Hands-on Workshop on Computational Biophysics

by

The Theoretical and Computational Biophysics Group (TCBG)

and

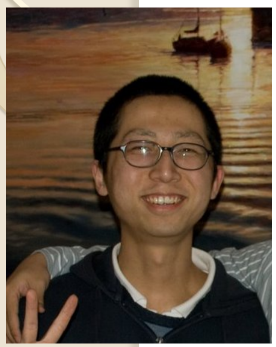
The National Center for Multiscale Modeling of Biological Systems (MMBioS)

ProDy

Protein Dynamics Analysis in Python

Cihan Kaya

Dr. Indira Shrivastava
Assist Prof, DCSB, Pitt



Dr. Ying Liu

Dr. Timothy R Lezon
Assistant Prof, DCSB, Pitt

Drs. Ahmet Bakan and Anindita Dutta

Dr. Chakra Chennubhotla
Assist Prof, DCSB, Pitt

Reference:

Bakan A, Meireles LM, **Bahar I**. (2011) ProDy: Protein dynamics inferred from theory and experiments *Bioinformatics* **27**:1575-7

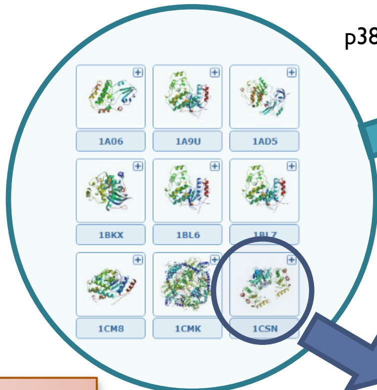
Bakan, A., Dutta, A., Whenzi, M., Liu, Y., Chennubhotla, C., Lezon, T.R., & Bahar, I. (2014) *Bioinformatics* in press.

ProDy for exploring conformational space

Protein Dynamics Analysis in Python

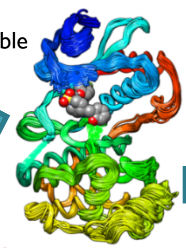
User inputs a protein sequence

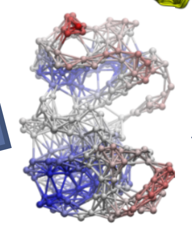
>1A9U:A|PDBID|CHAIN
GSSHHHHHHSSGLVPRGSHMSQERP
TFYRQELNKTIWEVPERYQNLSPVG
SGAYGSVCAAFDTKTGLRVAVKKLS
RPFQSIIHAKRTYRELRLLKHMKHEN
VIGLLDVFT......



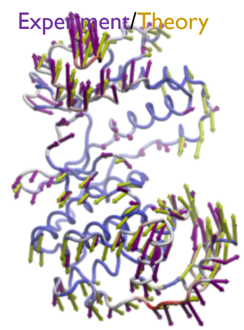
ProDy *identifies*, *retrieves*, *aligns*, and *analyzes (PCA)* structures that match the input sequence

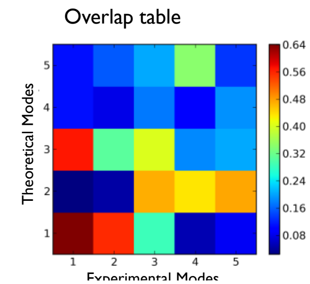
p38 ensemble (PCA)



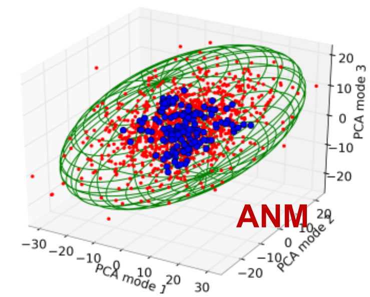


p38 network model (ANM)



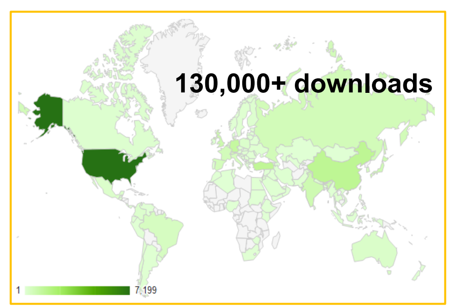


User can compare experimental and theoretical models

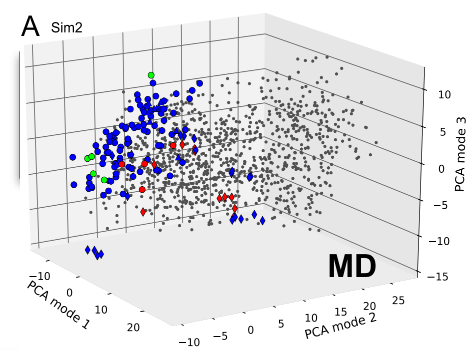


Growth of Source Code and Usage

	Releases	Downloads	Visits[2]	Unique[3]
Nov '10 - Oct '11	19	8,530	8,678	2,946
Nov '11 - Oct '12	15	35,108	16,472	6,414
Nov '12 - Oct '13	8*	87,909	19,888	8,145
Total	42	131,547	45,038	17,505



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

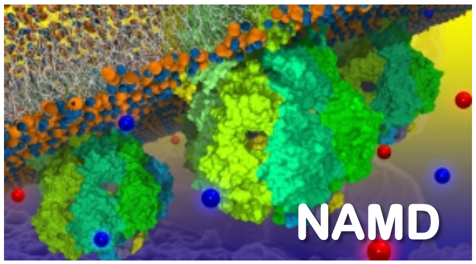


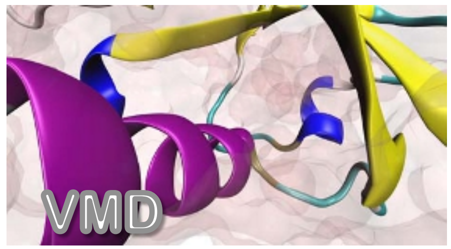
Bakan & Bahar, PSB 2011, 181-192

Tutorials

Days 1-3

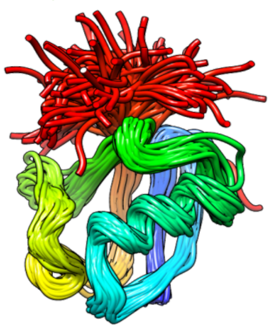
http://www.ks.uiuc.edu/Training/Tutorials/



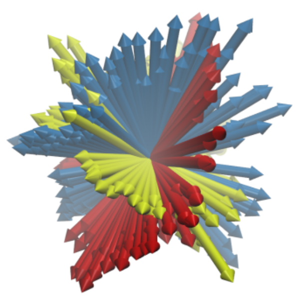


Days 4-5

http://www.csb.pitt.edu/prody/#tutorials



ProDy



NMWiz

Biomedical Technology Research Center (BTRC)

High Performance Computing

for

Multiscale Modeling of Biological Systems

Overarching biological theme:

- **Spatial organization**
- **Temporal evolution**

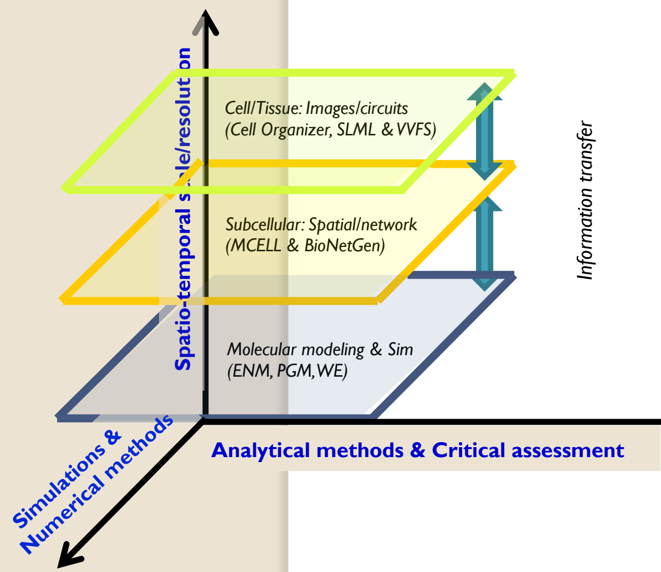
of (neuro)signaling systems/events

Acknowledgment: NIH - 5 P41 GM10371202

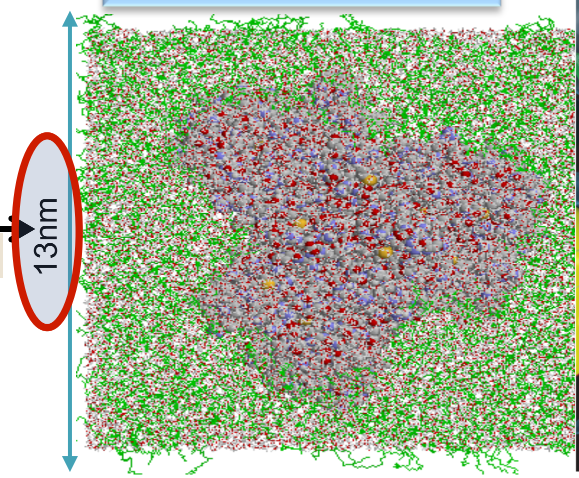
GOAL: TO GENERATE DATA FOR MESOSCOPIC SCALE

Developing integrated *methodology to enable information transfer across scales*

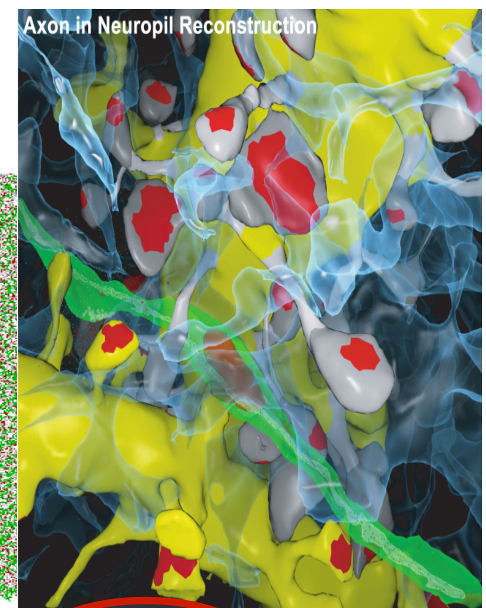
Microphysiological simulations



from molecules



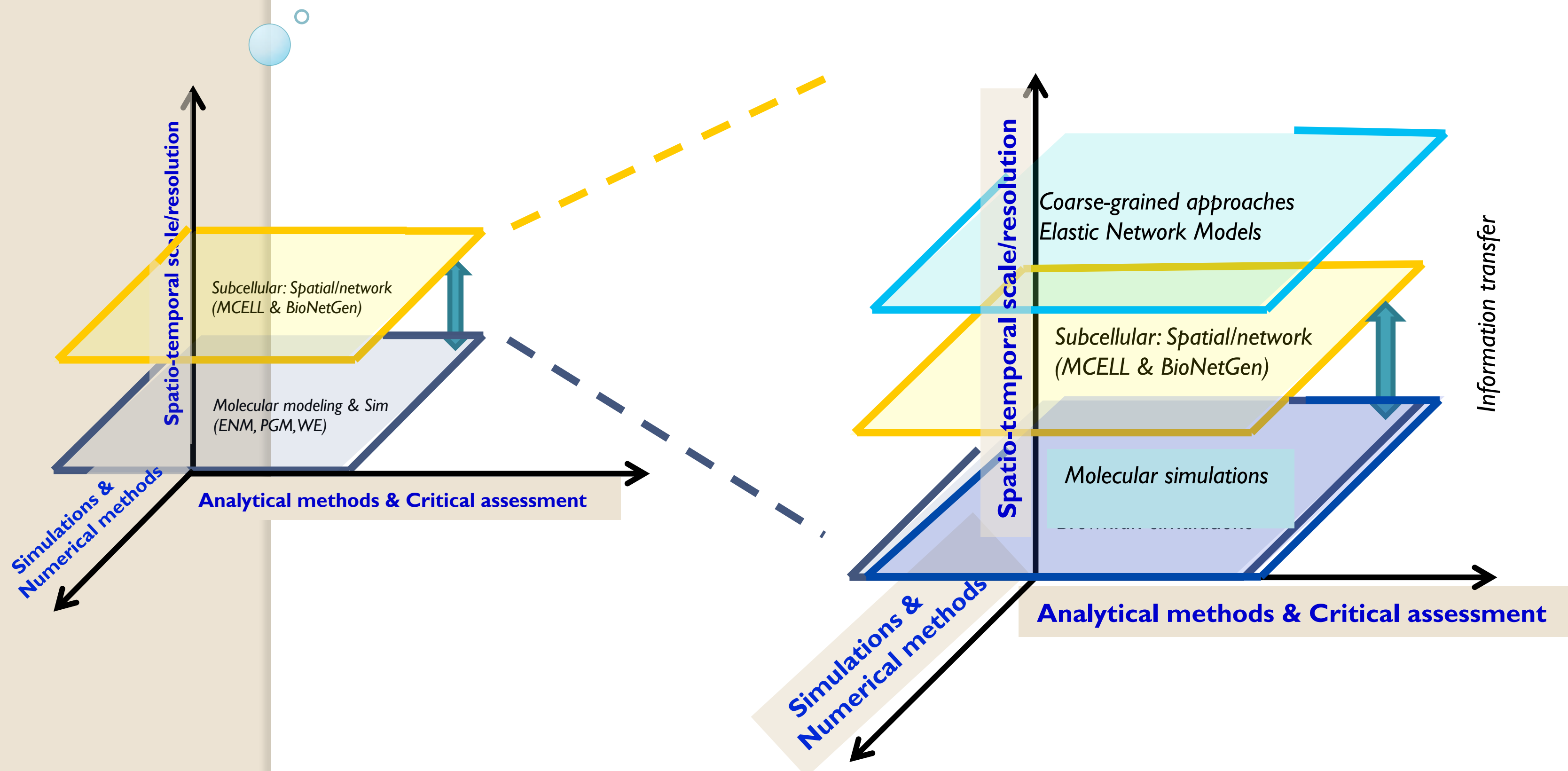
to subcellar events



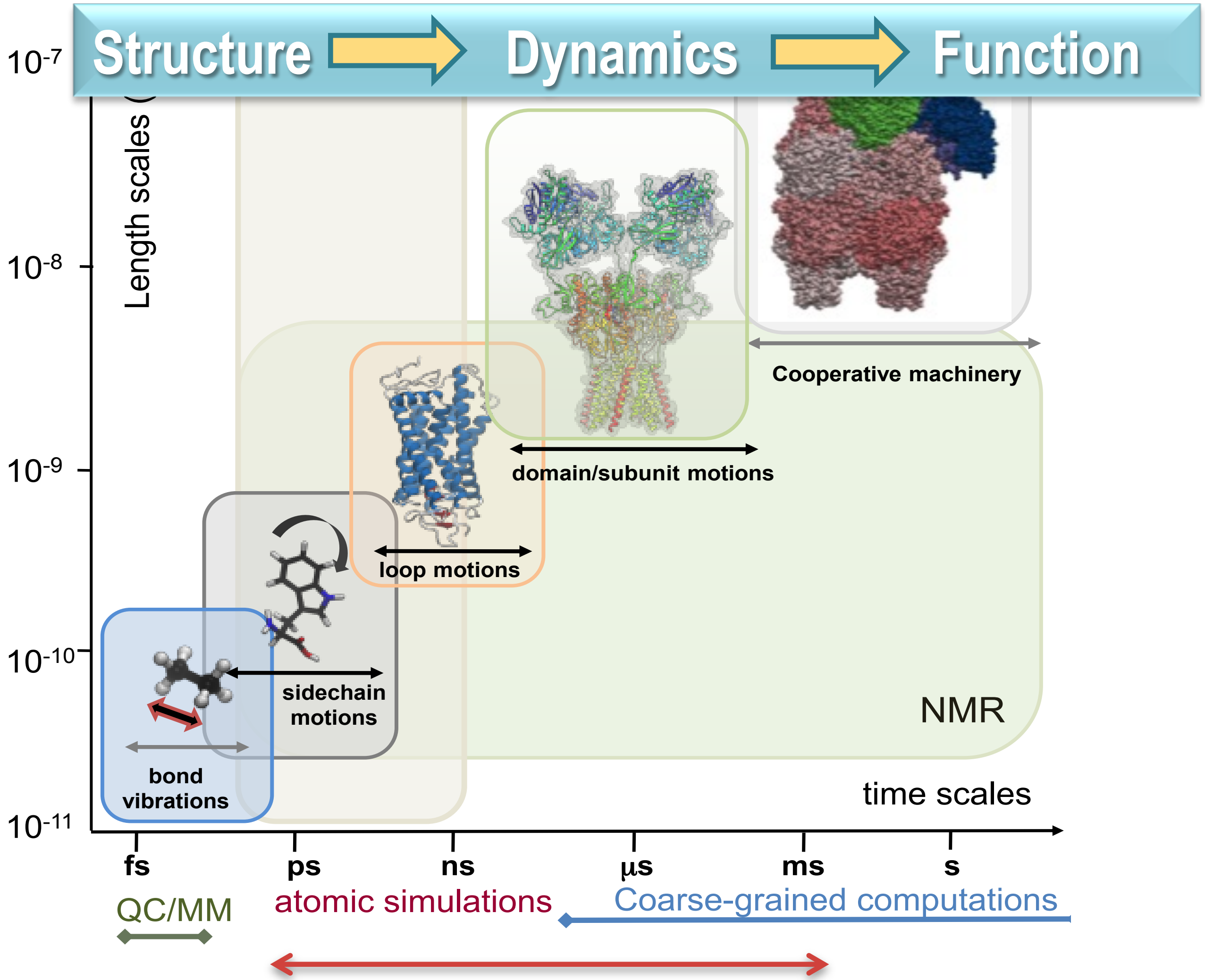
from 6 x 6 x 5 μm^3 sample of adult rat hippocampal stratum radiatum neuropil

Goal: to generate data for mesoscopic scale

Developing integrated *methodology for complex systems dynamics, to enable information transfer across scales*



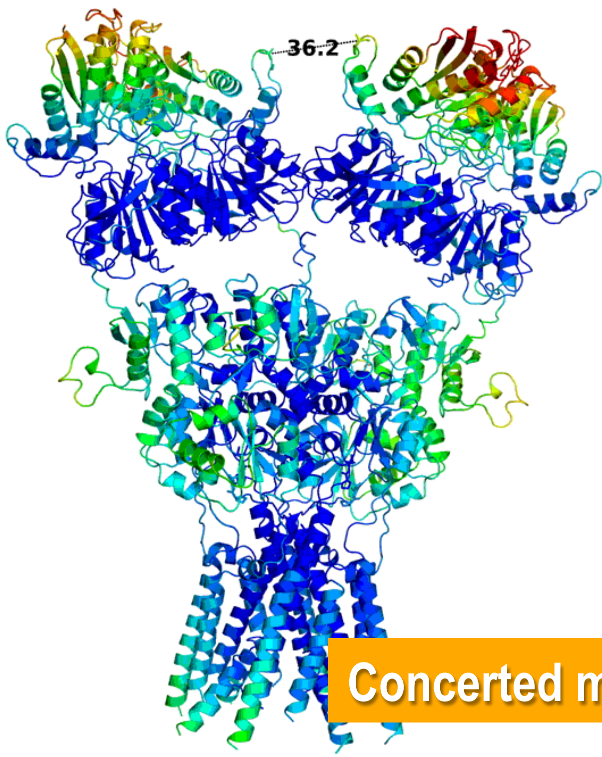
Each structure encodes a unique dynamics



Each structure encodes a unique dynamics

Structure → Dynamics → Function

Signaling dynamics of AMPARs and NMDARs

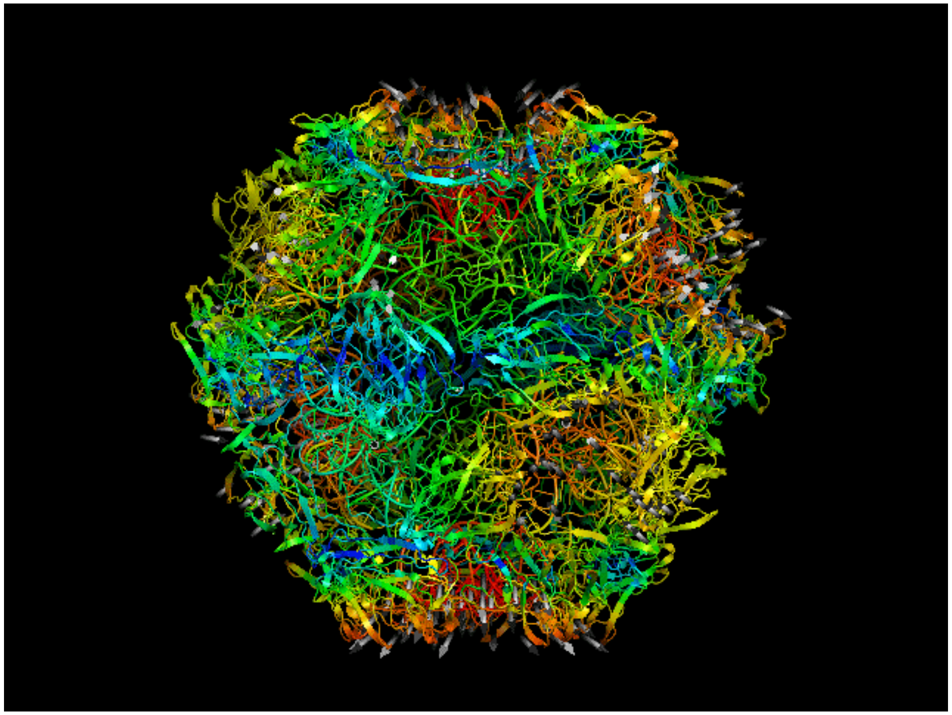


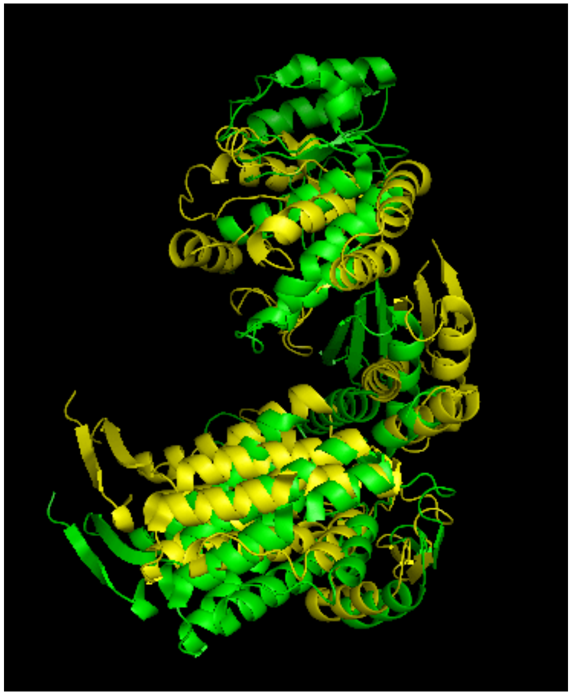
domain/subunit motions

Concerted movements of signaling molecules

Many proteins are molecular machines

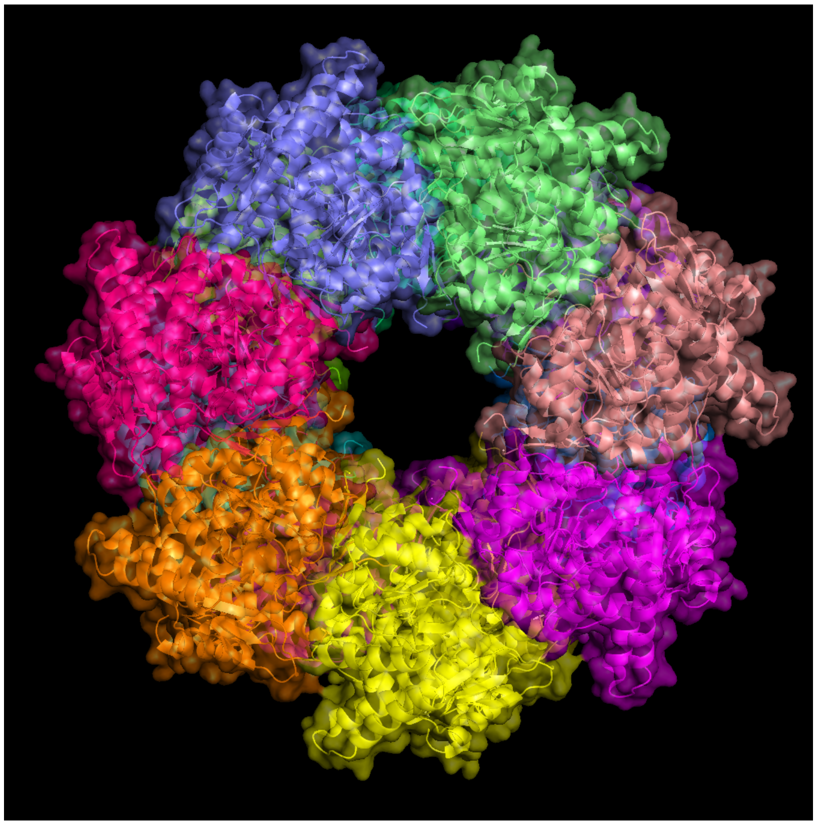
And mechanical properties become more important in complexes/assemblies





STMV dynamics (Zheng Yang)

Representation of structure as a network



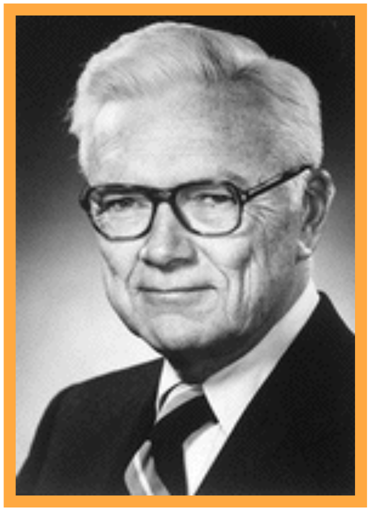
Why network models?

