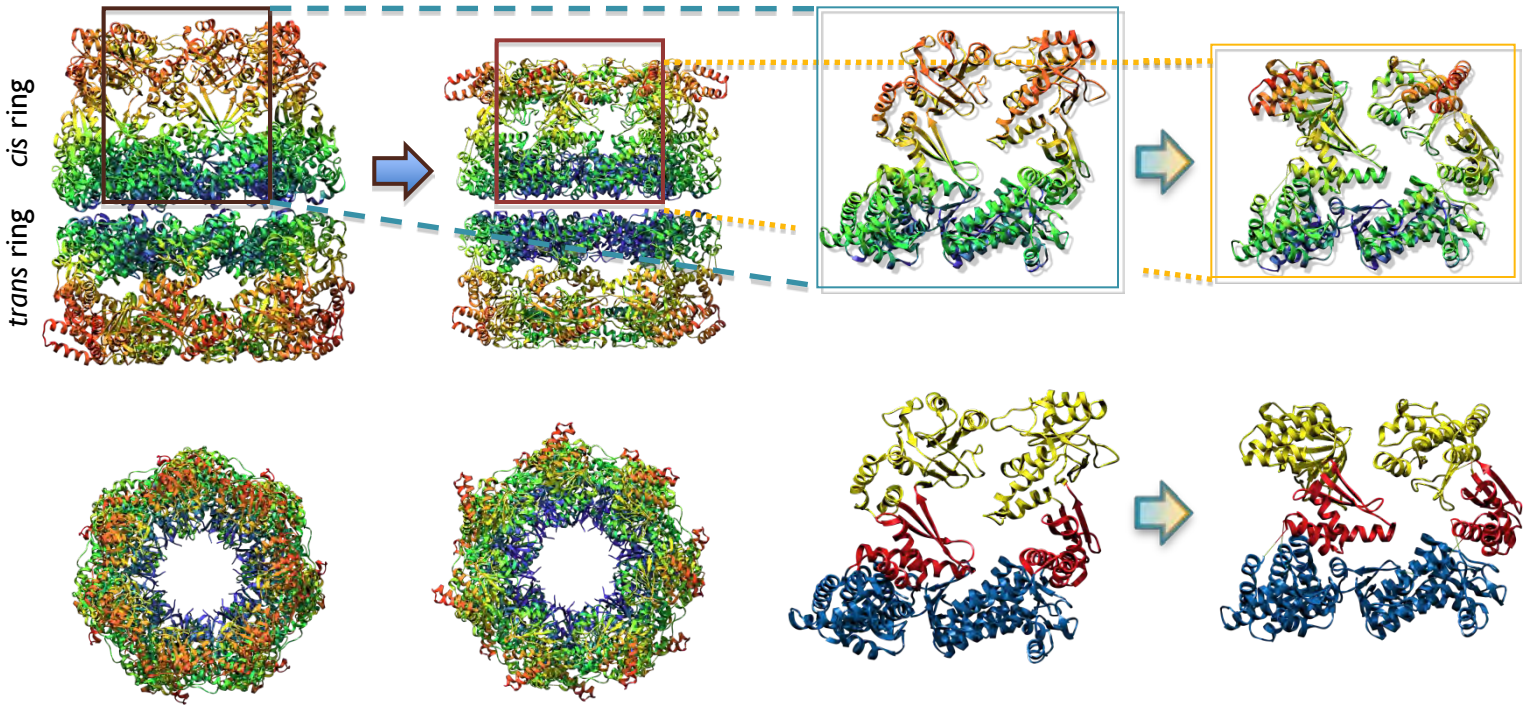
E461 mutant is a deformed structure along mode 1



