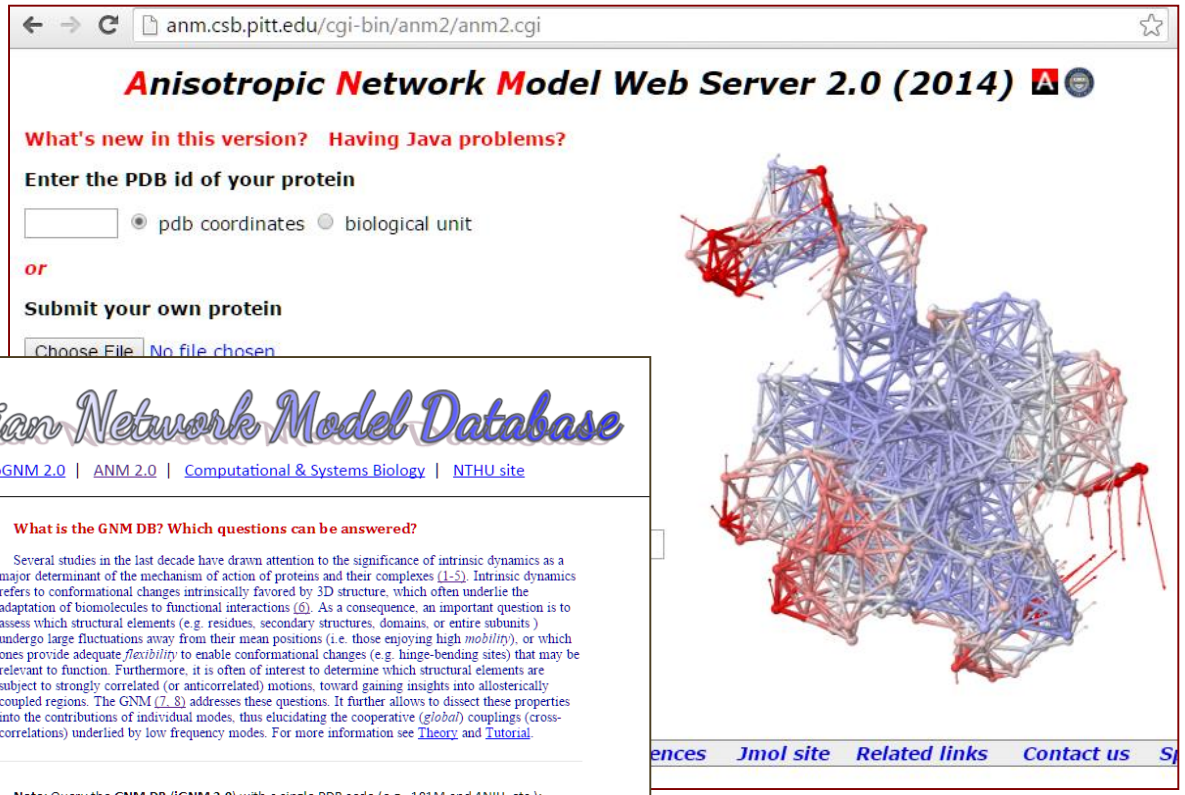
Collective Dynamics of Biomolecules using ProDy & Elastic Network Models

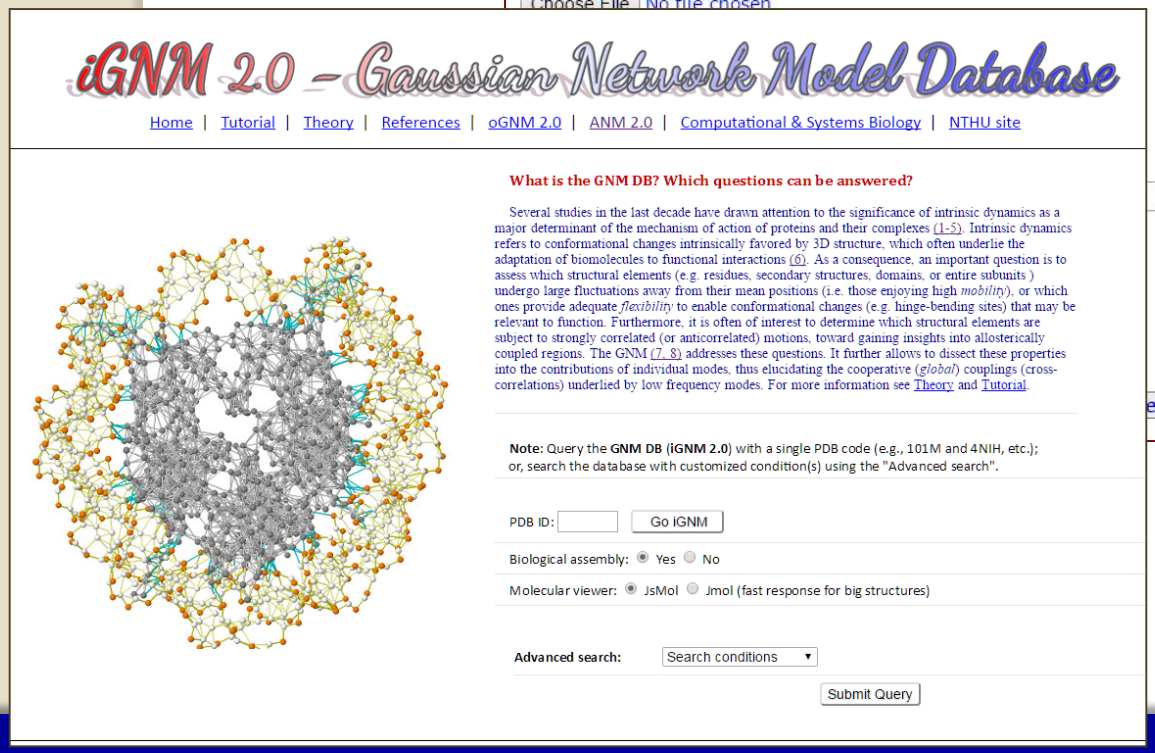
Ivet Bahar

Department of [1]Computational and Systems Biology
School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260

MMBioS Resources

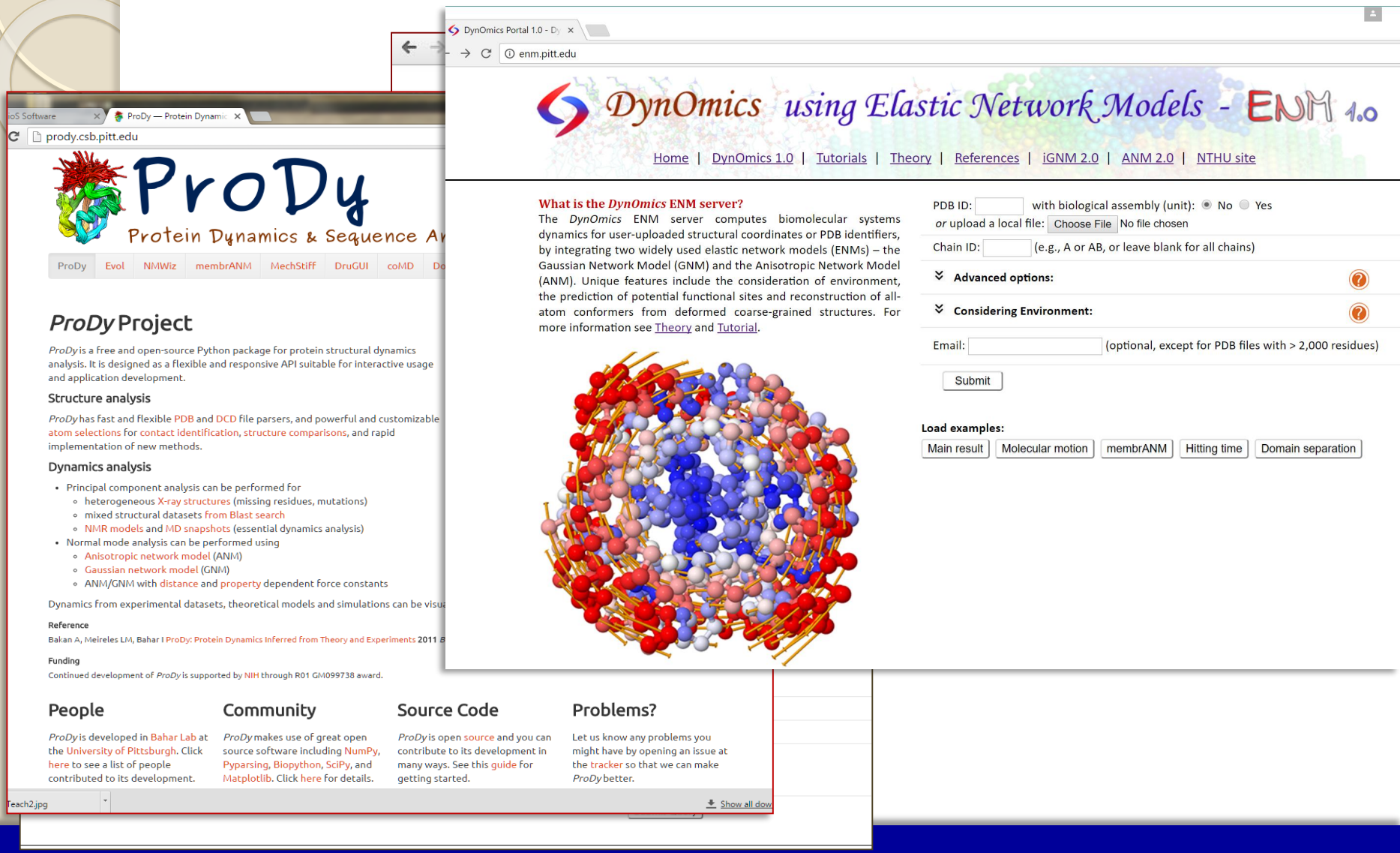


Eyal et al., *Bioinformatics* 2015



Li et al., *Nucleic Acids Res* 2016

MMBioS Resources



Bakan et al., *Bioinformatics* 2011; 2014

Li et al *Nucleic Acids Res* 2016

ProDy

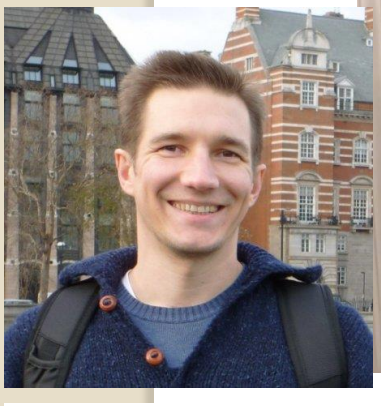
Protein Dynamics Analysis in Python

NMWiz

Cihan Kaya

She (John) Zhang

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 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.

ProDy: Usage and dissemination statistics

Date	Releases	Downloads[1]	Visits[2]	Unique[3]	Pageviews[2]	Countries[5]
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
Total	**53+**	**464,491+**	**148,978**	**68,757**	**639,782**	**136**

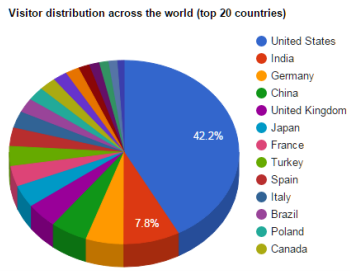
* Indicates software release made during the grant period.

[1] Download statistics retrieved from PyPI (https://pypi.python.org/pypi/ProDy) using (https://pypi.python.org/pypi/vanity).

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

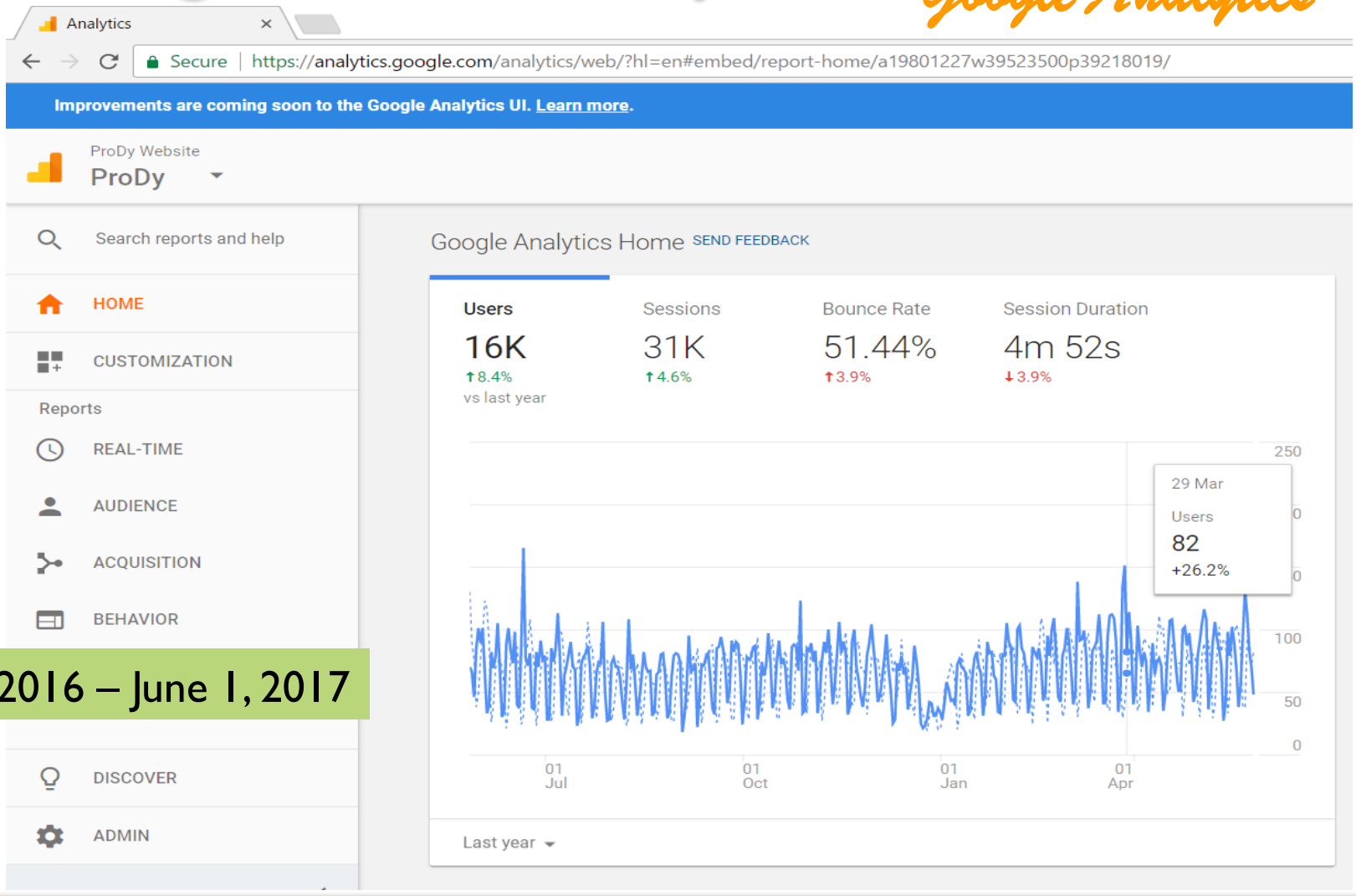
[3] Unique indicates number of unique visitors;

[5] Country of origin for visits.



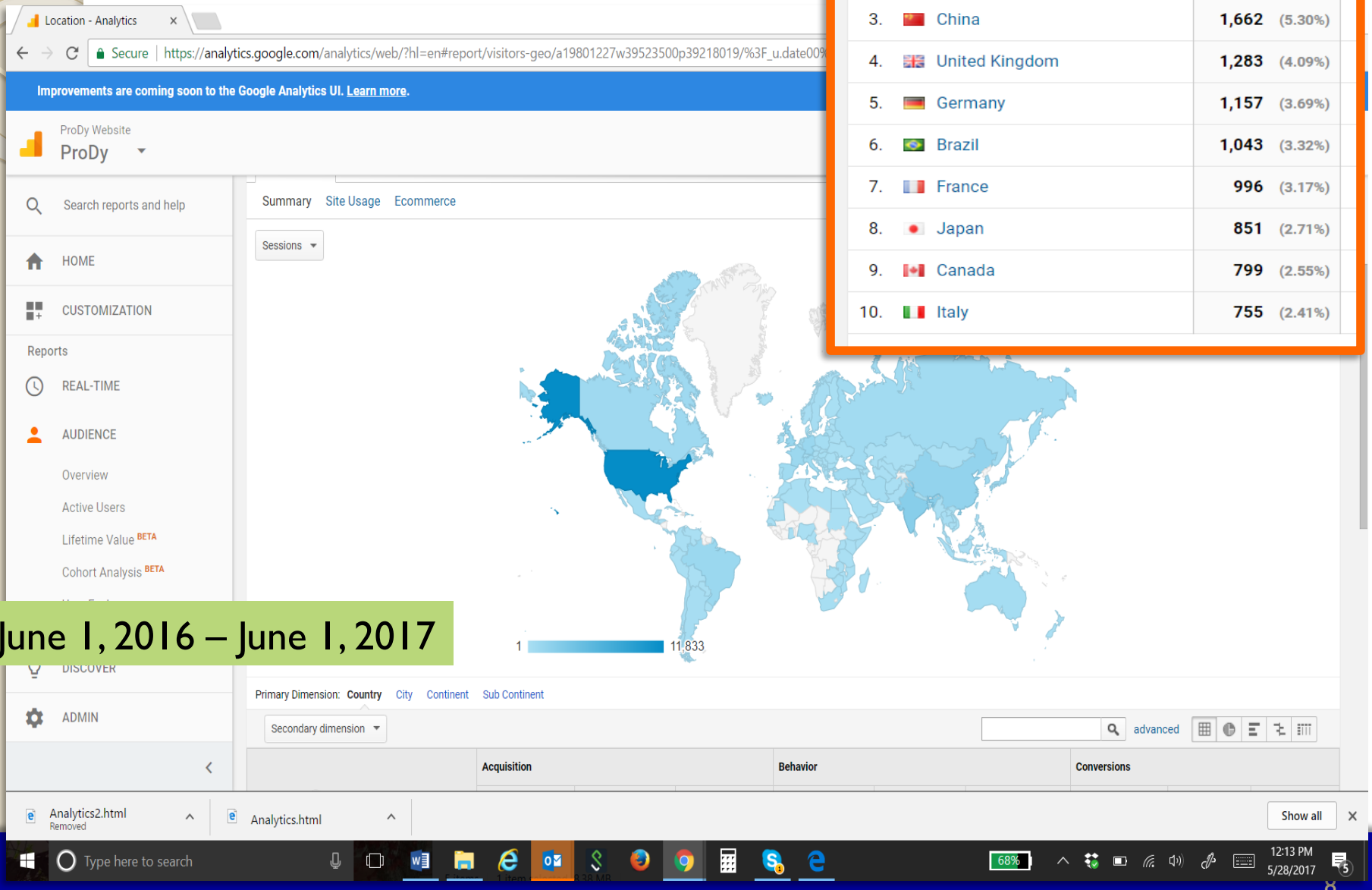
Usage in the last year

Google Analytics



June 1, 2016 – June 1, 2017

Who? Where?



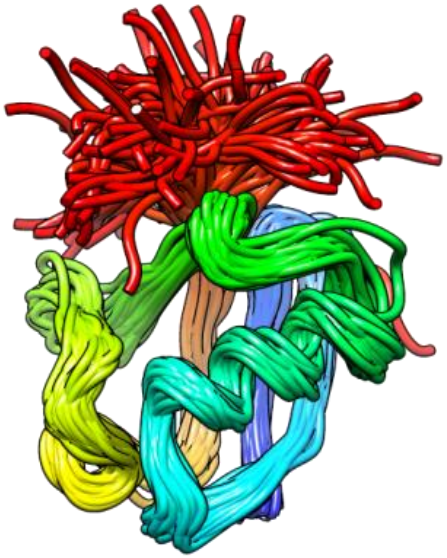
		31,374 % of Total: 100.00% (31,374)
1.	United States	11,833 (37.72%)
2.	India	2,415 (7.70%)
3.	China	1,662 (5.30%)
4.	United Kingdom	1,283 (4.09%)
5.	Germany	1,157 (3.69%)
6.	Brazil	1,043 (3.32%)
7.	France	996 (3.17%)
8.	Japan	851 (2.71%)
9.	Canada	799 (2.55%)
10.	Italy	755 (2.41%)

