Day 1 – Lecture 2

Bridging Structure and Function, Experiments and Computations

Ivet Bahar and Pemra Doruker

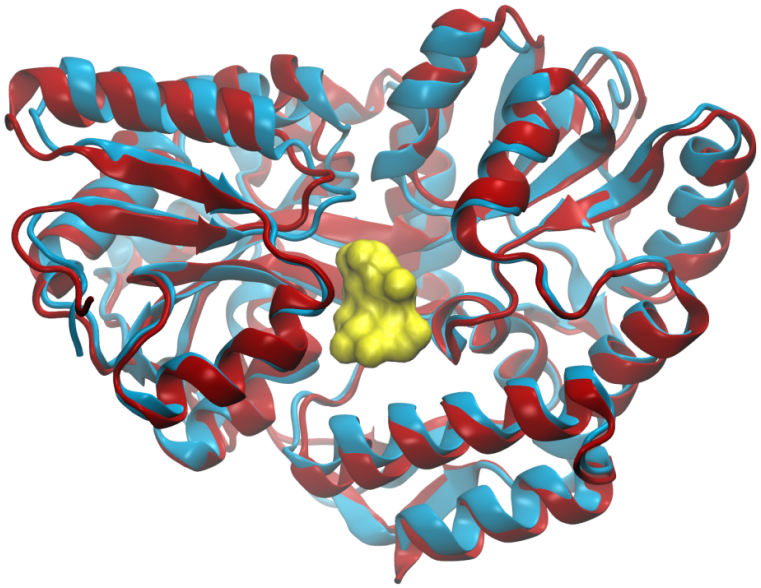
Department of Computational and Systems Biology
School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260

Summary

1. Theory
   a. Gaussian Network Model (GNM)
   b. Anisotropic Network Model (ANM)
   c. Resources/Servers/Databases (ProDy, DynOmics)
2. Bridging Sequence, Structure and Function
   a. Ensemble analysis and functional modes of motion
   b. Combining sequence and structure analyses – signature dynamics
   c. Modeling membrane proteins and lipid environment with ANM
3. Allostery and druggability
   a. Essential site scanning and allosteric pocket prediction
   b. Druggability simulations

Proteins exploit pre-existing soft modes for their interactions

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

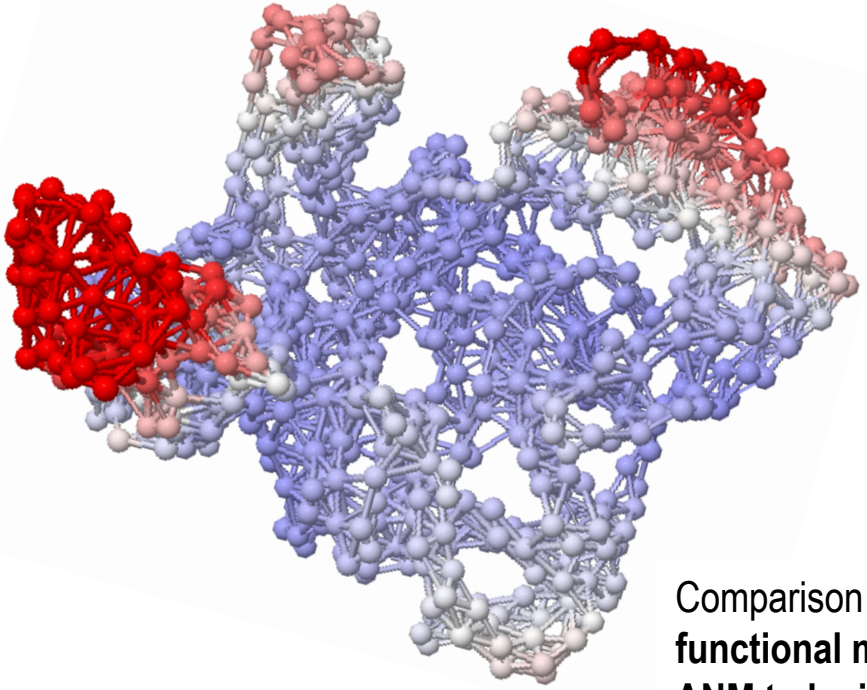


maltodextrin binding protein
Unbound/Bound

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

Allosteric changes in conformation

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

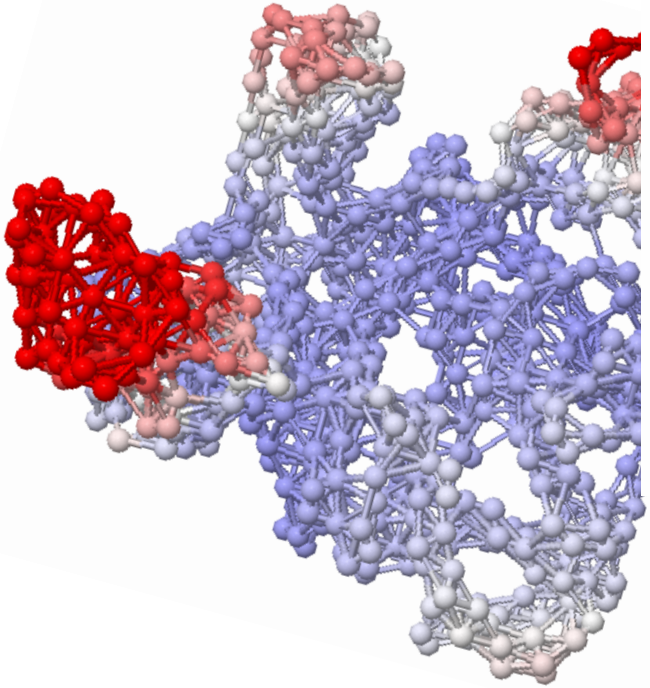


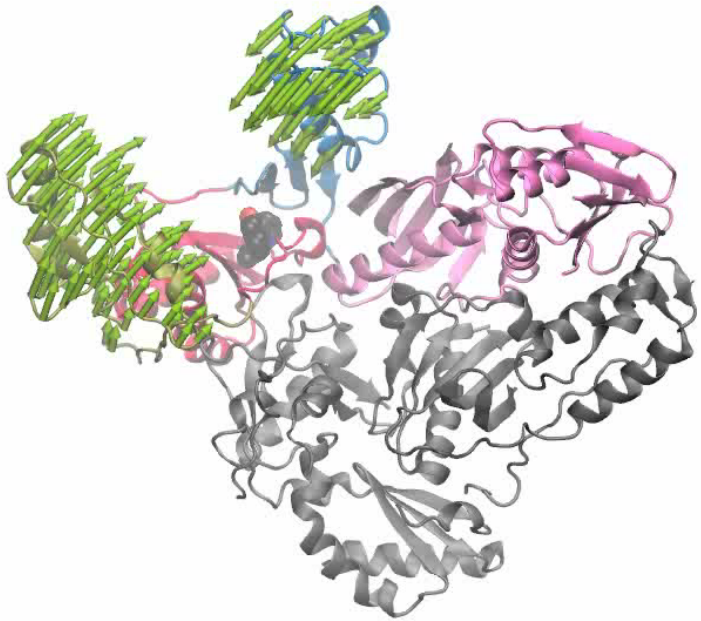
HIV Reverse Transcriptase (RT)
Red: most mobile
Blue: most constrained

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

Allosteric changes in conformation

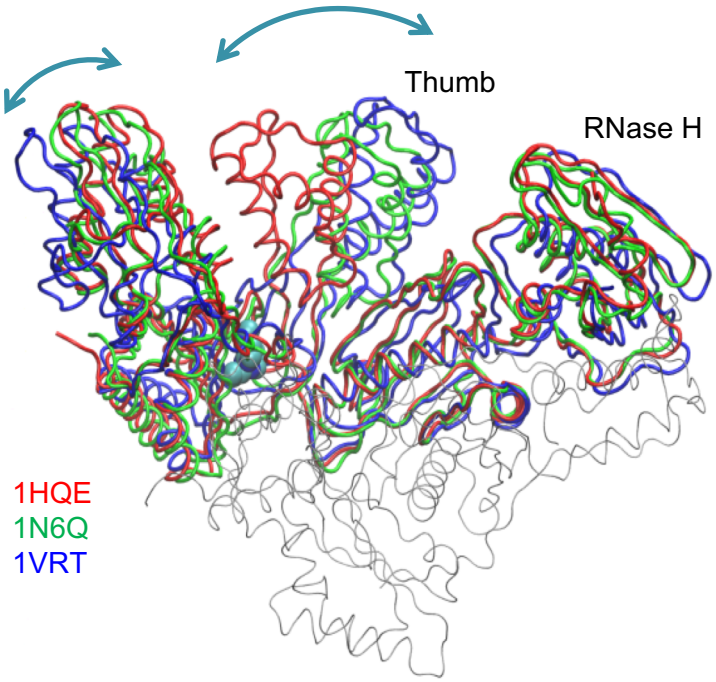
Elastic Network Models are particularly useful for exploring the cooperative motions of large multimeric structures



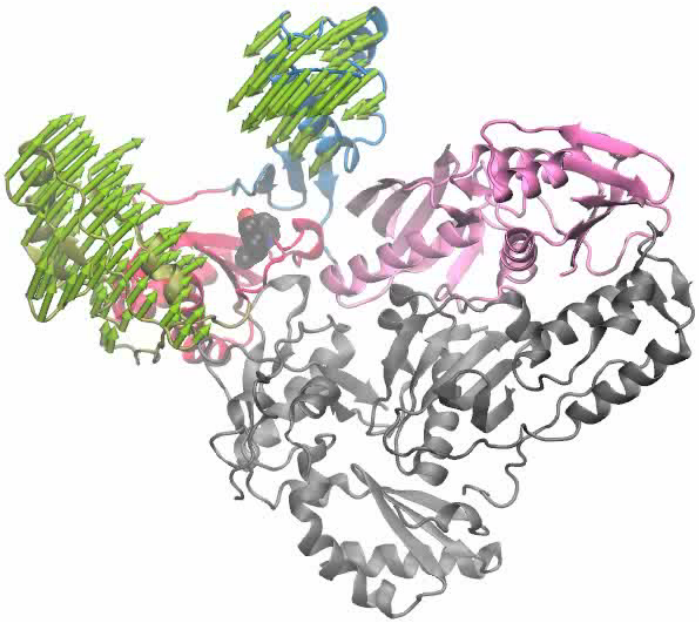


Comparison with experimental data shows that **the functional movements are those predicted by the ANM to be intrinsically encoded by the structure**

Induced Dynamics or Intrinsic Dynamics?