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

Yang et al. *Mol Biosyst* 2008

Passage between the R and T state of GroEL

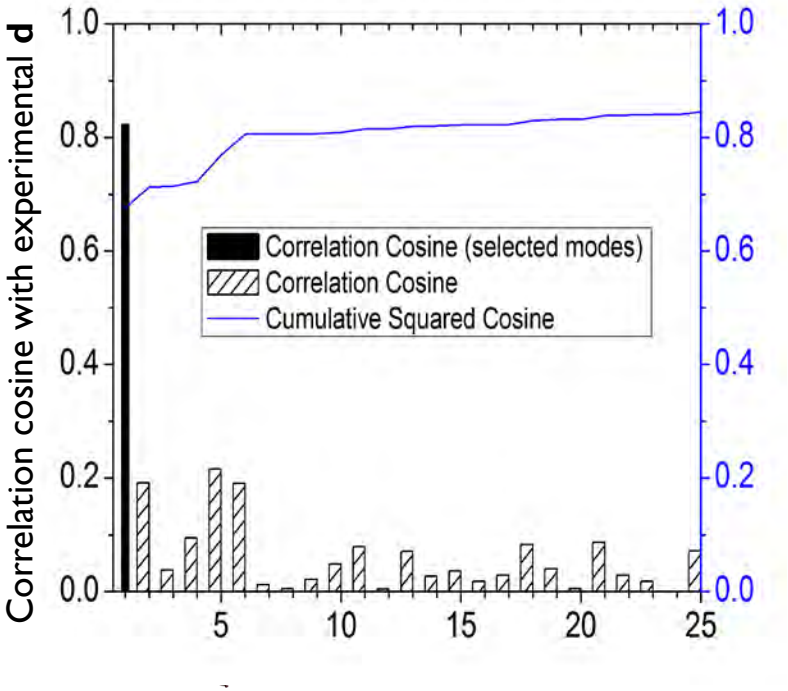


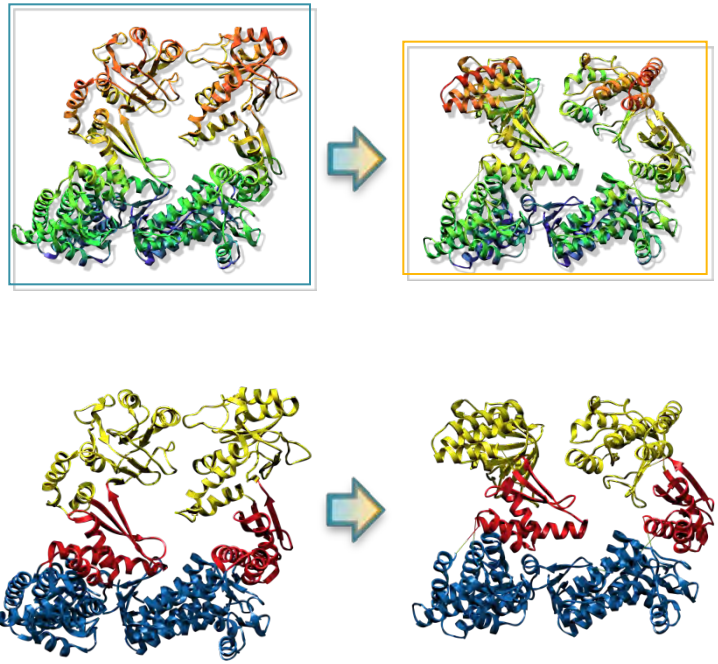
See...

Z Yang, P Marek and I Bahar, *PLoS Comp Biology* 2009

The softest mode enables the passage R → T

(with a correlation of 0.81)