June 1, 2016 – June 1, 2017

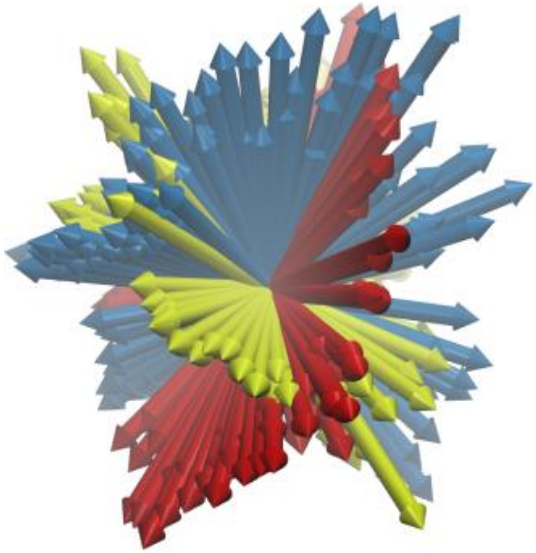
Tutorials

Day 1

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

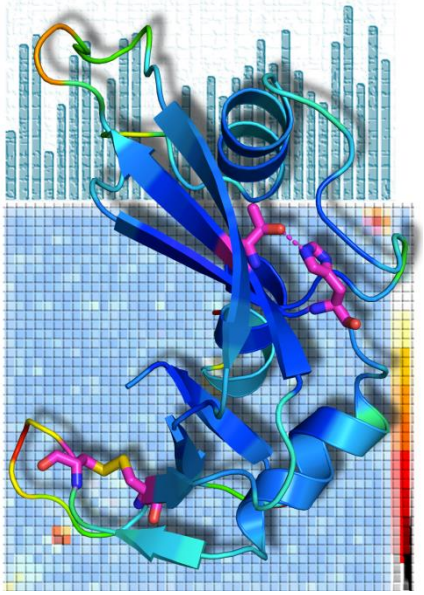


ProDy

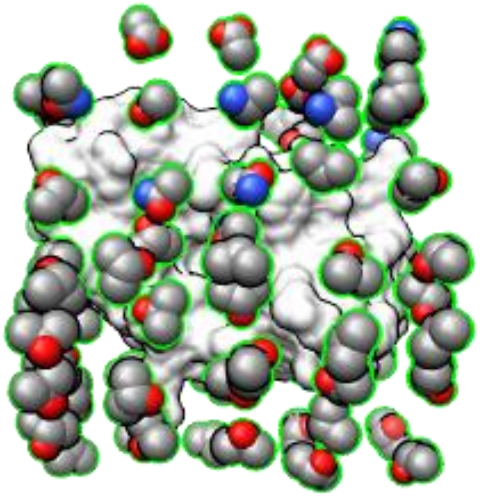


NMWiz

Day 2

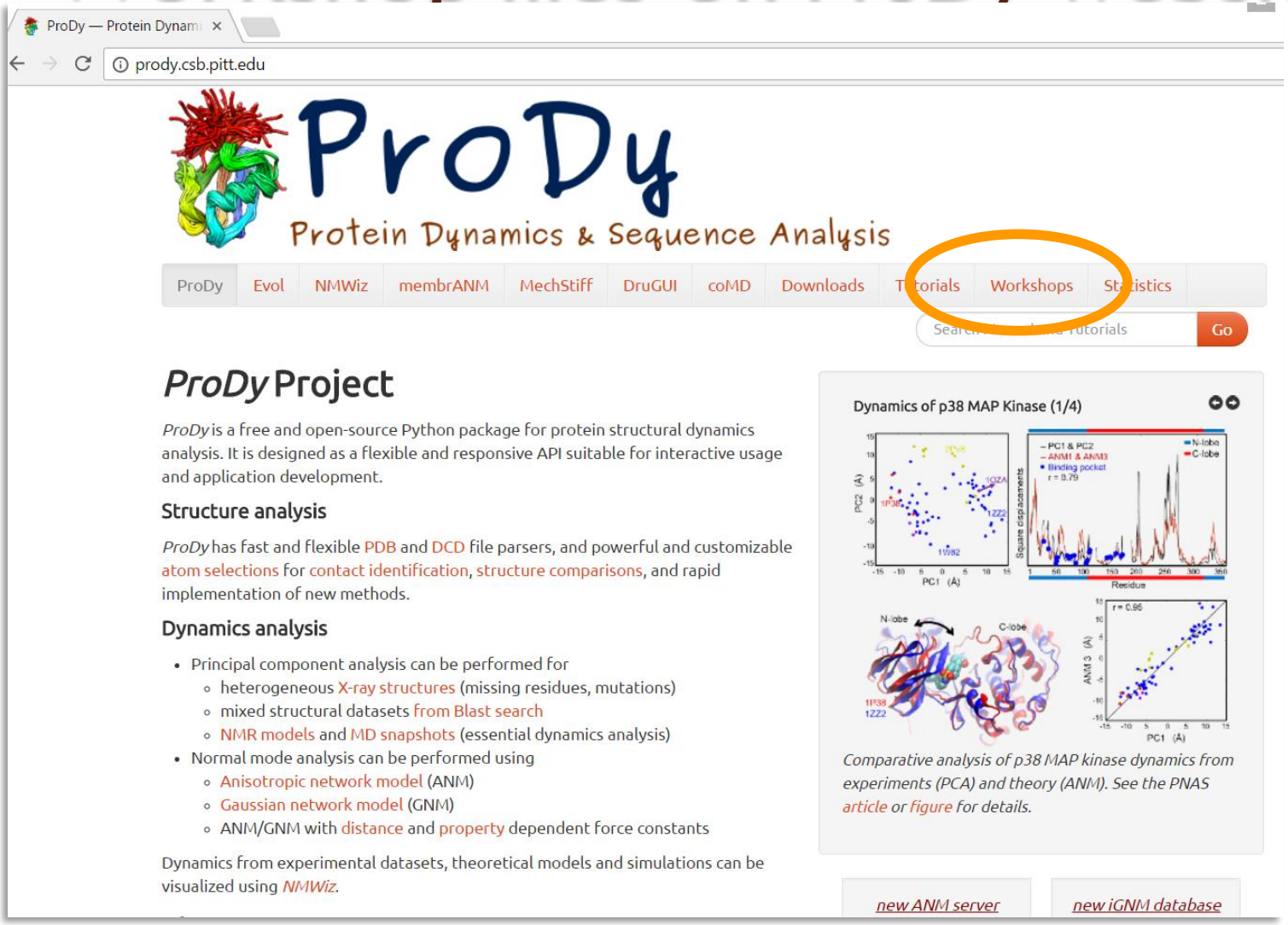


Evol

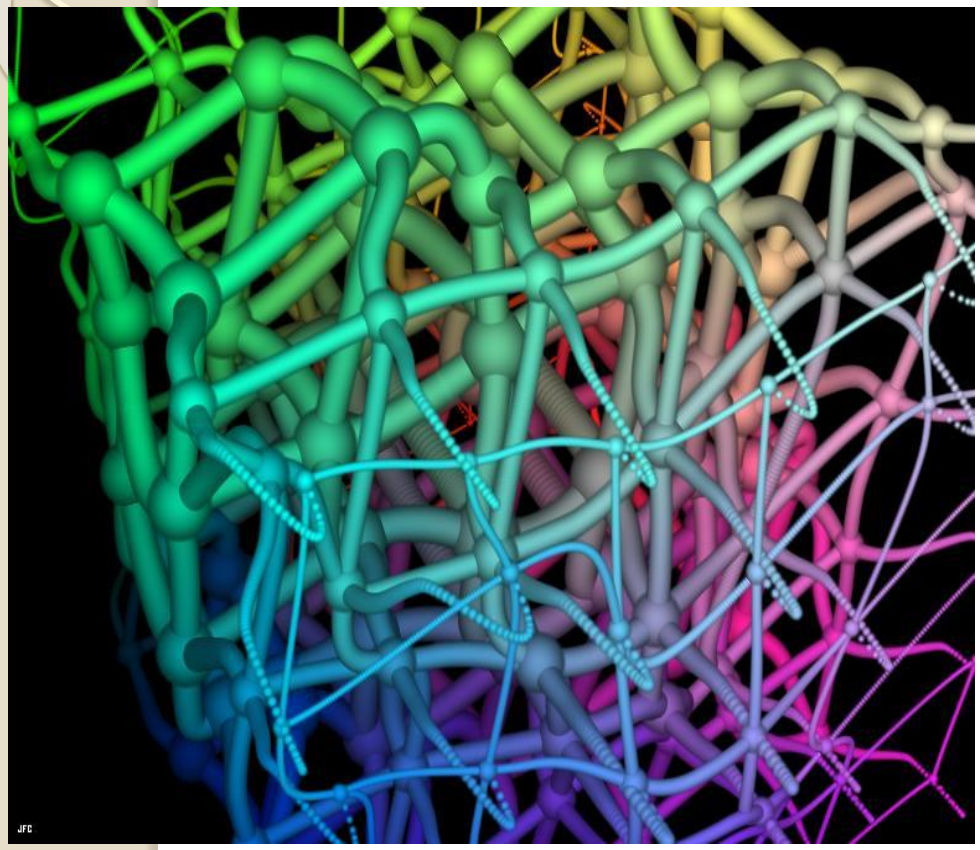


Druggability

Workshop files on *ProDy* website



Representation of structure as a network

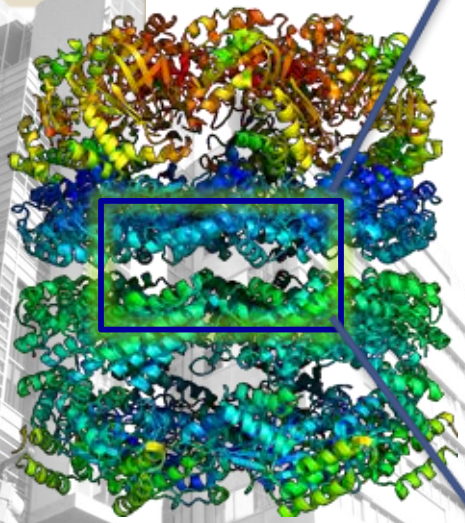


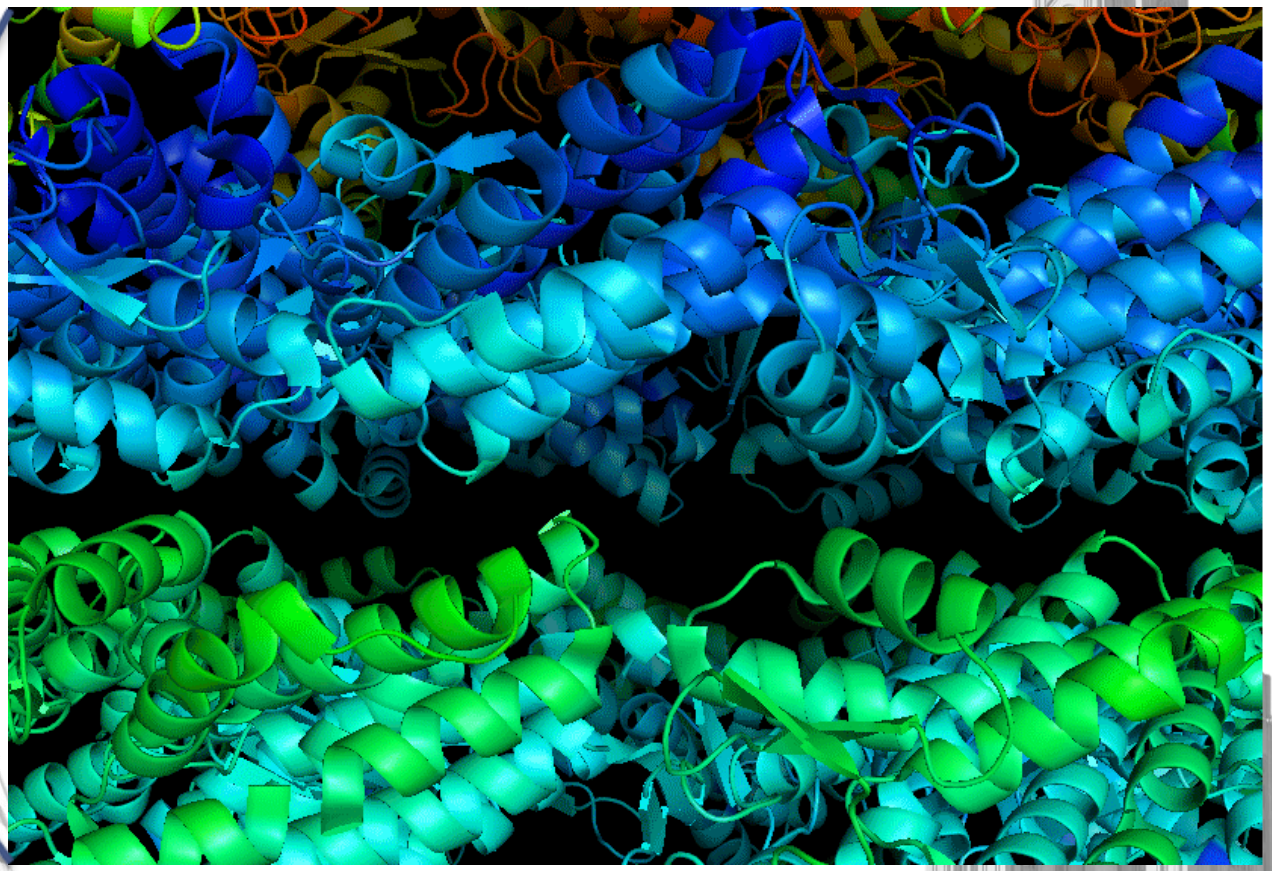
http://www.lactamme.polytechnique.fr/

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.

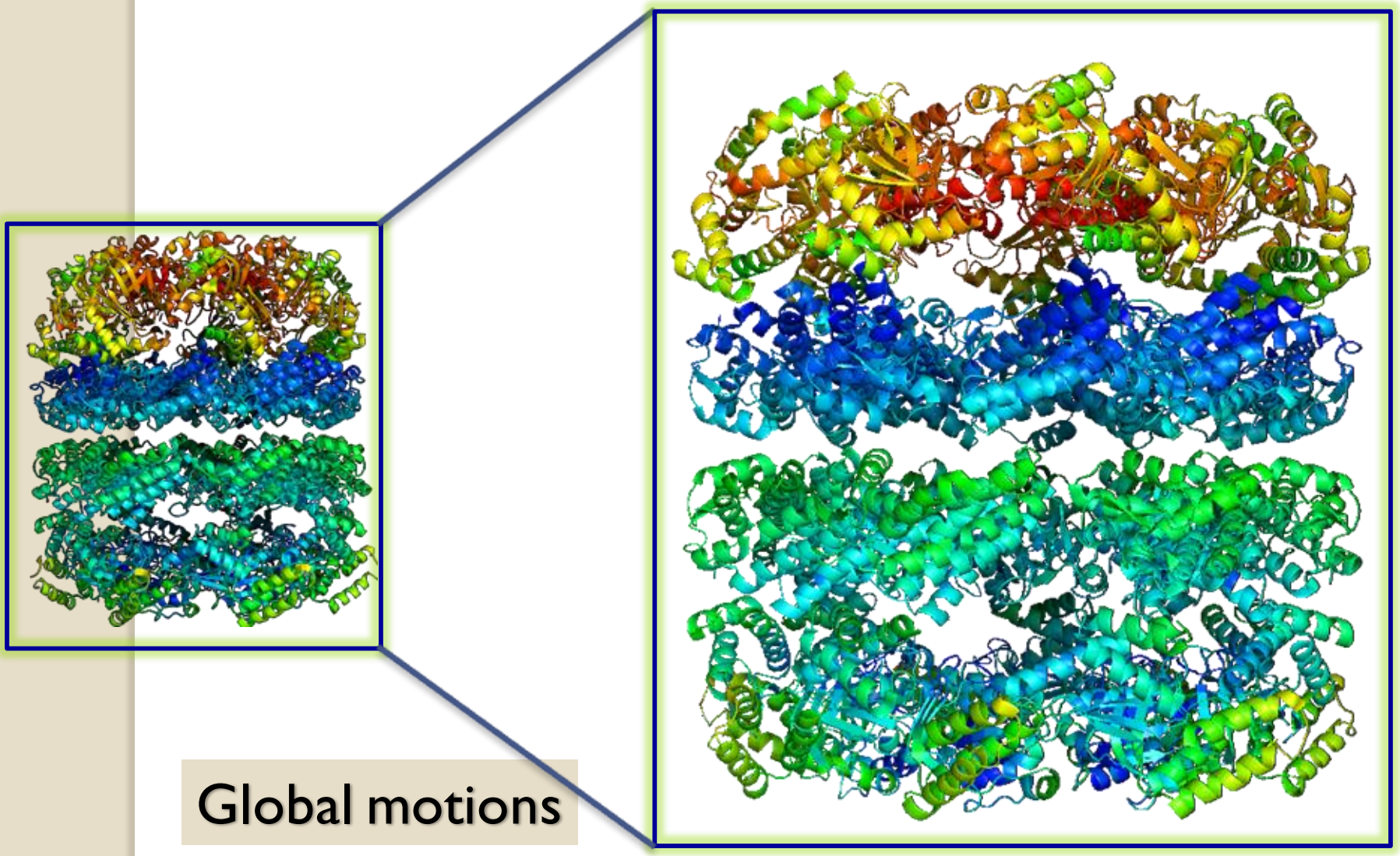
Proteins are not static:
They move, breath, work, dance, interact with each other





Local motions

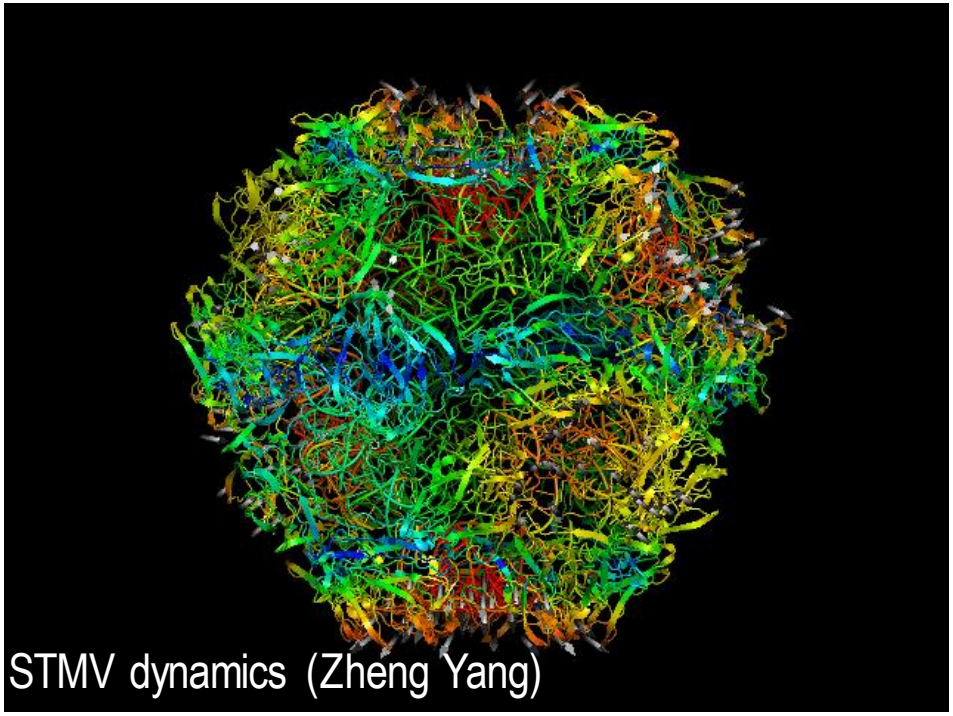
Proteins are not static:
They move, breath, work, dance, interact with each other

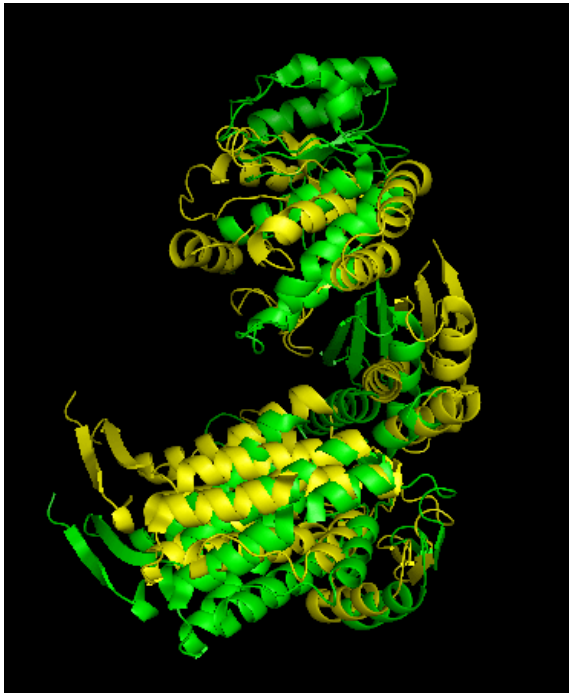


Global motions

Many proteins are molecular machines

And mechanical properties become more important in complexes/assemblies

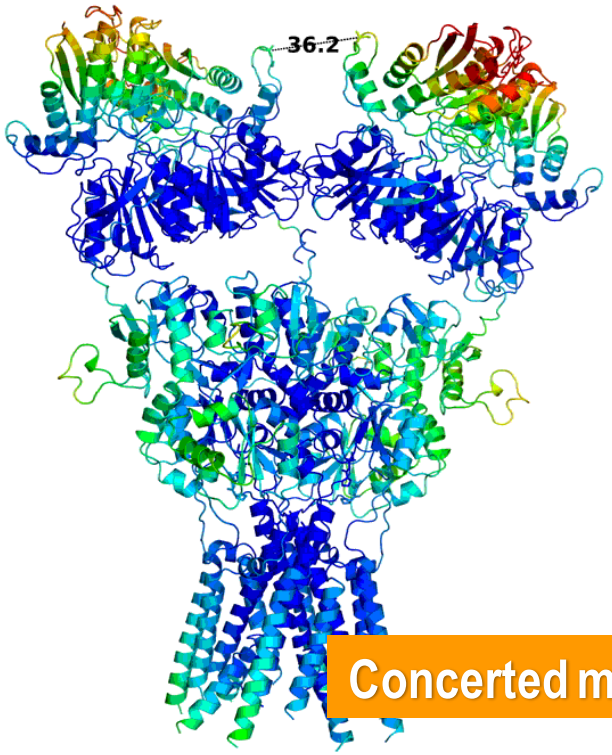


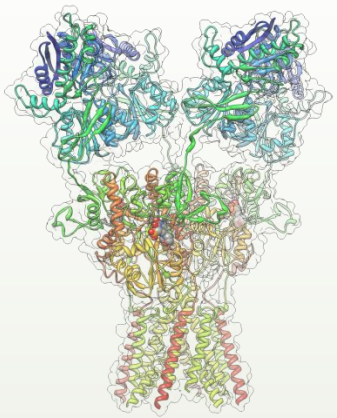


Each structure encodes a unique dynamics

Structure → **Dynamics** → **Function**

Signaling dynamics of AMPARs and NMDARs





domain/subunit motions

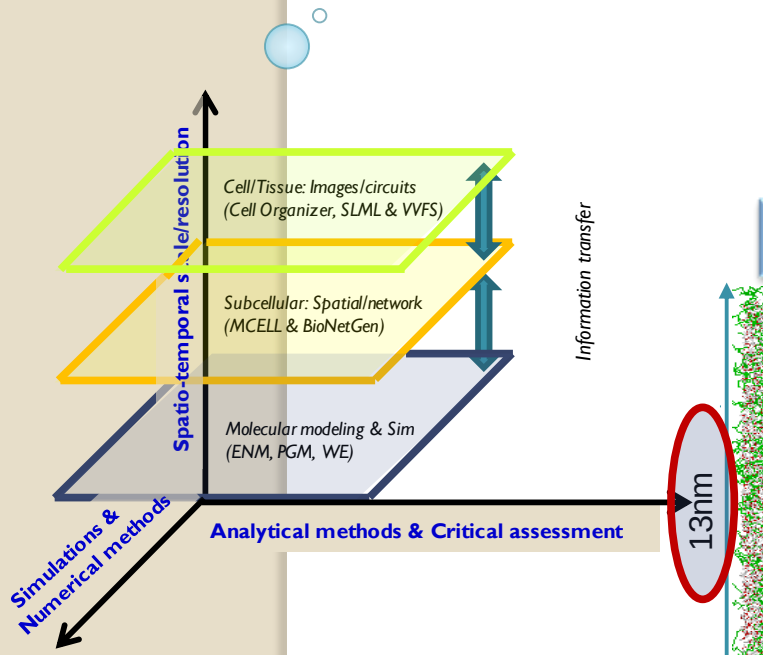
Concerted movements of signaling molecules

Dutta A, Krieger J, Lee JY, Garcia-Nafria J, Greger IH, Bahar I (2015) Cooperative Dynamics of Intact AMPA and NMDA Glutamate Receptors: Similarities and Subfamily-Specific Differences *Structure* **23**: 1692-170