- 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

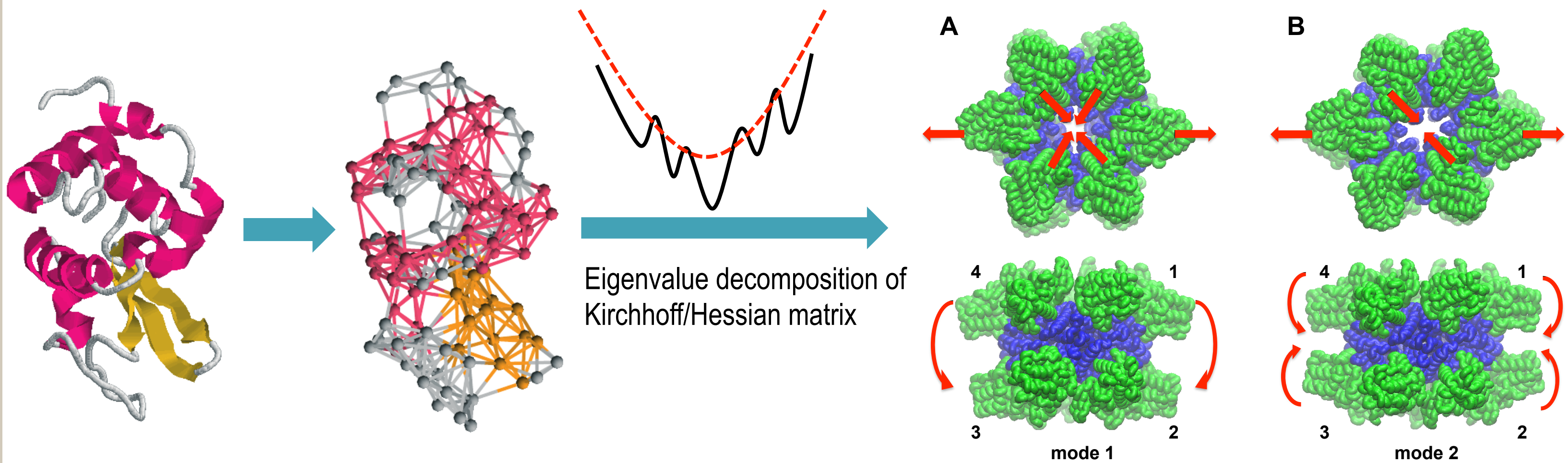
Elastic Network Model for Proteins



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

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

Collective motions
using elastic network models (ENM)

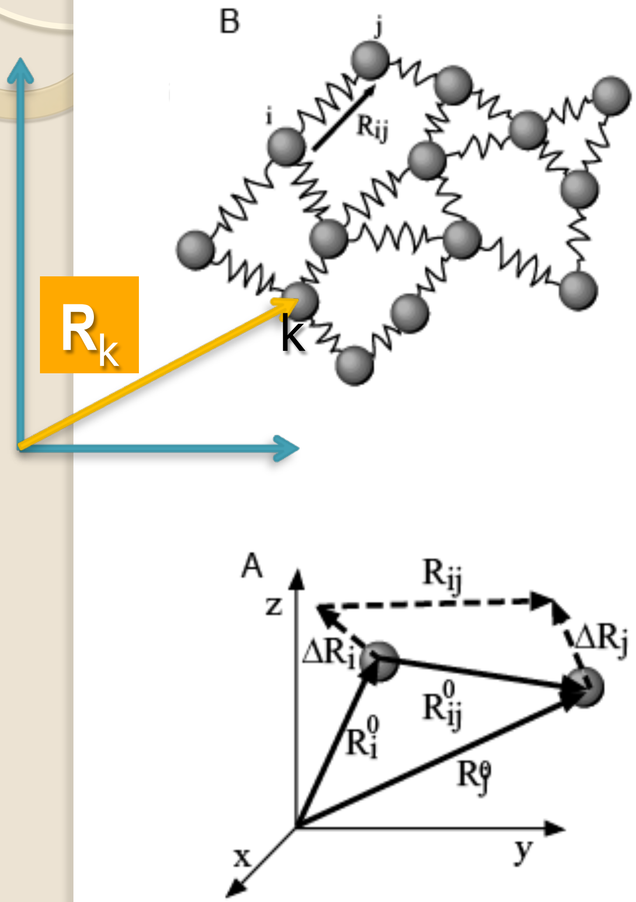


GNM: Bahar et al *Fold & Des* 1996; Haliloglu et al. *Phys Rev Lett* 1997
ANM: Doruker et al. *Proteins* 2000; Atilgan et al, *Biophys J* 2001

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

Bahar, Lezon, Yang & Eyal (2010) Global Dynamics of Proteins: Bridging Between Structure and Function *Annu Rev Biophys* **39**: 23-42

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)

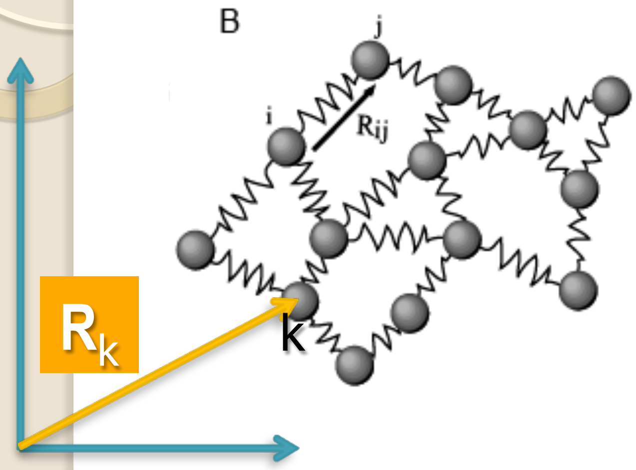
→ Nodes are subject to **Gaussian fluctuations** ΔR_i
→ Inter-residue distances R_{ij} also undergo Gaussian fluctuations

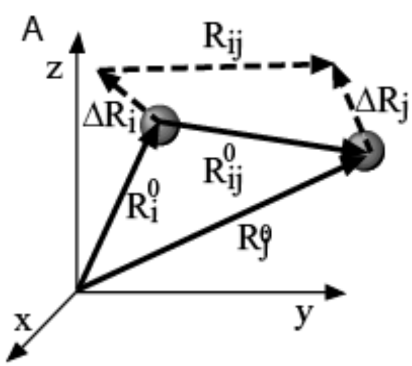
$$\rightarrow \quad \Delta \mathbf{R}_{ij} = \Delta \mathbf{R}_j - \Delta \mathbf{R}_i$$

Fluctuations in residue positions

Bahar, Atilgan & Erman, *Fold & Des* 1997

Gaussian network model (GNM)





Fluctuation vector:

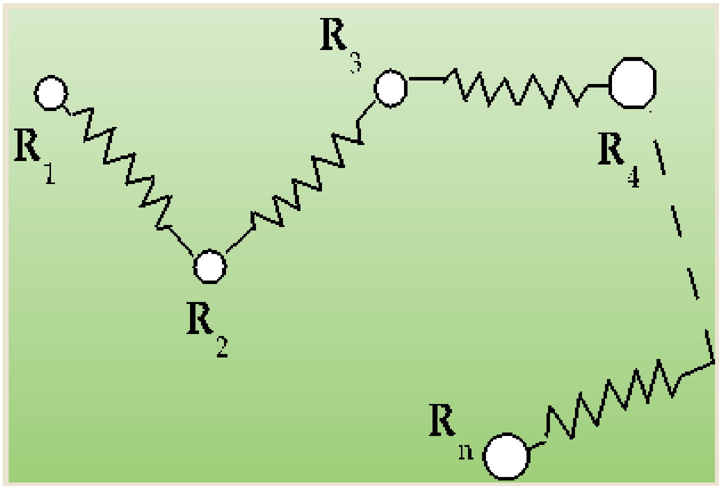
$$\rightarrow \Delta \mathbf{R} = \begin{bmatrix} \Delta \mathbf{R}_1 \\ \Delta \mathbf{R}_2 \\ \Delta \mathbf{R}_3 \\ \Delta \mathbf{R}_4 \\ .. \\ .. \\ .. \\ .. \\ \Delta \mathbf{R}_N \end{bmatrix}$$

Fluctuations in residue positions

Bahar, Atilgan & Erman, *Fold & Des* 1997

Rouse model for polymers

Classical bead-and-spring model



Kirchhoff matrix

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

Force constant

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

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

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

Rouse model for polymers

Kirchhoff matrix

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

Force constant

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

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

Rouse model for polymers

Fluctuation vector

Kirchhoff matrix

$$(\gamma/2)\,[\Delta\mathbf{R}_1 \;\; \Delta\mathbf{R}_2 \;\; \Delta\mathbf{R}_3 \;....\; \Delta\mathbf{R}_N] \begin{bmatrix} 1 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & .. & ... & & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{bmatrix} \begin{bmatrix} \Delta\mathbf{R}_1 \\ \Delta\mathbf{R}_2 \\ \Delta\mathbf{R}_3 \\ \\ \\ \end{bmatrix} =$$

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

Force constant

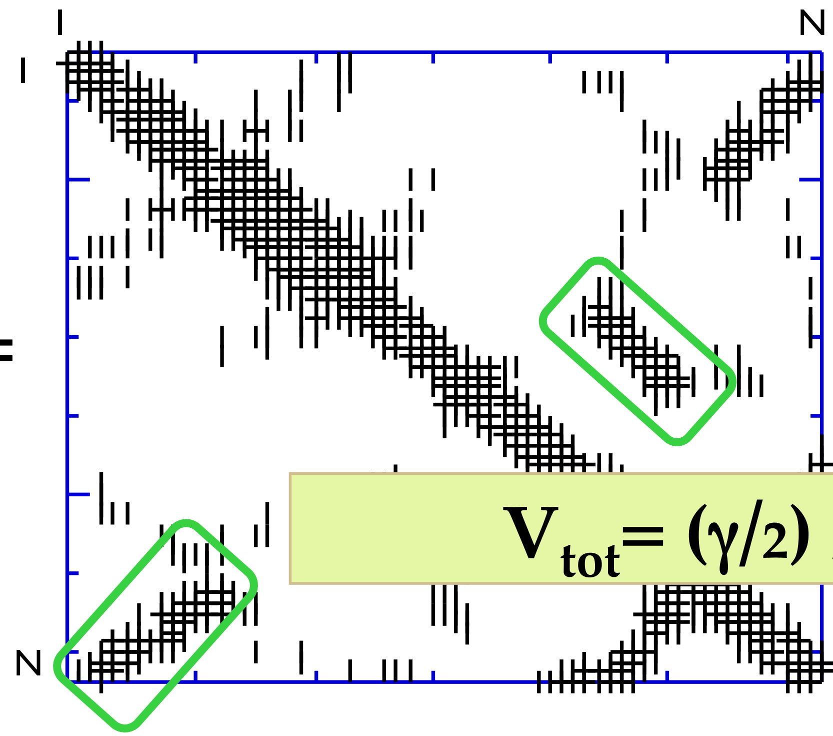
$$V_{tot} = (\gamma/2)\,[\,(\Delta R_{12})^2 + (\Delta R_{23})^2 + \ldots\ldots\ldots\; (\Delta R_{N\text{-}1,N})^2\,]$$

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

Kirchhoff matrix for inter-residue contacts

For a protein of N residues

$\Gamma =$



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

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

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

Γ provides a complete description of contact topology!

Statistical mechanical averages

For a protein of N residues

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

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

$\boldsymbol{\Gamma}$ provides a complete description of contact topology!

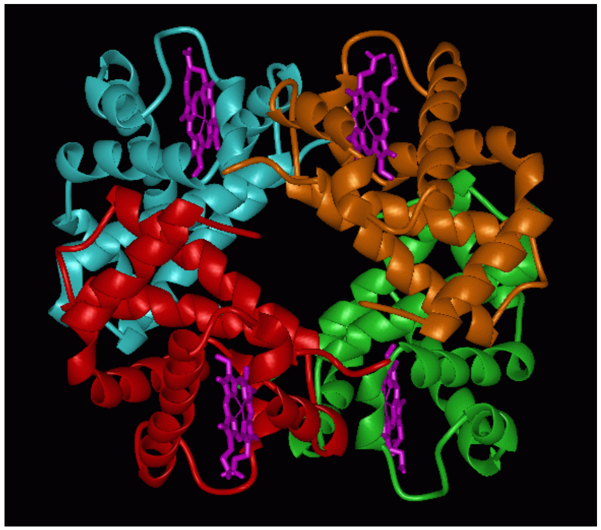
Kirchhoff matrix determines the mean-square fluctuations

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

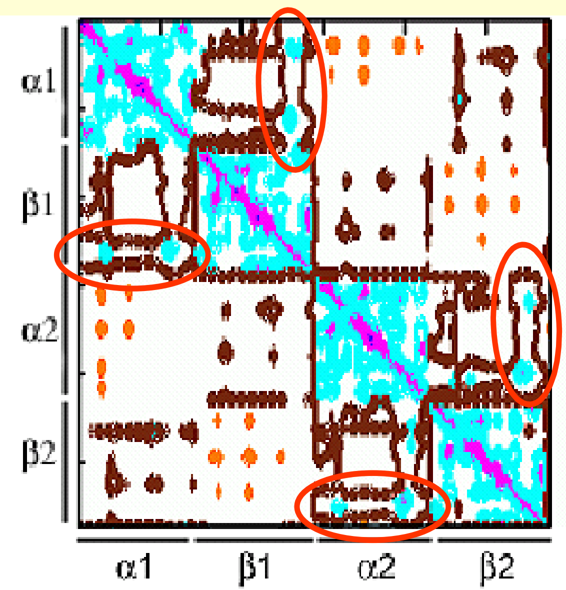
And cross-correlations between residue motions