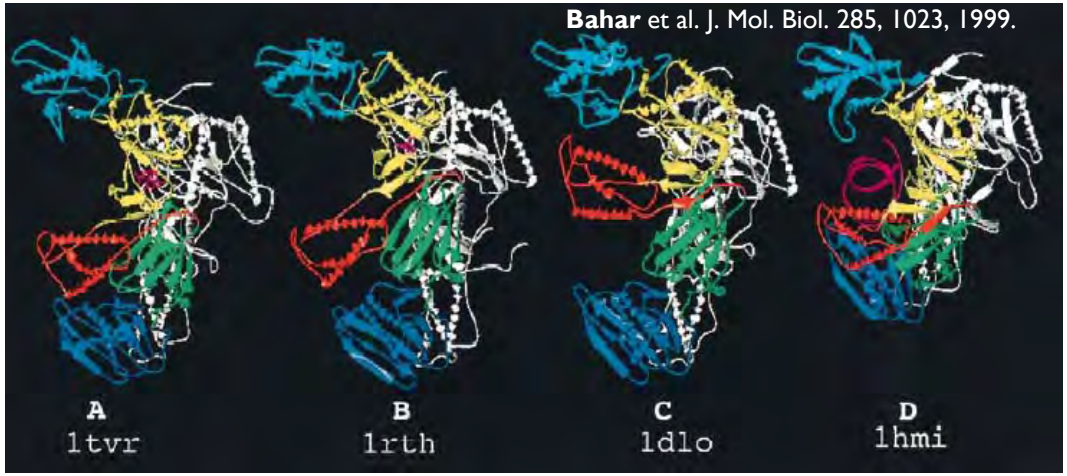
$$\mathbf{d} = [\Delta x_1 \;\; \Delta y_1 \;\; \Delta z_1 \;\; \ldots \;\; \Delta z_N]^T$$

See...