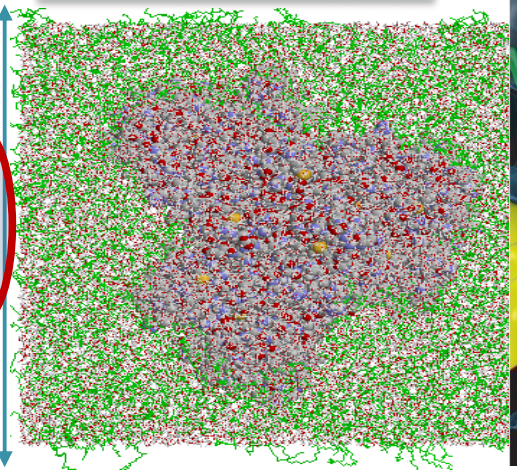
GOAL: TO GENERATE DATA FOR MESOSCOPIC SCALE

Developing integrated *methodology to enable information transfer across scales*

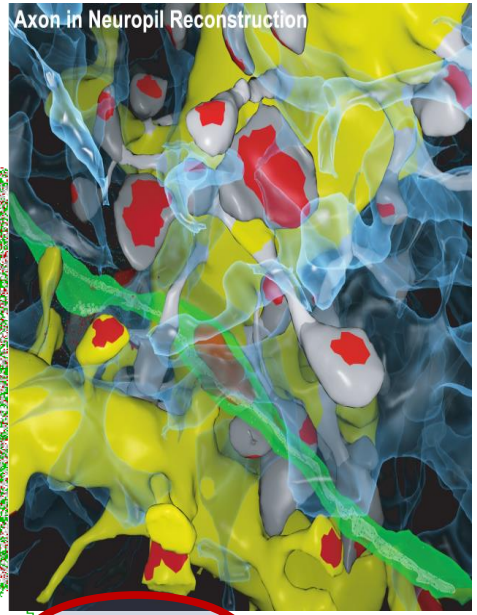
Microphysiological simulations



from molecules



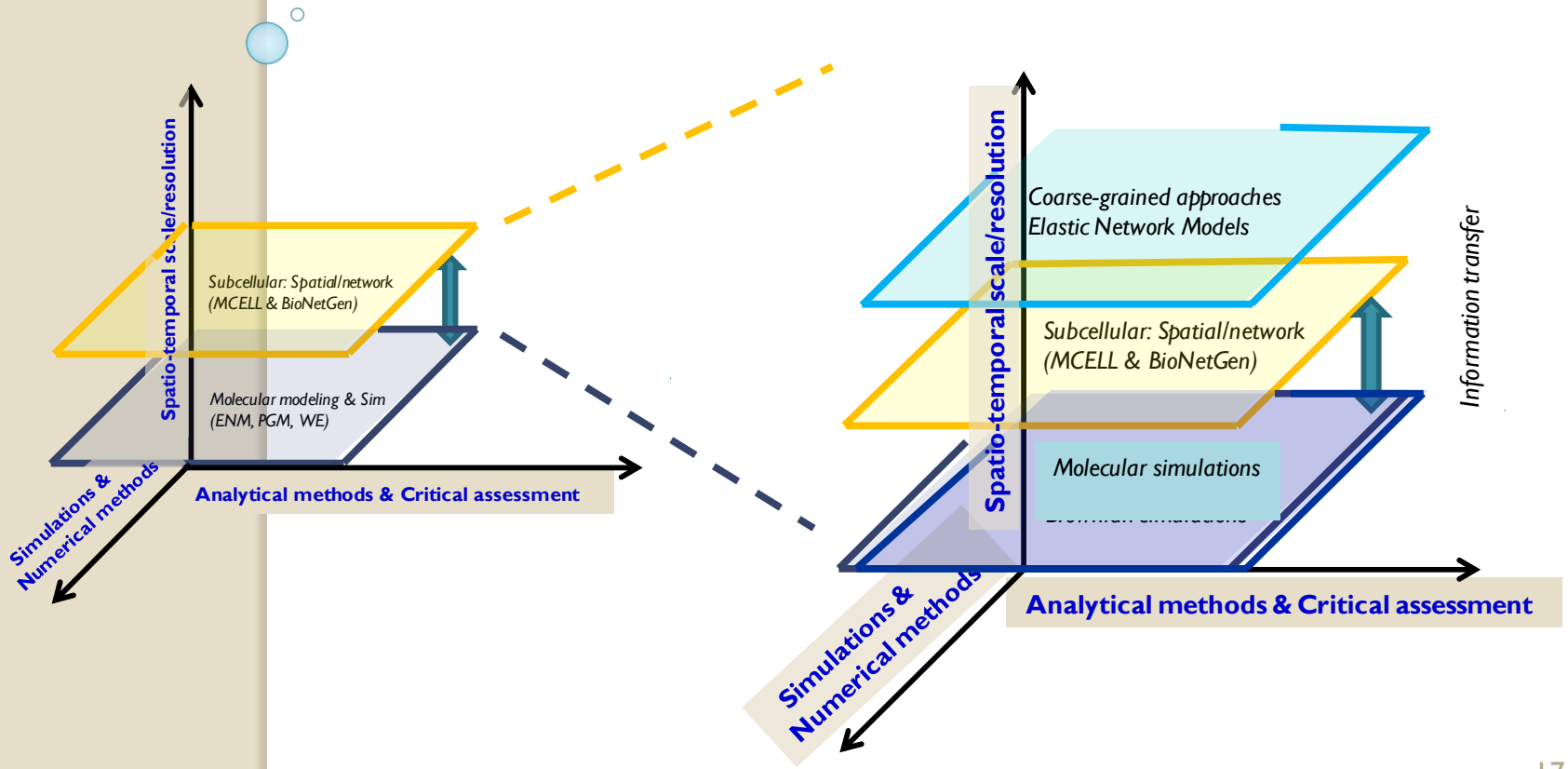
to subcellar events



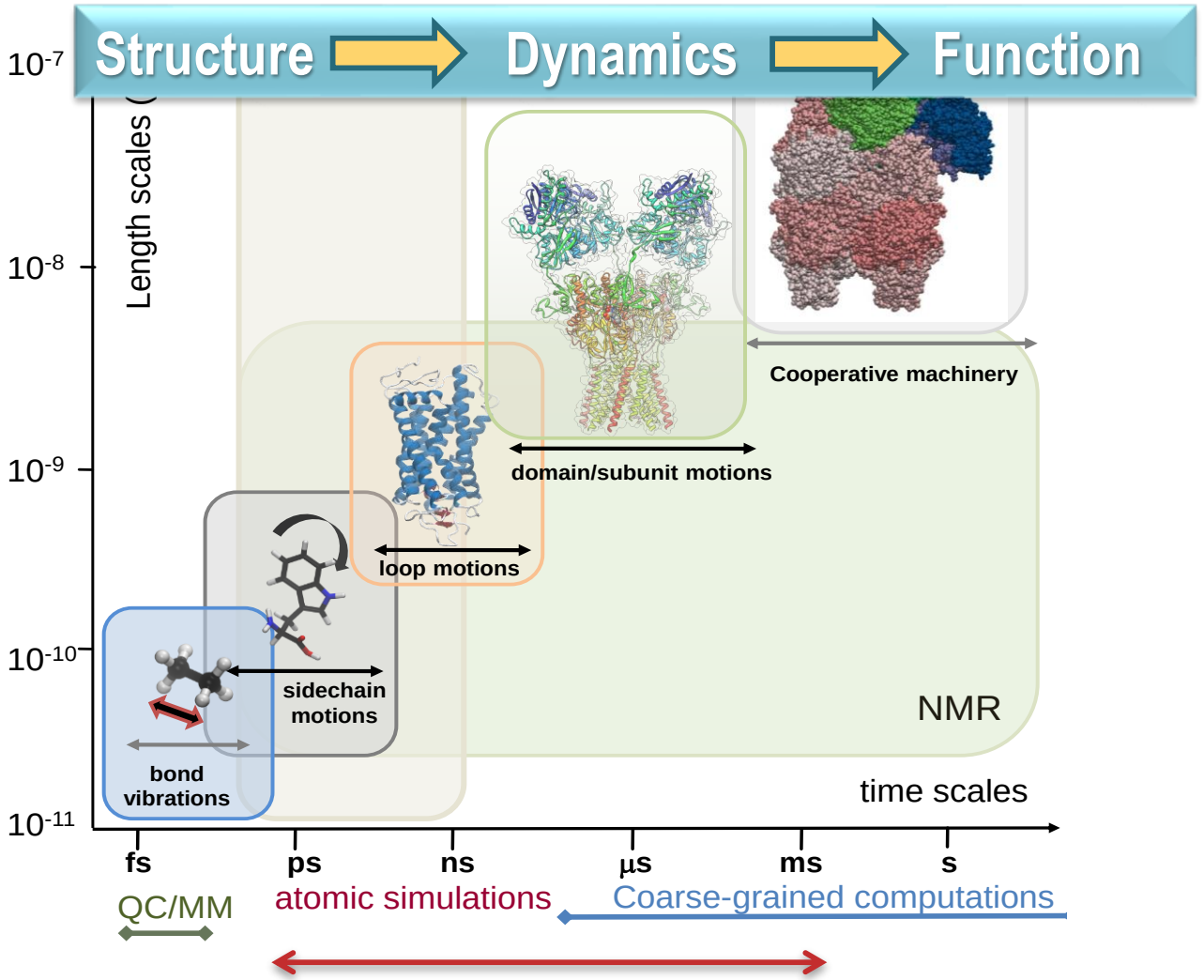
from 6 x 6 x 5 μm^3 sample of adult rat hippocampal stratum radiatum neuropil

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



Summary

1. Theory
   a. Gaussian Network Model (GNM)
   b. Anisotropic Network Model (ANM)
   c. Resources/Servers/Databases (ProDy, DynOmics)
2. Allosteric Changes in Structure
3. Ensemble analysis. Experiments *vs* Predictions
   Adaptability/evolution
4. Recent Extensions and Applications
   a. Membrane Proteins
   b. AMPA Receptor
   c. Chromatin

Two elastic network models:

Gaussian Network Model (GNM)

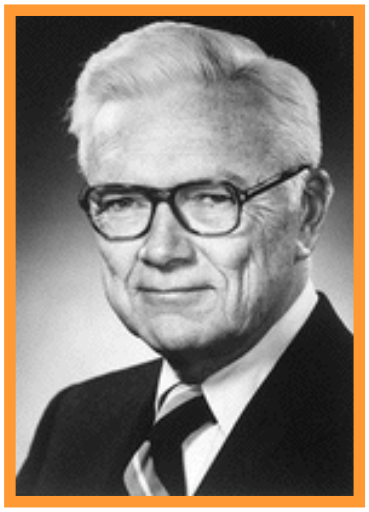
- Li H, Chang YY, Yang LW, Bahar I (2016) iGNM 2.0: the Gaussian network model database for bimolecular structural dynamics *Nucleic Acids Res* **44**: D415-422
- Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in protein *Folding & Design* **2**: 173-181.

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



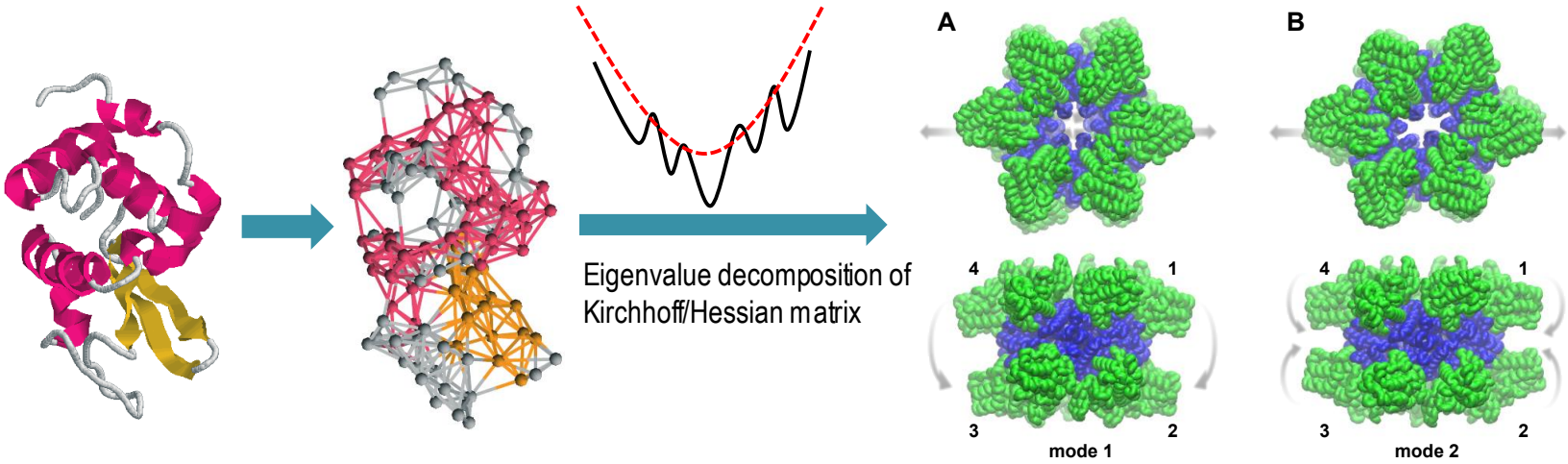
Paul J. Flory (1910-1985)
Nobel Prize in Chemistry 1974

Elastic Network Model for Proteins

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

Collective motions

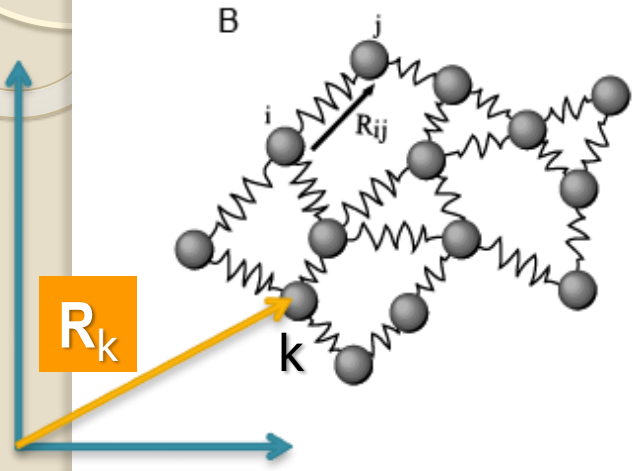
using elastic network models (ENM)

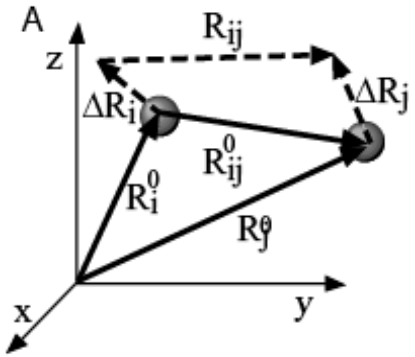


GNM: Bahar et al *Fold & Des* 1996; Haliloglu et al. *Phys Rev Lett* 1997
ANM: Doruker et al. *Proteins* 2000; Atilgan et al, *Biophys J* 2001

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

→ Nodes are subject to **Gaussian fluctuations** ΔR_i

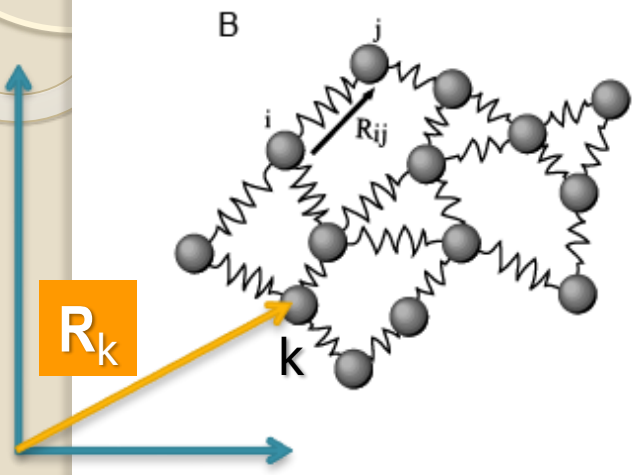
→ Inter-residue distances R_{ij} also undergo Gaussian fluctuations

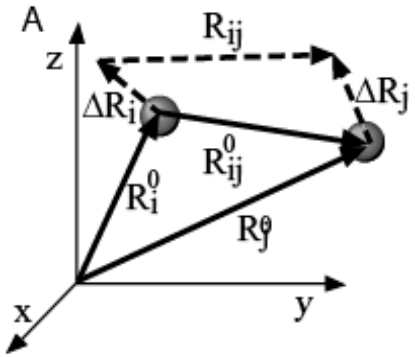
→ $\Delta \mathbf{R}_{ij} = \Delta \mathbf{R}_j - \Delta \mathbf{R}_i$

Fluctuations in residue positions

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

Gaussian Network Model (GNM)

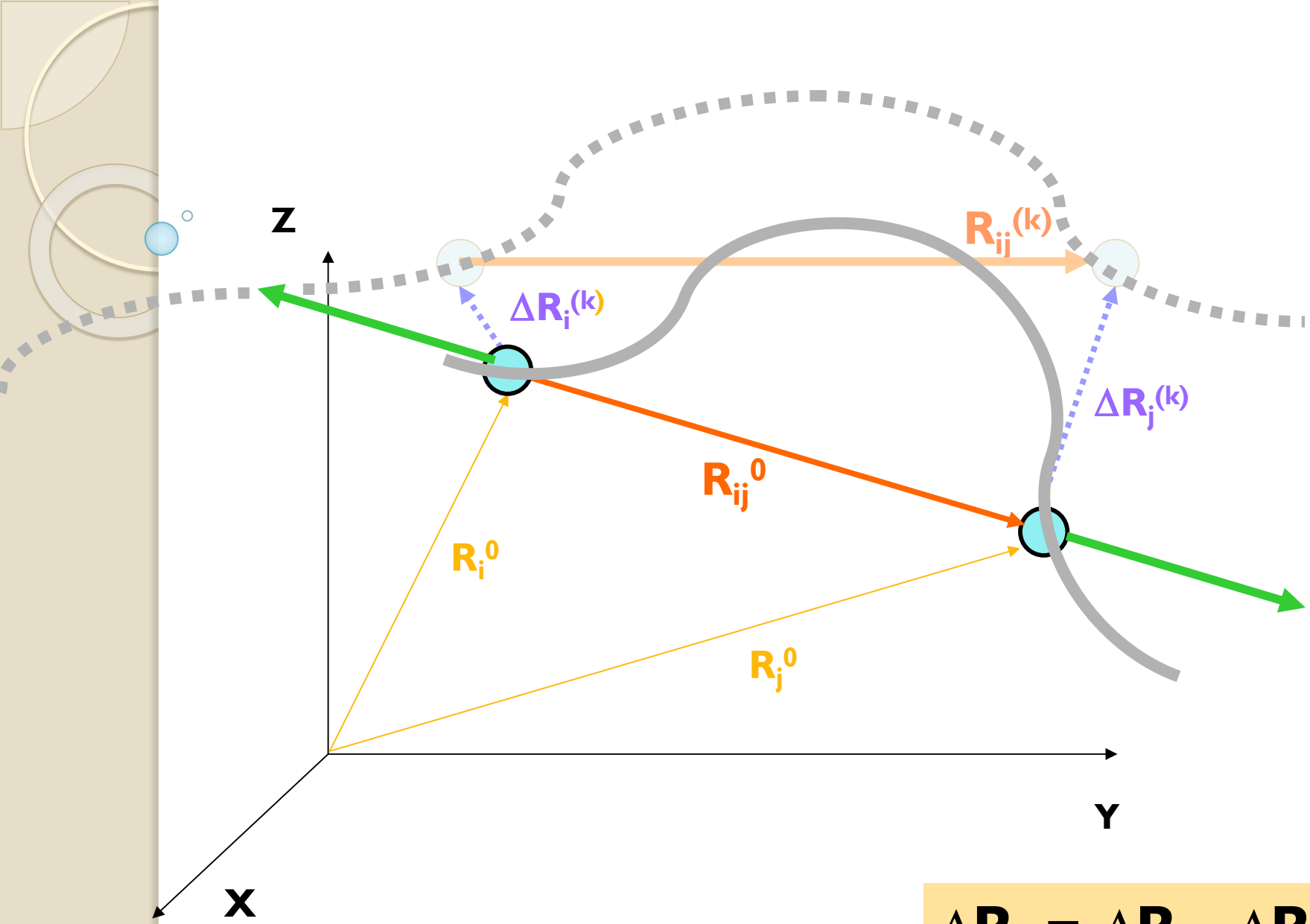




Fluctuation vector:

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

Fluctuations in residue positions



$$\Delta R_{ij} = \Delta R_j - \Delta R_i$$

Fluctuation

with respect to starting structure R(0)

Instantaneous deviation for atom i

$$\Delta \mathbf{R}_i(t_k) = \mathbf{R}_i(t_k) - \mathbf{R}_i(0)$$

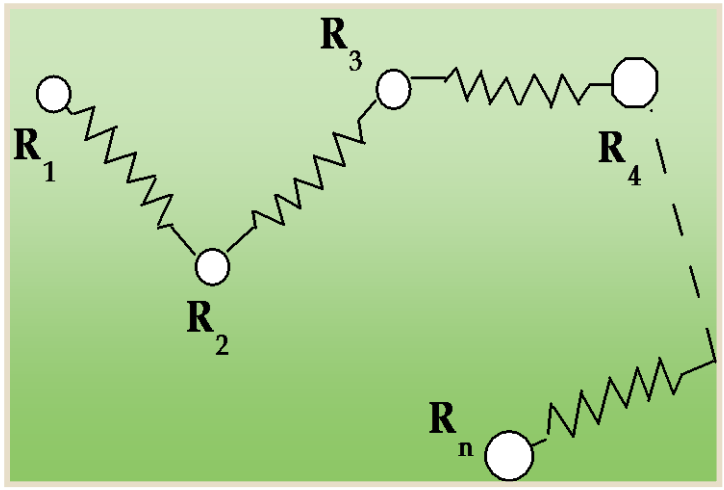
Under equilibrium conditions:

Average displacement from equilibrium: $< \Delta \mathbf{R}_i(t_k) > = 0$

But the mean-square fluctuation (MSF), $< (\Delta \mathbf{R}_i(t_k))^2 > \neq 0$

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

$$\Delta R_{12} = R_{12} - R_{12}^{0}$$

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

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

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

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

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

Rouse model for polymers

Fluctuation vector

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

Kirchhoff matrix

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

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

Force constant

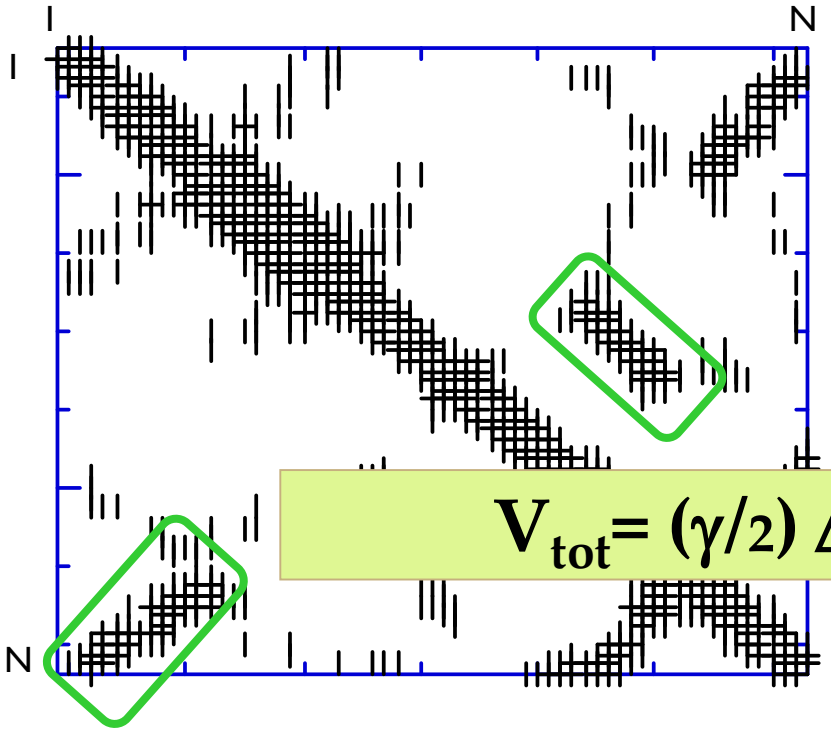
$$V_{tot} = (\gamma/2)\ [\ (\Delta R_{12})^2 + (\Delta R_{23})^2 + ... (\Delta R_{N-1,N})^2\ ]$$

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

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

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

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

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

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

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,

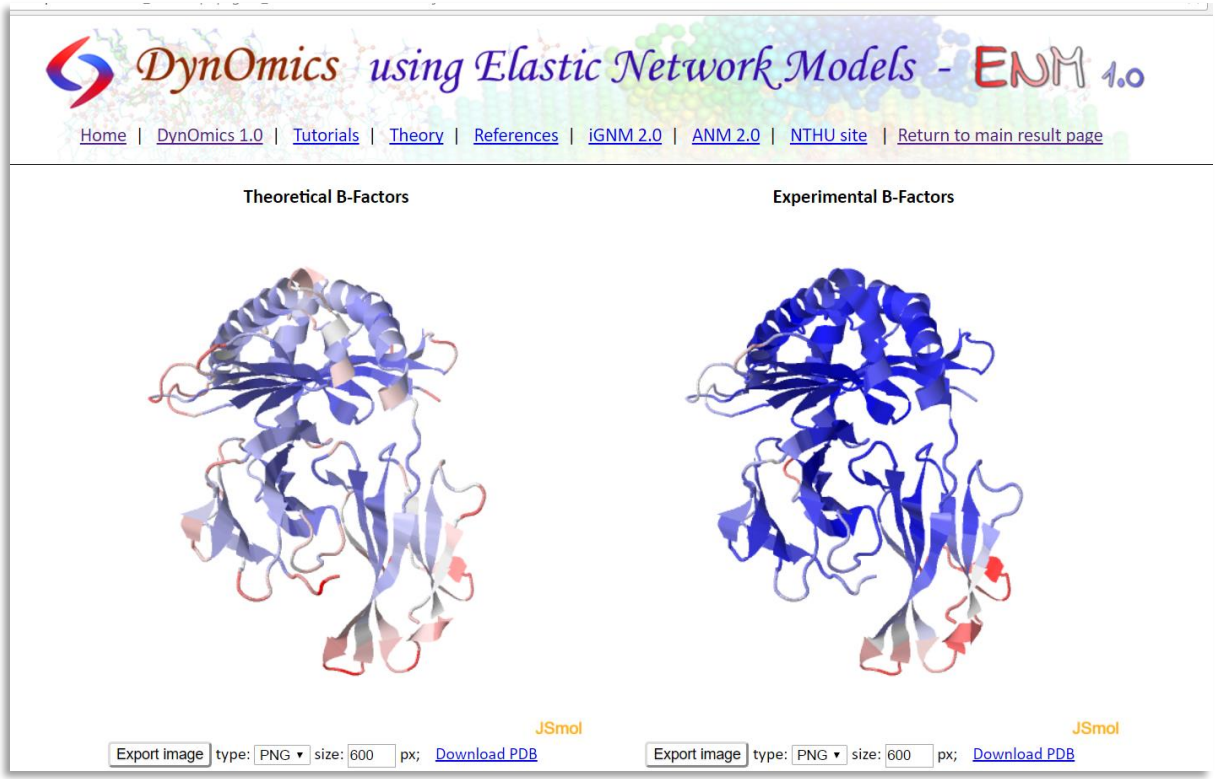
$$B_i = [8\pi^2/3] \langle(\Delta \mathbf{R}_i)^2\rangle$$

How do residue MSFs compare with the B-factors?