Experiments



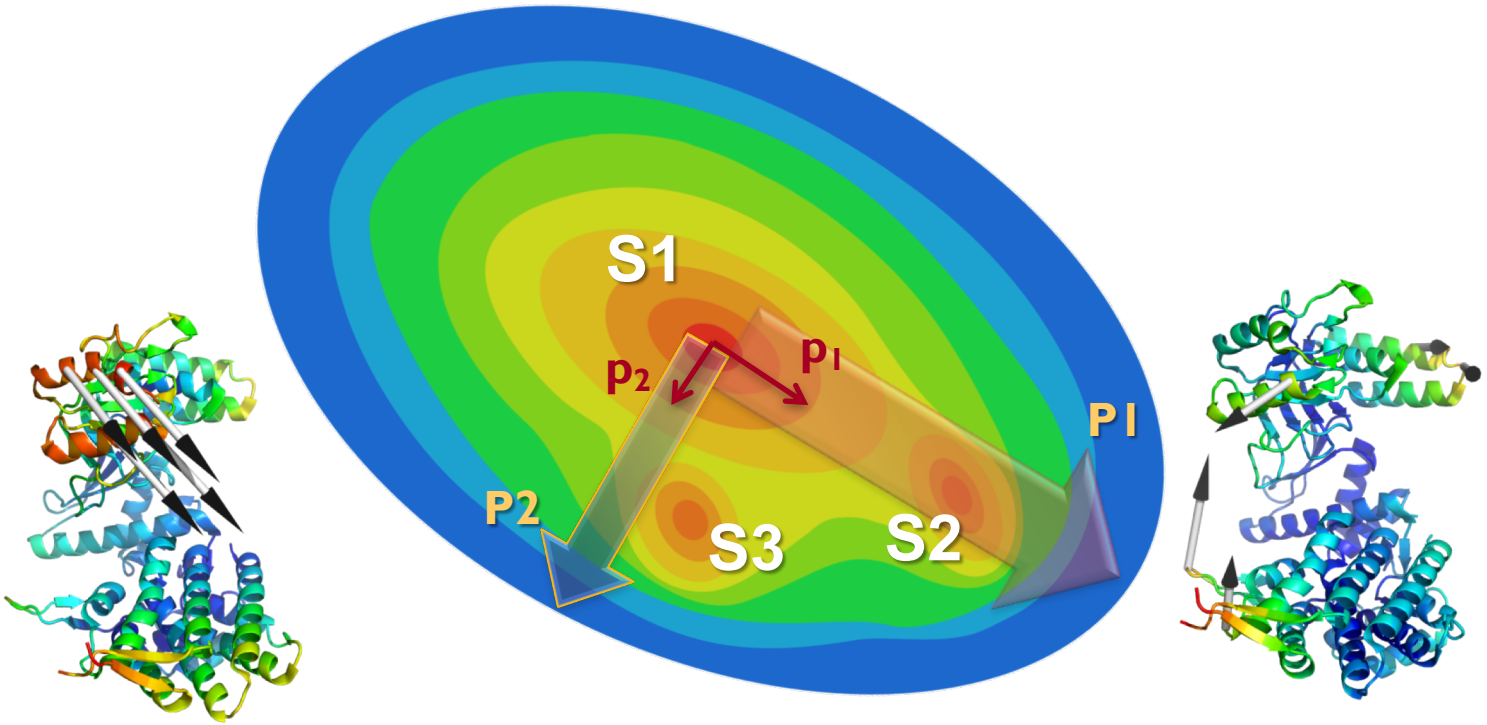
Theory

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

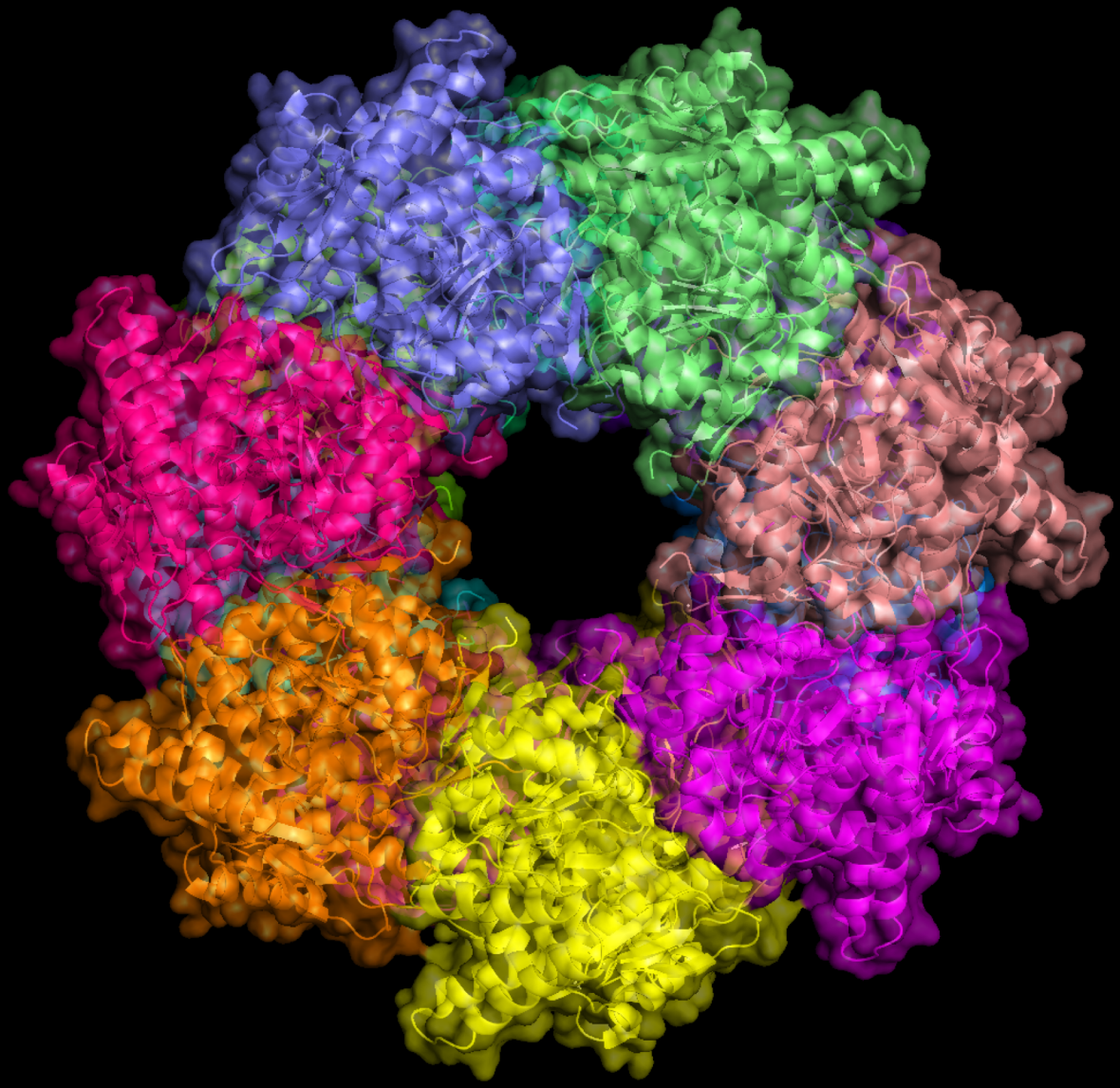
References:

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

Substates may be identified along soft modes

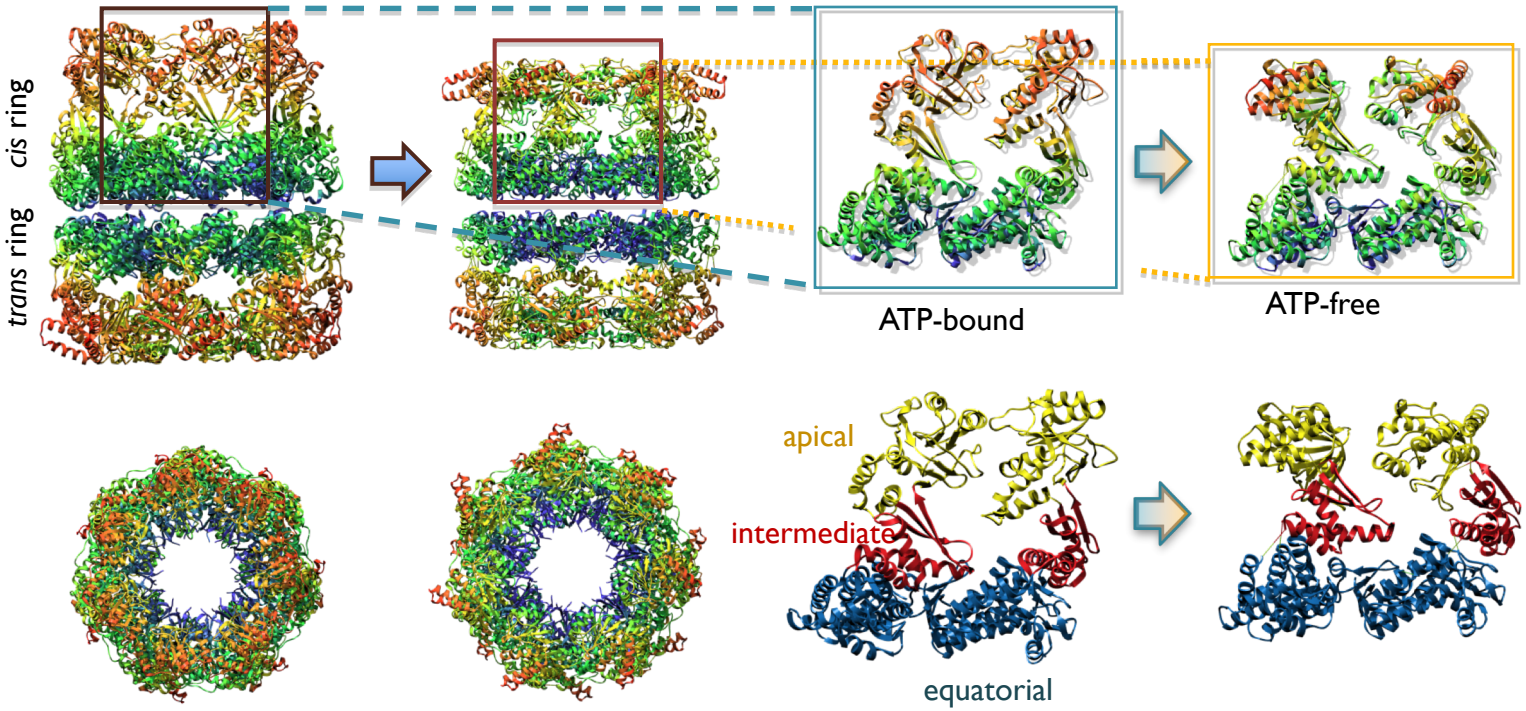


Bacterial chaperonin GroEL: an allosteric machine



GroEL Allosteric Dynamics

Passage between the R and T states



See...