Z Yang, P Marek and I Bahar, *PLoS Comp Biology* 2009

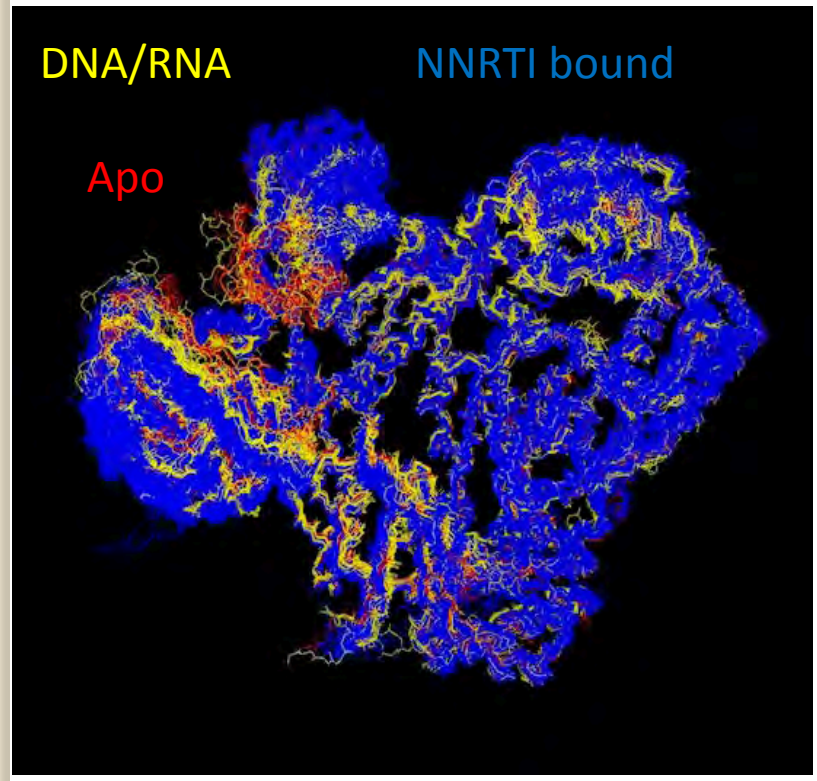
Dynamics inferred from known structures

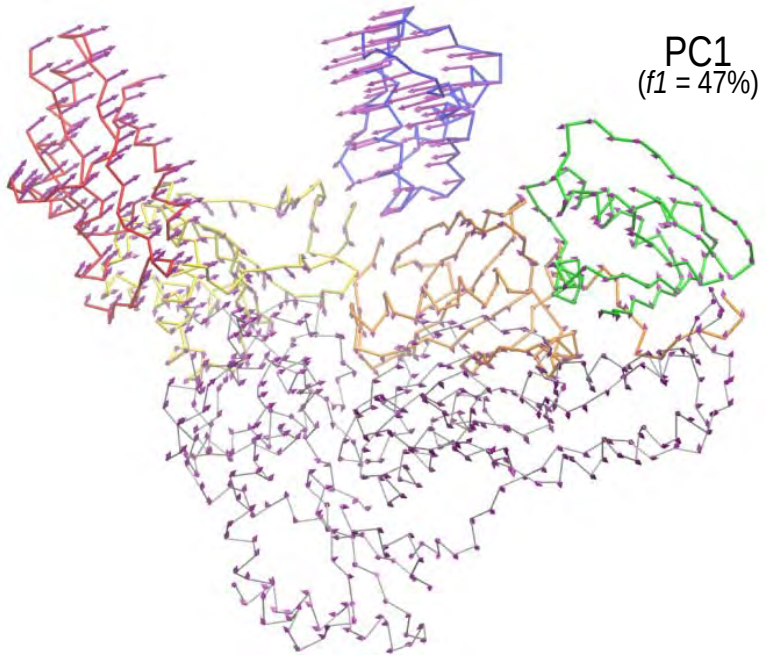
Comparison of static structures available in the PDB for the same protein in different form has been widely used is an **indirect** method of inferring dynamics.



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

Principal Component Analysis (PCA)

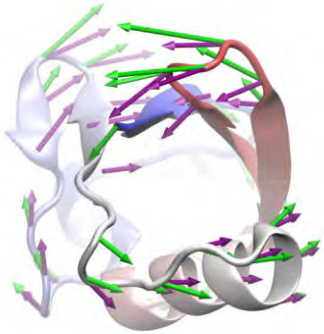


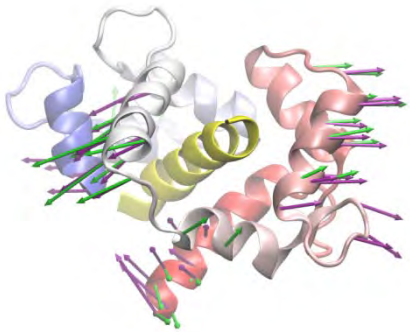


$$\mathbf{C}^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta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

$$\mathbf{C} = \mathbf{PSP}^T = \sum_{i=1}^{3N} \sigma_i \, p^i \, p^{iT}$$

Global motions inferred from theory and experiments

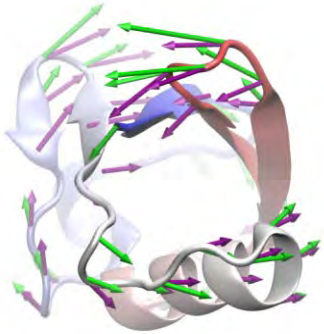


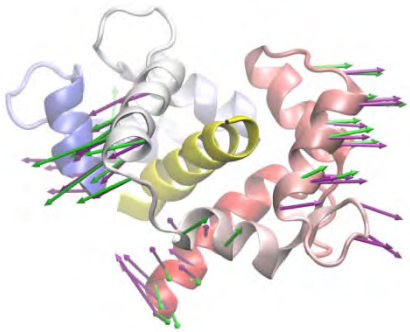


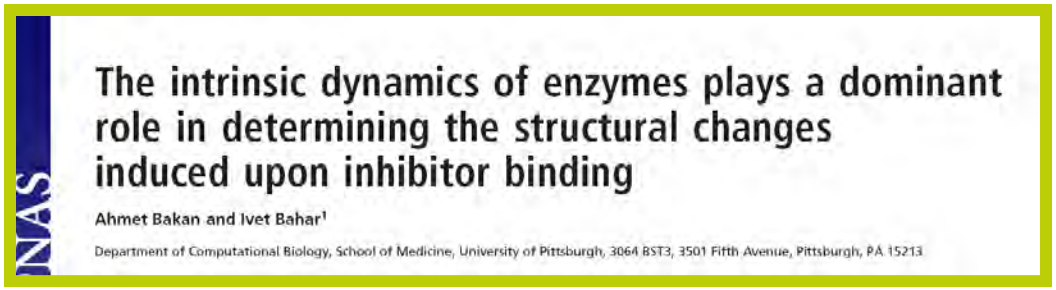
→ PCA of the ensemble of resolved structures