Output from DynOmics

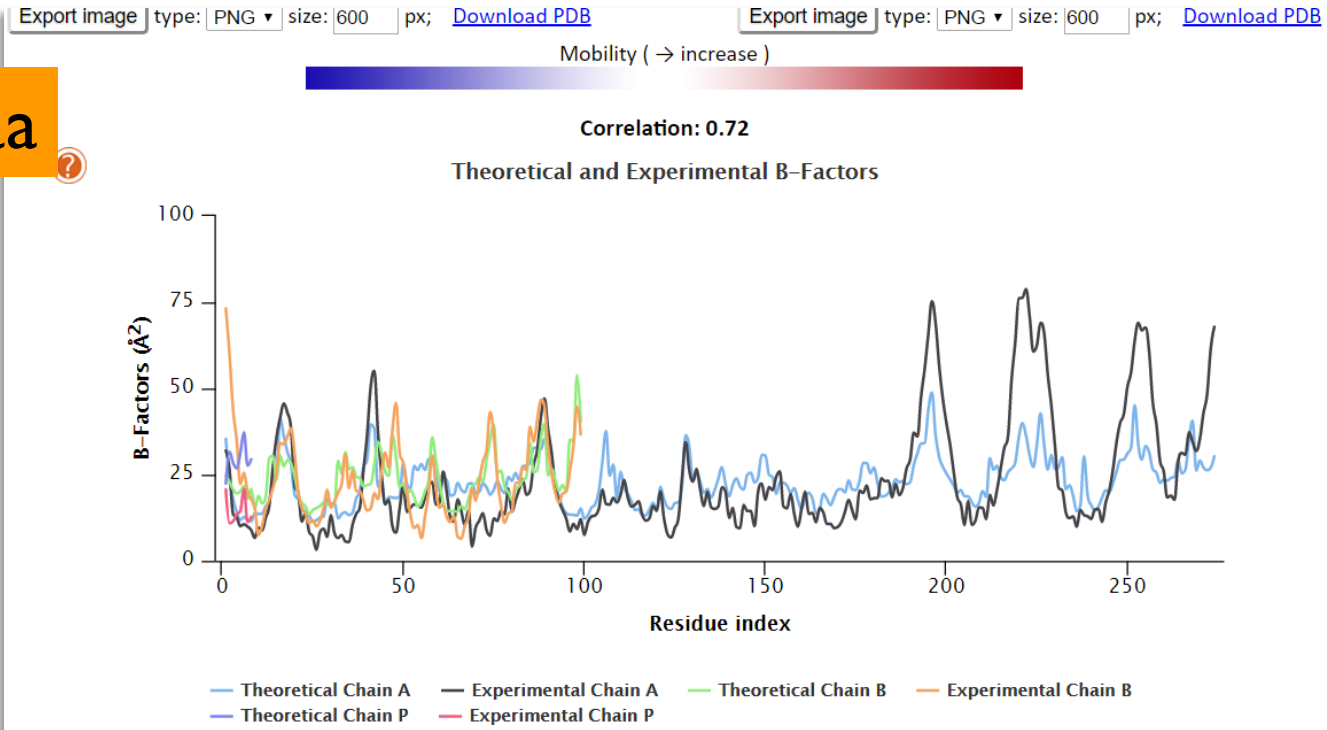
Example: 1vaa

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

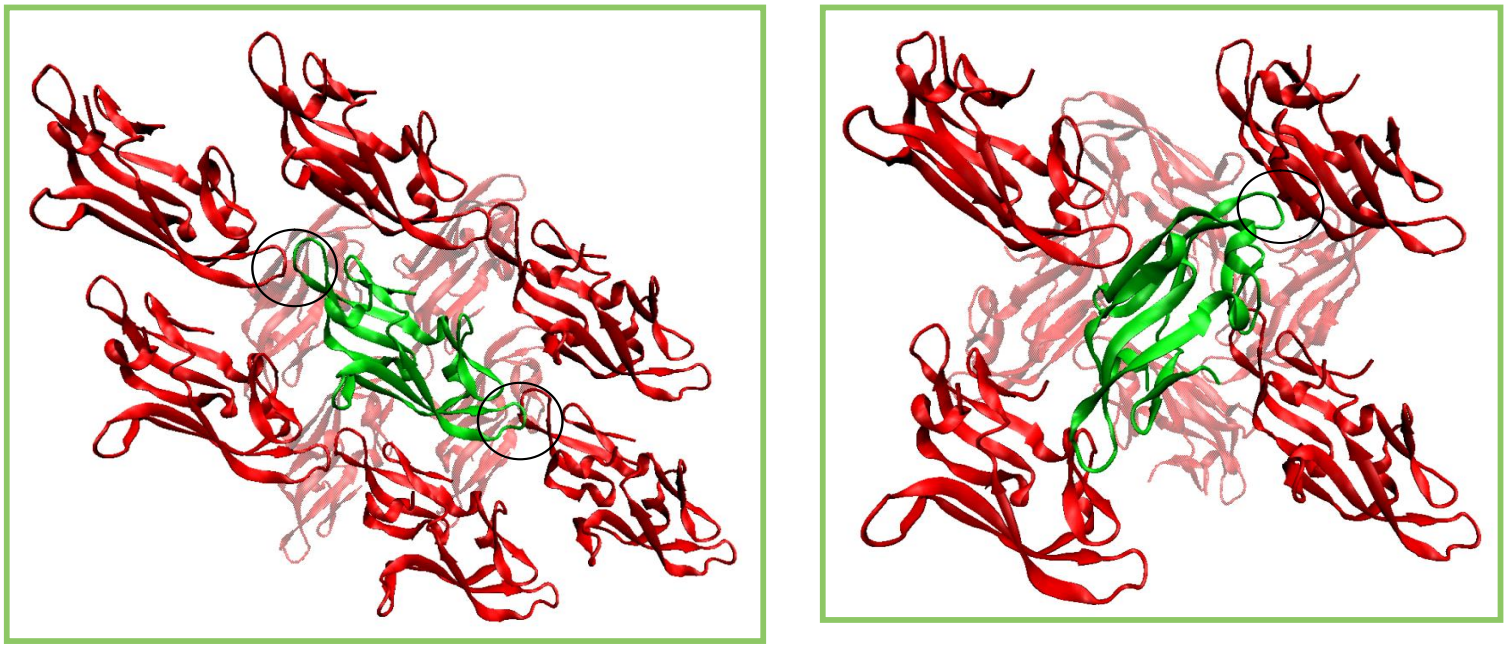


Output from DynOmics

1vaa



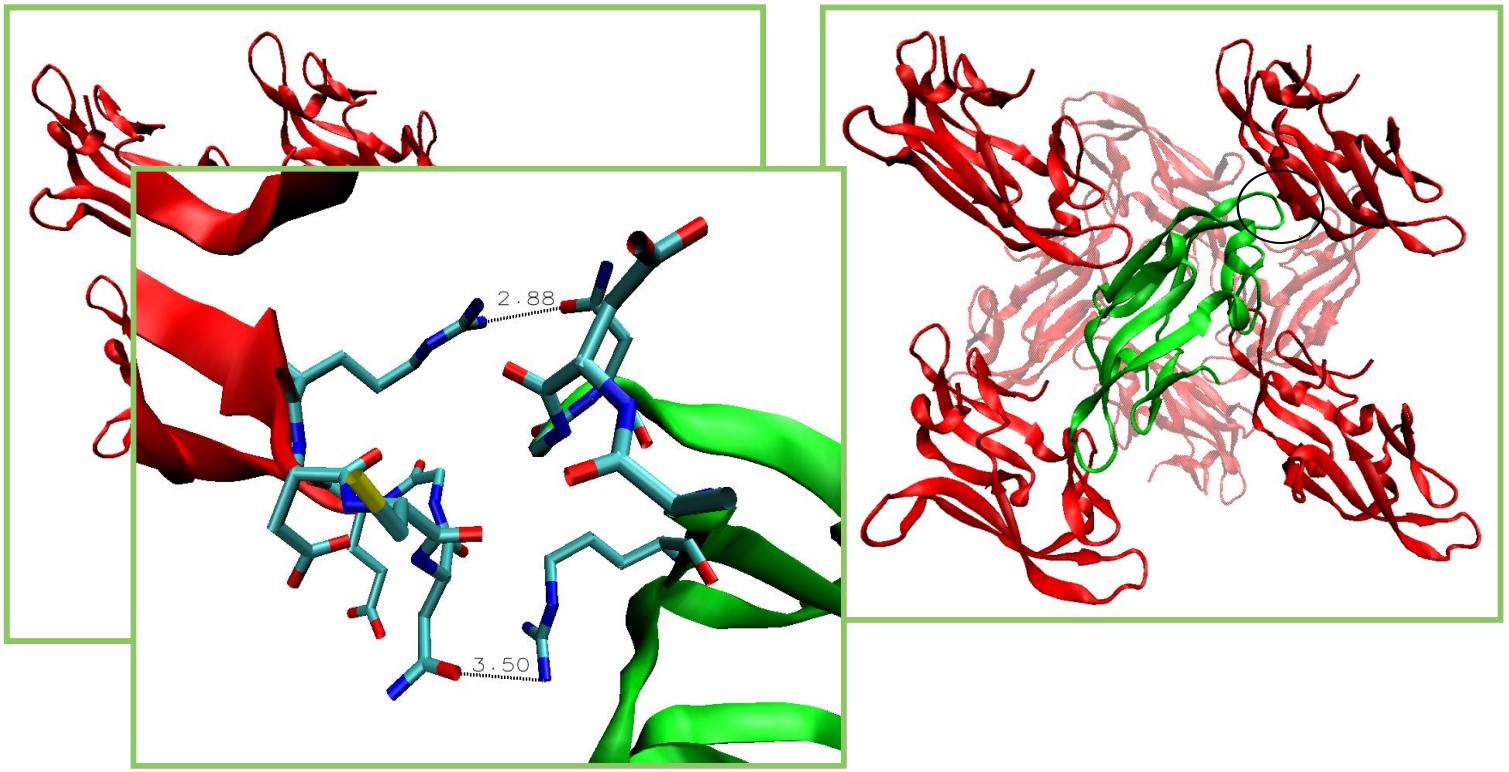
B-factors are affected by crystal contacts



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

1

B-factors are affected by crystal contacts



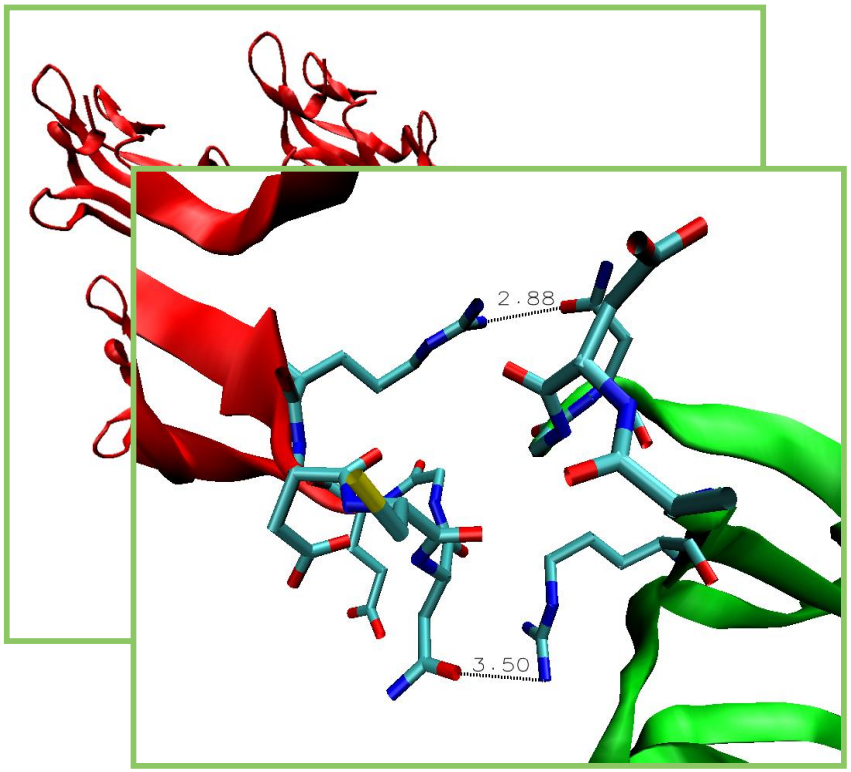
Particular loop motions are curtailed by intermolecular contacts in the crystal environment causing a discrepancy between theory and experiments

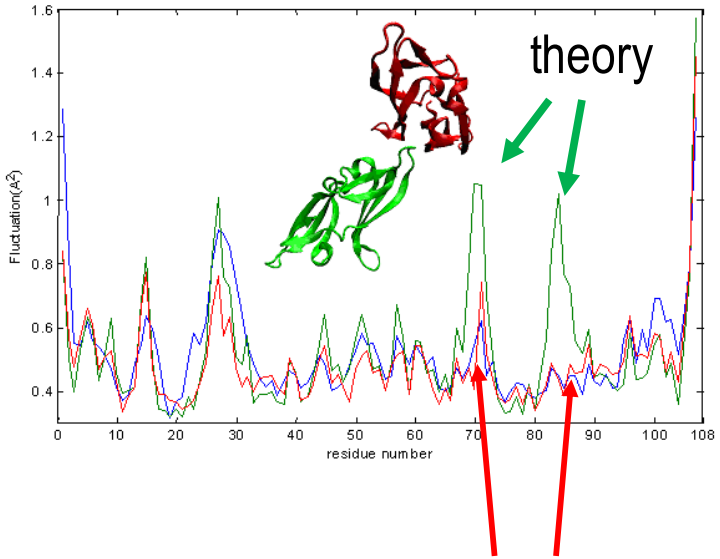
FOR MORE INFO...

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

1

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



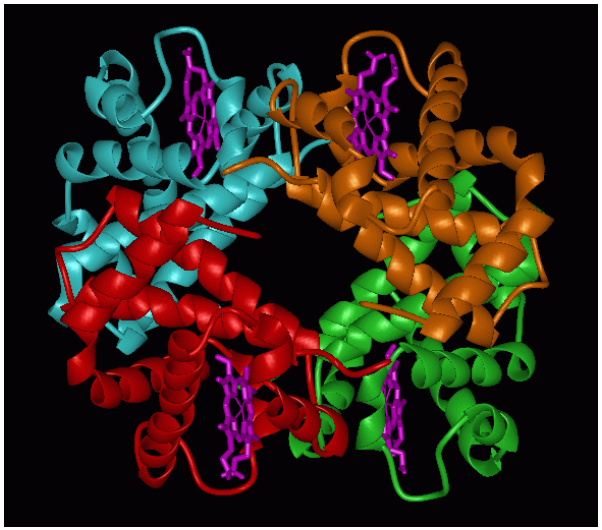


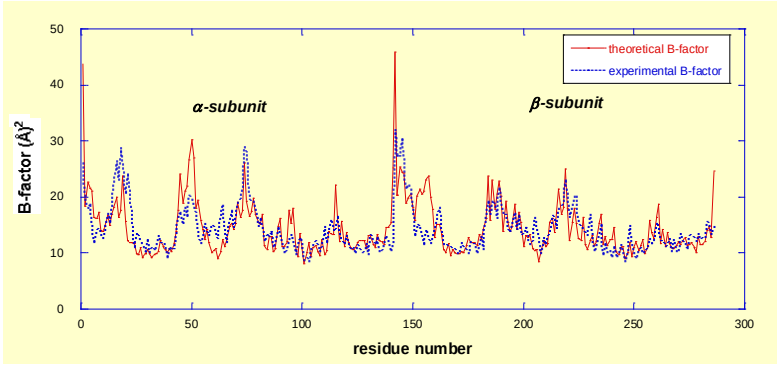
Particular loop motions are curtailed by intermolecular contacts in the crystal environment causing a discrepancy between theory and experiments

FOR MORE INFO...

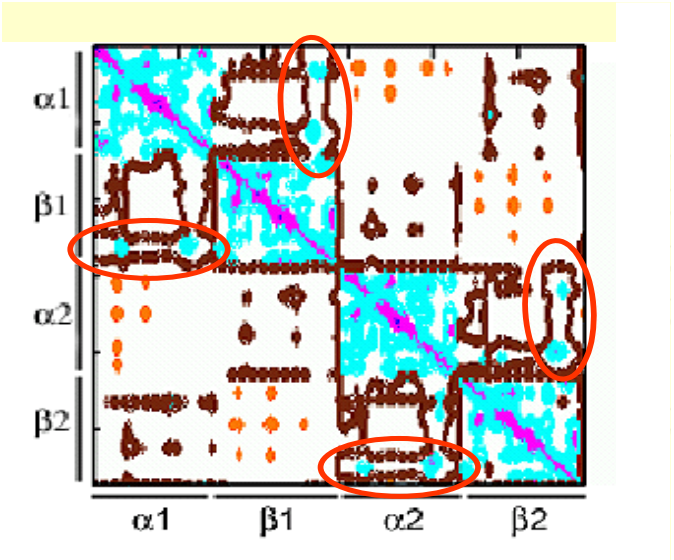
Liu, Koharudin, Gronenborn & Bahar (2009) *Proteins 77, 927-939.*

Application to hemoglobin





B- factors – Comparison with experiments



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

Cross-correlations

- Provide information on the relative movements of pairs of residues

- Purely orientational correlations (correlation cosines) are obtained by normalizing cross-correlations as

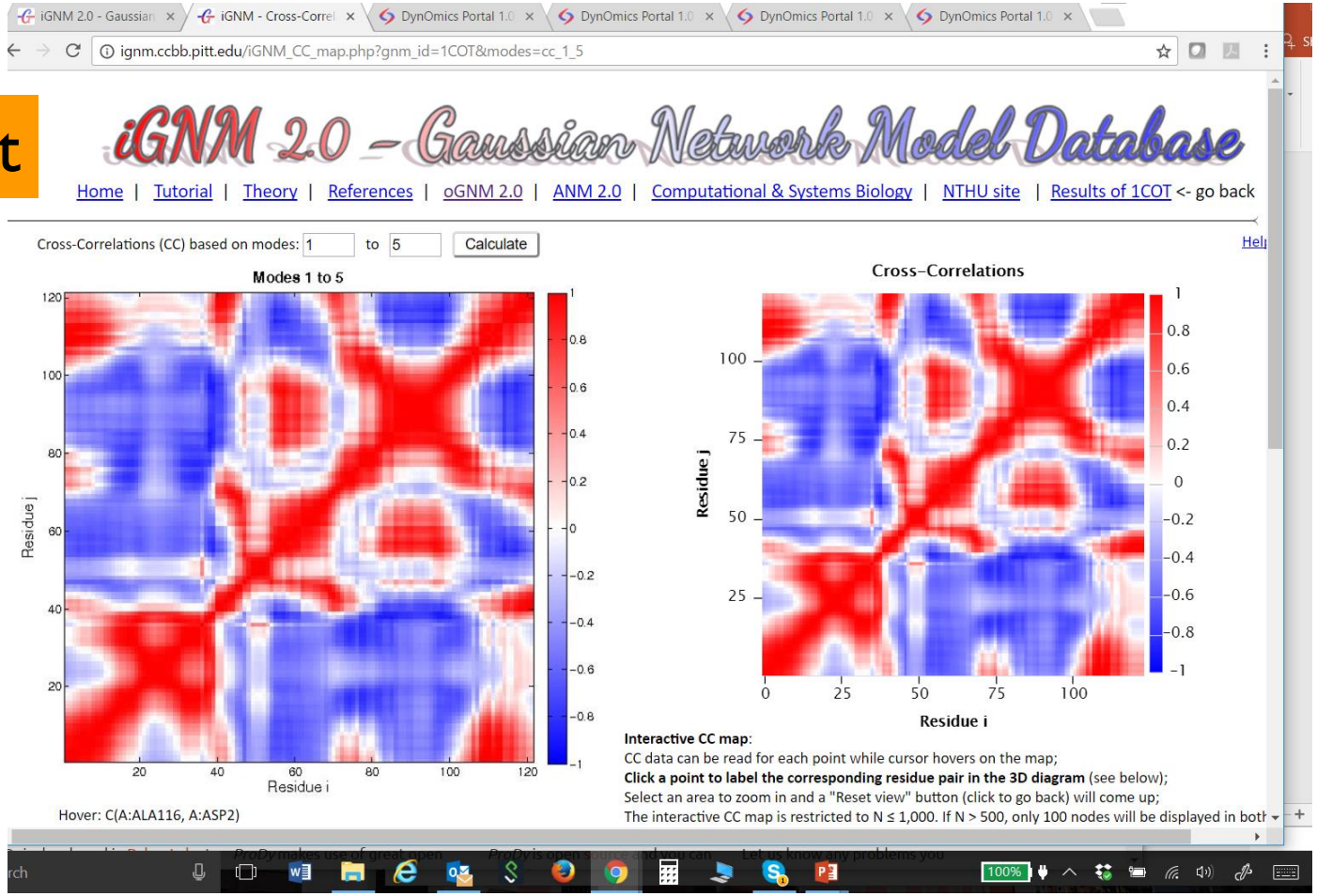
$$-1 \leq \frac{\langle(\Delta\mathbf{R}_i \,.\, \Delta\mathbf{R}_j)\rangle}{[\langle(\Delta\mathbf{R}_i)^2\rangle \, \langle(\Delta\mathbf{R}_i)^2\rangle]^{1/2}} \leq 1$$

Fully anticorrelated

Fully correlated

Output from iGNM

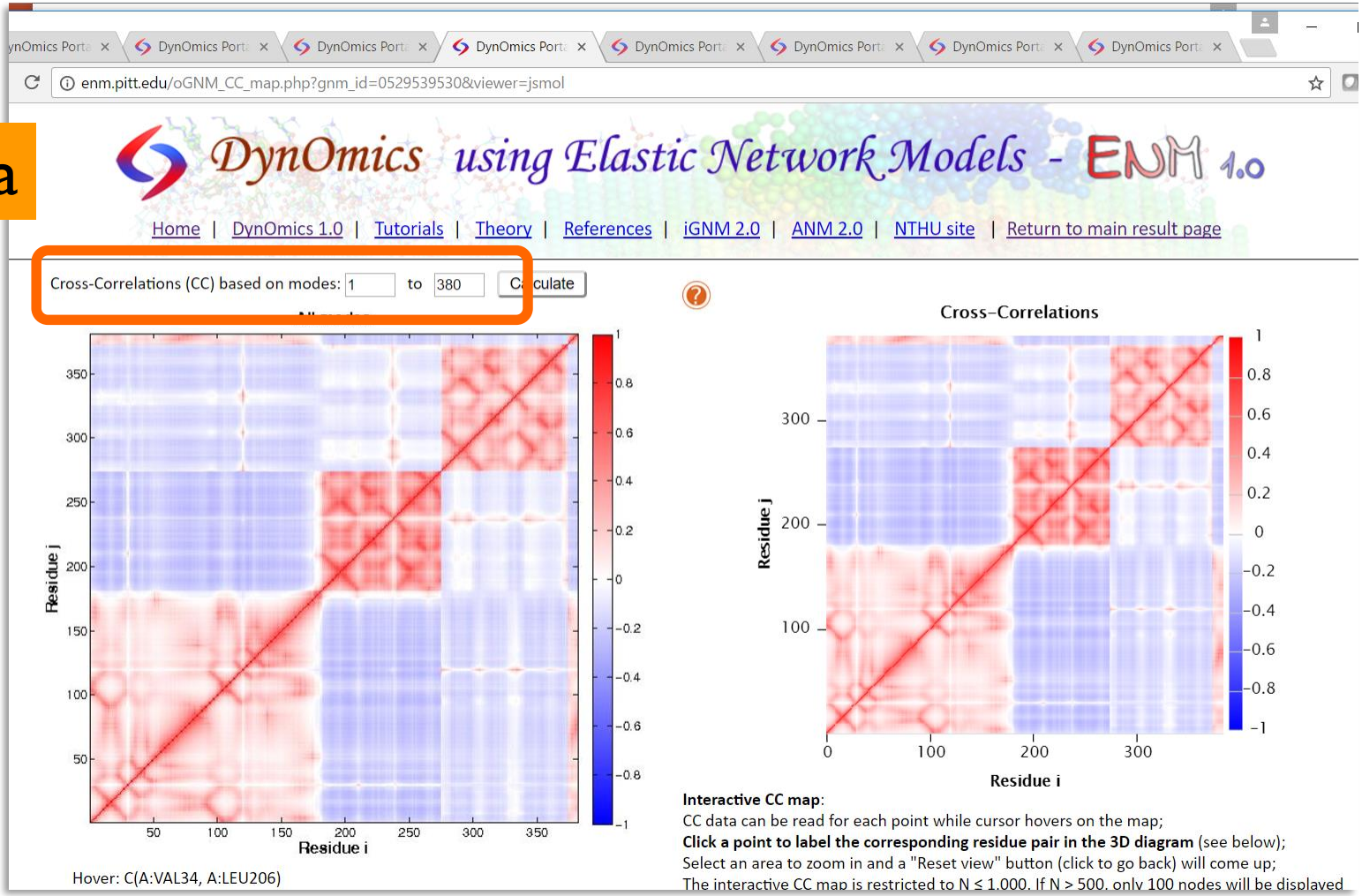
1cot



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

Output from DynOmics - ENM

1vaa



Li et al (2017) *Nucleic Acids Res* **44**: D415-422

Cross-Correlations

may be organized in a

Covariance Matrix **C**

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

Covariance scales with the inverse of the Kirchhoff matrix.

The proportionality constant is $3kT/\gamma$

Covariance matrix (NxN)

$$\mathbf{C} = \begin{array}{|c|c|c|c|c|} \hline \langle \Delta R_1 . \Delta R_1 \rangle & \langle \Delta R_1 . \Delta R_2 \rangle & \dots & \dots & \langle \Delta R_1 . \Delta R_N \rangle \\ \hline \langle \Delta R_2 . \Delta R_1 \rangle & \langle \Delta R_2 . \Delta R_2 \rangle & & & \\ \hline \dots & & & & \\ \hline \dots & & & & \\ \hline \langle \Delta R_N . \Delta R_1 \rangle & & & & \langle \Delta R_N . \Delta R_N \rangle \\ \hline \end{array} = \Delta \mathbf{R}\, \Delta \mathbf{R}^T$$

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

$\langle \Delta \mathbf{R}_I . \Delta \mathbf{R}_I \rangle$ = ms fluctuation of site I averaged over all *m* snapshots.

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

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

Contribution of mode k

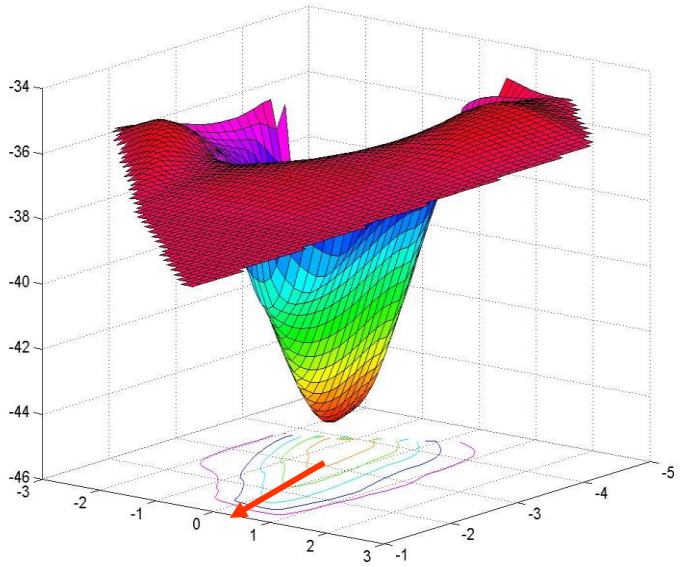
$$[\Delta\mathbf{R}_i \bullet \Delta\mathbf{R}_j]_k = (3k_BT/\gamma)\left[\lambda_k{}^{-1}\mathbf{u}_k\mathbf{u}_k{}^T\right]_{ij}$$

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

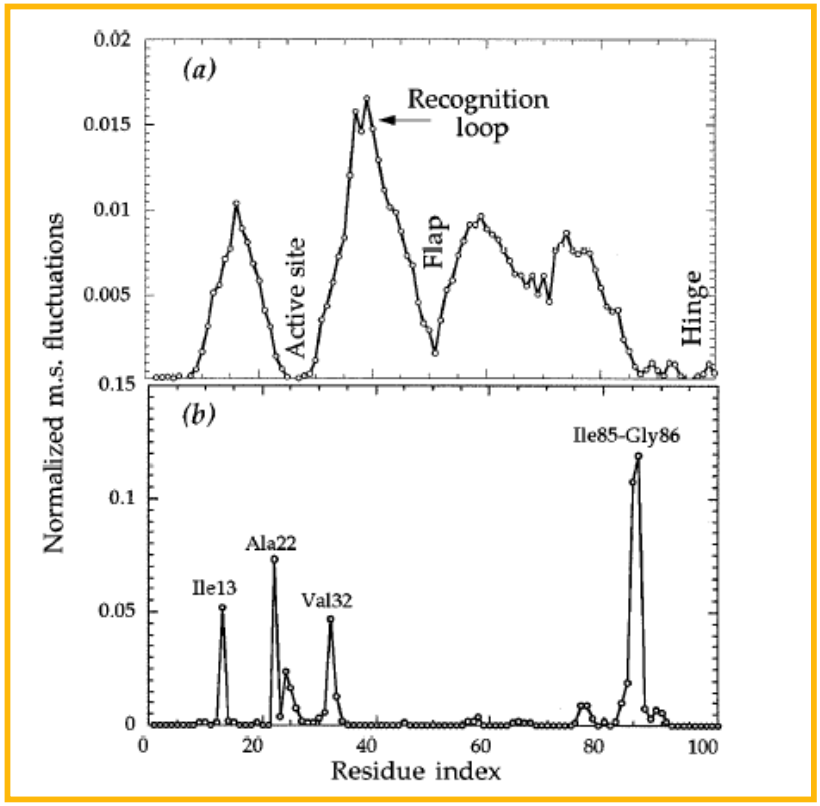
FOR MORE INFO...