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

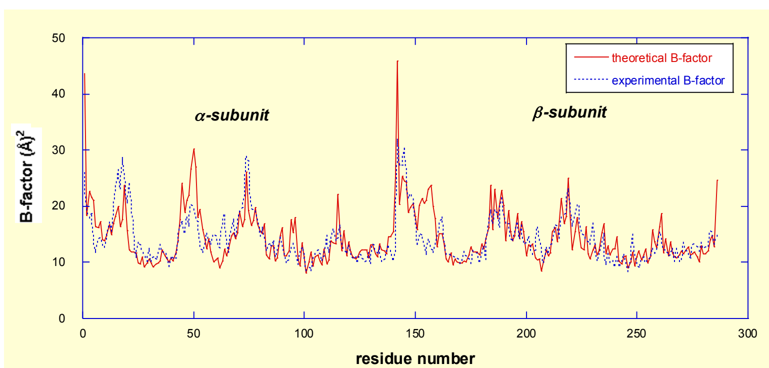
1. Application to hemoglobin



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



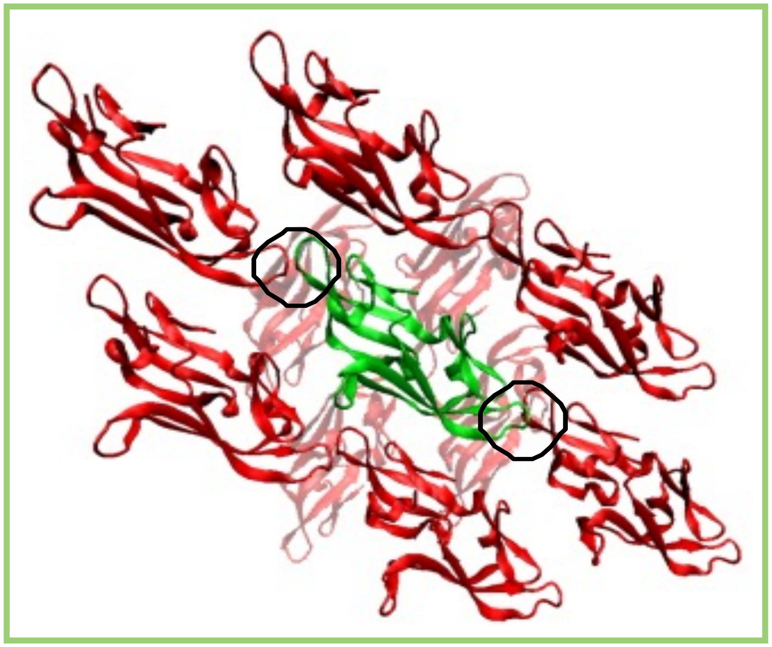
Intradimer cooperativity – Symmetry rule (Yuan et al. JMB 2002; Ackers et al. PNAS 2002.)

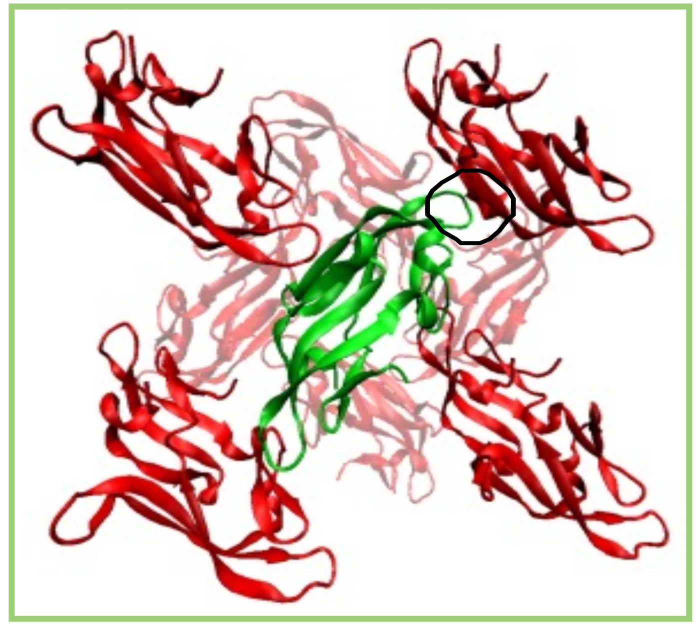


B- factors – Comparison with experiments

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

B-factors are affected by crystal contacts

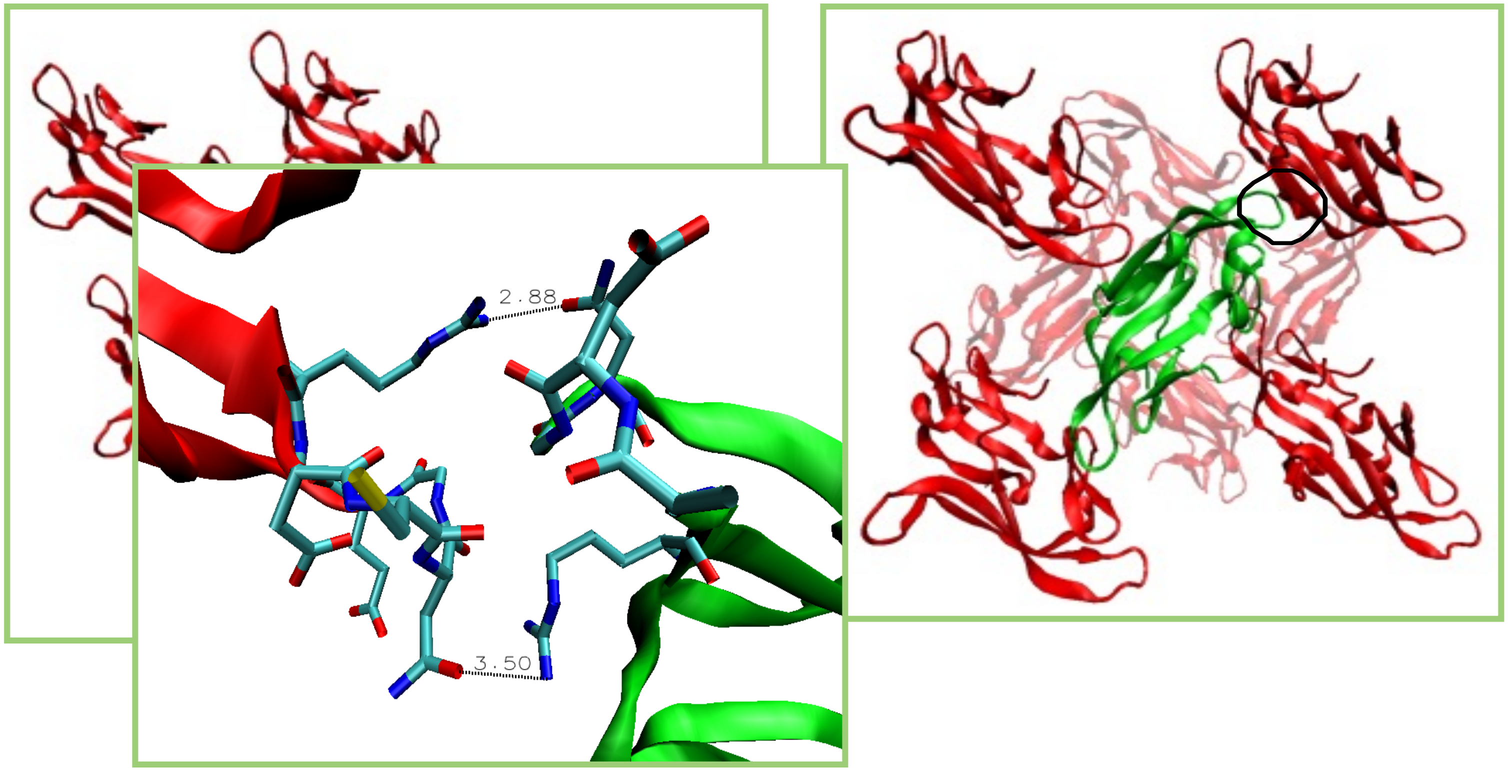




Two X-ray structures for a designed sugar-binding protein LKAMG

1

B-factors are affected by crystal contacts



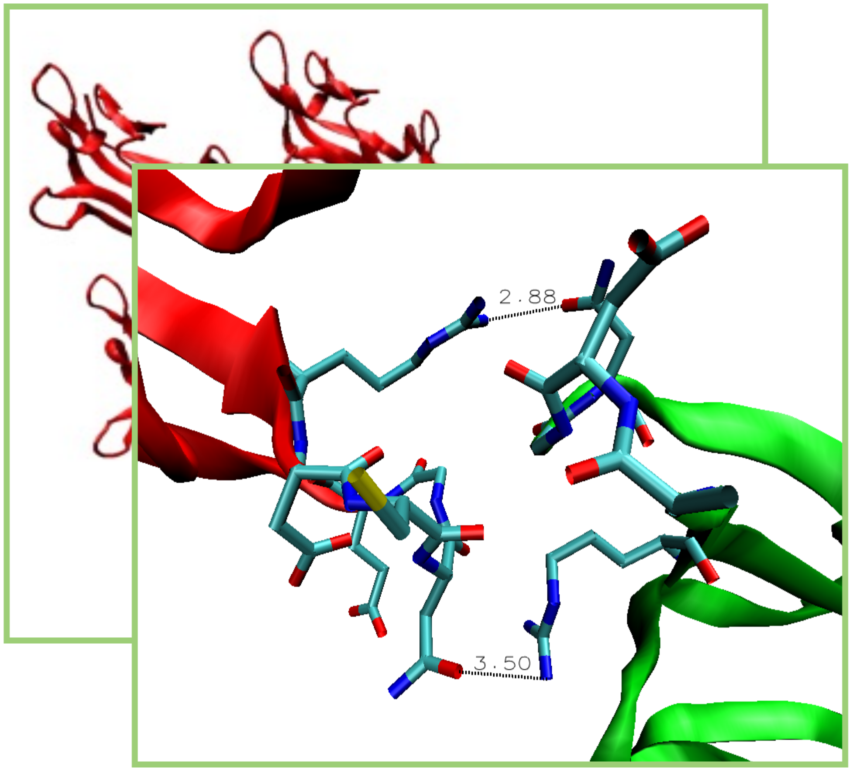
Particular loop motions are curtailed by intermolecular contacts in the crystal environment causing a discrepancy between theory and experiments

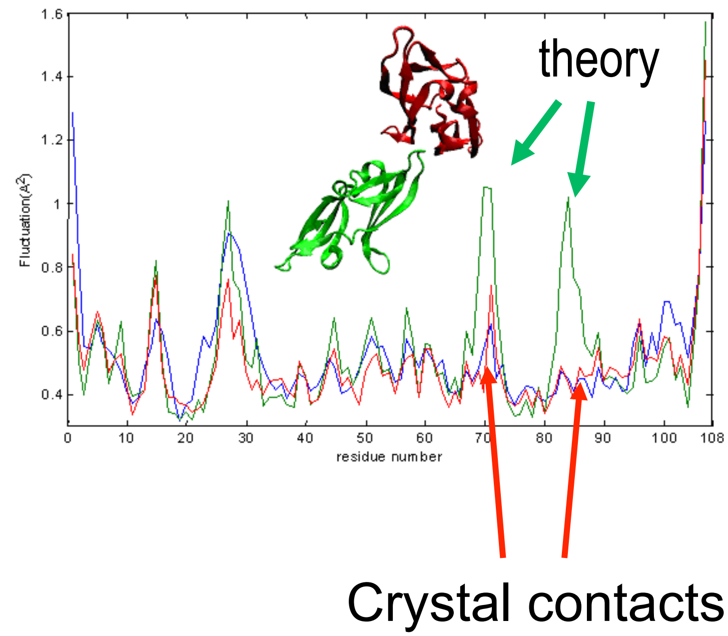
FOR MORE INFO...

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

1

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





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

Contribution of mode k

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

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

Contribution of mode k

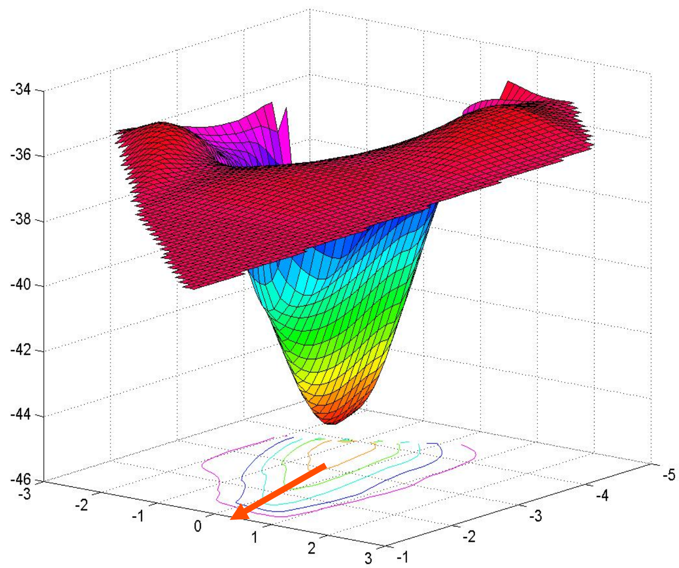
$$[\Delta\mathbf{R}_i \cdot \Delta\mathbf{R}_j]_k = (3k_B T / \gamma) \boxed{\left[\lambda_k^{-1} \mathbf{u}_k \mathbf{u}_k^T\right]_{ij}}$$

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

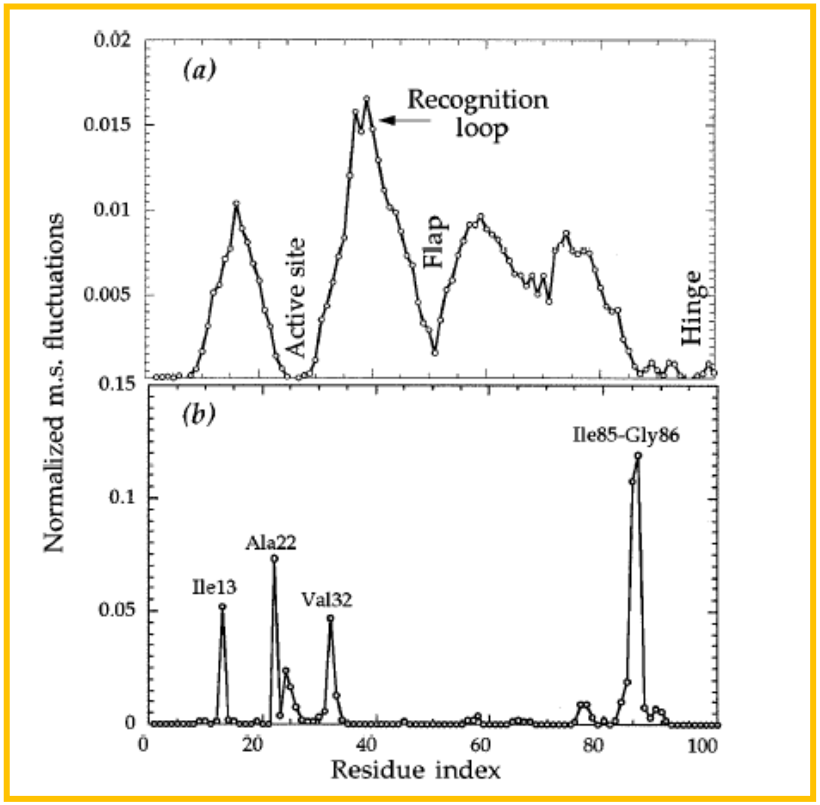
FOR MORE INFO...