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

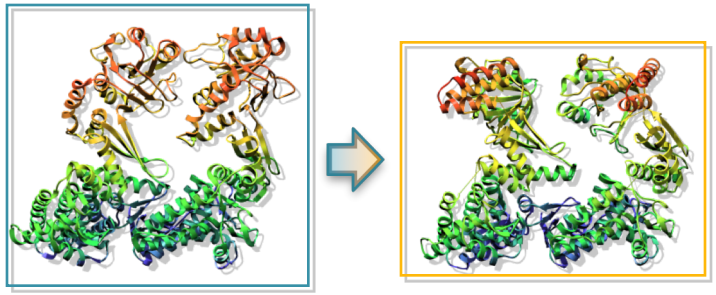
Computations

ANM yields a series of 3N dimensional **deformation vectors**

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

Given by ANM eigenvectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, …. $\mathbf{v}_{3N-6}$, with respective frequencies proportional to κ_1, κ_2, κ_3,…. κ_{3N-6}

Experiments



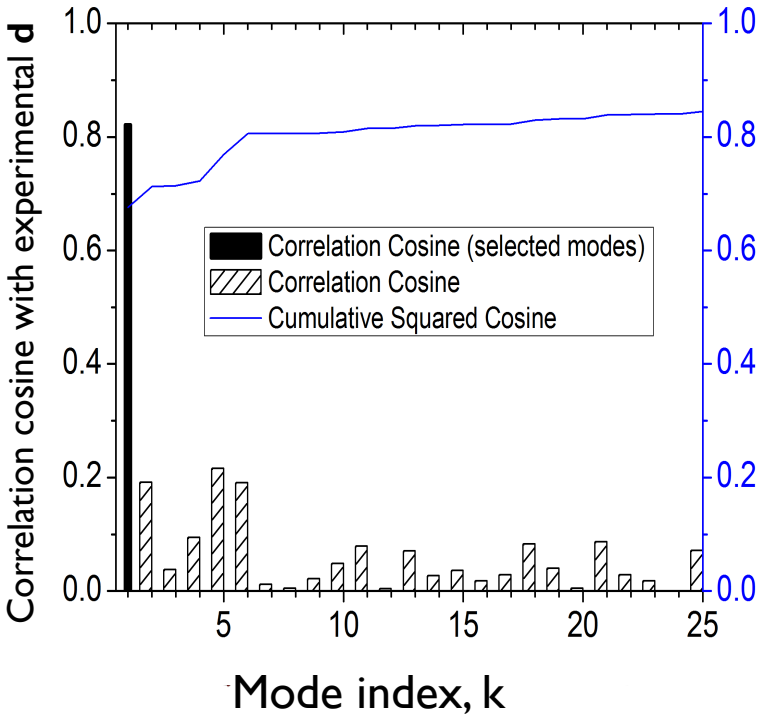
$$\mathbf{d} = [\Delta x_1 \;\; \Delta y_1 \;\; \Delta z_1 \;\ldots\;\; \Delta z_N]^T$$

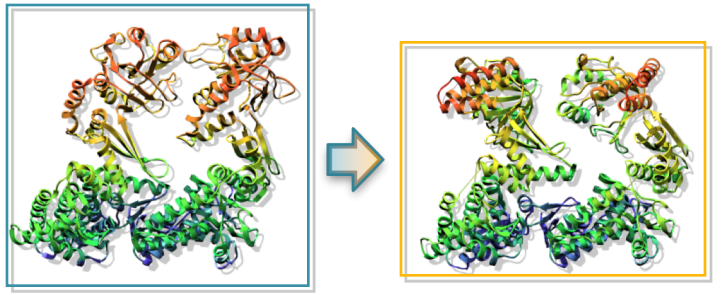
See...

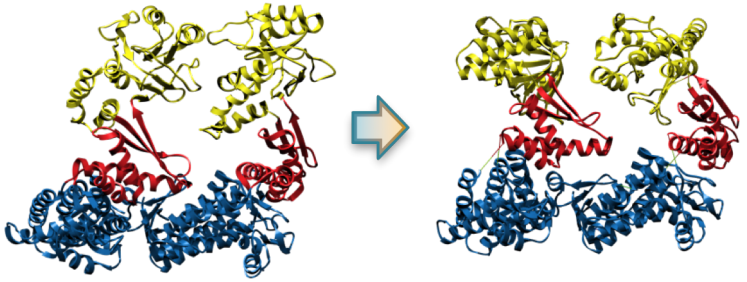
Z Yang, P Marek and I Bahar, *PLoS Comp Biology* 2009

What is the overlap between computations and experiments?

Correlation cosine between $\mathbf{v}_k$ and $\mathbf{d}$







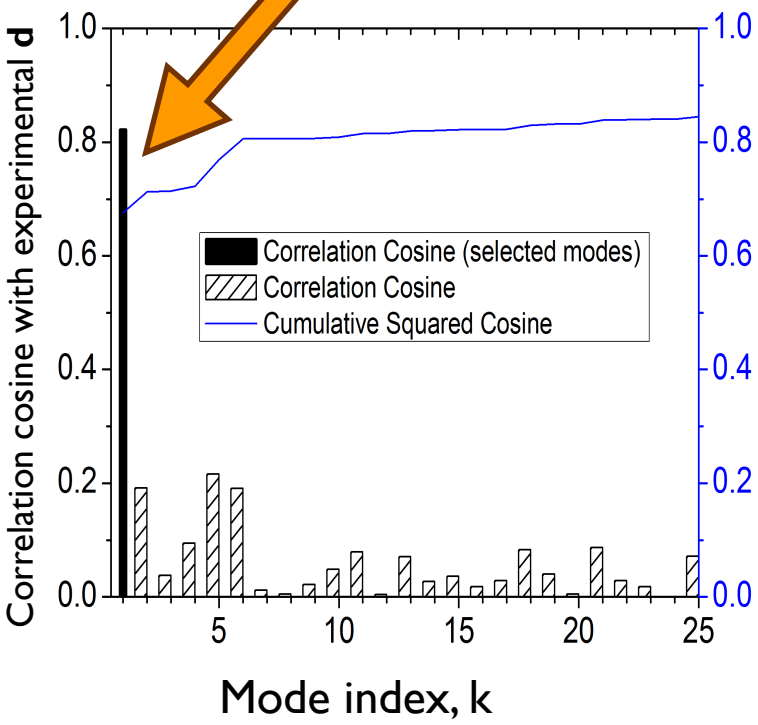
$$\mathbf{d} = [\Delta x_1 \;\; \Delta y_1 \;\; \Delta z_1 \;\; \ldots \;\; \Delta z_N]^T$$

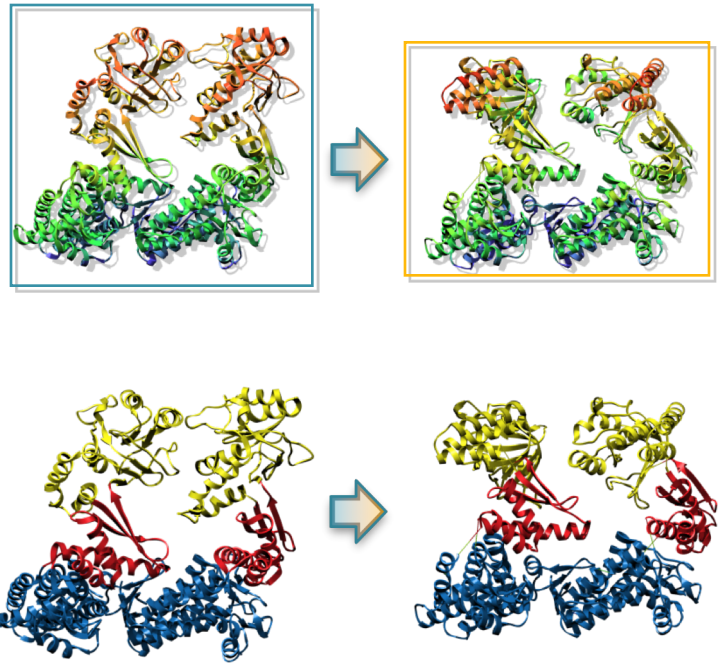
See...

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)





$$CO(m) = \left(\sum_{k=1}^{m} (\boldsymbol{v}_k \cdot \frac{\boldsymbol{d}}{|\boldsymbol{d}|})^2 \right)^{1/2}$$

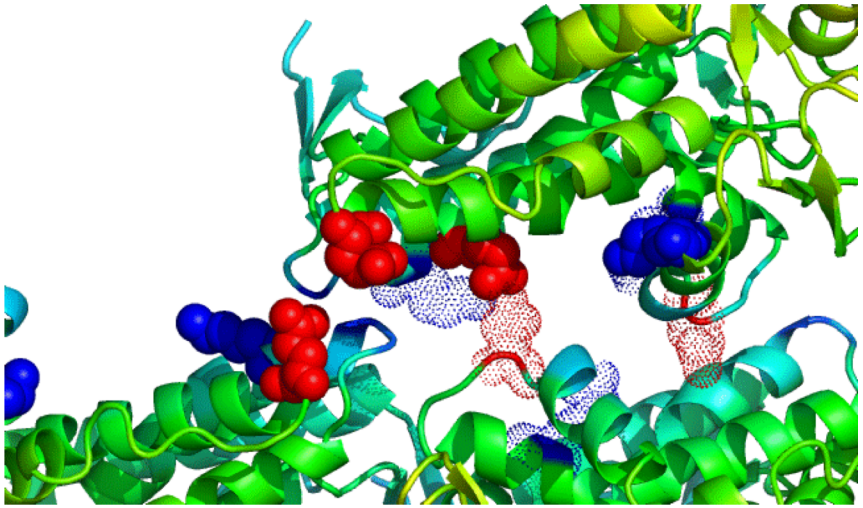
$$\mathbf{d} = [\Delta x_1 \;\; \Delta y_1 \;\; \Delta z_1 \;\ldots\;\; \Delta z_N]^T$$

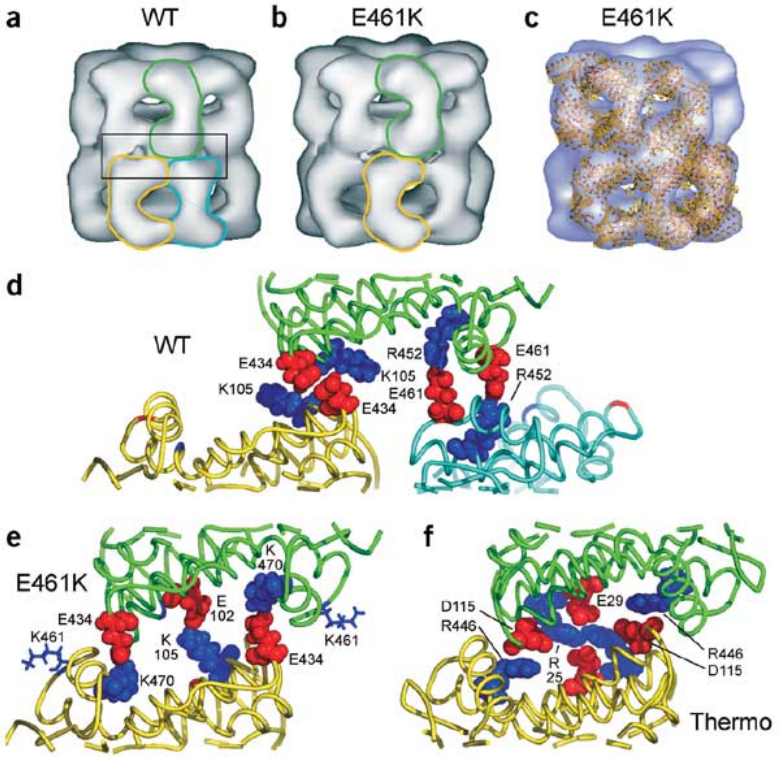
See...