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

Several modes contribute to dynamics



The first mode selects
the 'easiest' collective motion

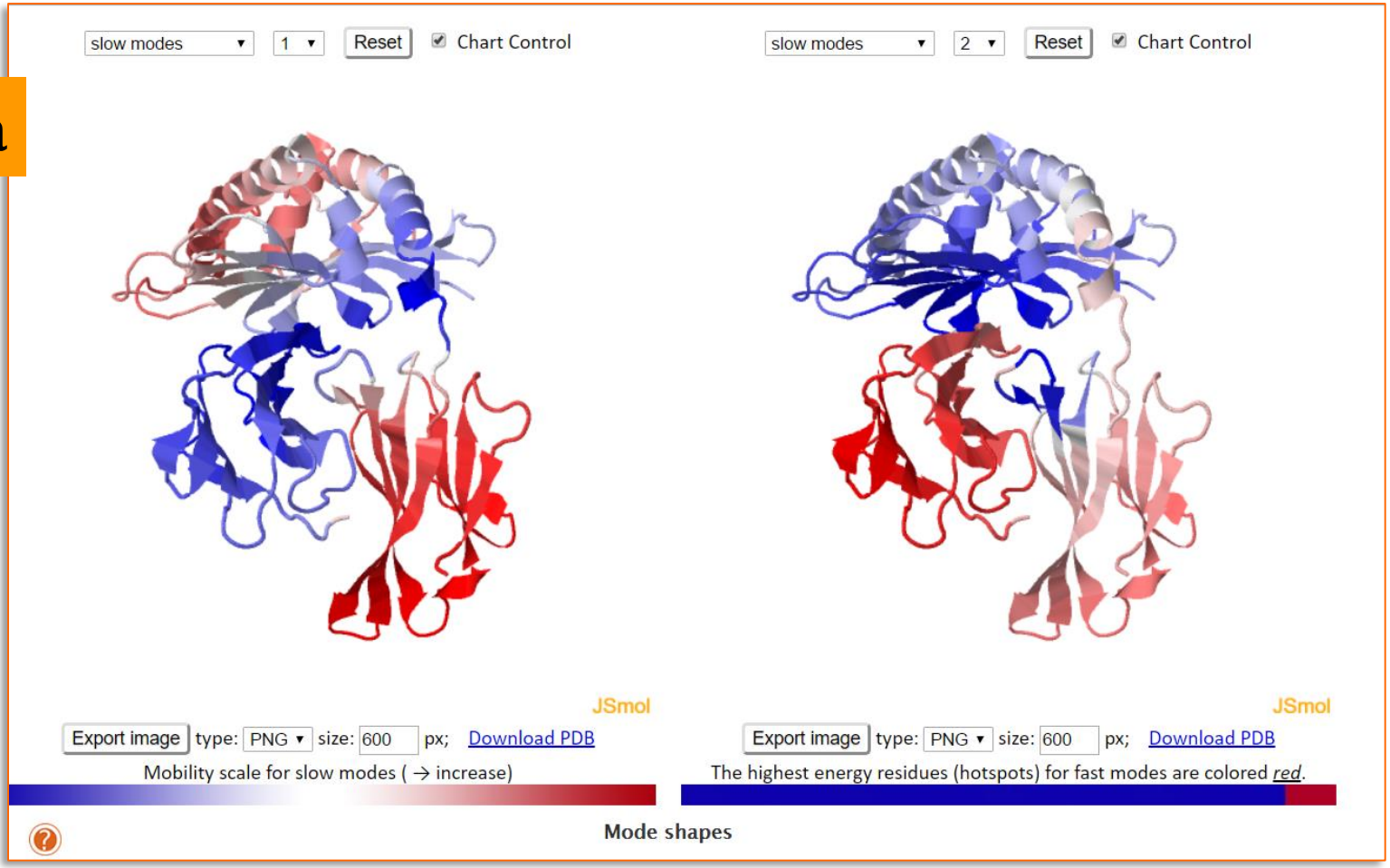


FOR MORE INFO...

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

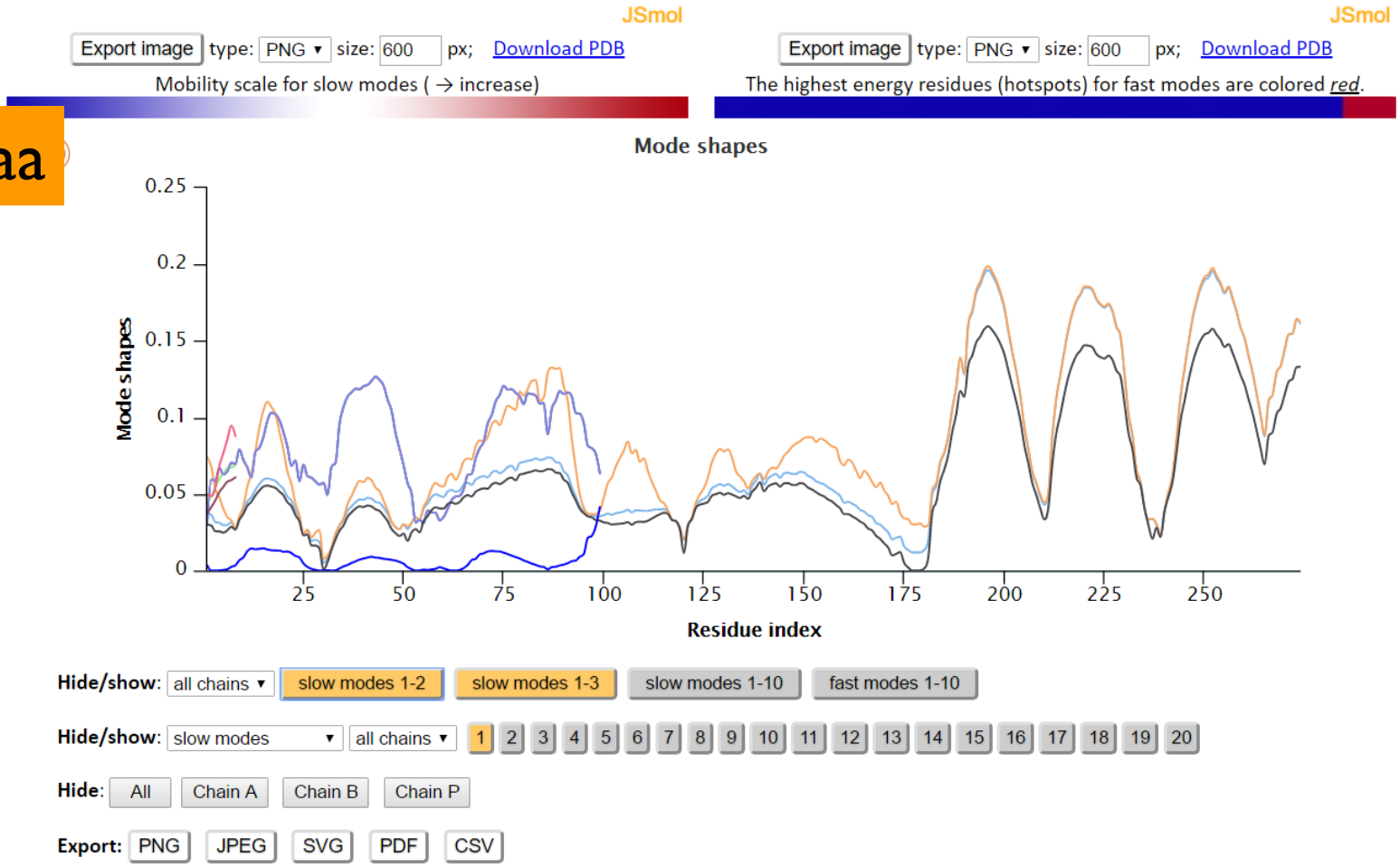
Output from DynOmics

1vaa

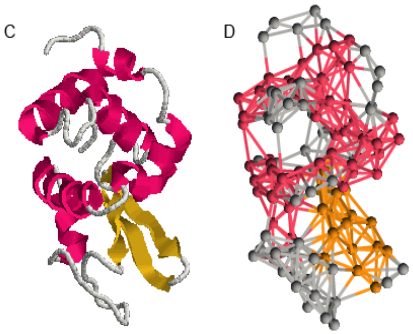


Output from DynOmics

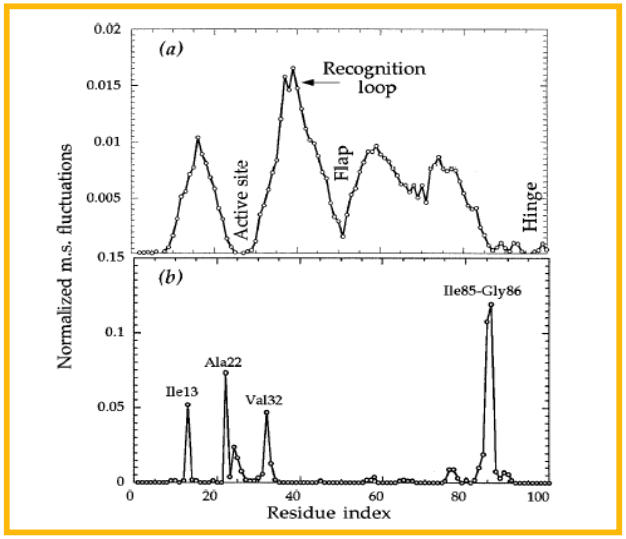
1vaa



Summary - Gaussian network model (GNM)



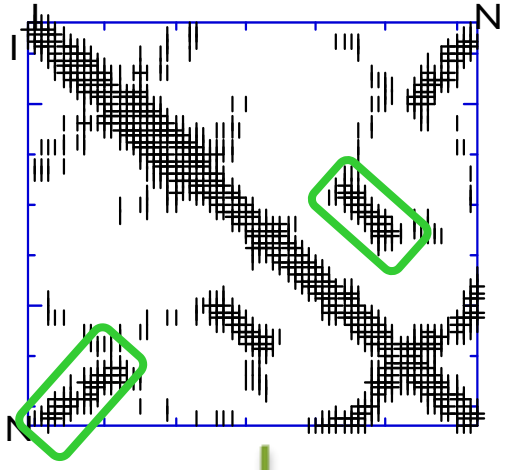
$$[\Delta \mathbf{R}_i \bullet \Delta \mathbf{R}_i]_k = (3k_B T/\gamma)\left[\lambda_k^{-1}\mathbf{u}_k\mathbf{u}_k^T\right]_{ii}$$



Kirchhoff matrix for inter-residue contacts

Contact: $R_{ij} < 10\text{Å}$

$\Gamma =$



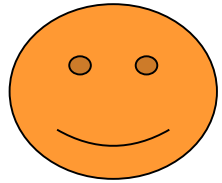
Several modes of motion contribute to dynamics

MSF of residue *i* = $<(\Delta R_i)^2>$

$$<(\Delta R_i)^2> = (3\,k_B\,T/\gamma)\left[\Gamma^{-1}\right]_{ii}$$

Recipe (GNM)

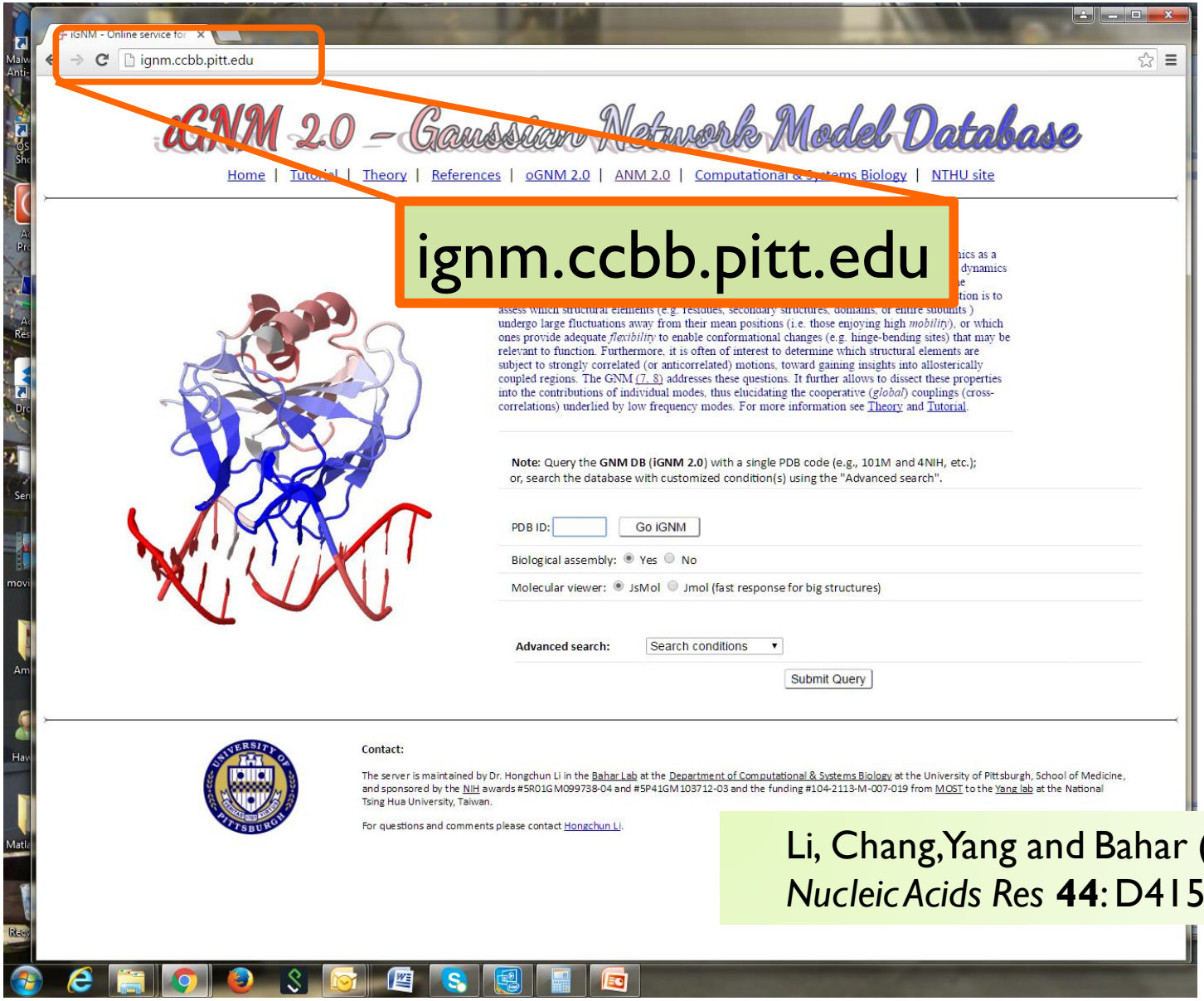
- Obtain the coordinates of network nodes from the PDB
- Write the corresponding Kirchhoff matrix Γ
- Eigenvalue decomposition of Γ yields
  the eigenvalues $\lambda_1, \lambda_2, \lambda_3, \ldots\ldots, \lambda_{N-1}$ (and $\lambda_0 = 0$)
  and eigenvectors $u_1, u_2, u_3, \ldots\ldots u_{N-1}$ (and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



Properties

- the eigenvalues scale with the frequency squared ($\lambda_i \sim w_i^2$)
- eigenvector u_k is an N-dim vectors
- the i^{th} element of u_k represents the displacement of node *i* in mode *k*
- the eigenvectors are normalized, i.e. $u_k \bullet u_k = 1$ 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 → low frequency modes have high weights

Database of GNM results



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

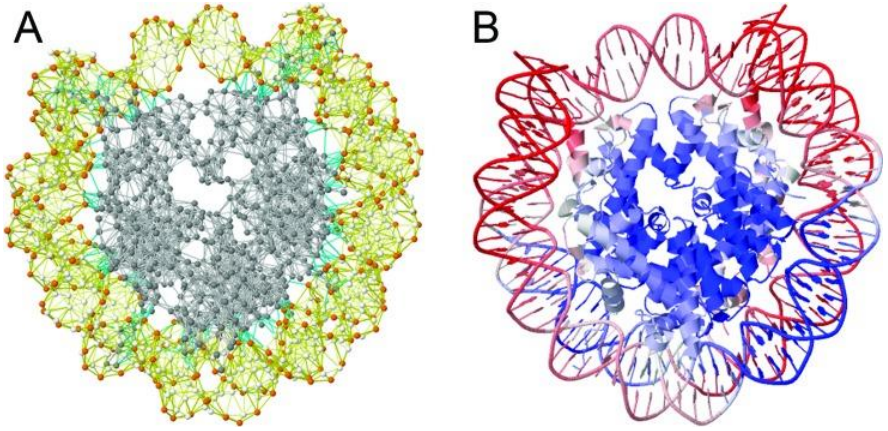
Why use iGNM2.0?

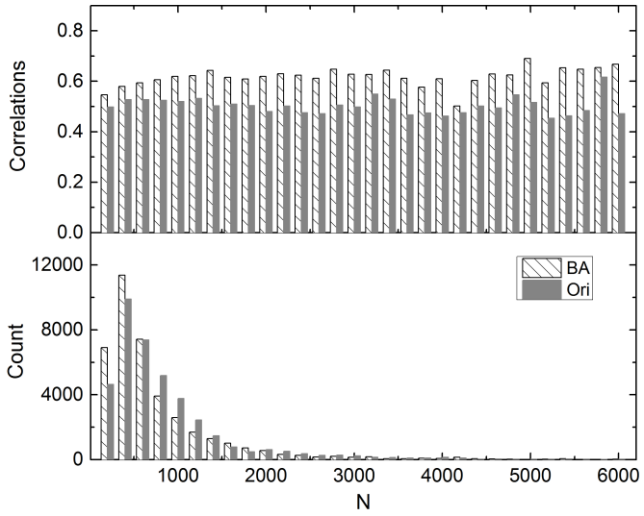
- Easy access to precomputed results for 95% of the PDB including
  - the largest structures beyond the scope of MD
  - protein-DNA/RNA complexes
  - biological assemblies (intact, biologically functional structures)

- Easy to understand, visualize, make functional inferences for any structure

13.9% of the structures in the *i*GNM 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)





Collective motions are functional

Collectivity (2D) for a given mode *k* is a measure of the degree of cooperativity (between residues) in that mode, defined as (*)

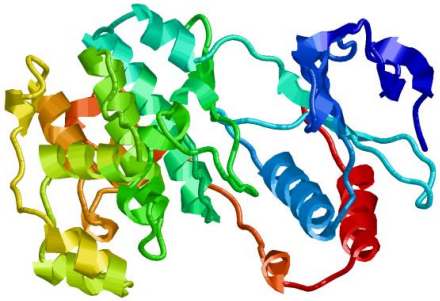
$$Collectivity_k = \frac{1}{N} e^{-\sum_i^N u_{k,i}^2 \ln u_{k,i}^2}$$

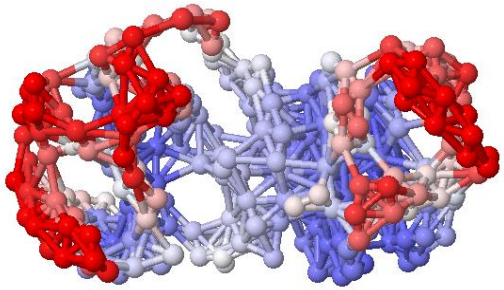
Information entropy associated with residue fluctuations in mode k

where, *k* is the mode number and *i* is the residue index. A larger collectivity value refers to a more distributive mode and *vice versa*. Usually soft modes are highly collective.

(*) Brüschweiler R. Collective protein dynamics and nuclear spin relaxation. J. Chem. Phys. 1995;102:3396–340

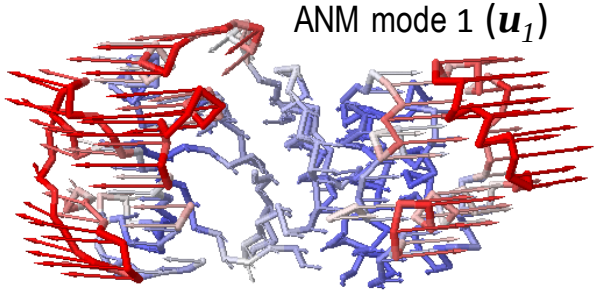
Anisotropic Network Model (ANM)

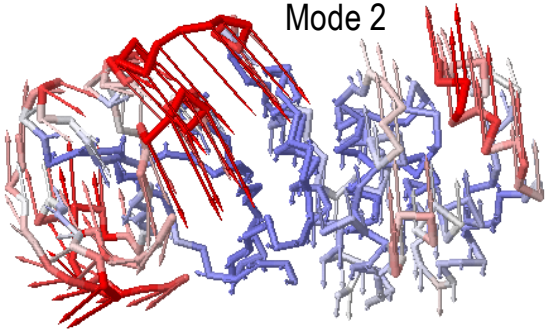


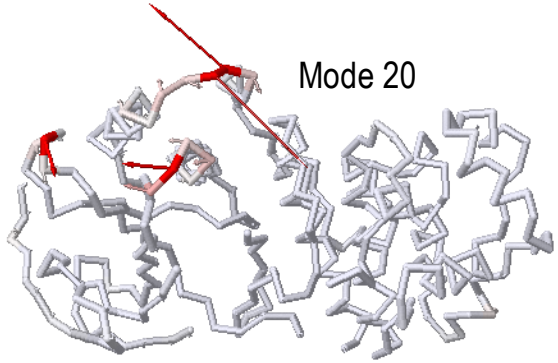


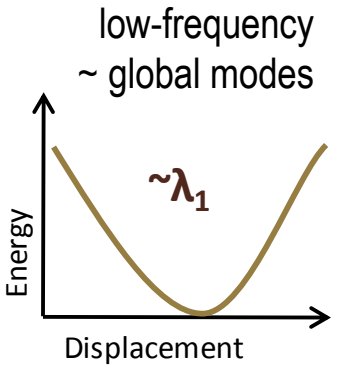
$$\mathbf{H} = \sum_{i=1}^{3N-6} \lambda_i \; \boldsymbol{u}_i \, \boldsymbol{u}_i^{\mathbf{T}}$$

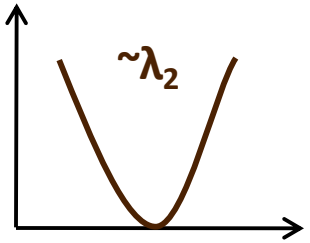
$$\mathbf{H}^{(ij)} = \frac{\gamma \, \mathbf{\Gamma}_{ij}}{\left(R_{ij}^{0}\right)^2} \begin{bmatrix} X_{ij}X_{ij} & X_{ij}Y_{ij} & X_{ij}Z_{ij} \\ Y_{ij}X_{ij} & Y_{ij}Y_{ij} & Y_{ij}Z_{ij} \\ Z_{ij}X_{ij} & Z_{ij}Y_{ij} & Z_{ij}Z_{ij} \end{bmatrix}$$



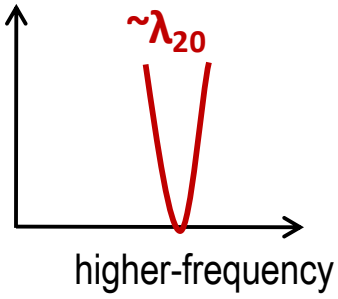








$\lambda_1 < \lambda_2 < \lambda_3 < \ldots$



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

Anisotropic Network Model

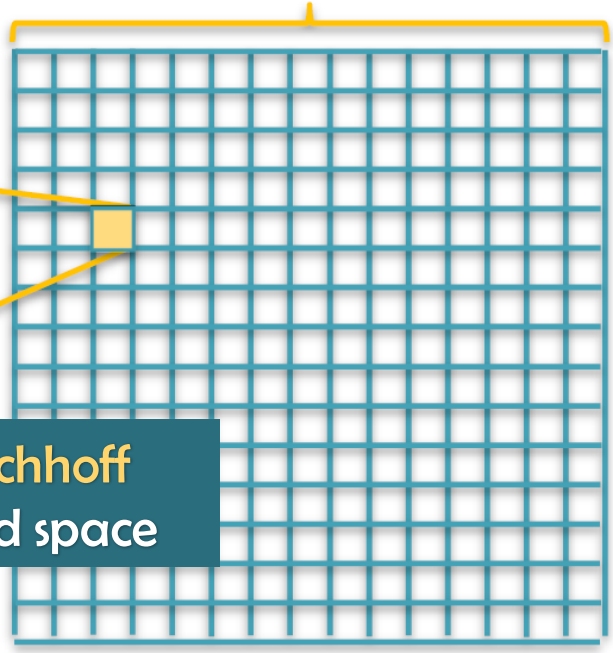
$$V(\mathbf{r}) = \frac{\gamma}{2}\sum_{i=1}^{N}\sum_{j>i}\left(\left|\mathbf{r}_{ij}\right| - \left|\mathbf{r}_{ij}^{0}\right|\right)^{2}\Theta\left(R_c - \left|\mathbf{r}_{ij}^{0}\right|\right)$$

Harmonic Step function

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

Hessian is calculated directly from structure

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



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

Doruker et al. Proteins 40 (2000).
Atilgan et al. Biophys J 80 (2001).