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

Several modes contribute to dynamics



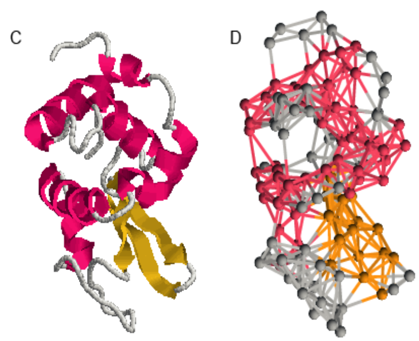
The first mode selects
the 'easiest' collective motion



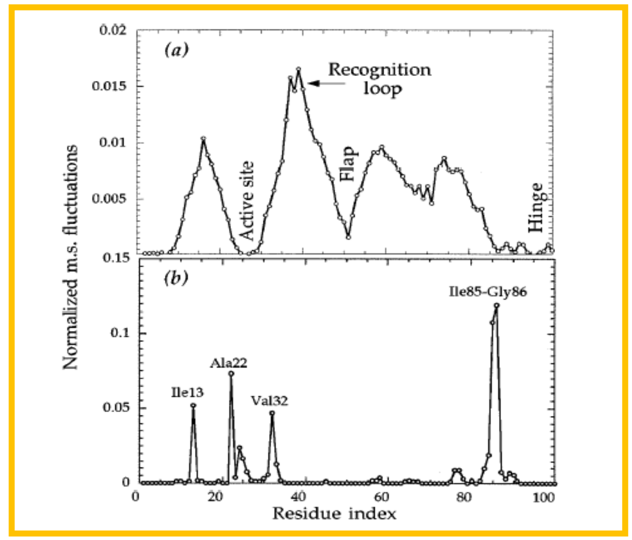
FOR MORE INFO...

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

Gaussian network model (GNM)

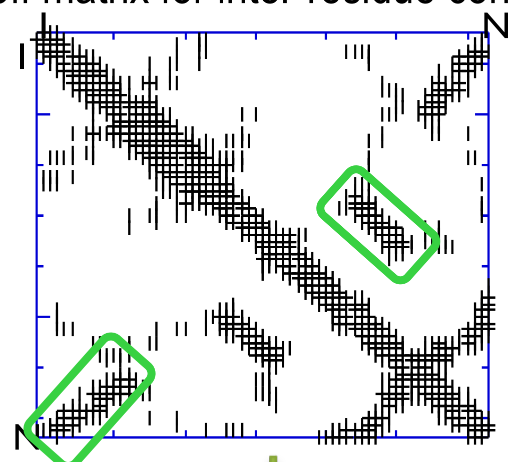


$$[\Delta \mathbf{R}_i \bullet \Delta \mathbf{R}_i]_k = (3k_B T/\gamma)\left[\lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^T\right]_i$$



Kirchhoff matrix for inter-residue contacts

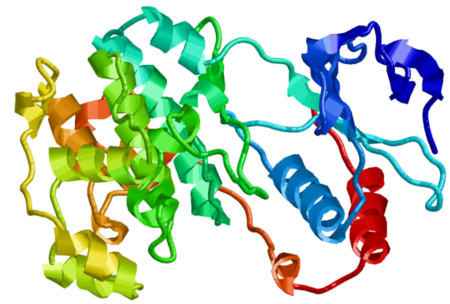
$\Gamma =$

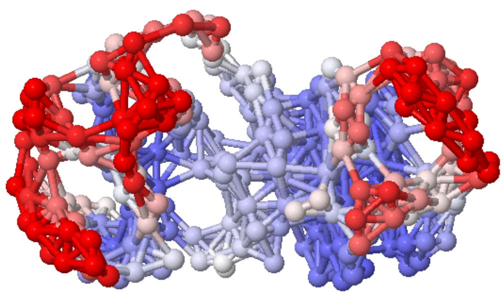


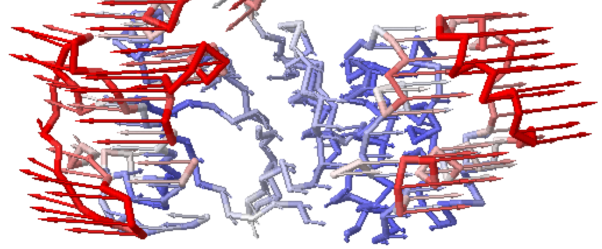
Several modes of motion contribute to dynamics

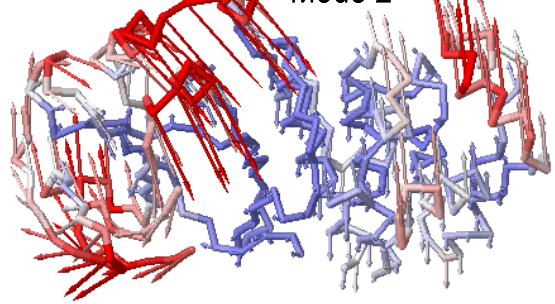
$$< \Delta \boldsymbol{R}_i \bullet \Delta \boldsymbol{R}_j >= (1/Z_N)\int (\Delta \boldsymbol{R}_i \bullet \Delta \boldsymbol{R}_j) e^{-V/k_B T} d\{\Delta \boldsymbol{R}\} = (3k_B T/\gamma)\left[\boldsymbol{\Gamma}^{-1}\right]_{ij}$$

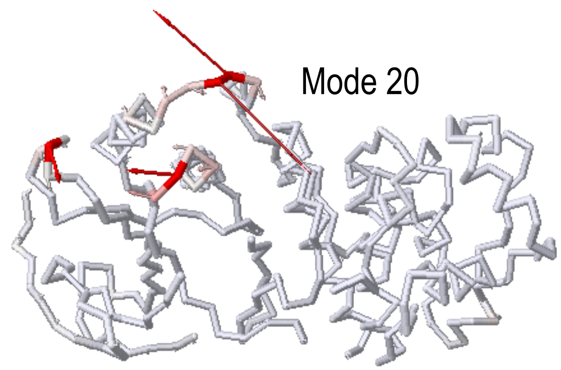
Anisotropic Network Model (ANM)



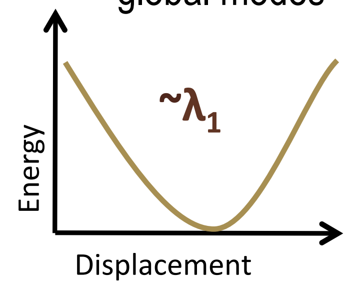


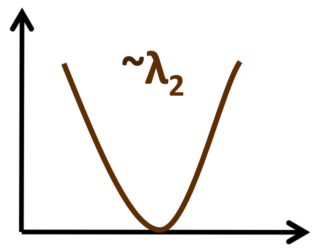




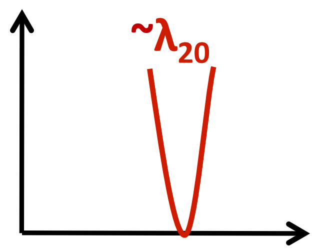


low-frequency
~ global modes





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



higher-frequency

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

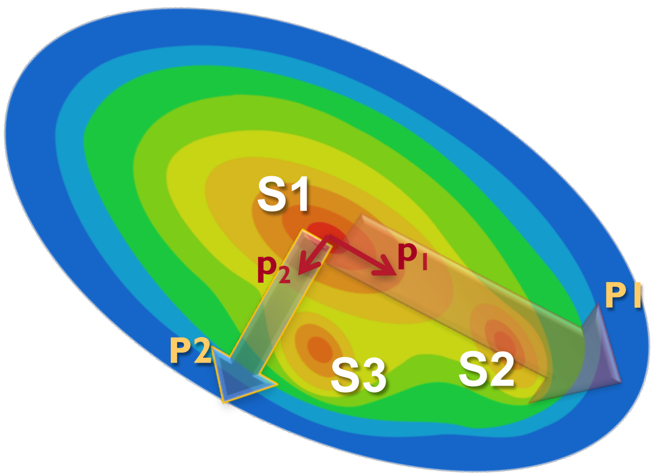
$$\mathbf{H}^{(ij)} = \frac{\gamma \, \Gamma_{ij}}{\left(R_{ij}^0\right)^2} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix}$$

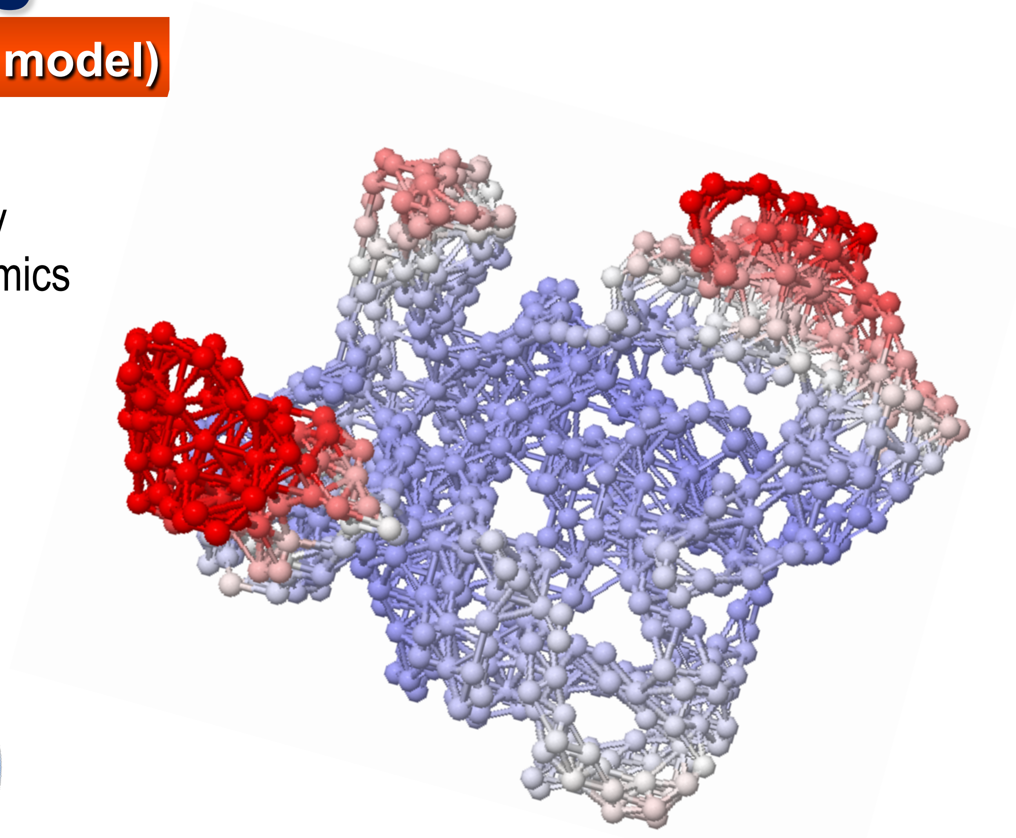
Doruker et al. (2000) *Proteins*; Atilgan AR et al. (2001) *Biophys J.*; Eyal et al. (2006) *Bioinformatics* 22, 2619

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**

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

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

- figure()
- showSqFlucts(anm)

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

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

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 $<(\Delta R_i \,.\, \Delta R_j)>$ 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