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





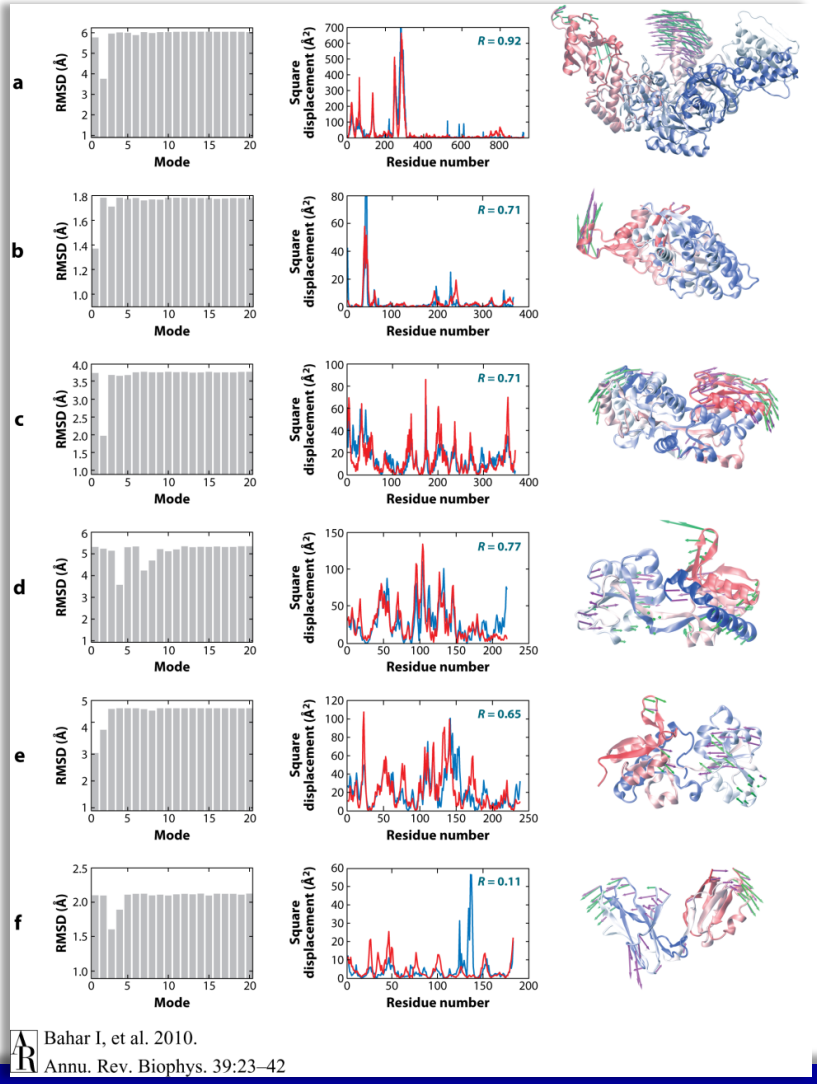
E461K mutation causes disruption of inter-ring transfer of ATP-induced signal (Sewell et al NSB 2004)

Yang et al. *Mol Biosyst* 2008

Experimentally observed structural changes are usually reconfigurations along soft modes

- Correlation cosine of 0.75 ± 0.15 between one of the softest modes and the experimentally observed change in structure

- Significant decrease in RMSD between the endpoints upon moving along a single soft mode (out of 3N-6 modes)



Bahar I, et al. 2010.
Annu. Rev. Biophys. 39:23–42

See...

A better comparison:

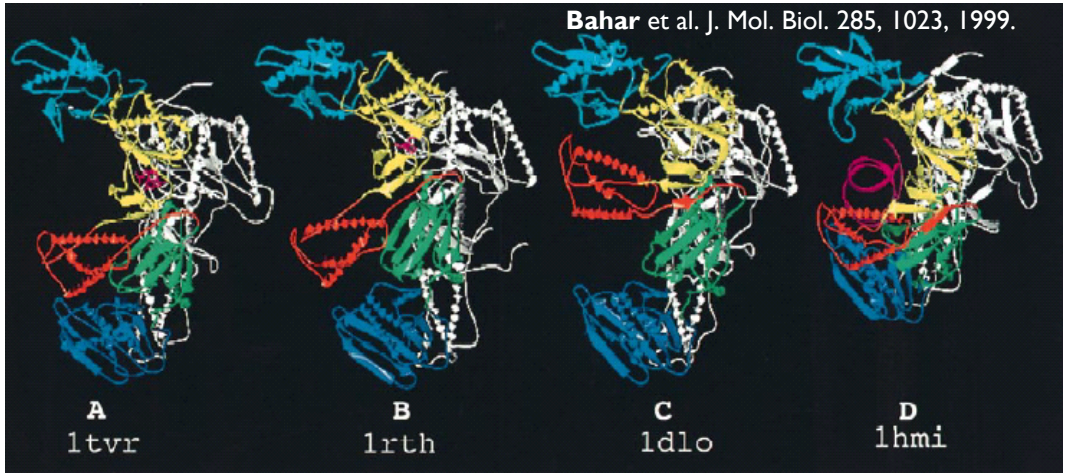
Consider more than 2 end points for a given structure, but all the known structures for a given protein, or the structurally resolved

Ensemble of structures

Bakan A, Bahar I (2009) The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc Natl Acad Sci USA* **106**: 14349-14354.

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.

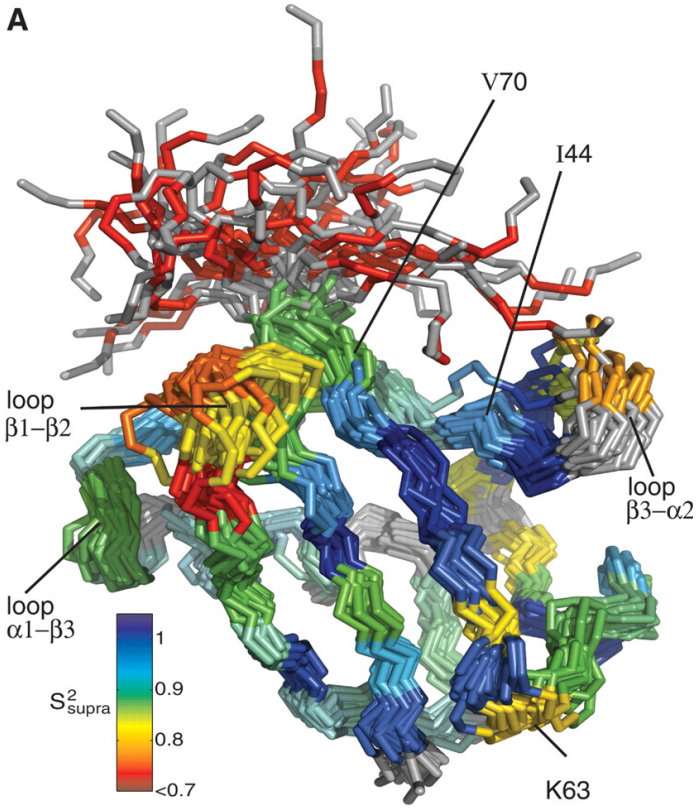


Different structures resolved for HIV-1 reverse transcriptase (RT)

RESEARCH ARTICLES

Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution

Oliver F. Lange, …, Jens Meiler, Helmut Grubmüller, Christian Griesinger, Bert L. de Groot



The ensemble covers the complete structural heterogeneity observed in 46 ubiquitin crystal structures, mostly complexes with other proteins.

- Conformational selection, rather than induced-fit explains the molecular recognition dynamics of ubiquitin.

- A concerted mode accounts for molecular recognition heterogeneity

Reference

Lange et al. (2008) *Science* 320:1471 - 1475

Ensembles of structures

- Structural changes accompanying substrate (protein) binding

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

Ubiquitin
140 structures
1732 models

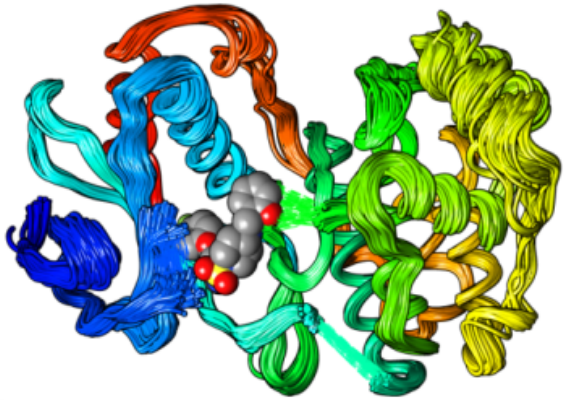
Ensembles of structures

- Structural changes accompanying substrate (protein) binding

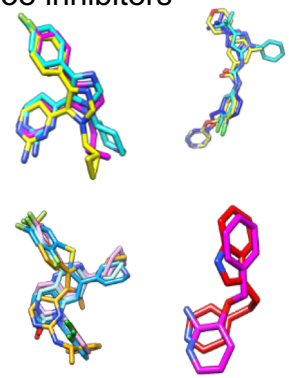
- Structural changes induced by, or stabilized upon, ligand binding

Ubiquitin
140 structures
1732 models

p38 MAP kinase
(**182 structures**)

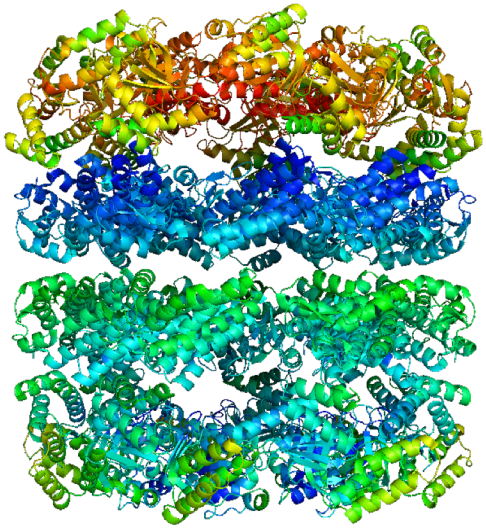


p38 inhibitors



Ensembles of structures

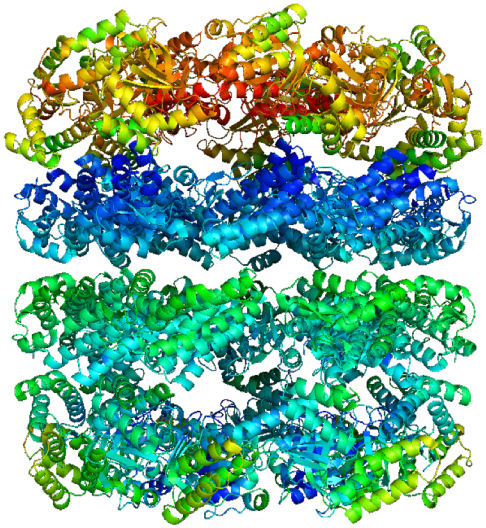
- Structural changes accompanying substrate (protein) binding
- Structural changes induced by, or stabilized upon, ligand binding
- Alternative conformations sampled during allosteric cycles



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 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 changes

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

What is Ensemble Analysis?