ANM covariance matrix (3Nx3N)

C_{3N} =

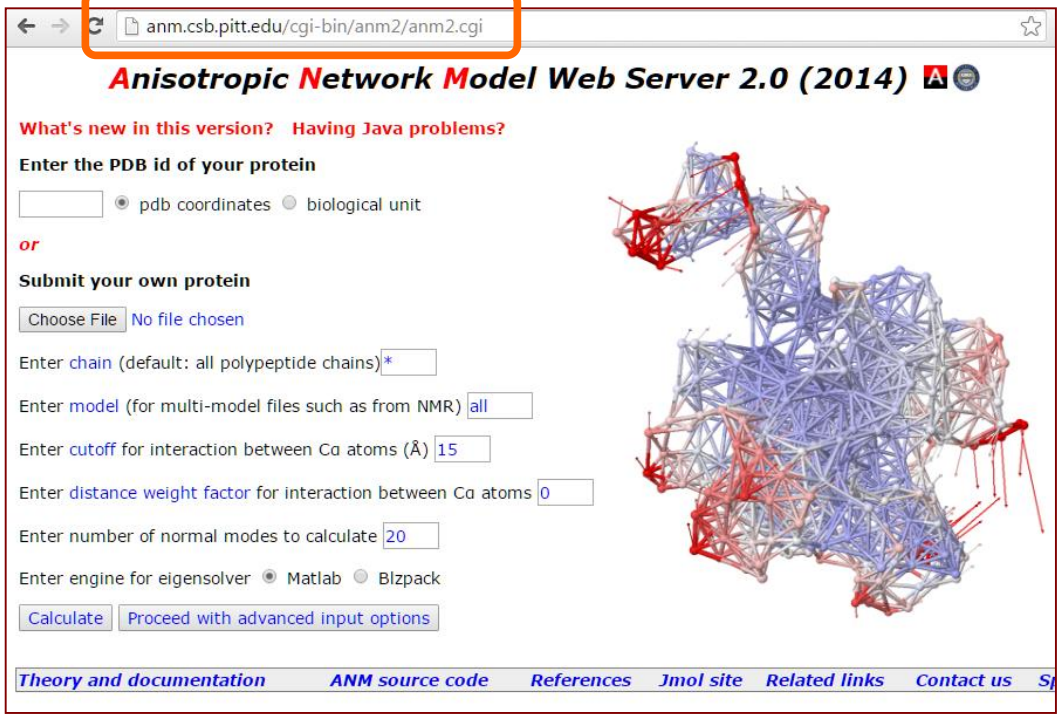
C_{11}	C_{21}	C_{13}		C_{1N}
C_{12}	C_{22}			
C_{N1}				C_{NN}

3N x 3N

$<\Delta X_1 \Delta X_2>$	$<\Delta X_1 \Delta Y_2>$	$<\Delta X_1 \Delta Z_2>$
$<\Delta Y_1 \Delta X_2>$	$<\Delta Y_1 \Delta Y_2>$	$<\Delta Y_1 \Delta Z_2>$
$<\Delta Z_1 \Delta X_2>$	$<\Delta Z_1 \Delta Y_2>$	$<\Delta Z_1 \Delta Z_2>$

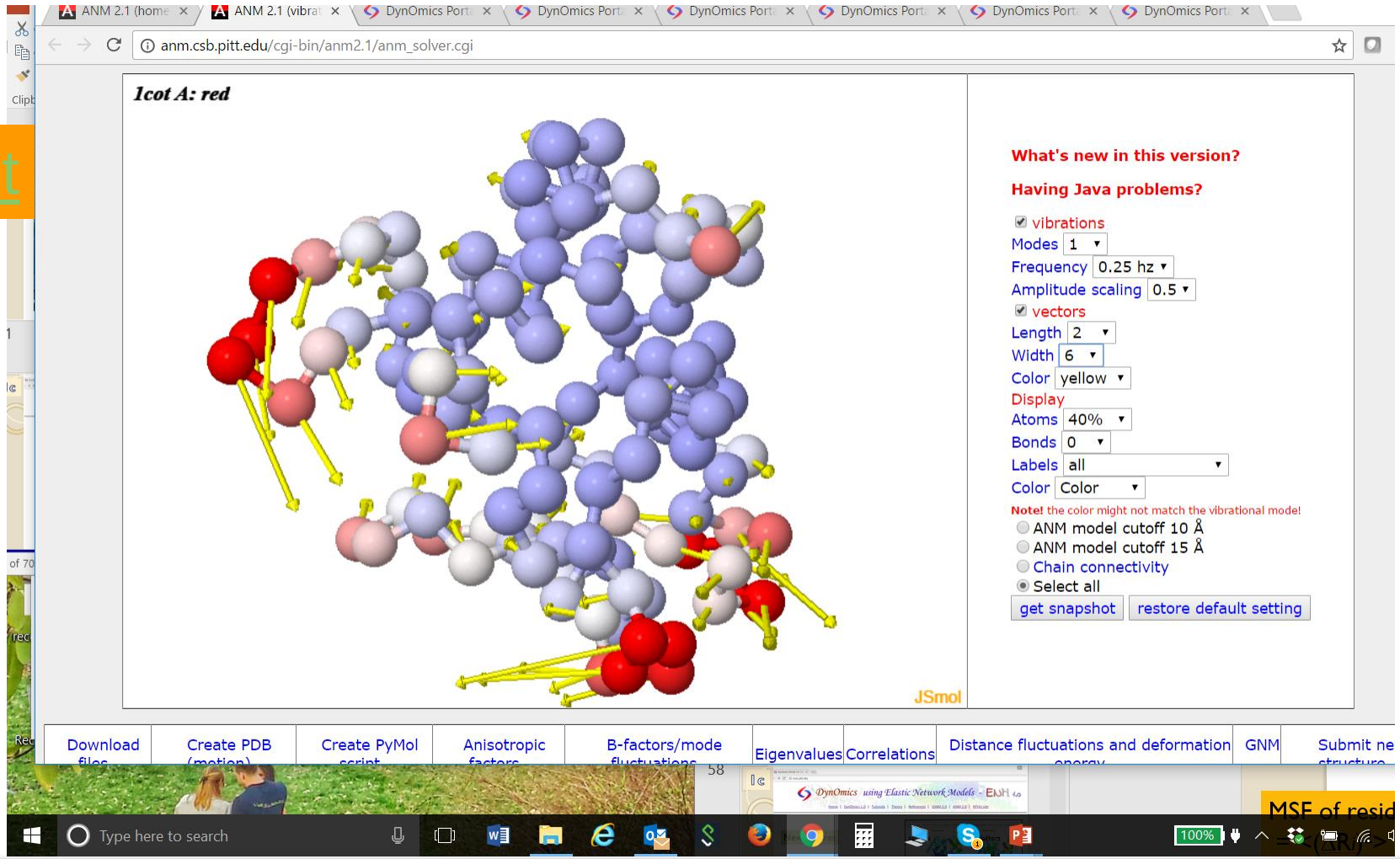
ANM server

http://anm.csb.pitt.edu/cgi-bin/anm2/anm2.cgi

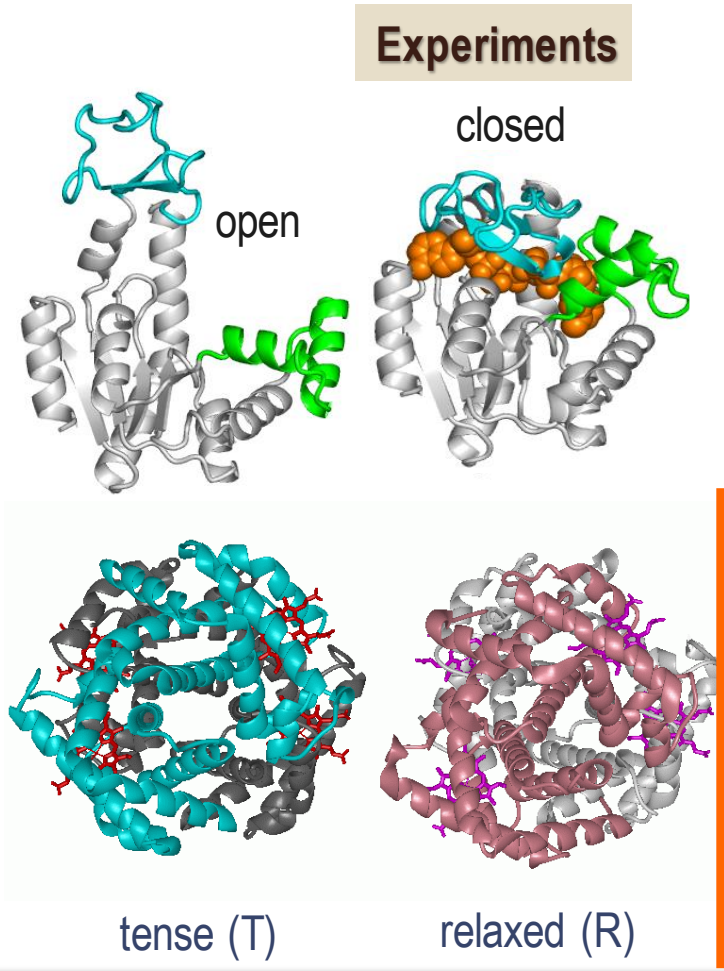


Output from ANM server

1cot



Softest modes are functional



E coli adenylate kinase dynamics: comparison of elastic network model modes with ^{15}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;

Ic

DynOmics Portal 1.0 - Dy

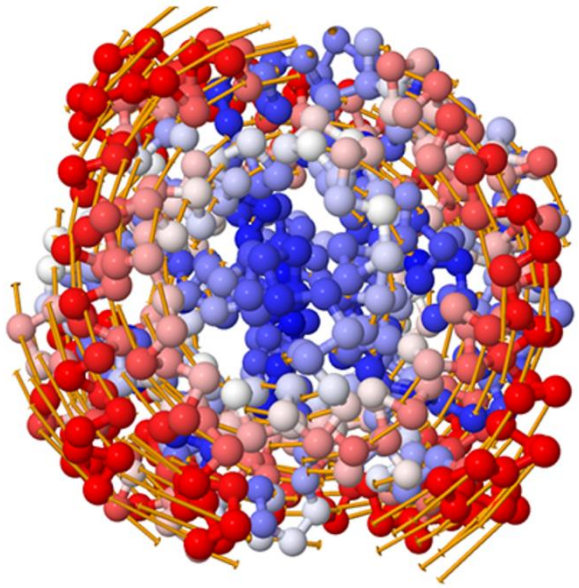
enm.pitt.edu

DynOmics using Elastic Network Models - ENM 1.0

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: Choose 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)

Submit

Load examples:

Main result | Molecular motion | membrANM | Hitting time | Domain separation

enm.pitt.edu

Li et al (2017) *Nucleic Acids Research (e-pub on* May 3).

Ic

DynOmics Portal 1.0 - Dy

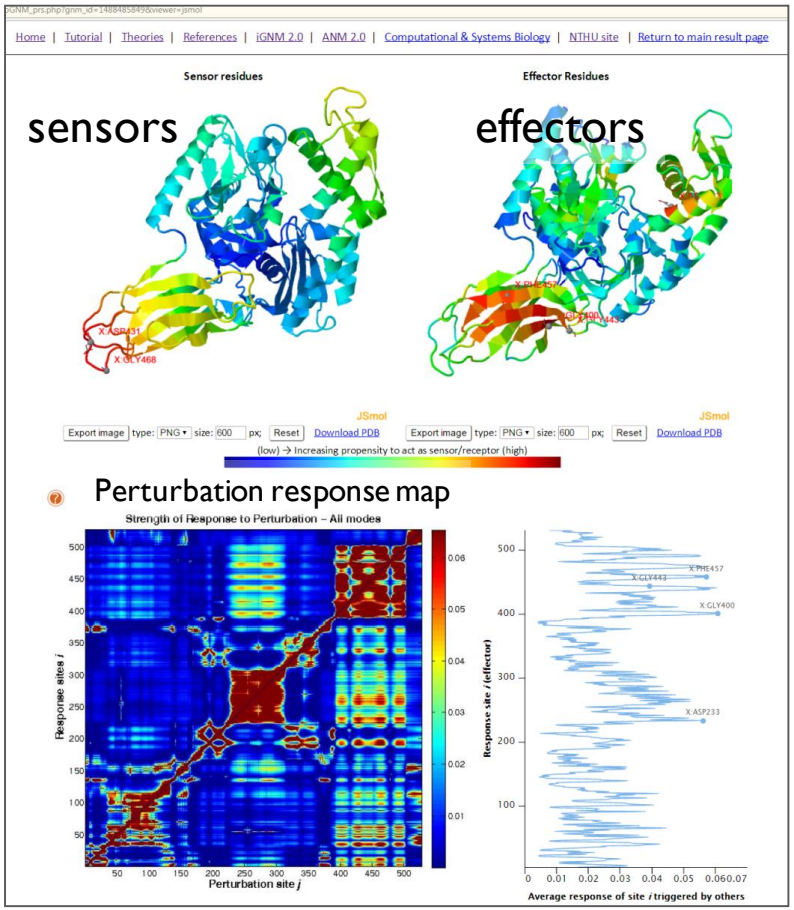
enm.pitt.edu

DynOmics using Elastic Network Models - ENM 1.0

Home | DynOmics 1.0 | Tutorials | Theory | References | iGNM 2.0 | ANM 2.0 | NTHU site

New features

- sensors and effectors
- first passage times for signaling
- mechanically functional sites
- effect of oligomerization
- coupling to membrane



Dynamics of Structural Proteomics and Beyond

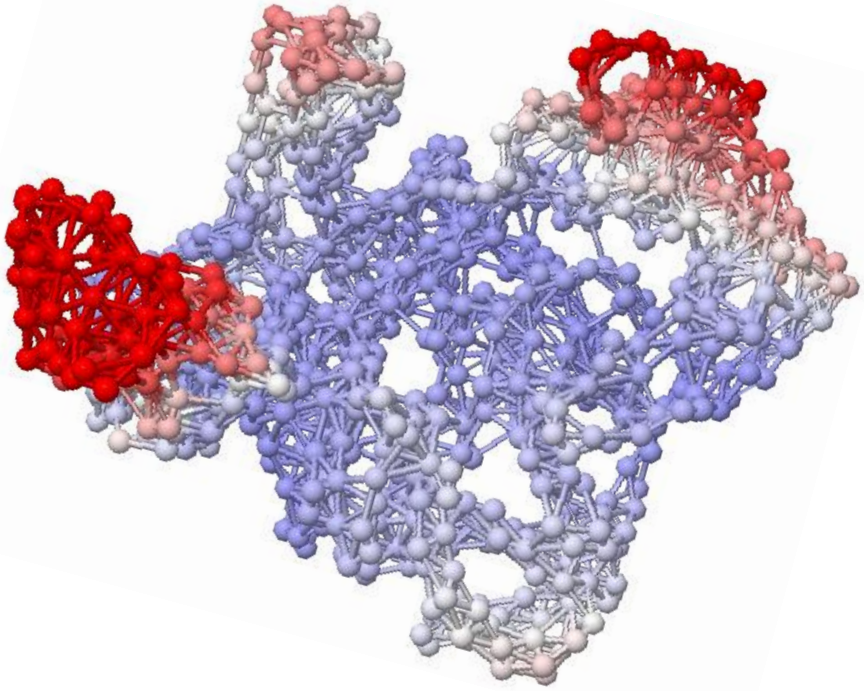
Li et al (2017) *Nucleic Acids Research (e-pub on* May 3).

Plan

1. Theory
   a. Gaussian Network Model (GNM)
   b. Anisotropic Network Model (ANM)
   c. Resources/Servers/Databases (ProDy, DynOmics etc)
2. Allosteric Changes in Structure
3. Ensemble analysis. Experiments vs Predictions
   Adaptability/evolution
4. Recent Extensions and Applications
   a. Membrane Proteins
   b. AMPA Receptors
   c. Chromatin

Allosteric changes in conformation

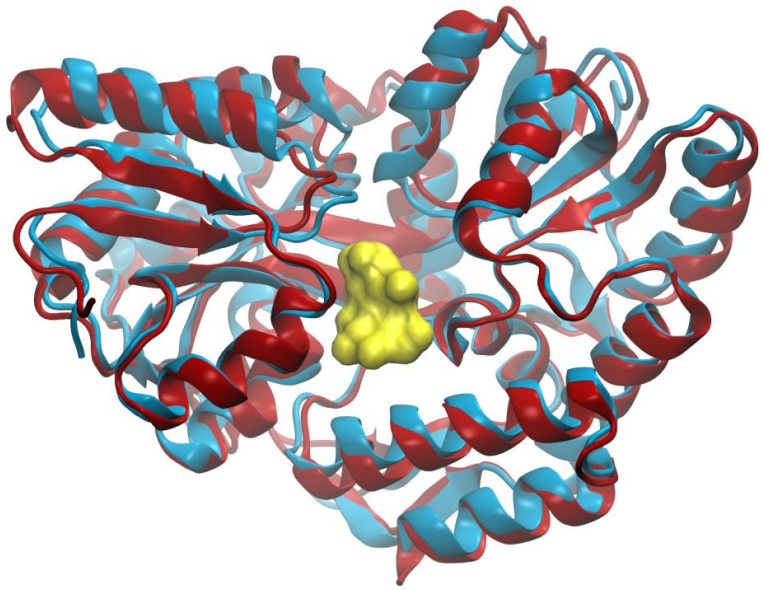
Elastic Network Models are particularly useful for exploring the cooperative motions of large multimeric structures



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

Proteins exploit pre-existing soft modes for their interactions

Structural changes involved in protein binding correlate with intrinsic motions in the unbound state

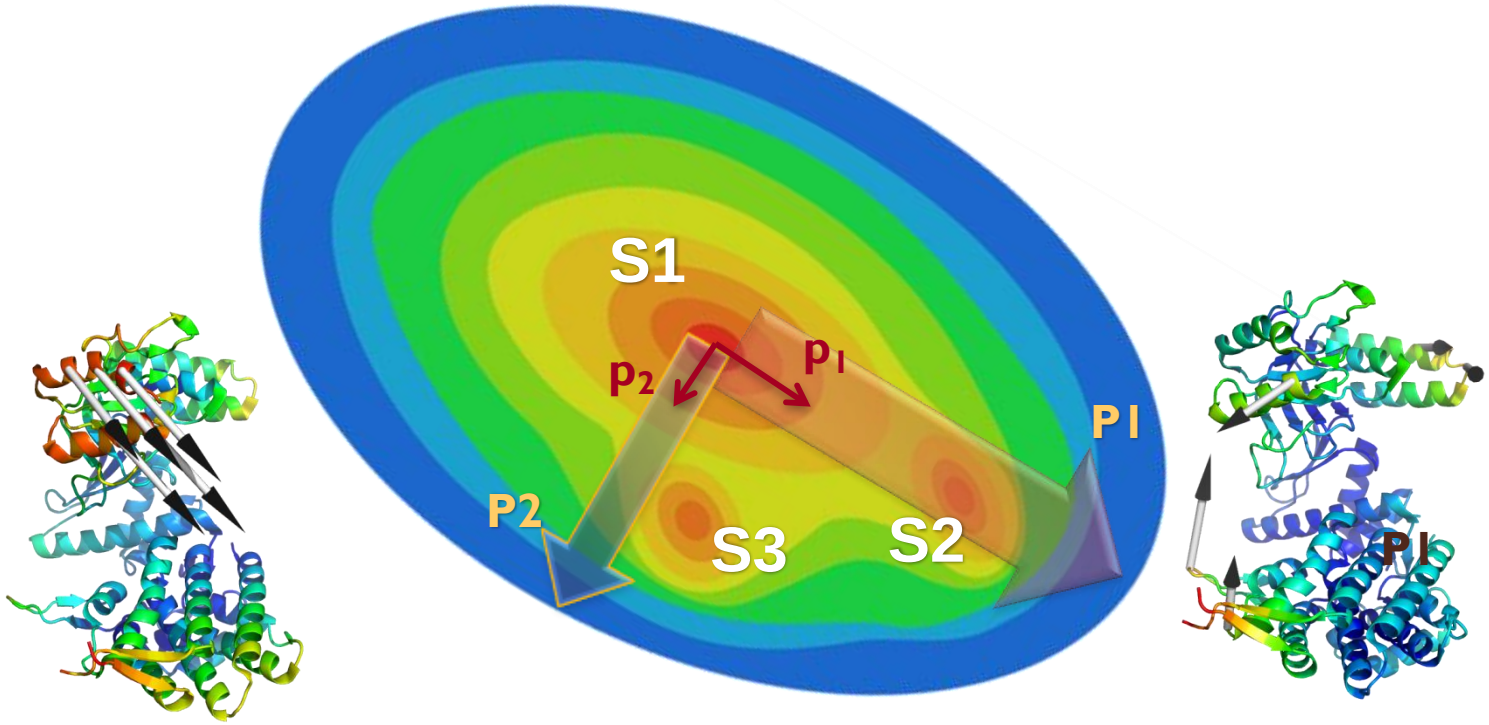


maltodextrin binding protein
Unbound/Bound

Dror Tobi & I. Bahar (2005) *PNAS* 102:18908-18913

2

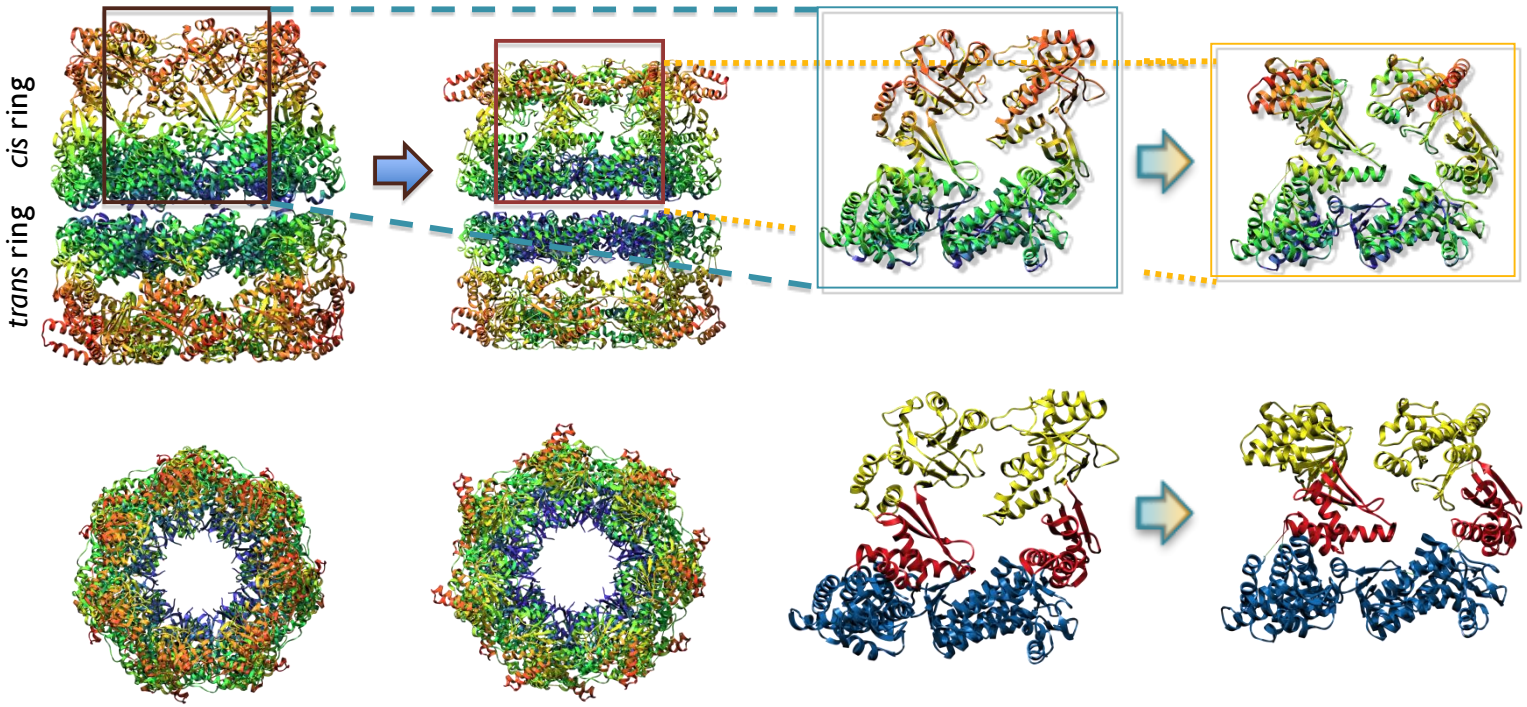
Substates may be identified along soft modes



Hybrid ANM/MD methods include atomic details and specificity

Allosteric dynamics of GroEL

Passage between the R and T states



See...

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

What is the overlap between computations and experiments?