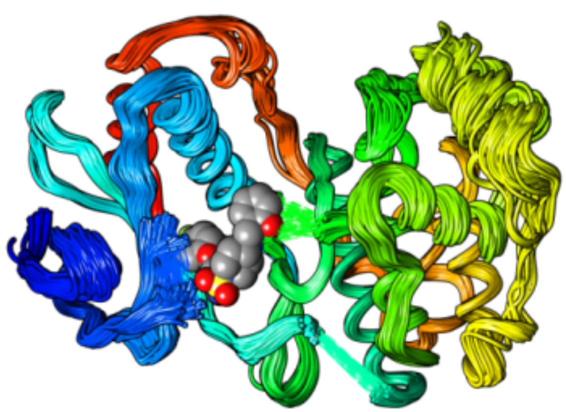
Ensembles of structures

- Structural changes accompanying substrate (protein) binding

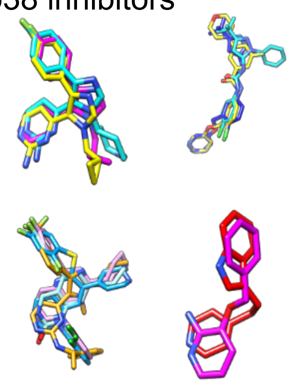
- Structural changes induced by, or stabilized upon, ligand binding

Ubiquitin
140 structures
1732 models

p38 MAP kinase
(**182 structures**)

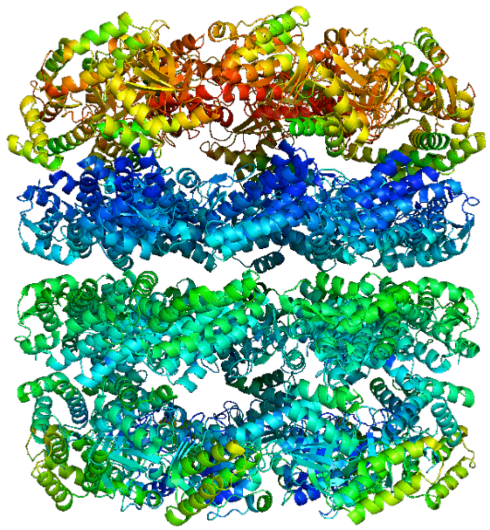


p38 inhibitors



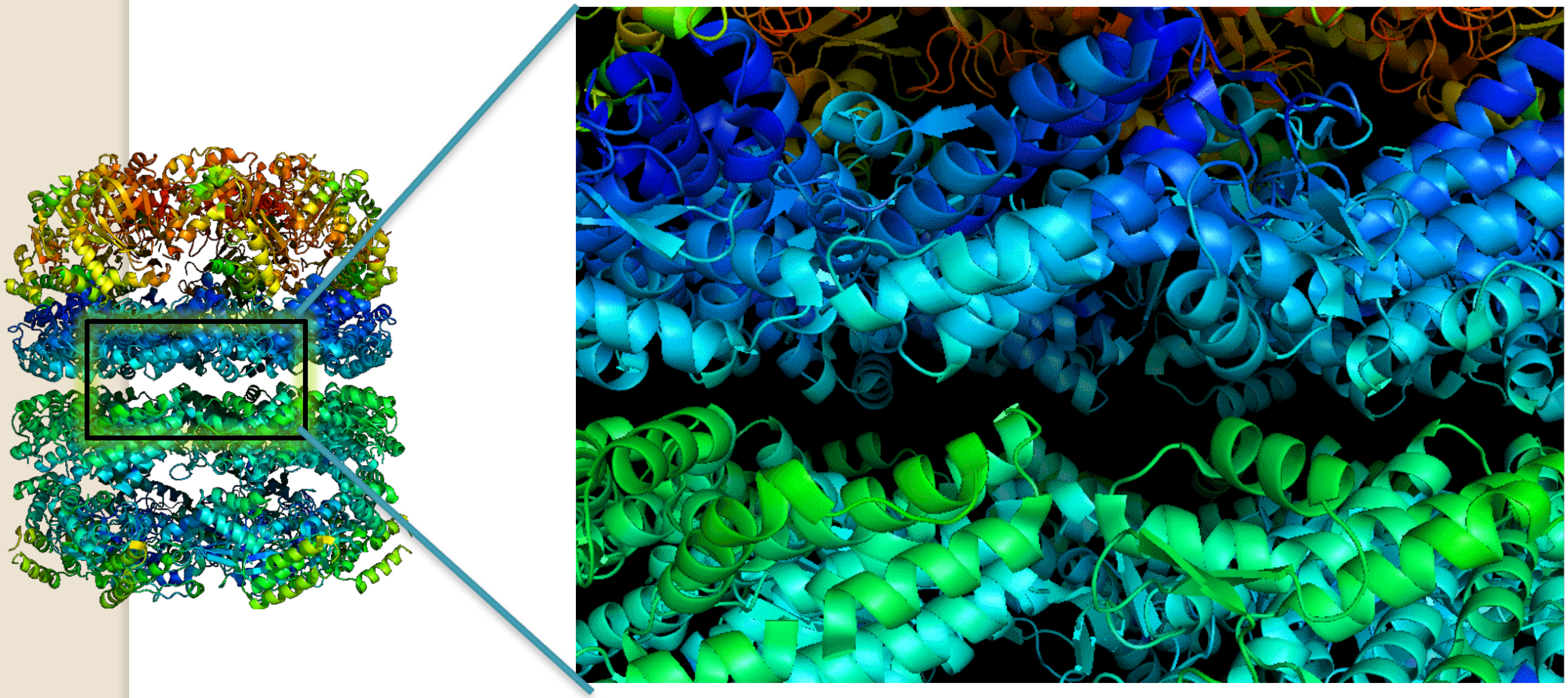
Ensembles of structures

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

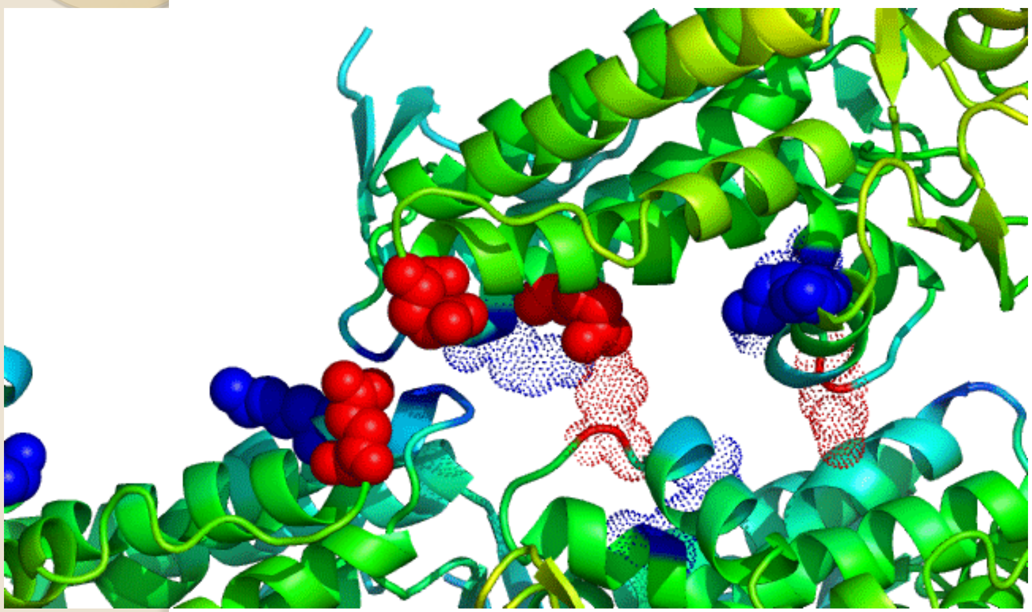


Yang et al. *PLoS Comp Biol* 2009

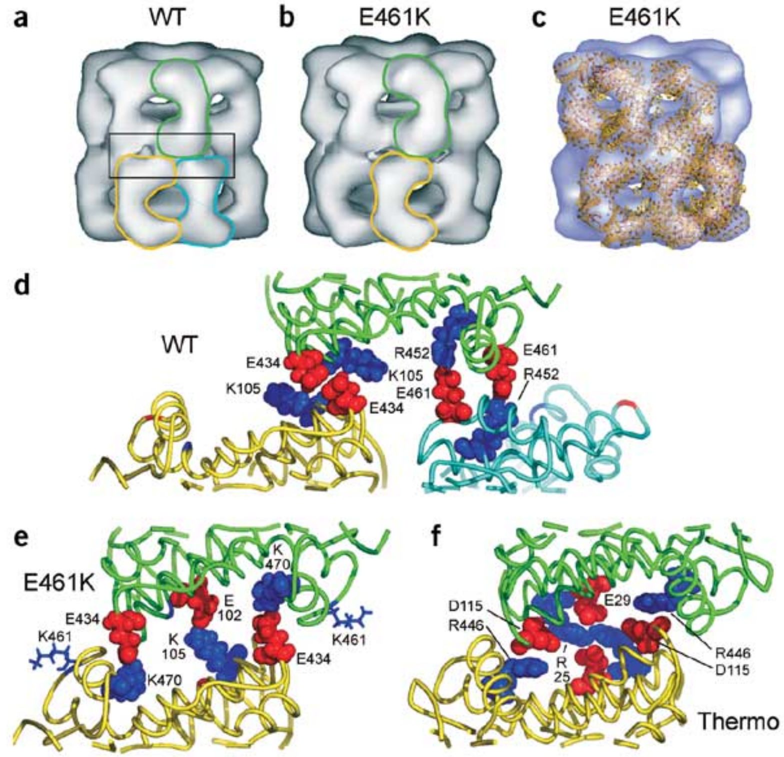
Redistribution of interactions at interfaces