→ ANM analysis of a single structure from the ensemble

Global motions inferred from theory and experiments



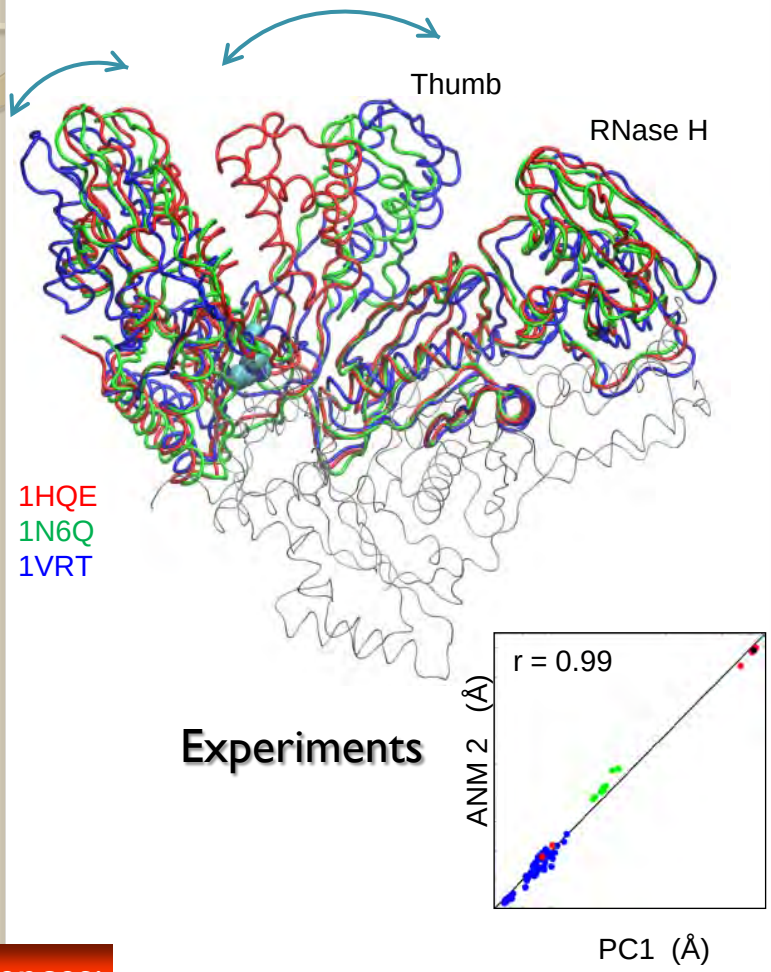




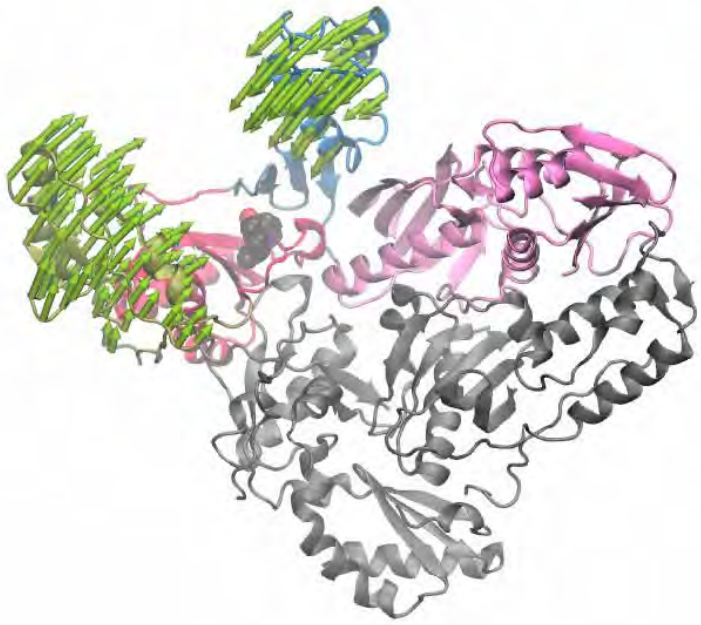
Reference:

*Bakan & Bahar (2009) PNAS **106**, 14349-54*

Induced Dynamics or Intrinsic Dynamics?