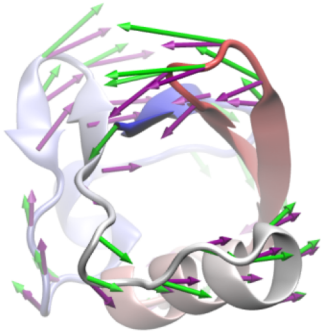
Principal component analysis

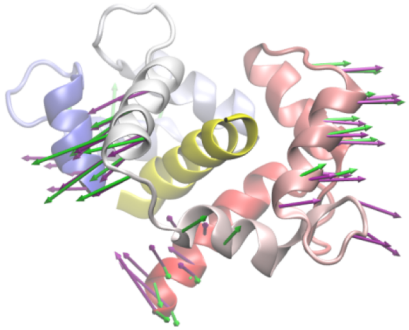
- ANM analysis
- Select a representative structure (e.g. with minimal RMSD from others)
- Theoretical
- Decompose either **H** or **C** into a series of modes (3N-6 eigenvectors)

PCA

- Superimpose/align the structures
- Experimental
- Decompose it into a series of modes of covariance (3N-6 eigenvectors)

Global motions inferred from theory and experiments

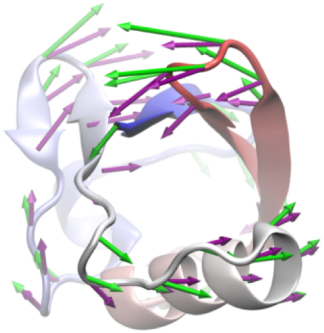


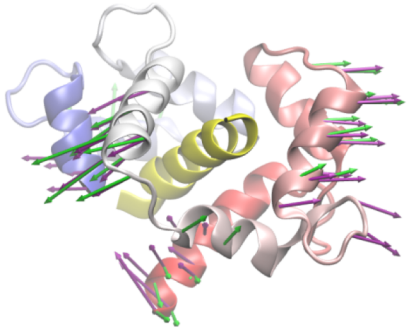


→ PCA of the ensemble of resolved structures

→ ANM analysis of a single structure from the ensemble

Global motions inferred from theory and experiments





PNAS

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*

Covariance matrix (NxN)

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

ΔR = N-dim vector of instantaneous fluctuations ΔR_i for all residues ($1 \leq i \leq N$)

$\langle \Delta R_1 \cdot \Delta R_1 \rangle$ = ms fluctuation of site 1 averaged over all *m* snapshots.

Covariance matrix (3Nx3N)

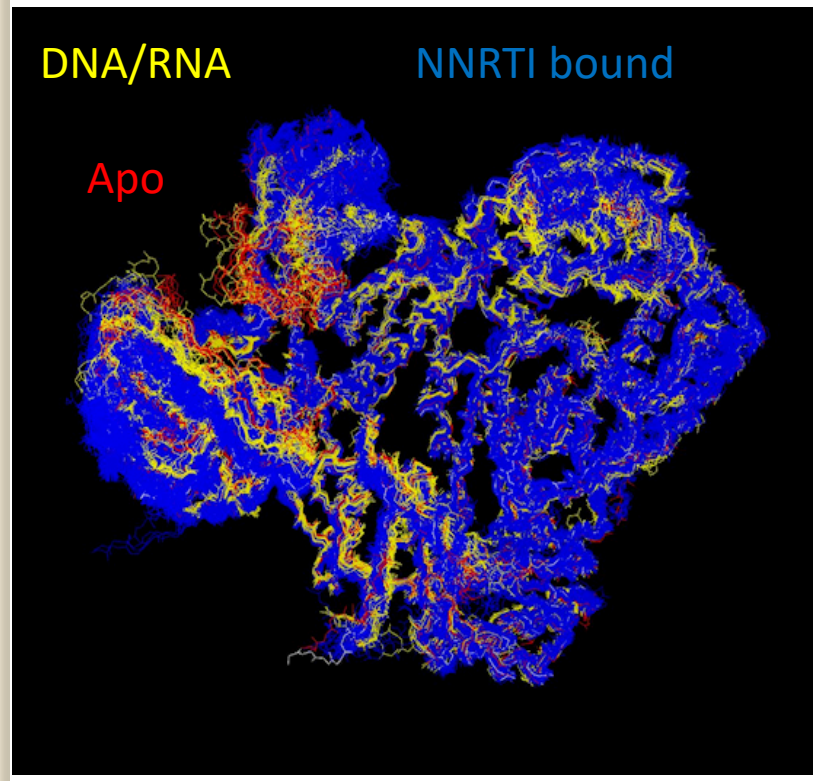
$C_{3N} =$

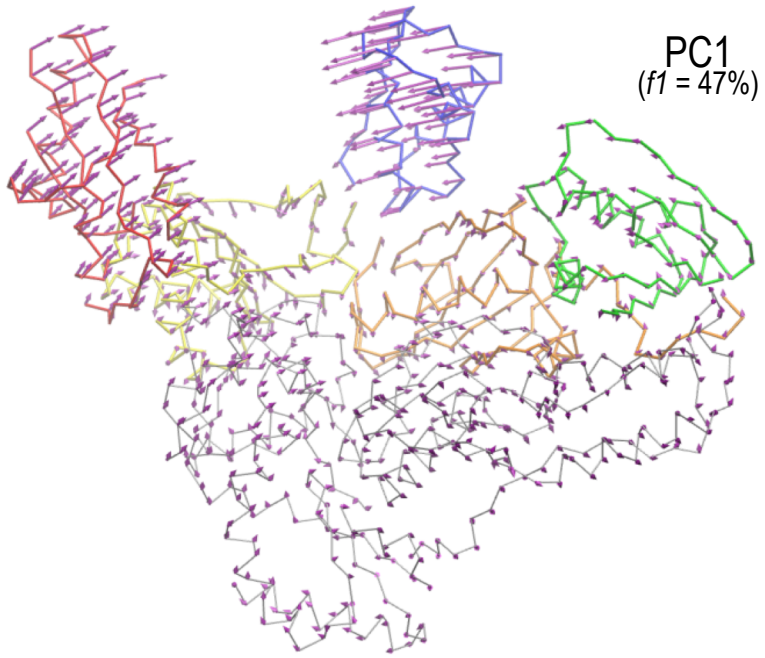
C_{11}	C_{21}	C_{13}		C_{1N}
C_{12}	C_{22}			
C_{N1}				C_{NN}

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)

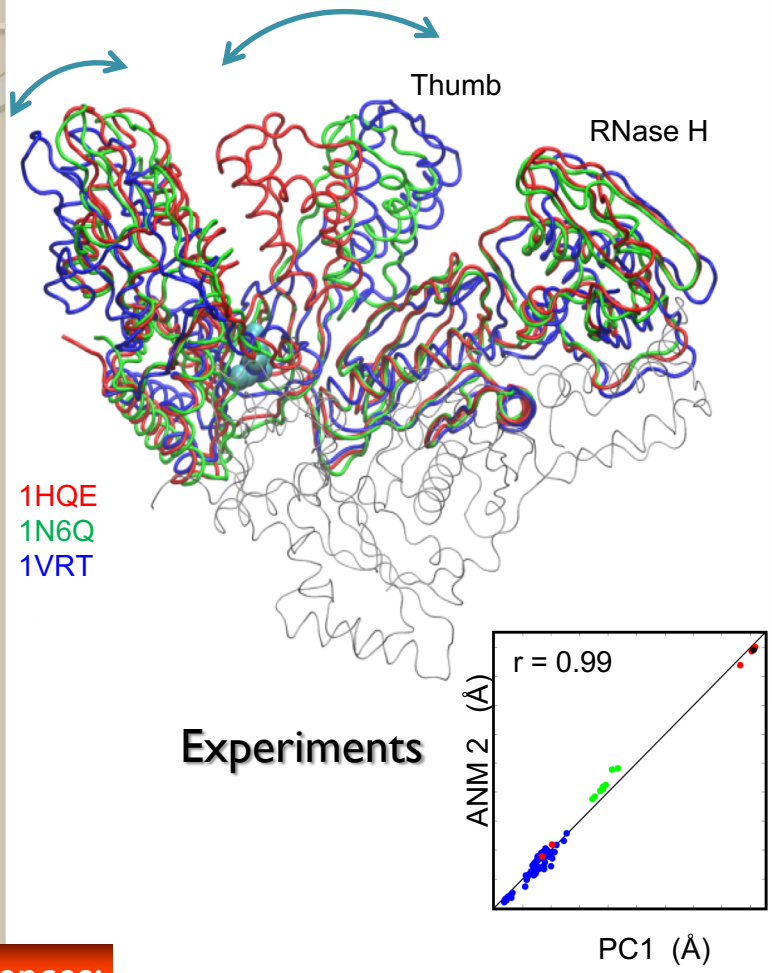




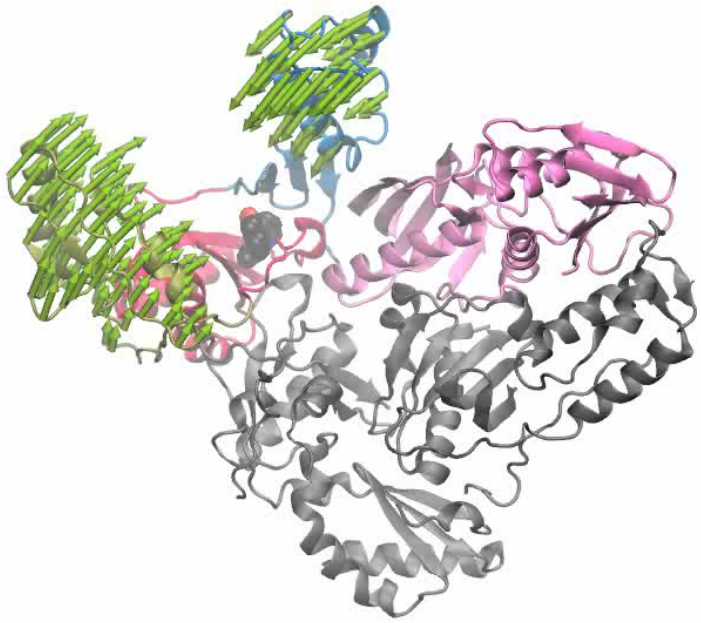
$$\mathbf{C}^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta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