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

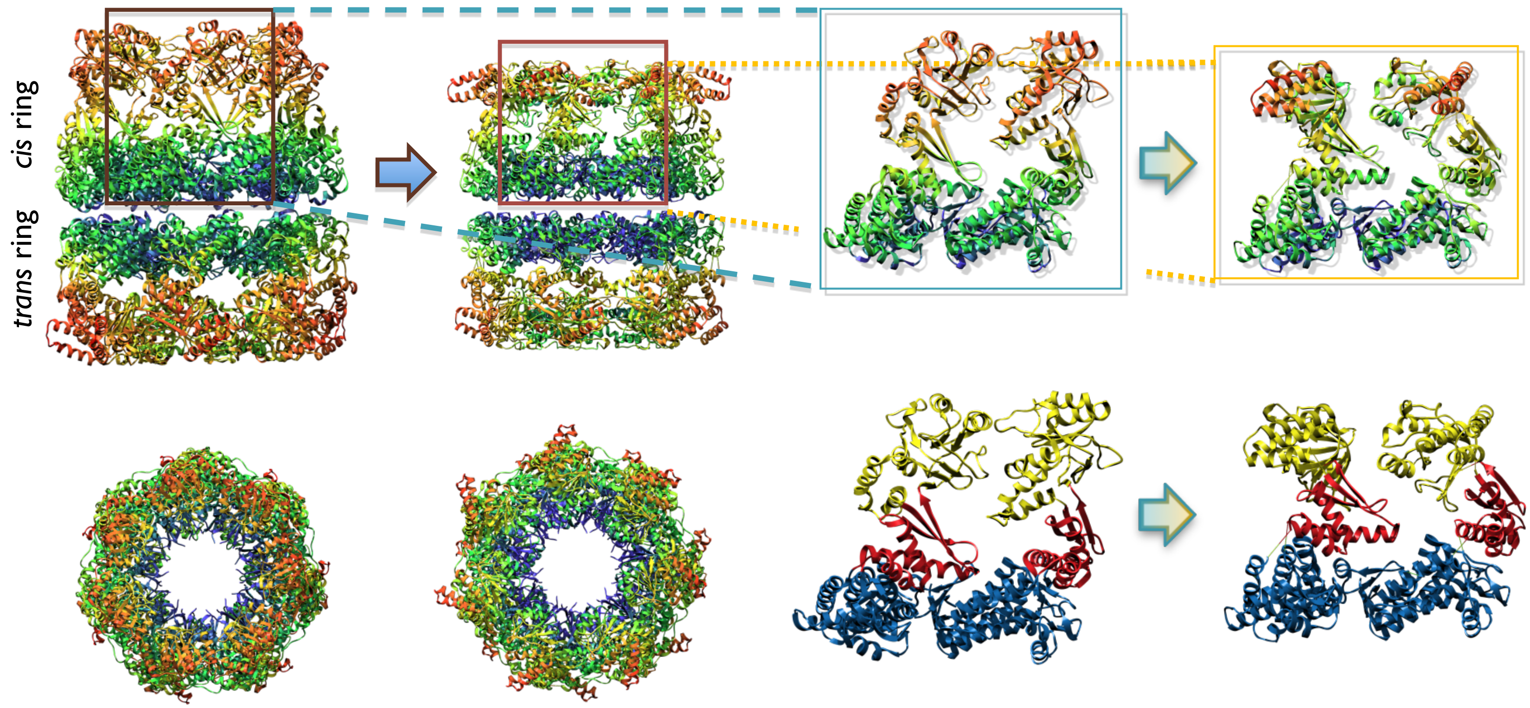


E461 mutant is a deformed structure along mode 1



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

Passage between the R and T state of GroEL

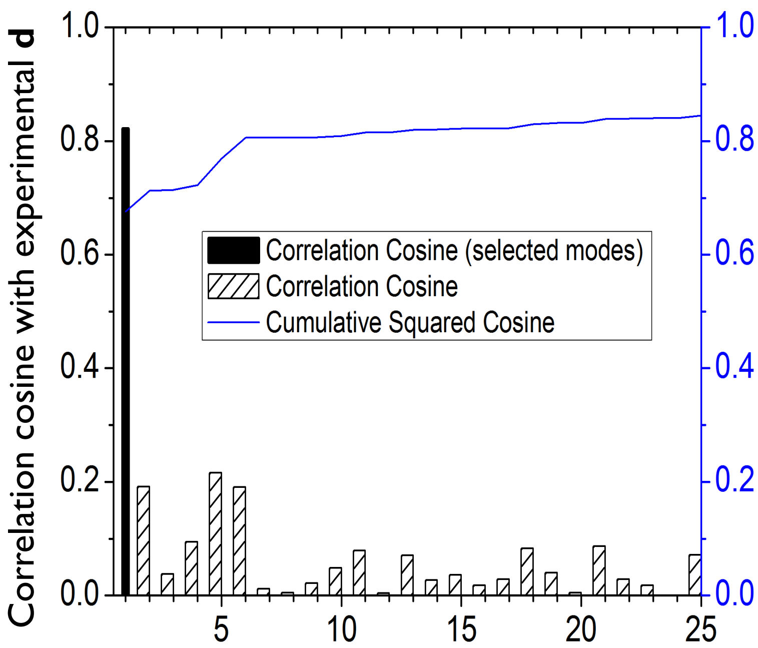


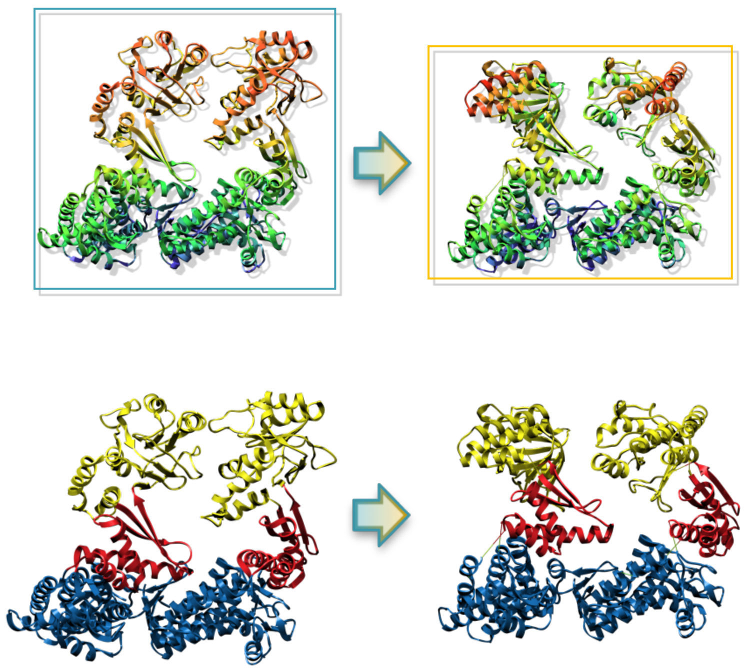
See…

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)





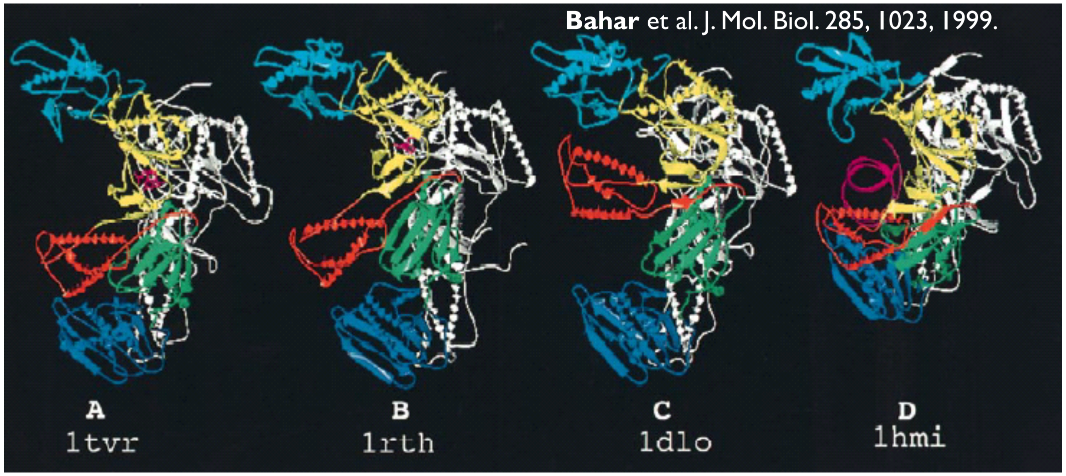
$\mathbf{d} = [\Delta x_1 \ \ \Delta y_1 \ \ \Delta z_1 \ \ldots \ \ \Delta z_N]^T$

See…

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

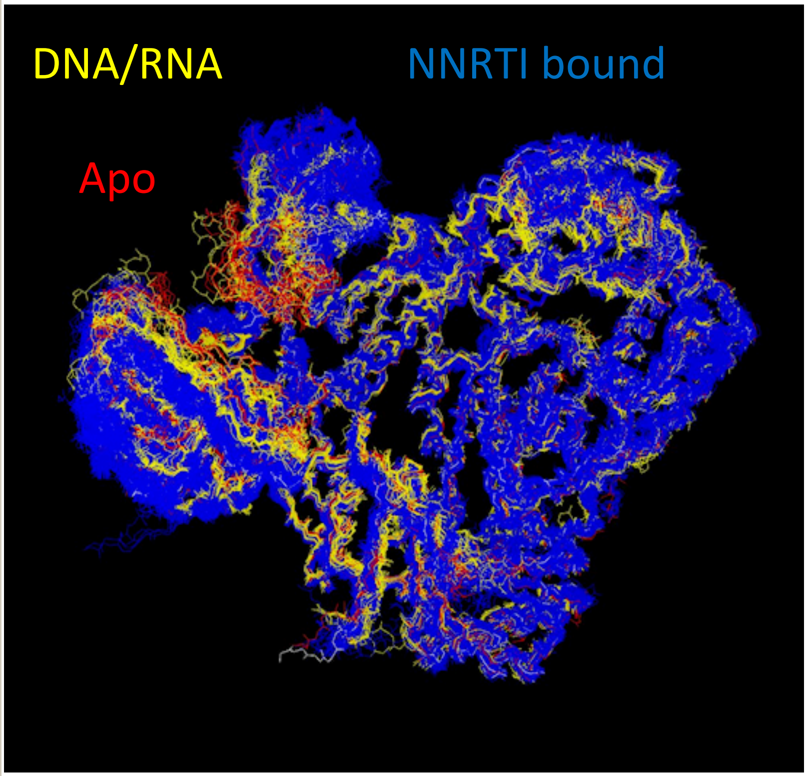
Dynamics inferred from known structures

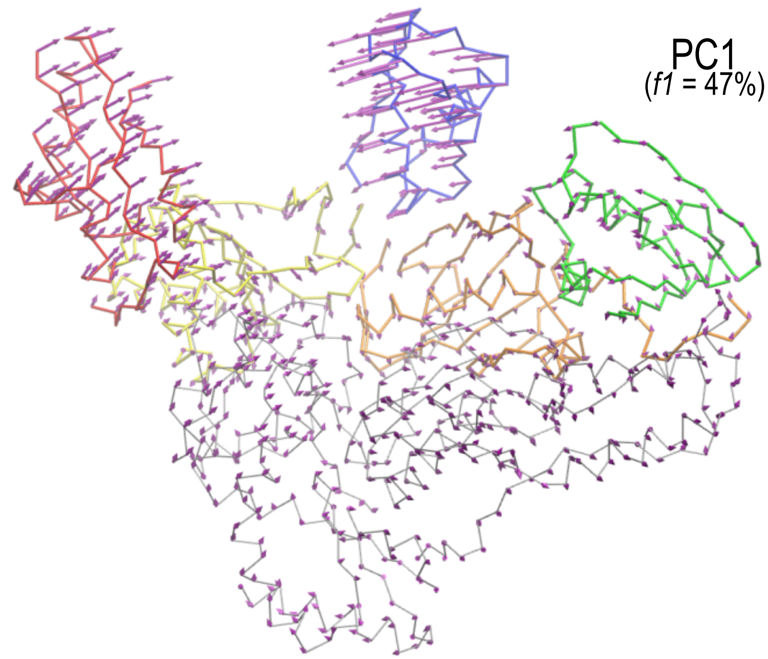
Comparison of static structures available in the PDB for the same protein in different form has been widely used is an **indirect** method of inferring dynamics.



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

Principal Component Analysis (PCA)

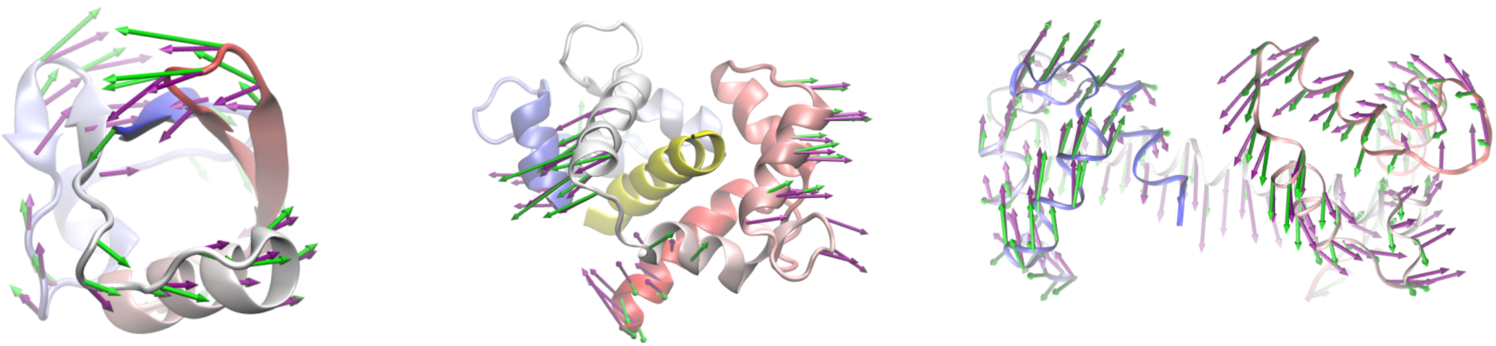




$$\mathbf{C}^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta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

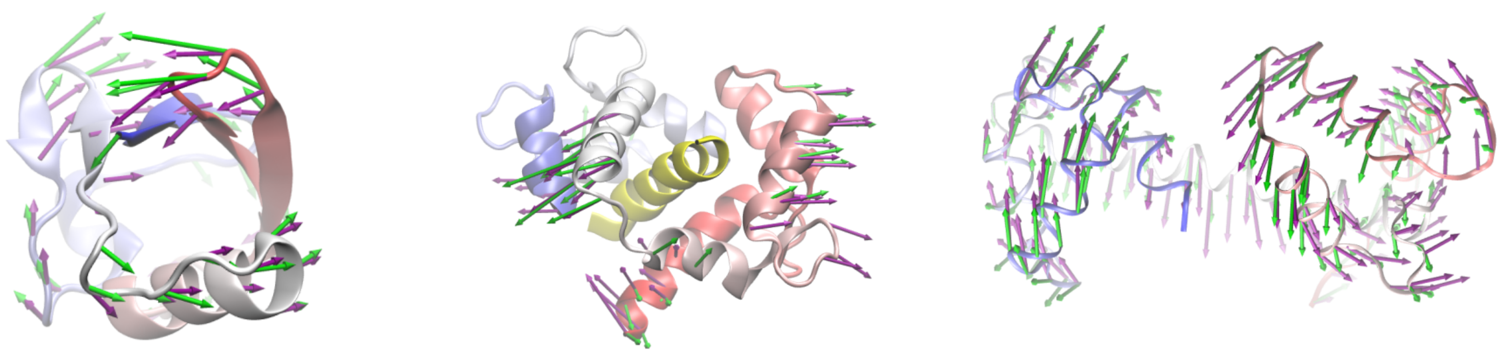
$$\mathbf{C} = \mathbf{PSP}^T = \sum_{i=1}^{3N} \sigma_i \, \boldsymbol{p}^i \, \boldsymbol{p}^{i\mathrm{T}}$$

Global motions inferred from theory and experiments



→ PCA of the ensemble of resolved structures
→ ANM analysis of a single structure from the ensemble

Global motions inferred from theory and experiments



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

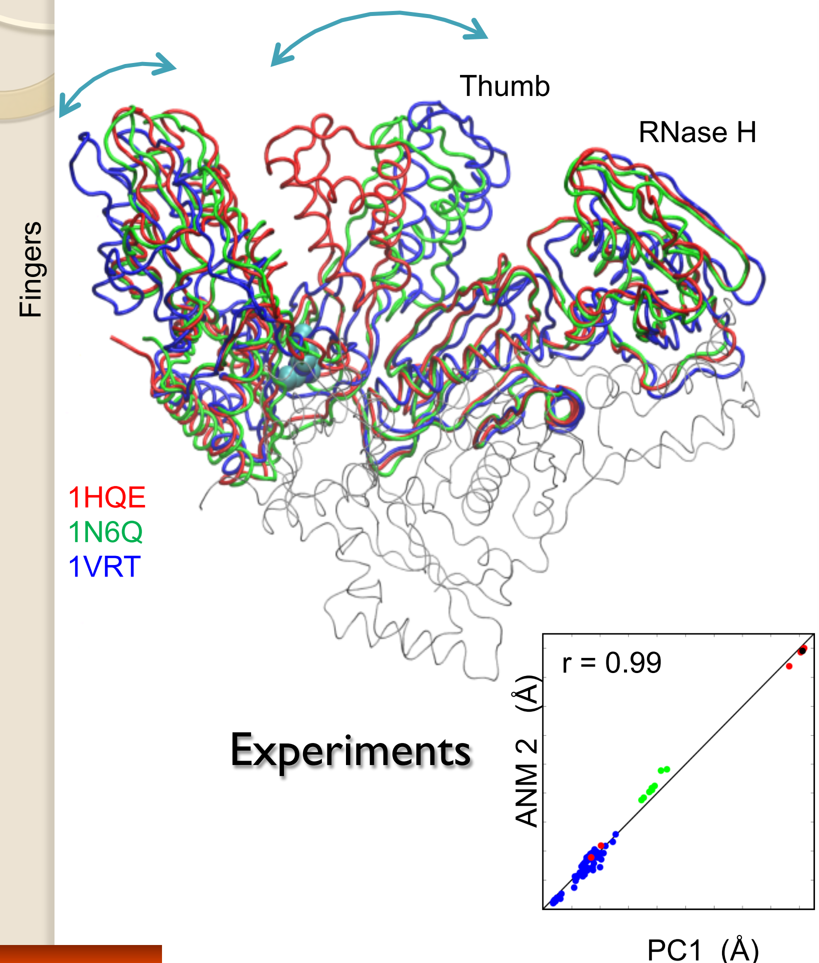
Ahmet Bakan and Ivet Bahar[1]