Experiments



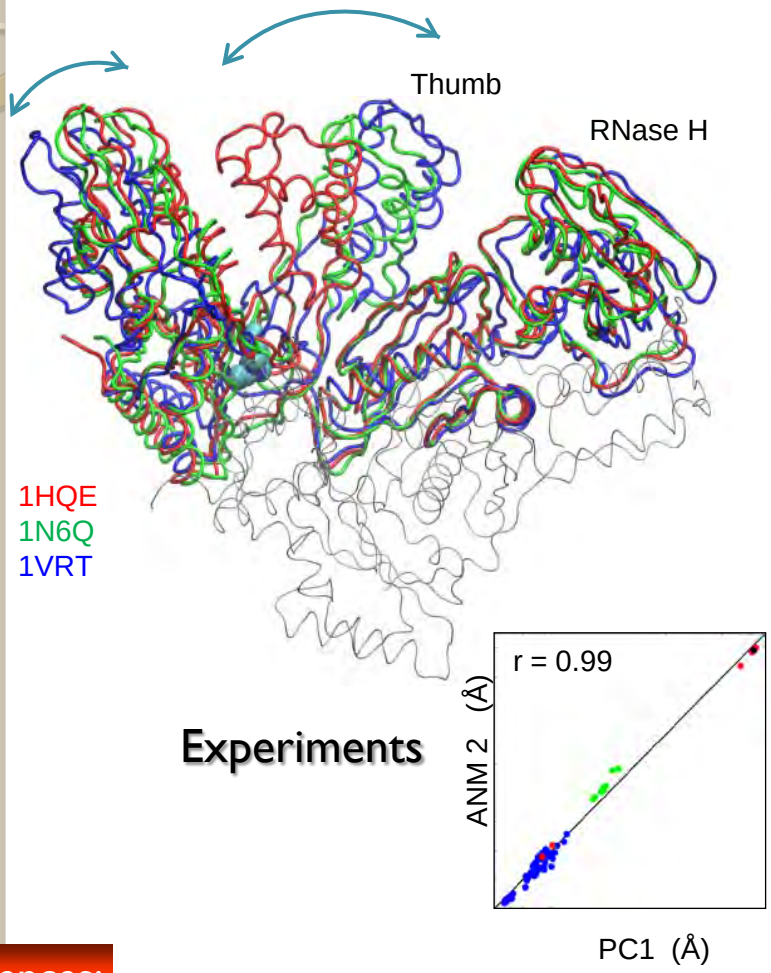
Theory

http://www.youtube.com/watch?v=1OUzdzm68YY

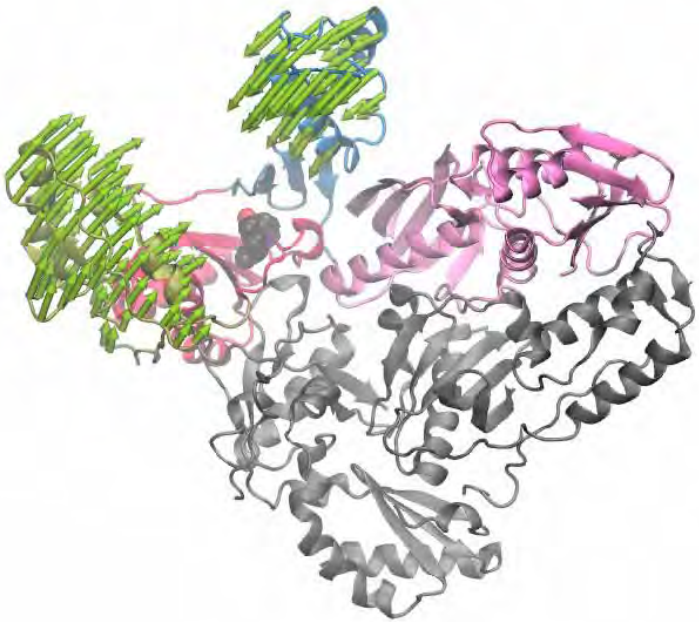
References:

Bakan & Bahar (2009) PNAS **106**, 14349-54.