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

Soft modes enable functional movements



Experiments



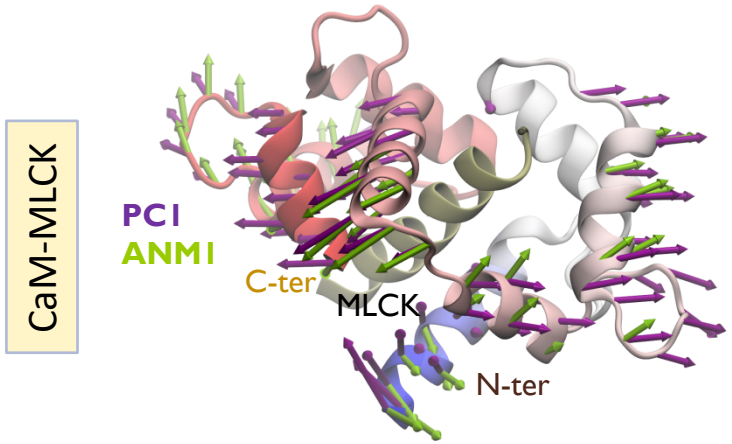
Theory

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

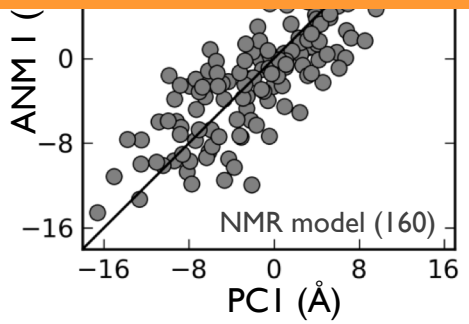
References:

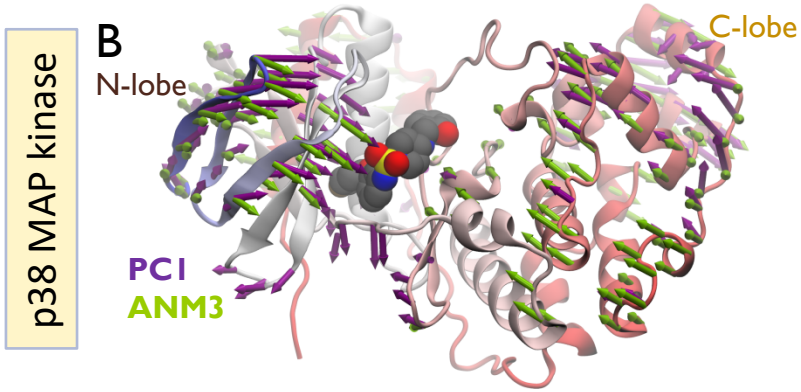
Bakan & Bahar (2009) PNAS **106**, 14349-54.

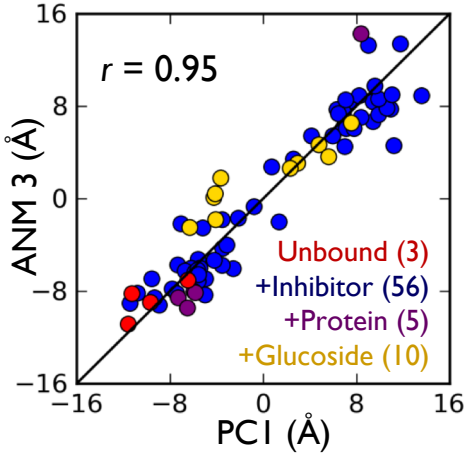
Experimental structures (for a given protein) are mainly variants along soft modes



Pre-existing paths







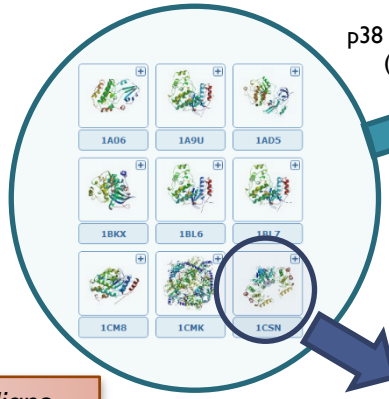
Meireles L, Gur M, Bakan A, Bahar I (2011) Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins *Protein Science* **20**: 1645-58.

ProDy for exploring conformational space

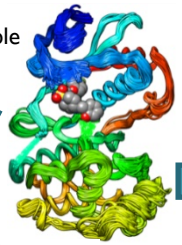
Protein Dynamics Analysis in Python

User inputs a protein sequence

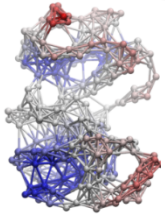
```
>1A9U:A|PDBID|CHAIN
GSSHHHHHHSSGLVPRGSHMSQER
PTFYRQELNKTIWEVPERYQNLSPV
GSGAYGSVCAAFDTKTGLRVAVKK
LSRPFQSIIHAKRTYRELRLLKHMKH
ENVIGLLDVFT.....
```



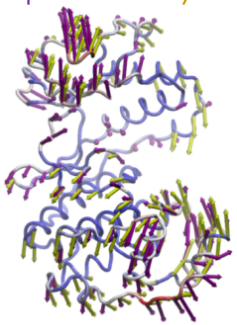
p38 ensemble (PCA)



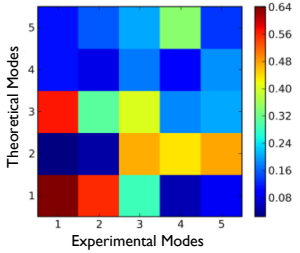
p38 network model (ANM)



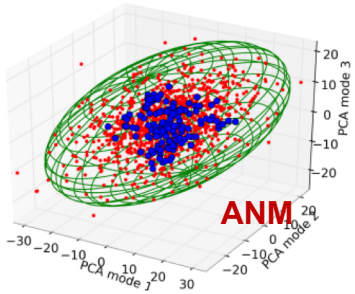
Experiment/Theory



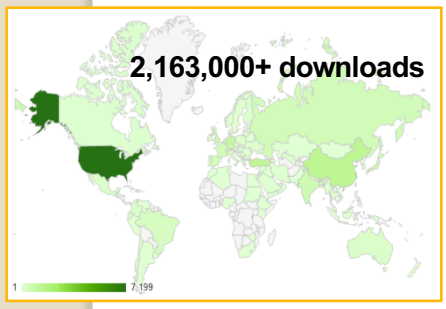
Overlap table



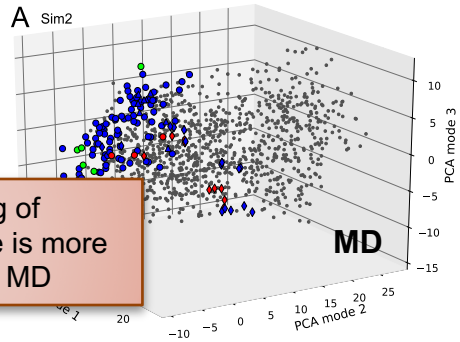
User can compare experimental and theoretical models



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



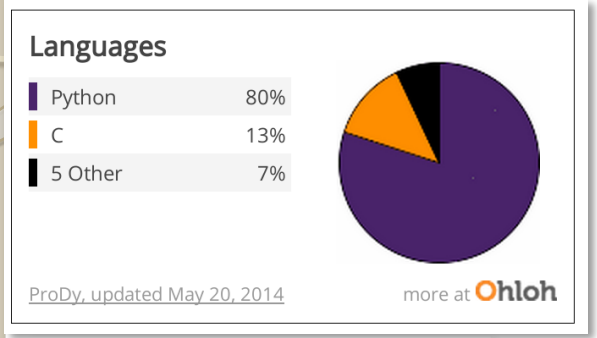
Source http://www.google.com/analytics/



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

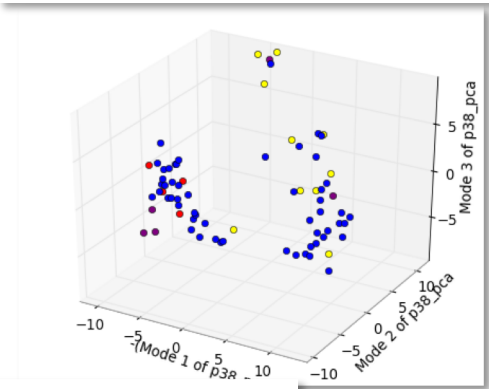
ProDy: An Interactive Tool

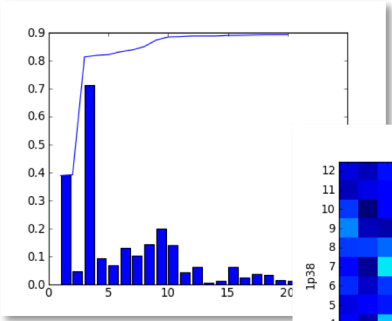


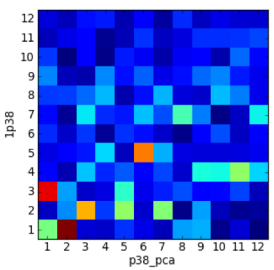
matplotlib

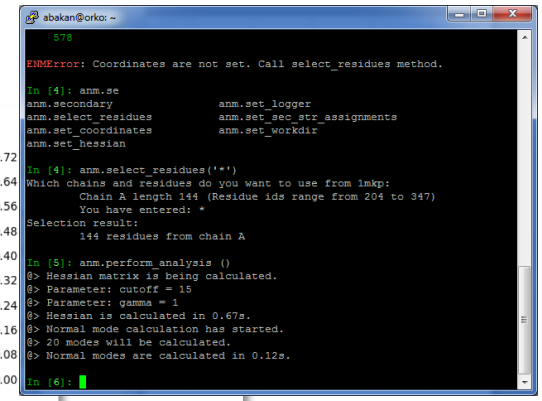
NumPy

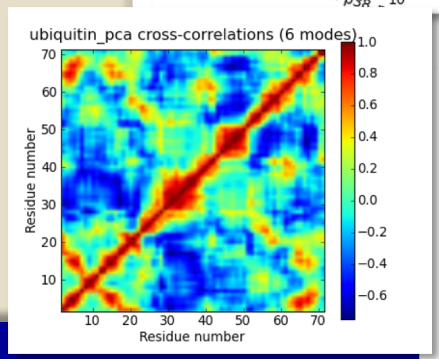
IP[y]: IPython
Interactive Computing

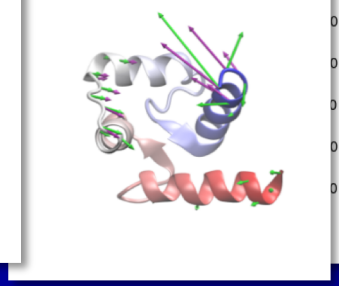


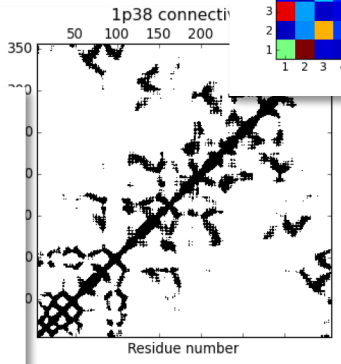


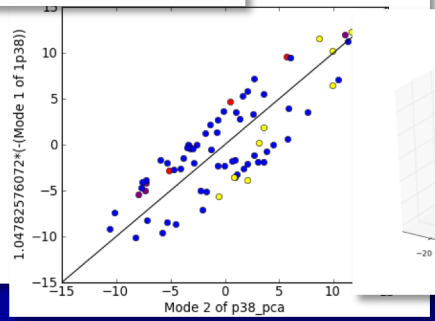


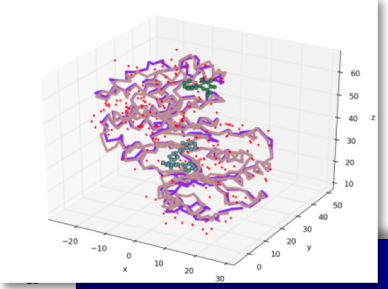








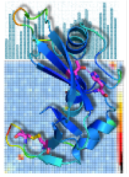




Suite of tools

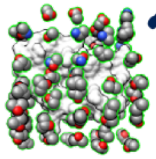
ProDy
Protein Dynamics & Sequence Analysis