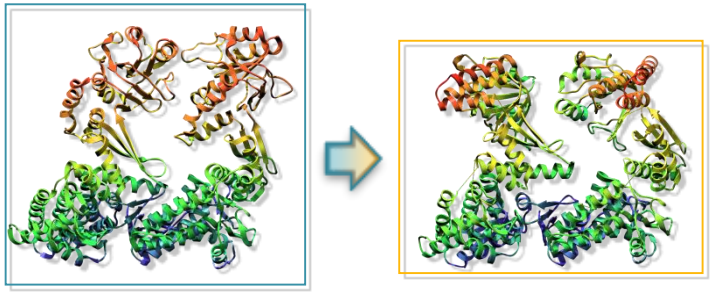
Computations

ANM yields a series of 3N dimensional deformation vectors

Mode 1 (slowest mode)
Mode 2
Mode 3
….
Mode 3N-6 (fastest mode)

Given by eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \ldots \mathbf{u}_{3N-6}$, with respective frequencies of $\lambda_1, \lambda_2, \lambda_3, \ldots \lambda_{3N-6}$

Experiments



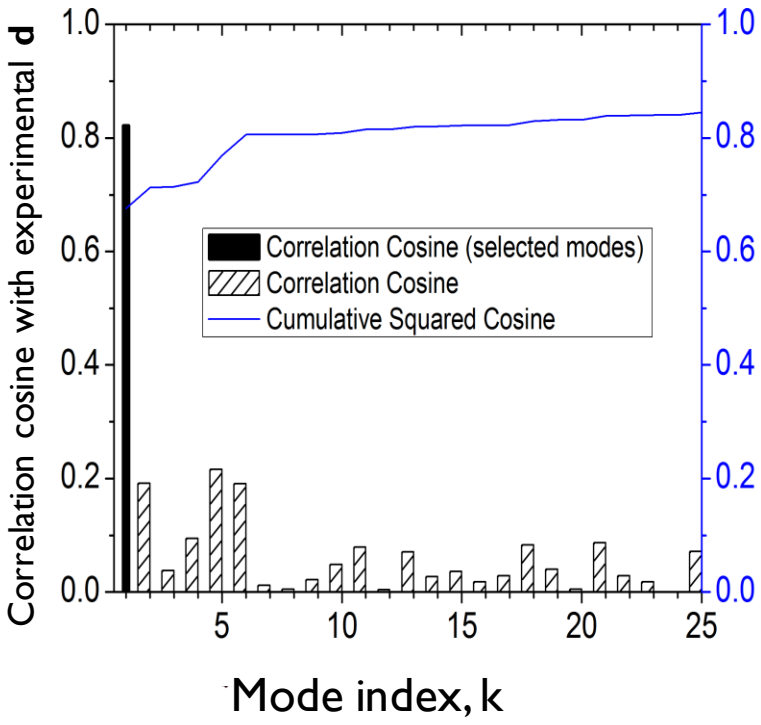
$$\mathbf{d} = [\Delta x_1 \;\; \Delta y_1 \;\; \Delta z_1 \;\ldots\;\; \Delta z_N]^T$$

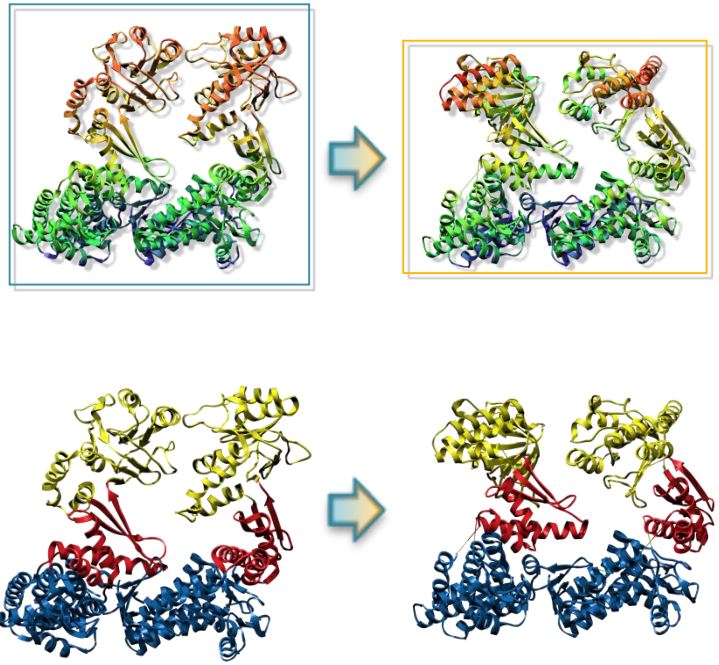
See...

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

What is the overlap between computations and experiments?

Correlation cosine between u_k and d





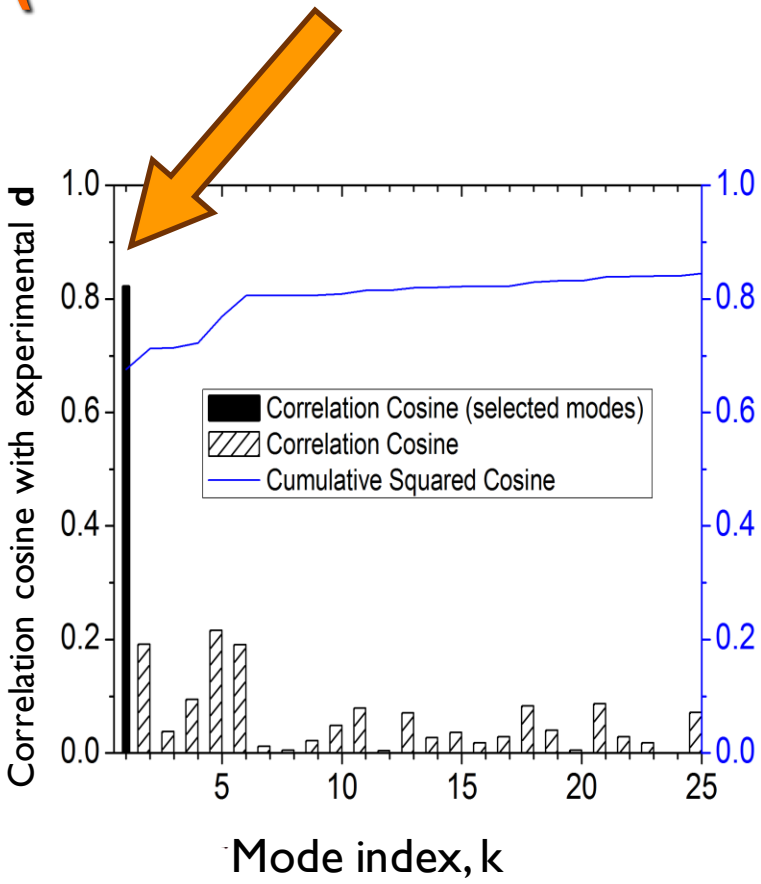
$\mathbf{d} = [\Delta x_1 \ \Delta y_1 \ \Delta z_1 \ \ldots \ \Delta z_N]^T$

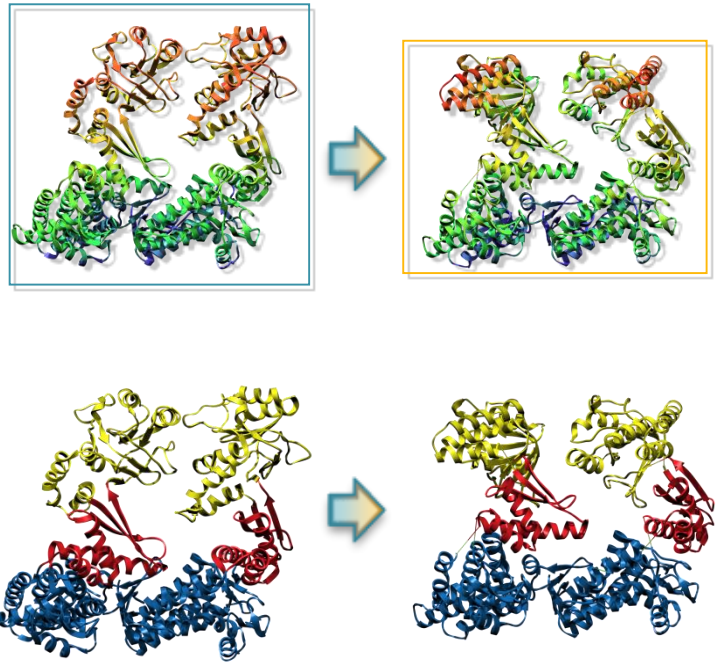
See...

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

The softest mode enables the passage R → T

(with a correlation of 0.81)





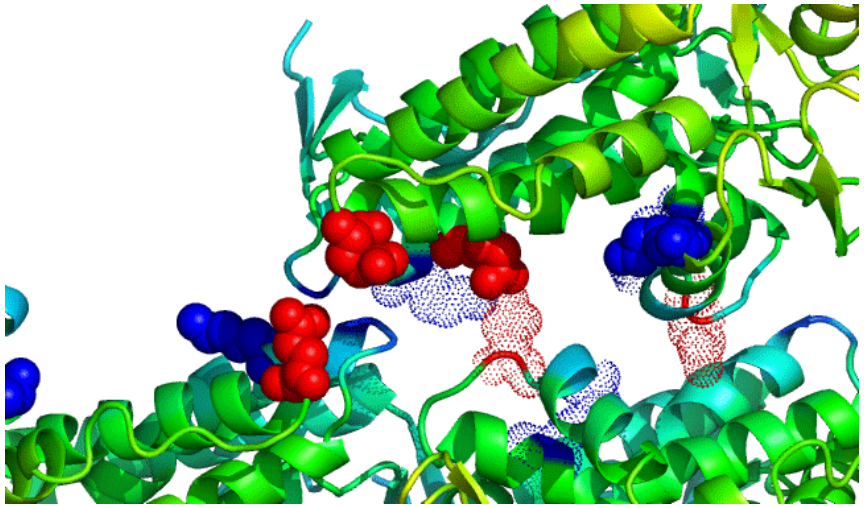
$\mathbf{d} = [\Delta x_1 \;\; \Delta y_1 \;\; \Delta z_1 \;\; \ldots \;\; \Delta z_N]^T$

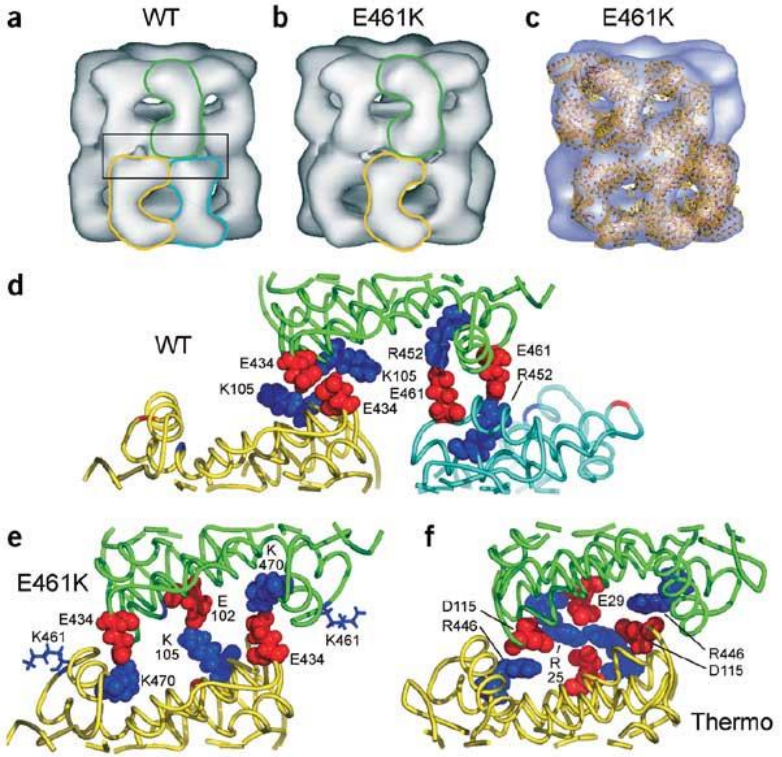
See...

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

Mutations may stabilize conformers along soft modes – which may be impair function

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

Plan

1. Theory
   a. Gaussian Network Model (GNM)
   b. Anisotropic Network Model (ANM)
   c. Resources/Servers/Databases (ProDy, DynOmics etc)
2. Allosteric Changes in Structure
3. Ensemble analysis. Experiments vs Predictions
   Adaptability/evolution
4. Recent Extensions and Applications
   a. Membrane Proteins
   b. AMPA Receptors
   c. Chromatin

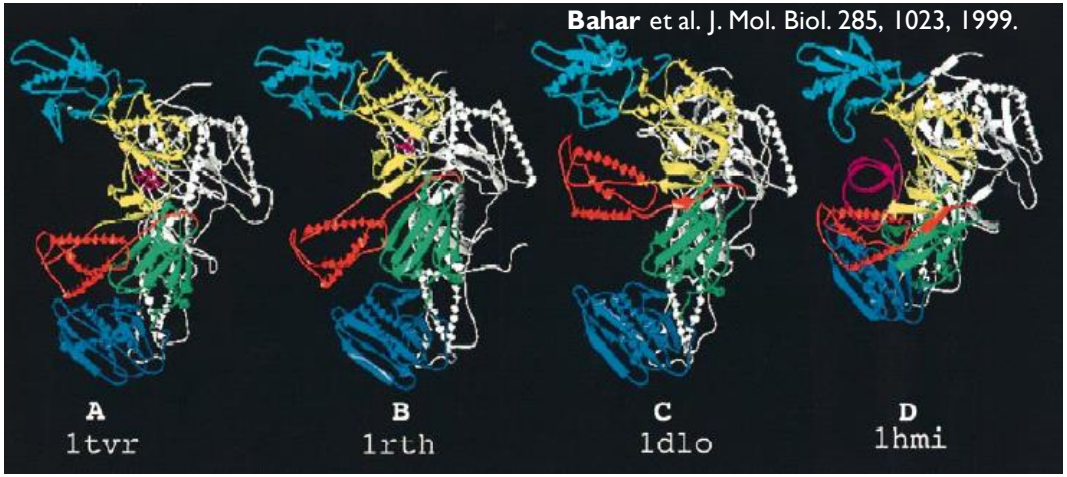
A better comparison:

Consider more than 2 end points for a given structure, but all the known structures for a given protein, or the structurally resolved

Ensemble of structures

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)

Ensembles of structures

- Structural changes accompanying substrate (protein) binding

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

Ubiquitin
140 structures
1732 models

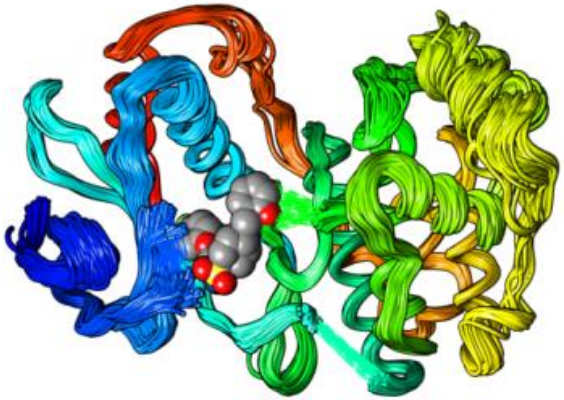
Ensembles of structures

- Structural changes accompanying substrate (protein) binding

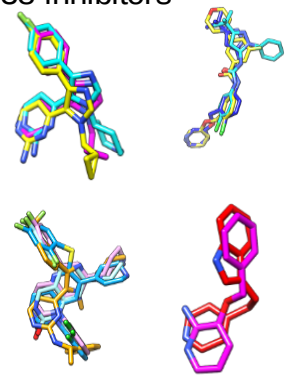
- Structural changes induced by, or stabilized upon, ligand binding

Ubiquitin
140 structures
1732 models

p38 MAP kinase
(**182 structures**)

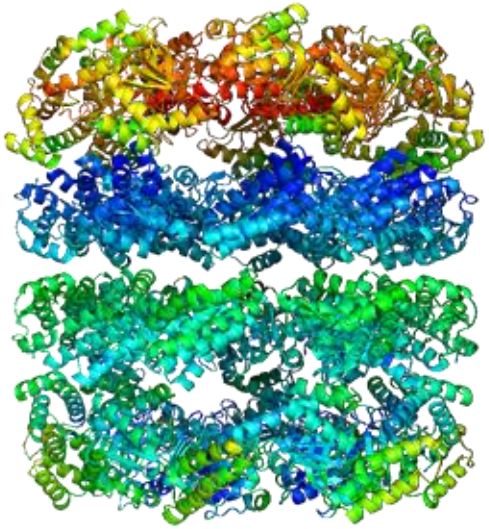


p38 inhibitors



Ensembles of structures

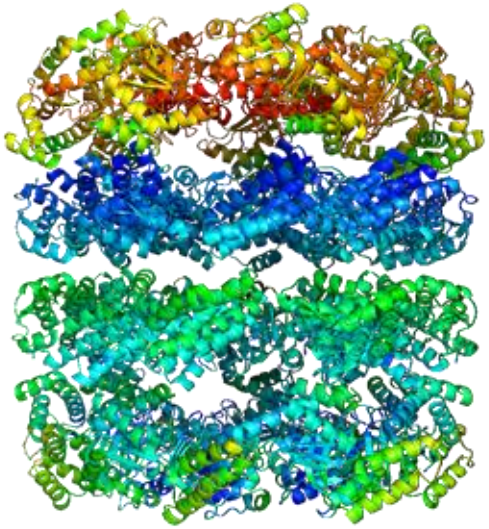
- Structural changes accompanying substrate (protein) binding
- Structural changes induced by, or stabilized upon, ligand binding
- Alternative conformations sampled during allosteric cycles



Yang et al. *PLoS Comp Biol* 2009

Ensembles of structures

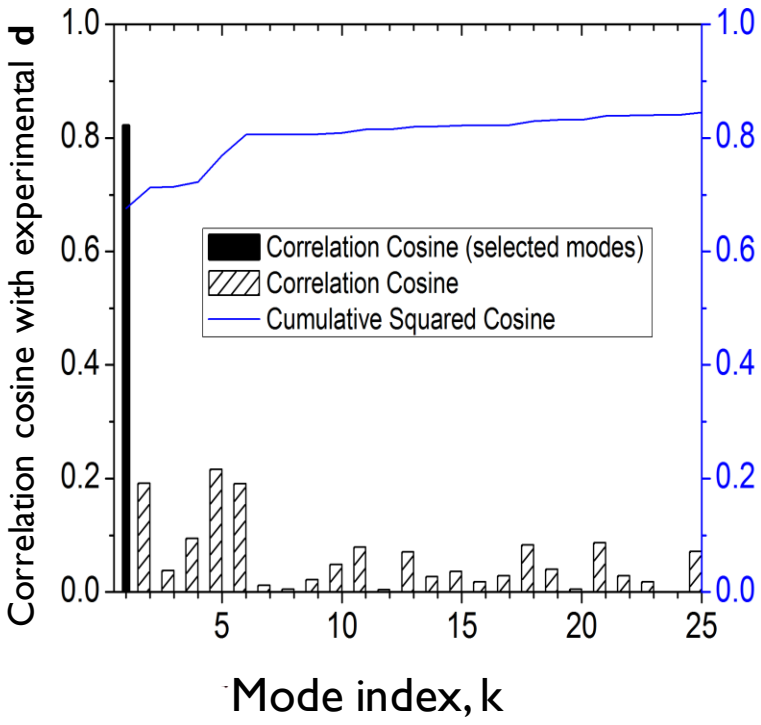
- Structural changes accompanying substrate (protein) binding
- Structural changes induced by, or stabilized upon, ligand binding
- Alternative conformations sampled during allosteric cycles

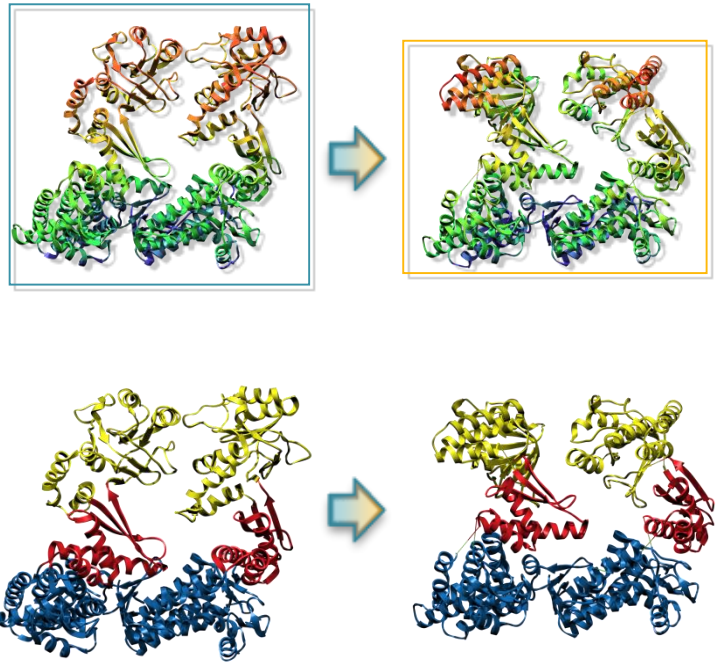


Yang et al. *PLoS Comp Biol* 2009

What is the overlap between computations and experiments?

Correlation cosine between u_k and d





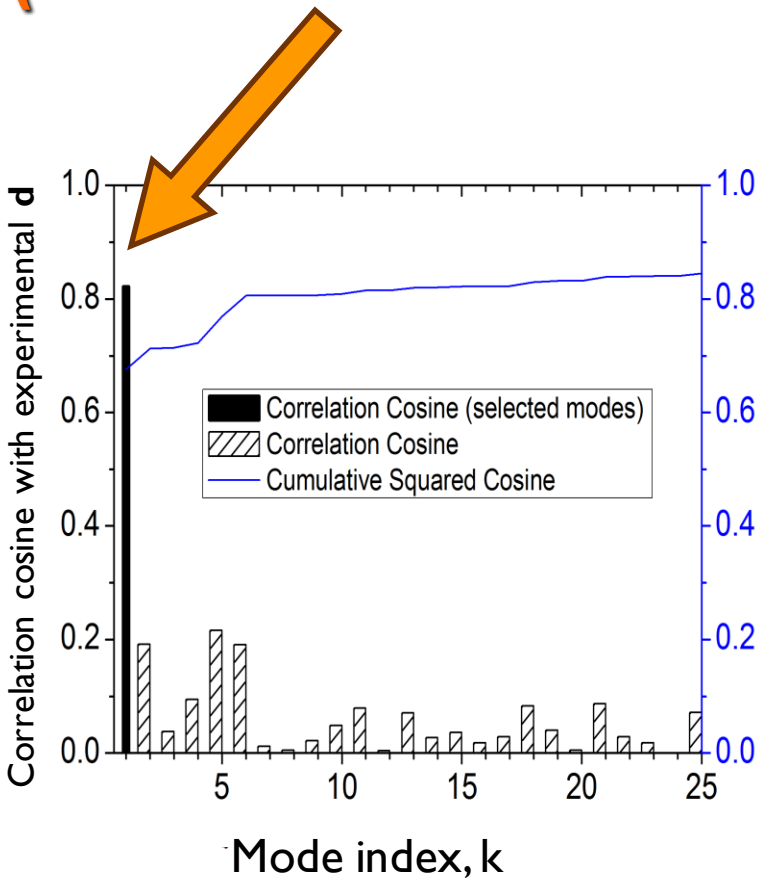
$\mathbf{d} = [\Delta x_1 \ \Delta y_1 \ \Delta z_1 \ \ldots \ \Delta z_N]^T$

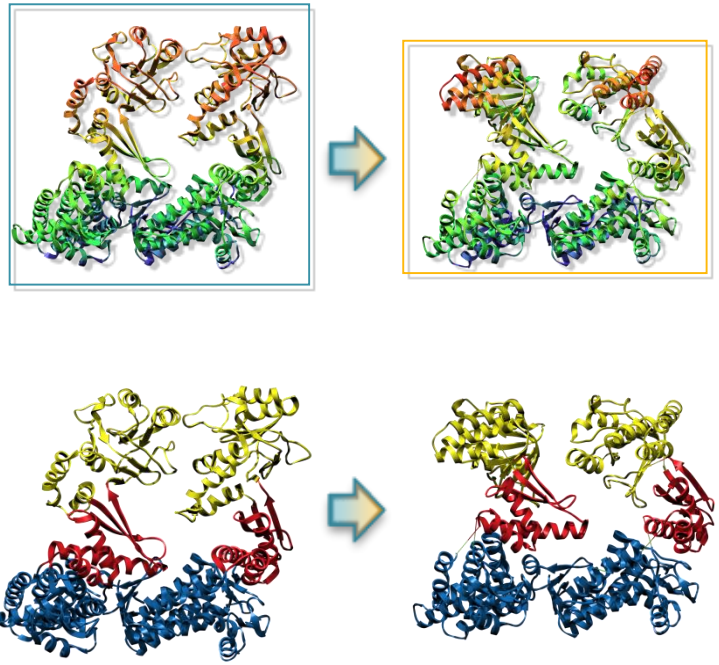
See...