Soft modes enable functional movements



Experiments



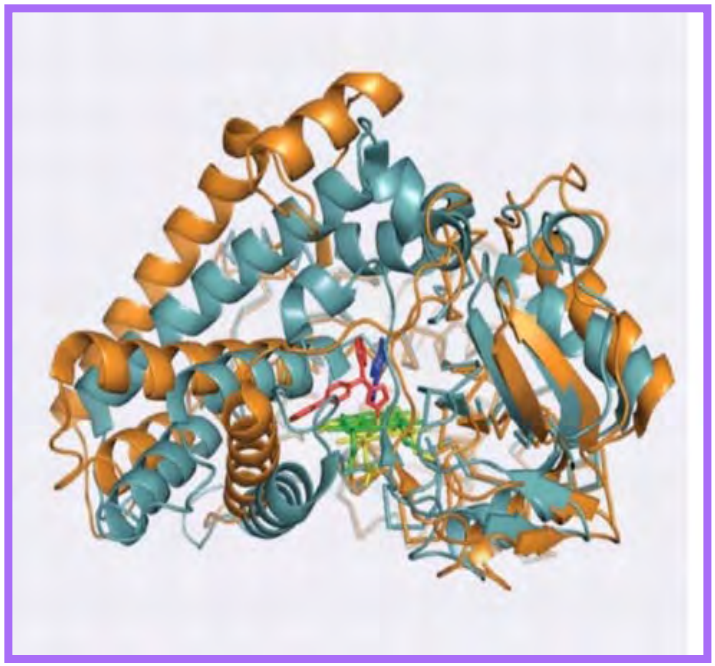
Theory

http://www.youtube.com/watch?v=1OUzdzm68YY

References:

Bakan & Bahar (2009) PNAS **106**, 14349-54.

Intrinsically accessible motions enable
Optimal binding of substrate or drugs



Conformational flexibility + sequence variability mediates substrate selectivity

- Two conformations of P450-CYP2B4:

  open (orange) with a large substrate (bifonazole, red), and

  closed (light blue) with the smaller substrate 4-(4-chlorophenyl) imidazole (blue)

See...

N. Tokuriki and D. S. **Tawfik** (2009) *Science* **324**: 203-207

ProDy for exploring conformational space

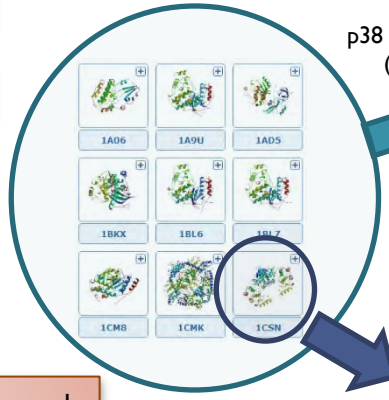
Protein Dynamics Analysis in Python

Usage example

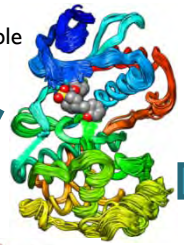
User inputs a protein sequence

>1A9U:A|PDBID|CHAIN
GSSHHHHHHSSGLVPRGSHMSQER
PTFYRQELNKTIWEVPERYQNLSPV
GSGAYGSVCAAFDTKTGLRVAVKK
LSRPFQSIIHAKRTYRELRLLKHMKH
ENVIGLLDVFT.....