Elastic Network Model (ANM/GNM) Analysis
Principal component analysis of experimentally resolved structures



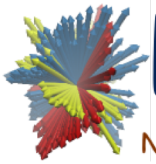
Evol
Bridging Evolution & Dynamics

Multiple Sequence Alignment
Sequence conservation
Correlated Mutations



DruGUI
Druggability Suite for VMD

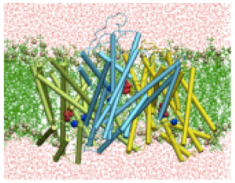
Computational Drug Discovery
Binding Site Prediction
Affinity Estimation



NMWiz
Normal Mode Wizard for VMD

A VMD plugin
Visualization of collective motions
Animations/movies

Suite of tools

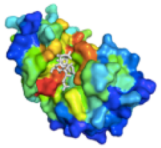


membrANM
Membrane Anisotropic Network Model

Modeling coupled protein-lipid dynamics
Useful for membrane proteins

PRS
Perturbation-Response Scanning

Propagation allosteric signals
Effector and sensor residues



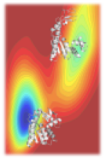
ESSA
Essential Site Scanning Analysis

Residues that can alter protein's essential dynamics upon binding
Prediction of allosteric pockets

SignDy
Signature Dynamics of Families

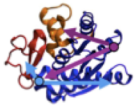
Shared global ENM mode profiles and departures from them, dynamics-based trees

Suite of tools



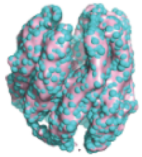
ClustENMD
Conformational Sampling

Unbiased conformational sampling of flexible and large biomolecular systems at atomic resolution and with high efficiency



coMD
Collective Molecular Dynamics

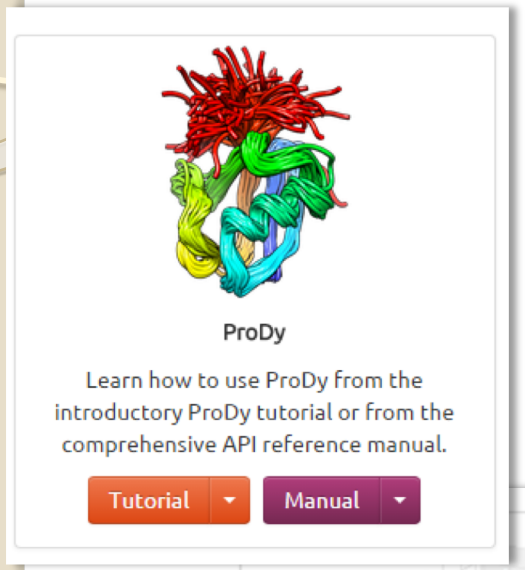
ENM guided MD simulations
Efficient sampling of energy landscape and transitions

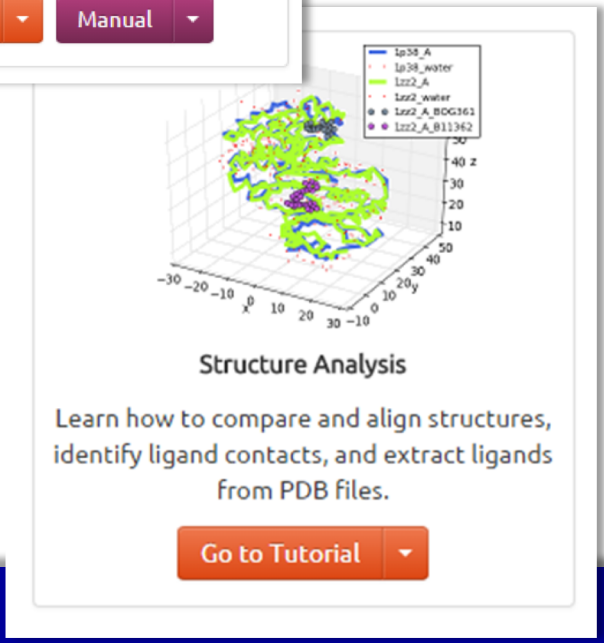


CryoDy
Dynamics from Cryo-EM Maps

dynamics of protein complexes resolved by cryo-EM

Tutorials: ProDy & Structure Analysis





- Retrieving PDB Files
- BLAST-Searching the PDB
- Constructing Biomolecular Assemblies
- Determining functional motions
- Aligning and Comparing Structures
- Identifying Intermolecular Contacts

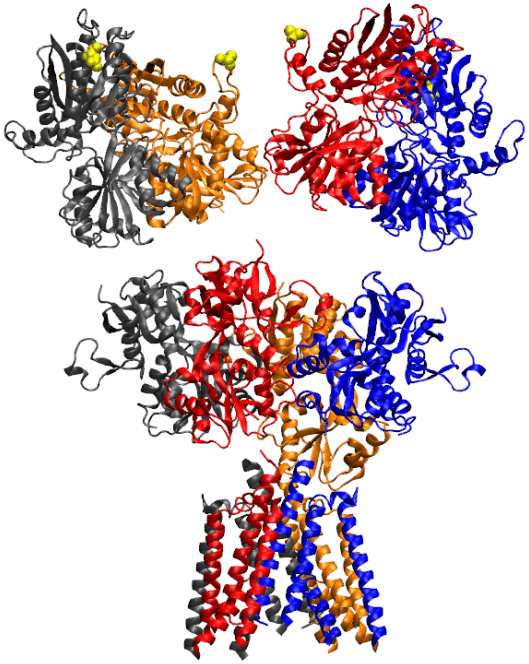
Major advantages of ProDy:

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

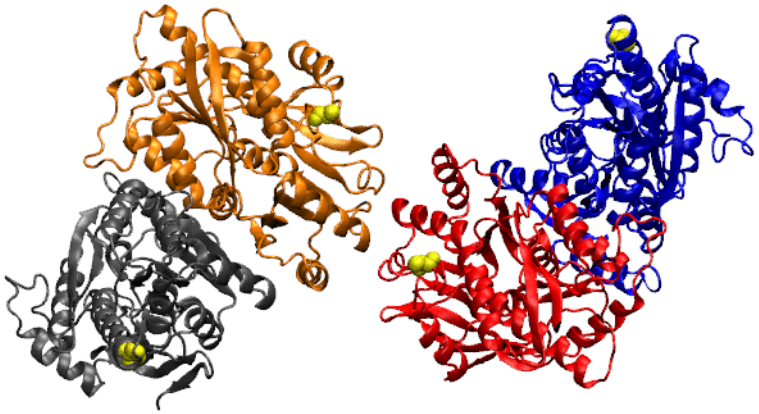
Caveats

- Low resolution approach
- No specific interactions
- Lack of atomic details
- Linear theory – applicable near an energy minimum
- not a tool for structure prediction (could be used for refinement)

Allosteric transition of AMPAR



Side view

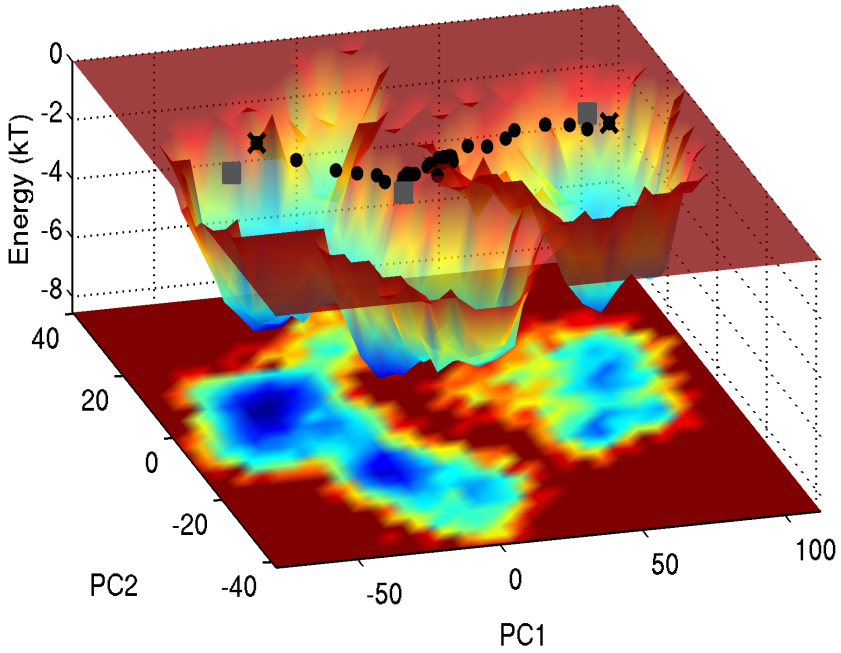


Top view

The trajectory was generated with adaptive-ANM (aANM) using the first 30 modes
Initial: N-shaped (PDB id: 4uqj) → Target: O-shaped (PDB id: 5ide) AMPAR