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

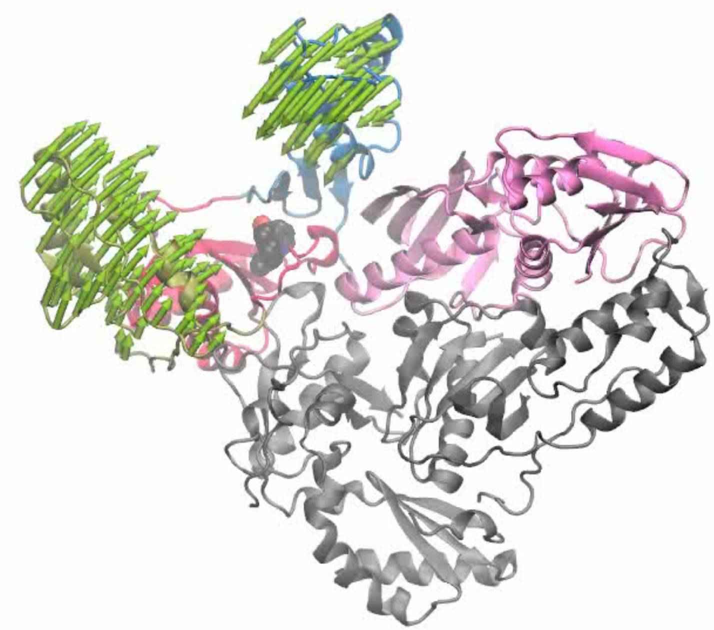
Reference:

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

Induced Dynamics or Intrinsic Dynamics?



Experiments



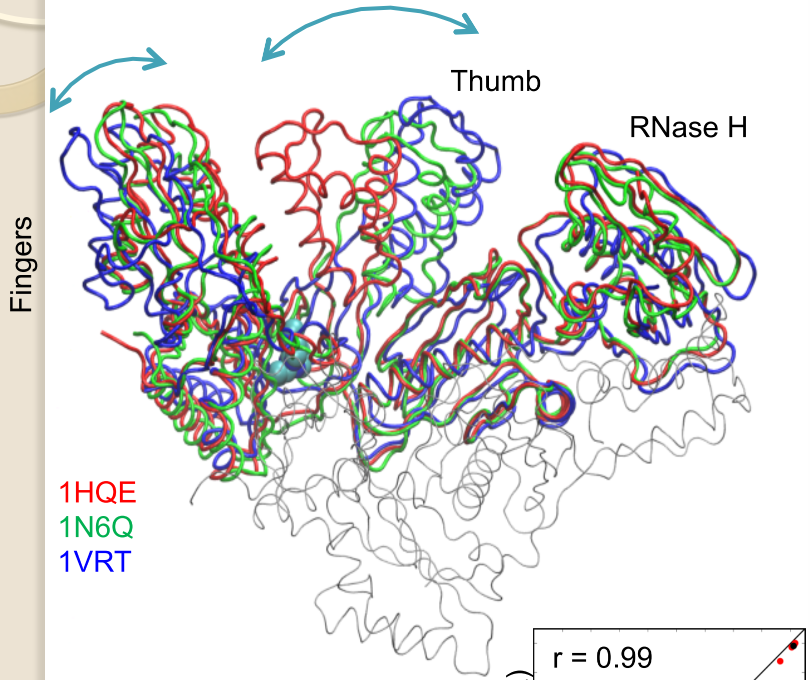
Theory

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

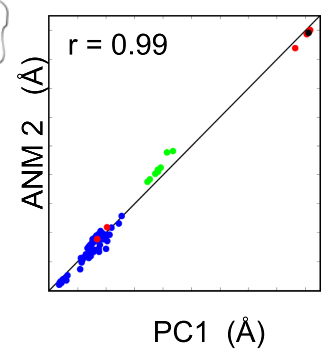
References:

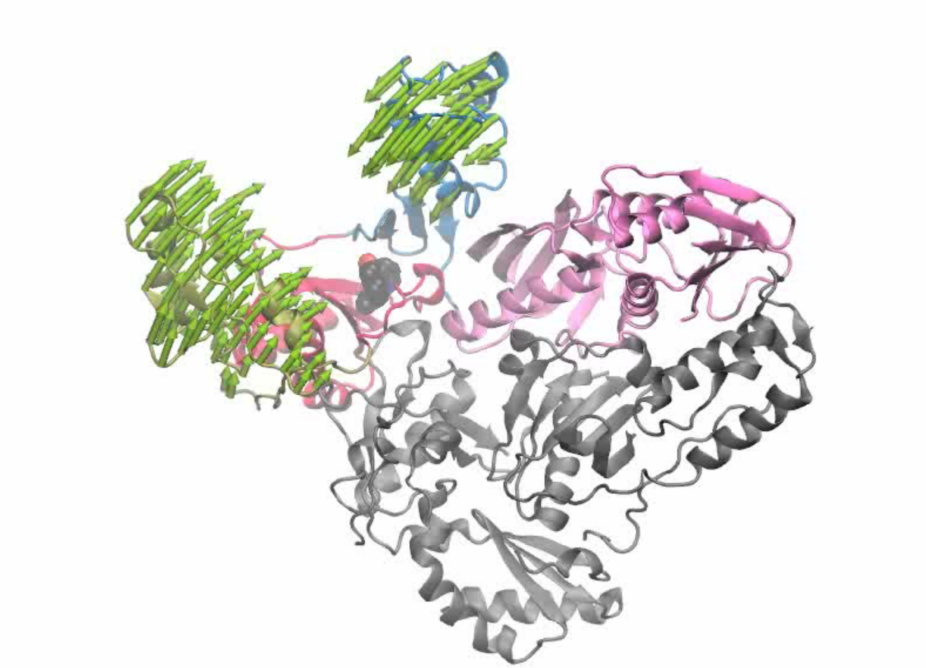
Bakan & Bahar (2009) PNAS **106**, 14349-54.

Soft modes enable functional movements



Experiments





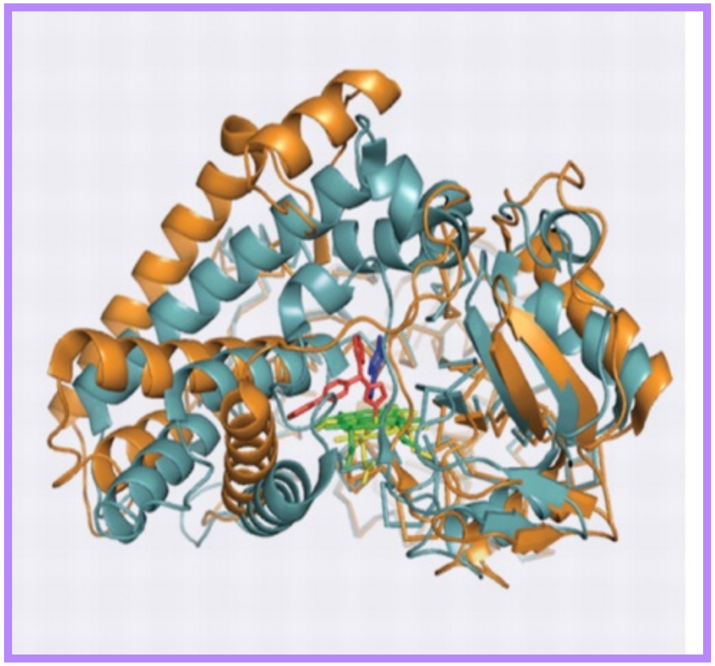
Theory

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...

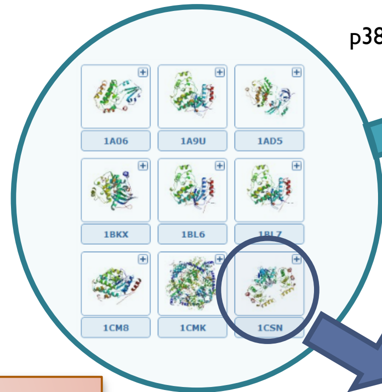
N. Tokuriki and D. S. **Tawfik** (2009) *Science* **324**: 203-207

ProDy for exploring conformational space

Protein Dynamics Analysis in Python

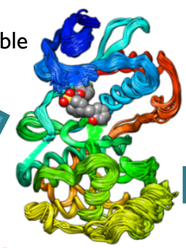
User inputs a protein sequence

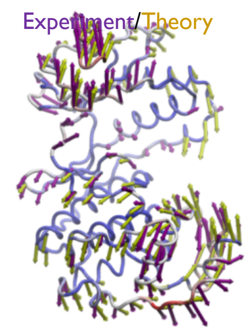
>1A9U:A|PDBID|CHAIN
GSSHHHHHHSSGLVPRGSHMSQERP
TFYRQELNKTIWEVPERYQNLSPVG
SGAYGSVCAAFDTKTGLRVAVKKLS
RPFQSIIHAKRTYRELRLLKHMKHEN
VIGLLDVFT......

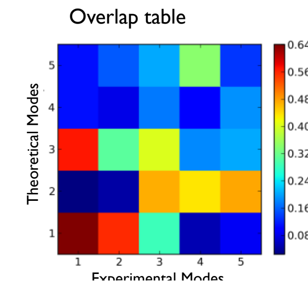


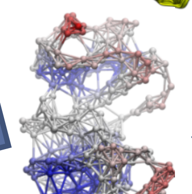
ProDy *identifies*, *retrieves*, *aligns*, and *analyzes (PCA)* structures that match the input sequence

p38 ensemble (PCA)



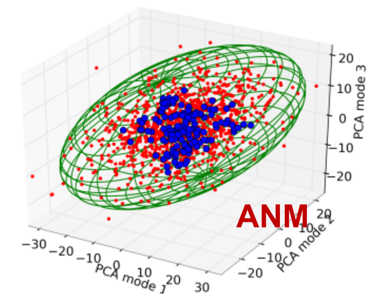






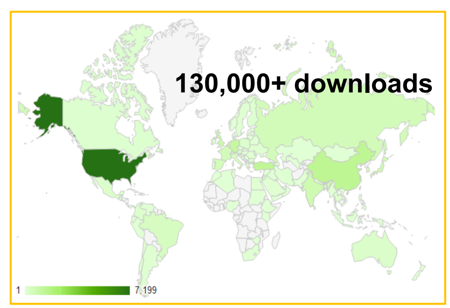
p38 network model (ANM)

User can compare experimental and theoretical models

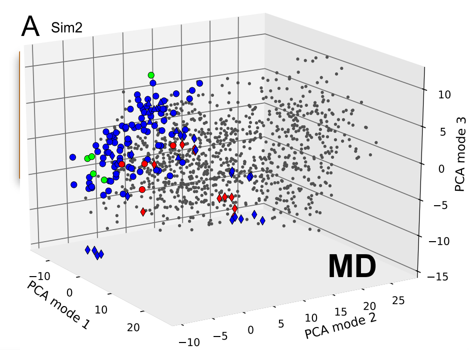


Growth of Source Code and Usage

	Releases	Downloads	Visits[2]	Unique[3]
Nov '10 - Oct '11	19	8,530	8,678	2,946
Nov '11 - Oct '12	15	35,108	16,472	6,414
Nov '12 - Oct '13	8*	87,909	19,888	8,145
Total	42	131,547	45,038	17,505



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



Bakan & Bahar, PSB 2011, 181-192

Acknowledgment

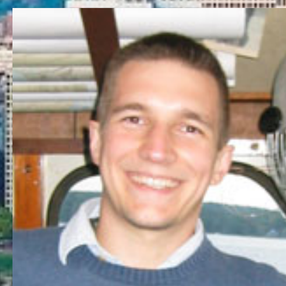
Support from NIGMS, NLM, NIDDK & NIAID

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

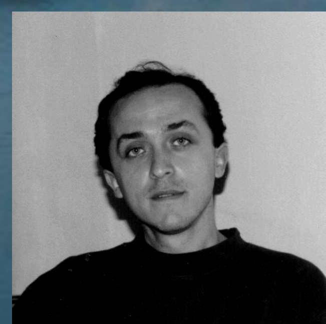


Dr. Tim Lezon
Comp & Systems Biol,
U of Pittsburgh

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

Dr. Indira Shrivastava
Assist Prof. DCSB, Pitt