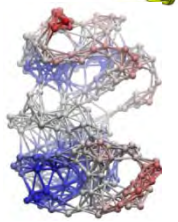
ProDy *identifies*, *retrieves*, *aligns*, and *analyzes (PCA)* structures that match the input sequence



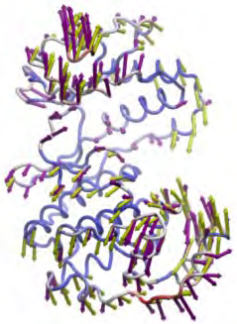
p38 ensemble (PCA)



p38 network model (ANM)

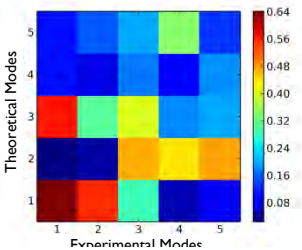


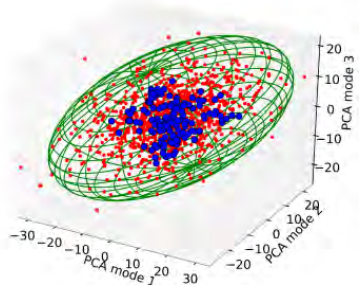
Experiment/Theory



User can compare experimental and theoretical models

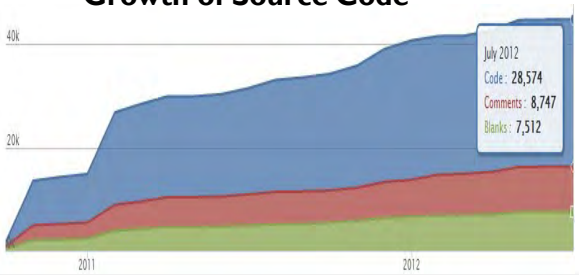
Overlap table

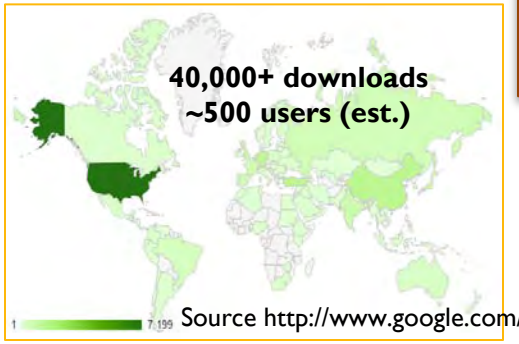




User can sample an ensemble of conformations along ANM modes for docking simulations

Growth of Source Code





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

A. Bakan, L.M. Meireles, I. Bahar, *Bioinformatics* 2011.

Acknowledgment

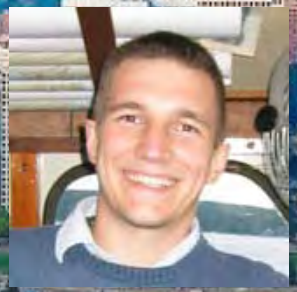
Support from NIGMS, NLM, NIDDK & NIAID

Group members
- Elia Zomot
- Anindita Dutta
- Ahmet Bakan
- Ignacio General
- **Murat Can Cobanoglu**
- Tim Lezon
- Mary Cheng
- Filippo Pulara
- **Indira Shrivastava**
- Mert Gur
- Kaitlyn Hu

Former members
- Lee-Wei Yang
- **Eran Eyal**
- **Dror Tobi**
- Basak Isin
- AJ Rader
- Chakra Chennubhotla
- Enrique Marcus
- Zheng Yang
- Enrique Markus
- **Ying Liu**
- **Lin Liu**
- Lidio Meireles
- Divesh Bhatt

Collaborators
- Angela Gronenborn
- Lila Gierasch
- Benoit Roux
- Michael Tsang
- John Lazo
- Andreas Vogt
- Mike Widom
- Andrej Sali
- Klaus Schulten
- Susan Amara
- Pemra Doruker

Dr. Ahmet Bakan
Comp & Systems Biology
U of Pittsburgh



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

Anindita Dutta
CMU/Pitt PhD Program

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

Burak Erman
Koc University, Istanbul

Ali Rana Atilgan
Sabanci University, Istanbul

Turkan Haliloglu
Bogazici University, Istanbul