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

The softest mode enables the passage R → T

(with a correlation of 0.81)



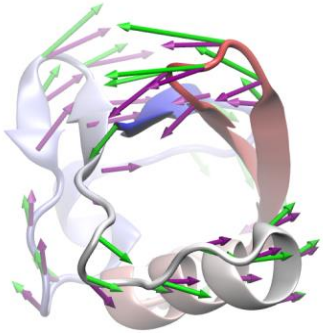


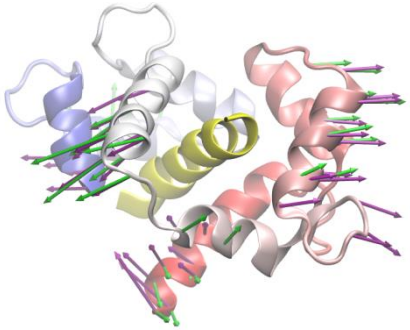
$\mathbf{d} = [\Delta x_1 \;\; \Delta y_1 \;\; \Delta z_1 \;\; \ldots \;\; \Delta z_N]^T$

See...

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

Global motions inferred from theory and experiments

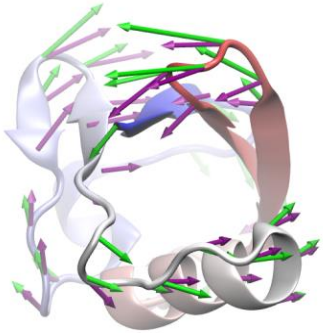


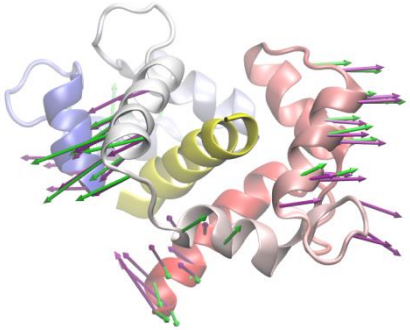


→ PCA of the ensemble of resolved structures

→ 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*

What is Ensemble Analysis?

Principal component analysis

Input:

An ensemble of structures for a given protein

- NMR models (~40)
- X-ray structures resolved under different conditions (ligand-bound/unbound, different stages of molecular machinery or transport cycle
- MD snapshots/frames

Output:

Principal modes of conformational

- variations/differences between NMR models
- rearrangements/changes under different functional states
- dynamics/fluctuations observed in simulations

What is Ensemble Analysis?

Principal component analysis

- **Method:**
- Superimpose of the structures
- Evaluate the covariance matrix (differences between individual coordinates and mean coordinates)
- Decompose it into a series of modes of covariance (3N-6 eigenvectors)

Output:

Principal modes of conformational

- variations/differences between NMR models
- rearrangements/changes under different functional states
- dynamics/fluctuations observed in simulations

Covariance matrix (NxN)

$\mathbf{C} =$

$\langle \Delta R_1 . \Delta R_1 \rangle$	$\langle \Delta R_1 . \Delta R_2 \rangle$	…	…	$\langle \Delta R_1 . \Delta R_N \rangle$
$\langle \Delta R_2 . \Delta R_1 \rangle$	$\langle \Delta R_2 . \Delta R_2 \rangle$			
…				
…				
$\langle \Delta R_N . \Delta R_1 \rangle$				$\langle \Delta R_N . \Delta R_N \rangle$

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

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

$\langle \Delta \mathbf{R}_I . \Delta \mathbf{R}_I \rangle$ = ms fluctuation of site I averaged over all *m* snapshots.

Covariance matrix (3Nx3N)

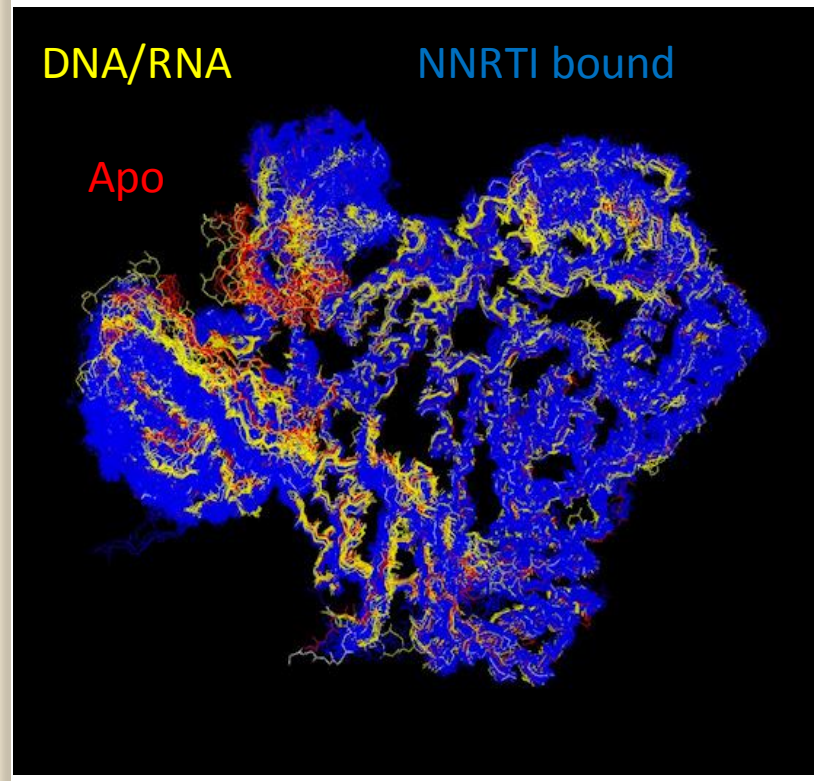
C_{3N} =

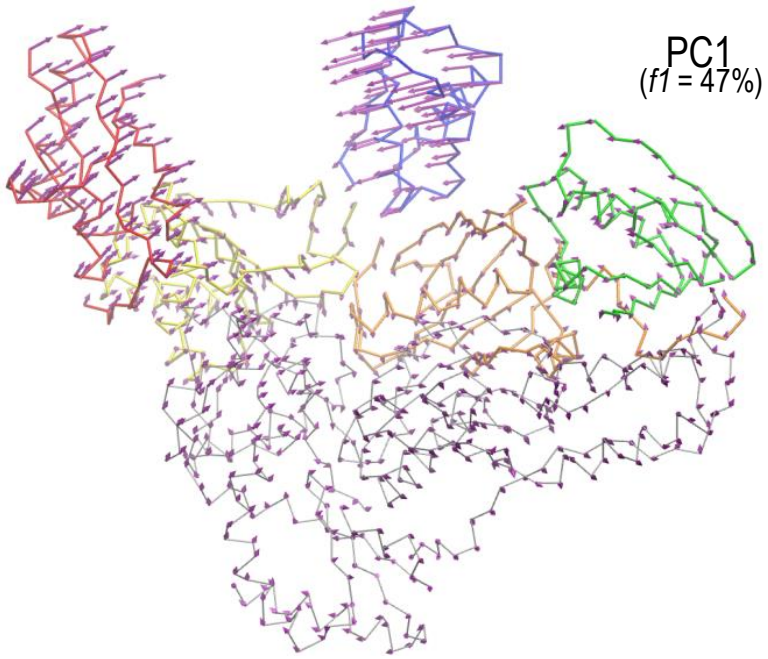
C_{11}	C_{21}	C_{13}		C_{1N}
C_{12}	C_{22}			
C_{N1}				C_{NN}

3N x 3N

$\langle \Delta X_1 \Delta X_2 \rangle$	$\langle \Delta X_1 \Delta Y_2 \rangle$	$\langle \Delta X_1 \Delta Z_2 \rangle$
$\langle \Delta Y_1 \Delta X_2 \rangle$	$\langle \Delta Y_1 \Delta Y_2 \rangle$	$\langle \Delta Y_1 \Delta Z_2 \rangle$
$\langle \Delta Z_1 \Delta X_2 \rangle$	$\langle \Delta Z_1 \Delta Y_2 \rangle$	$\langle \Delta Z_1 \Delta Z_2 \rangle$

Principal Component Analysis (PCA)

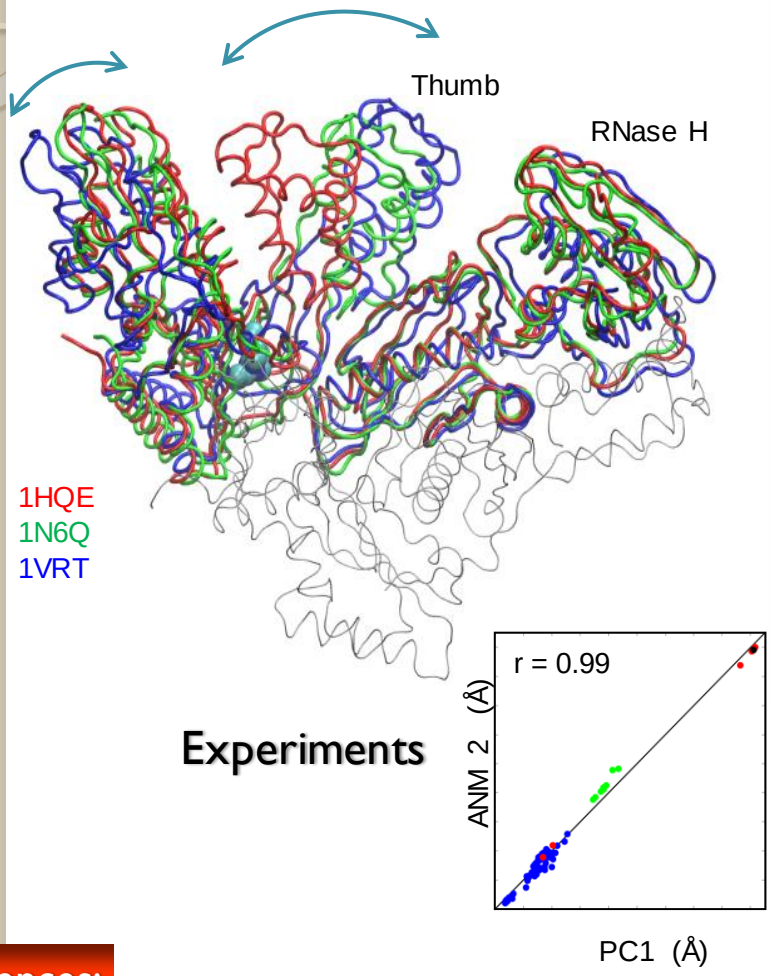




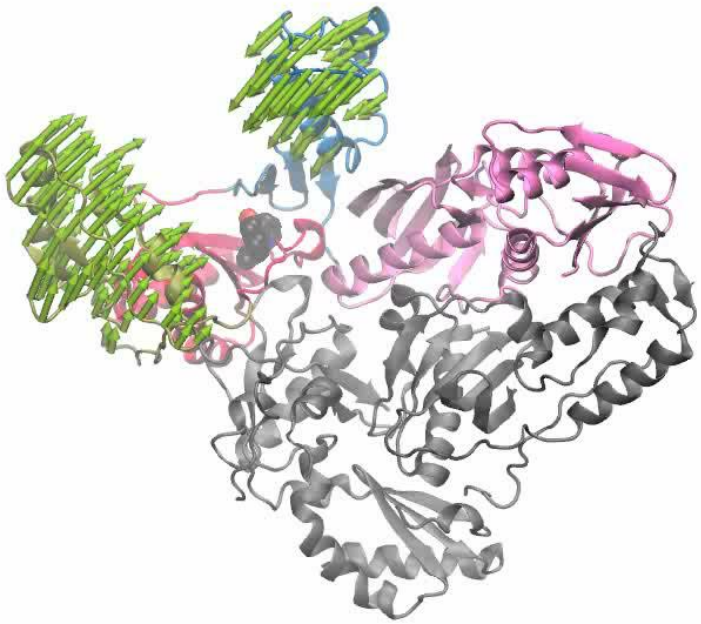
$$\mathbf{C}^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta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

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

Induced Dynamics or Intrinsic Dynamics?



Experiments



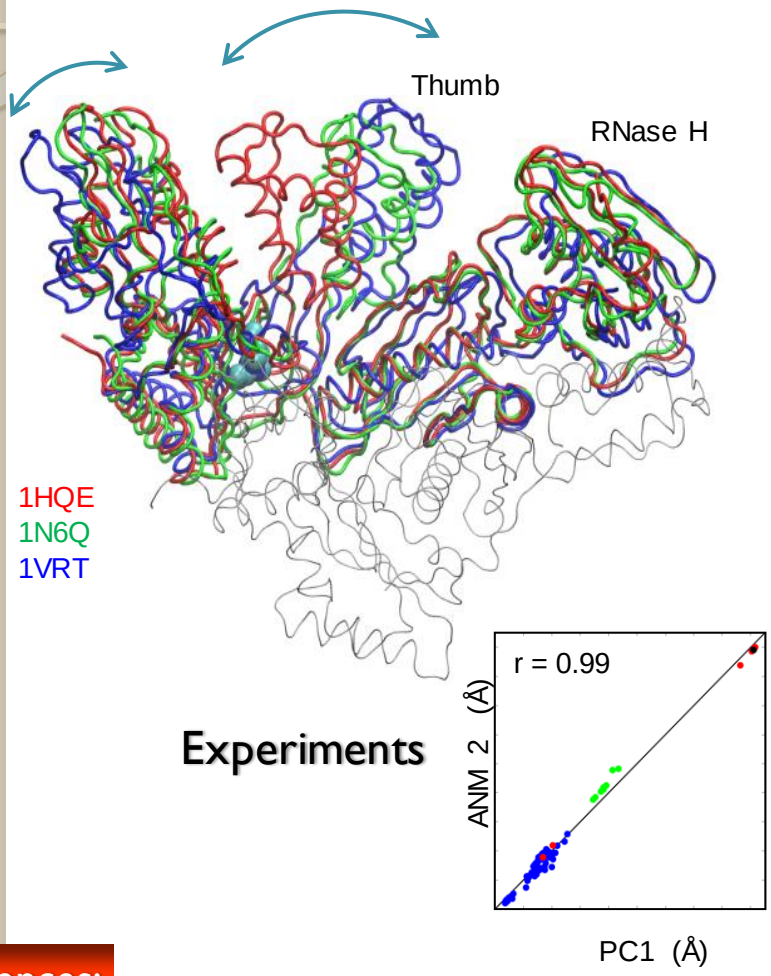
Theory

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

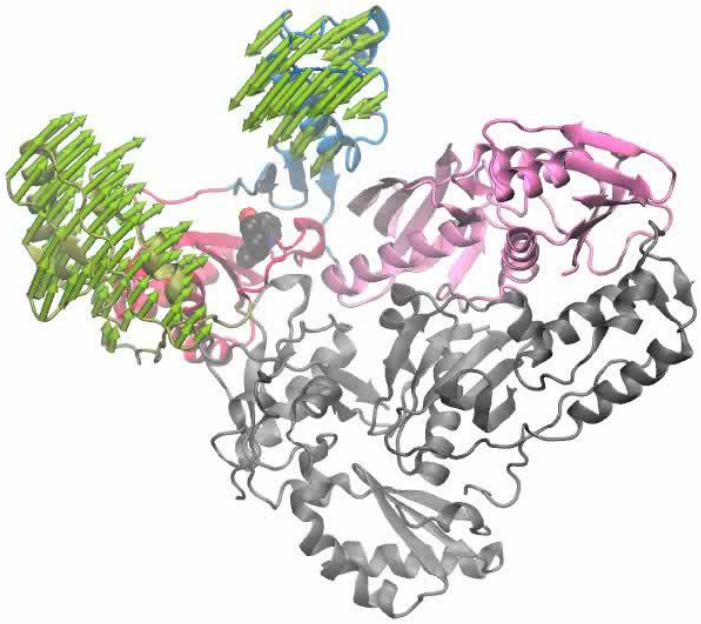
References:

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

Soft modes enable functional movements



Experiments



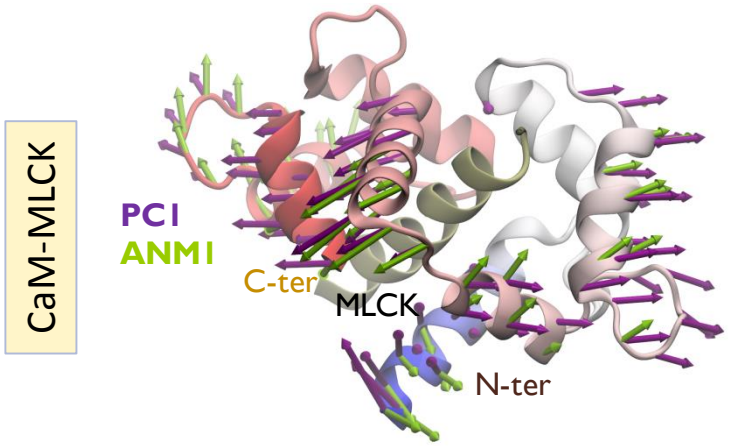
Theory

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

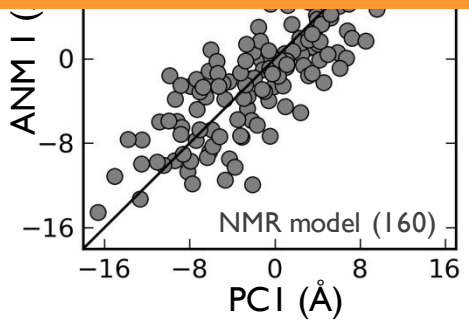
References:

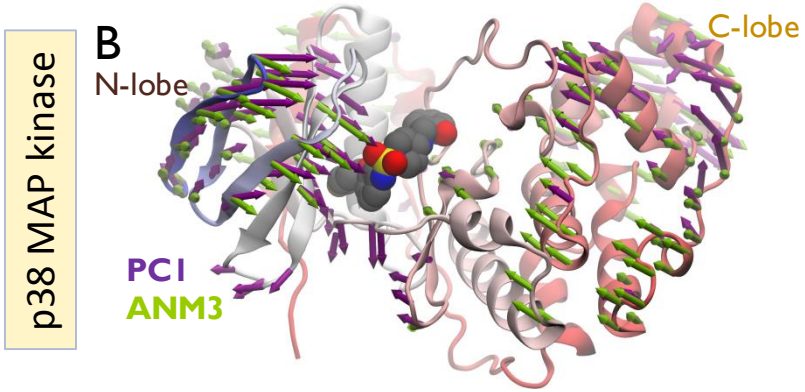
Bakan & Bahar (2009) PNAS **106**, 14349-54.

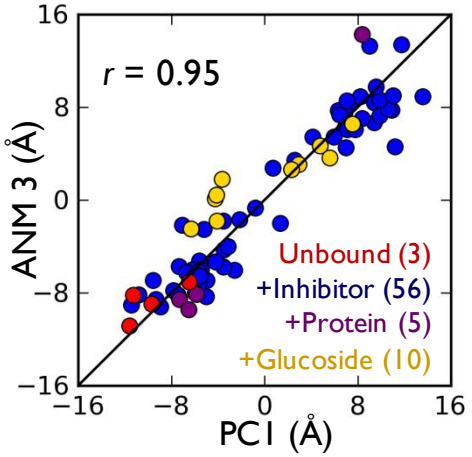
Experimental structures (for a given protein) are mainly variants along soft modes



Pre-existing paths







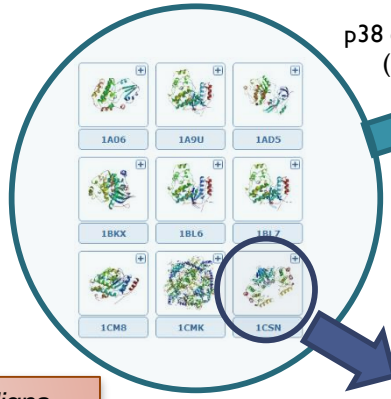
Meireles L, Gur M, Bakan A, Bahar I (2011) Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins *Protein Science* **20**: 1645-58.

ProDy for exploring conformational space

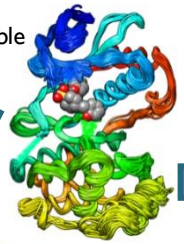
Protein Dynamics Analysis in Python

User inputs a protein sequence

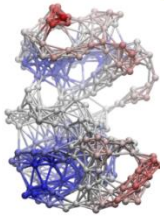
```
>1A9U:A|PDBID|CHAIN
GSSHHHHHHSSGLVPRGSHMSQER
PTFYRQELNKTIWEVPERYQNLSPV
GSGAYGSVCAAFDTKTGLRVAVKK
LSRPFQSIIHAKRTYRELRLLKHMKH
ENVIGLLDVFT.....
```



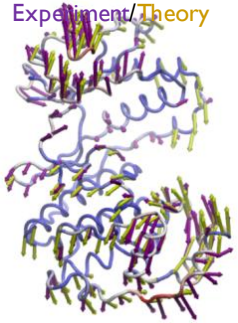
p38 ensemble (PCA)



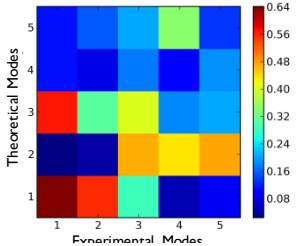
p38 network model (ANM)



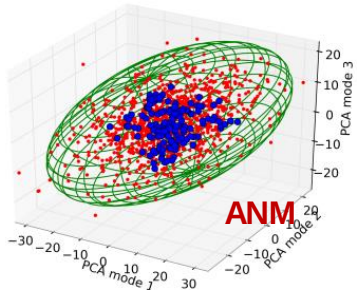
Experiment/Theory



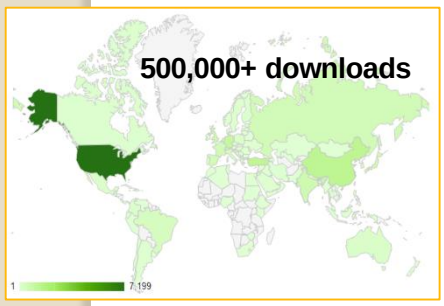
Overlap table



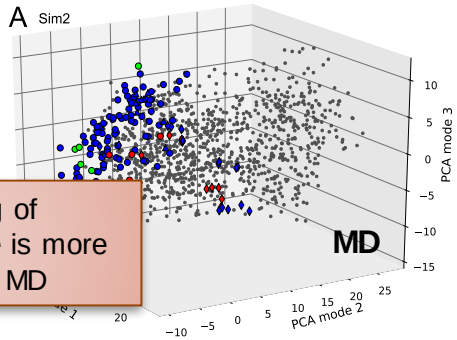
User can compare experimental and theoretical models



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



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



ProDy-ANM sampling of conformational space is more complete than that of MD

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

Major advantages of ProDy:

- Simplicity
- Visualizing the global dynamics
- Applicability to large systems
- Assessing cooperative motions
- Efficiency – immediate results
- Relevance to observables, to

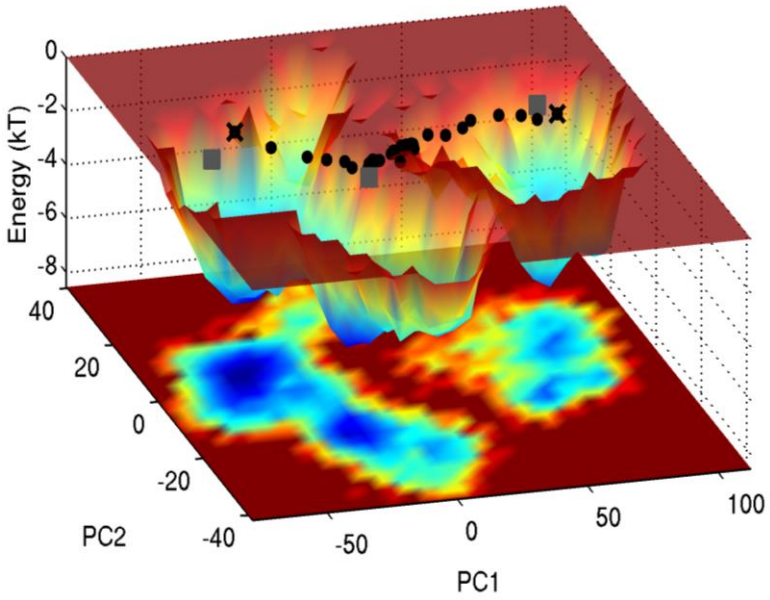
functional mechanisms & allostery

Disadvantages

- Low resolution approach
- No specific interactions
- Lack of atomic details
- Linear theory – applicable near energy minimum
- Requires structural data – not a tool for structure prediction

Co-MD: Guiding MD simulations by ANM modes

Dr. Mert Gur



ANM-guided transition pathways

- Isin B, Schulten K, Tajkhorshid E, Bahar I (2008) *Biophysical J* 95: 789-803.
- Yang Z, Májek P, Bahar I (2009) *PLoS Comput Biol* 5: e1000360.
- Gur M, Madura JD, Bahar I (2013) *Biophys J* 105:1643-52
- Das A, Gur M, Cheng MH, Jo S, Bahar I, Roux B (2014) *PLoS Comput Biol* 10: e1003521

coMD trajectories proceed along the minima of free energy landscape

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

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

Support from NIGMS, NLM, NIDDK & NIAID

Acknowledgment

Dr. Lee-Wei Yang
Assoc Prof, Tsinghua U, Taiwan

Dr. Ahmet Bakan
Kabbage, Inc.

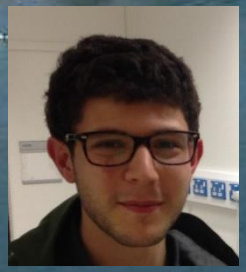
Dr. Ying Liu
Google, Inc

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

Anindita Dutta
Agilent Technologies

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

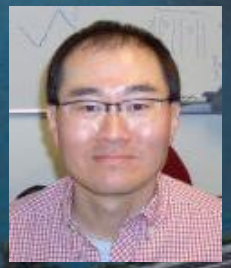
Cihan Kaya



Dr. James Krieger

Dr. Karolina Mikulska-Ruminska

Dr. Hongchun Li



Dr. JiYoung Lee

She (John)Zhang