Hybrid methods to overcome caveats ANM-guided atomistic simulations

coMD
Collective Molecular Dynamics

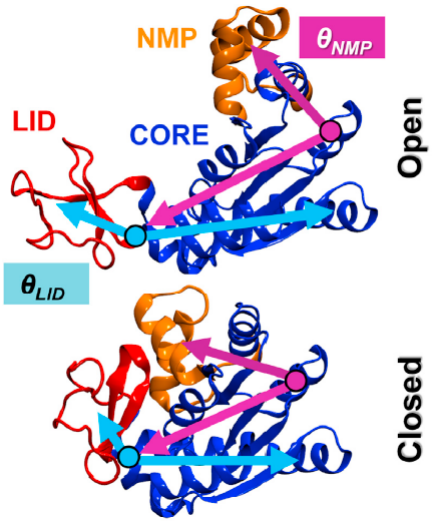


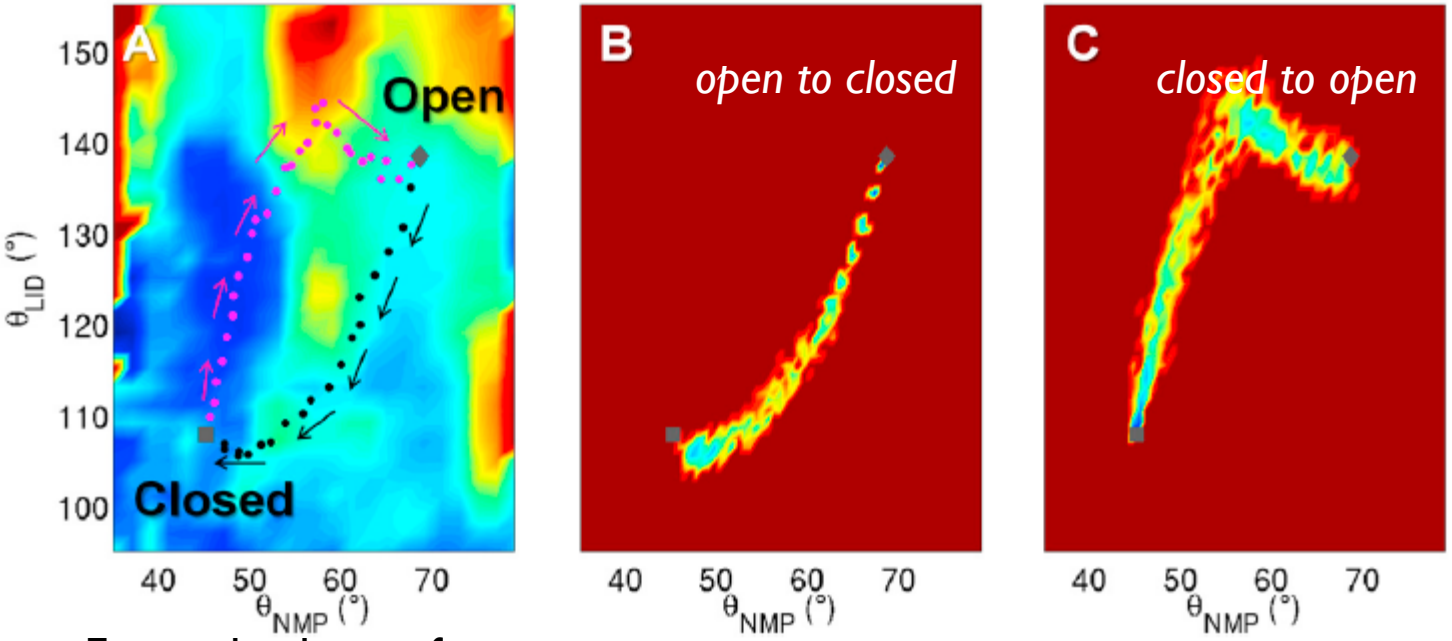
ANM-guided transition pathways

- 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

coMD transition pathways for adenylate kinase

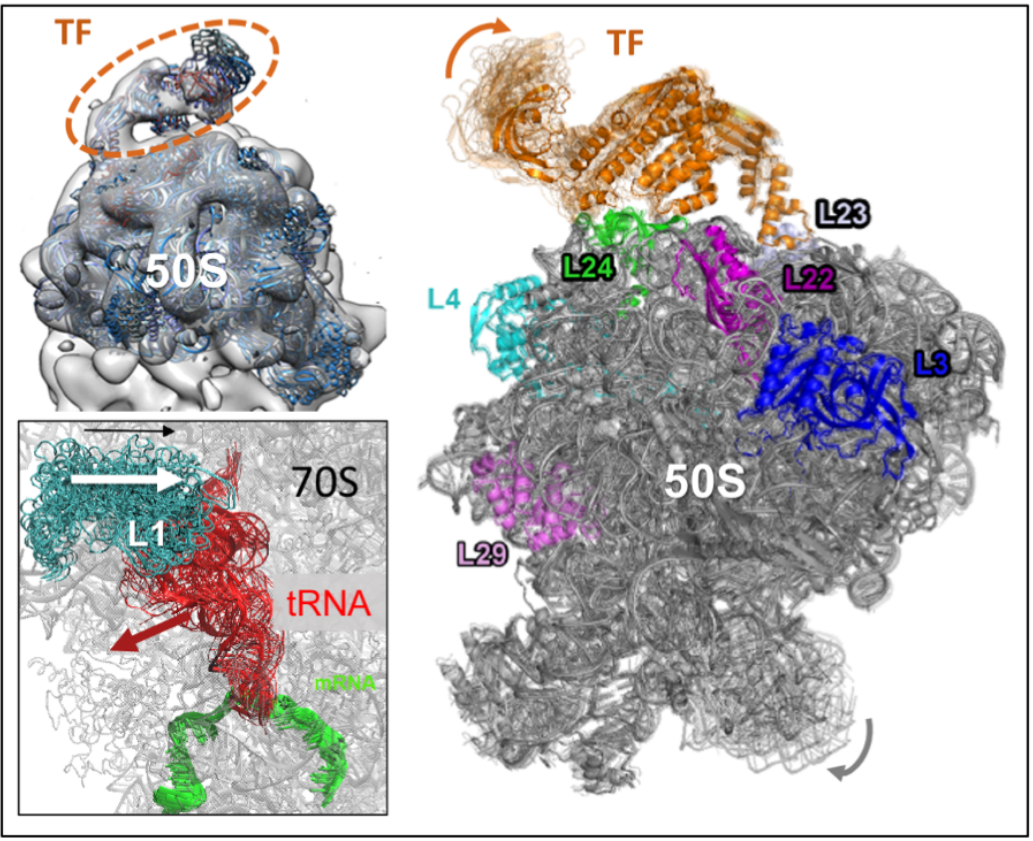




Energy landscape from Beckstein et al, JMB, 2009

Gur, Madura, Bahar, *Biophys J.* (2013) 105, 1643-1652

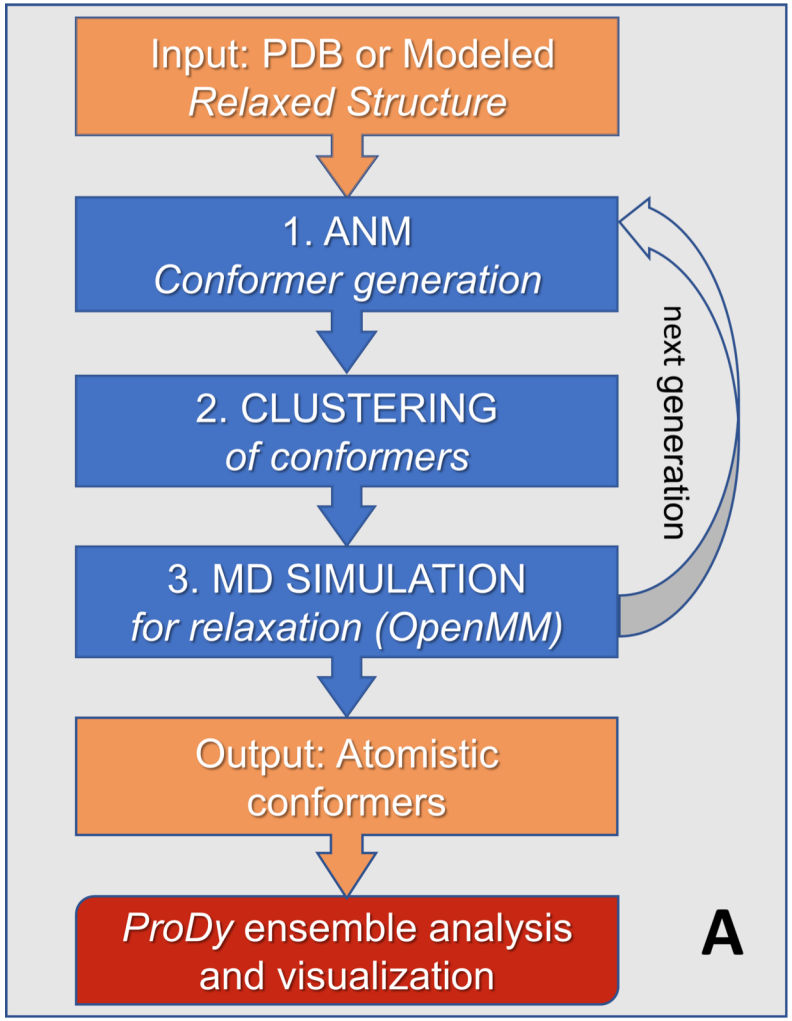
ClustENM for conformational sampling of flexible proteins and supramolecules

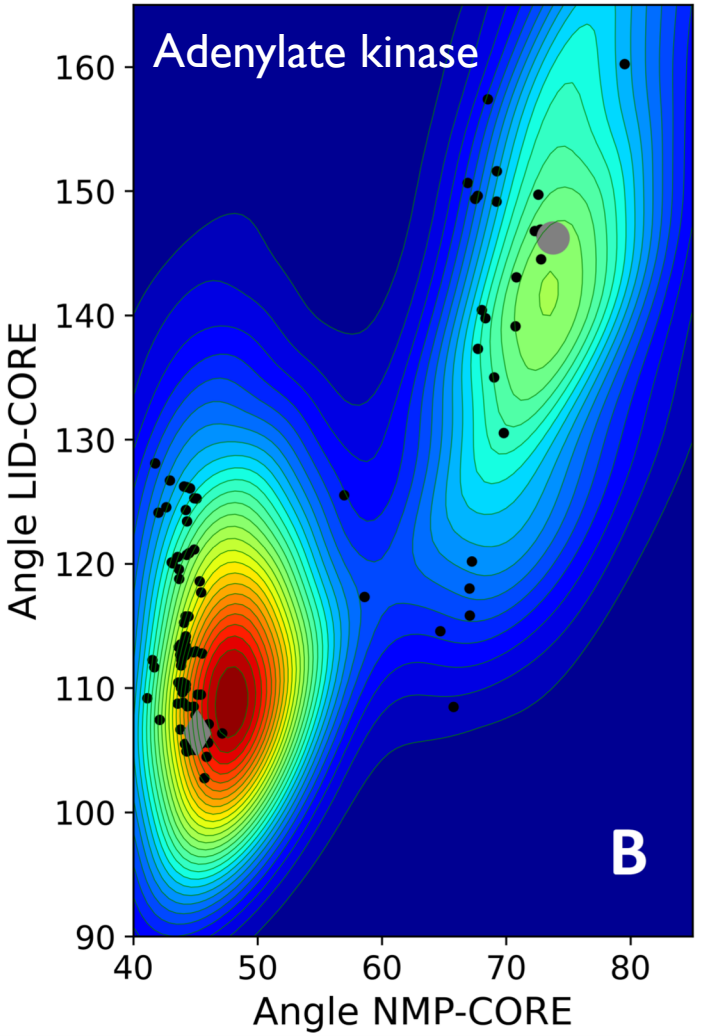


Trigger factor (TF)- ribosome (50S) ClustENM conformers aligned to the cryo-EM map. Conformers of 70S showing translocation of tRNA.

- **Original method:** Kurkcuoglu, Z.; Bahar, I.; Doruker, P., ClustENM: ENM-Based Sampling of Essential Conformational Space at Full Atomic Resolution. *J Chem Theory Comput* (2016) 12: 4549-62
- **Ensemble docking to flexible proteins:** Kurkcuoglu, Z.; Doruker, P., *PLoS One* (2016) 11, e0158063.
- **Trigger factor- ribosome dynamics:** Can, M.T.; Kurkcuoglu, Z.; Ezeroglu, G.; Uyar, A.; Kurkcuoglu, O.; Doruker, P., *PLoS One* (2017) 12, e0176262.
- **Protein-protein/DNA docking with HADDOCK**: Kurkcuoglu, Z.; Bonvin, A., *Proteins* (2020) 88: 292-306

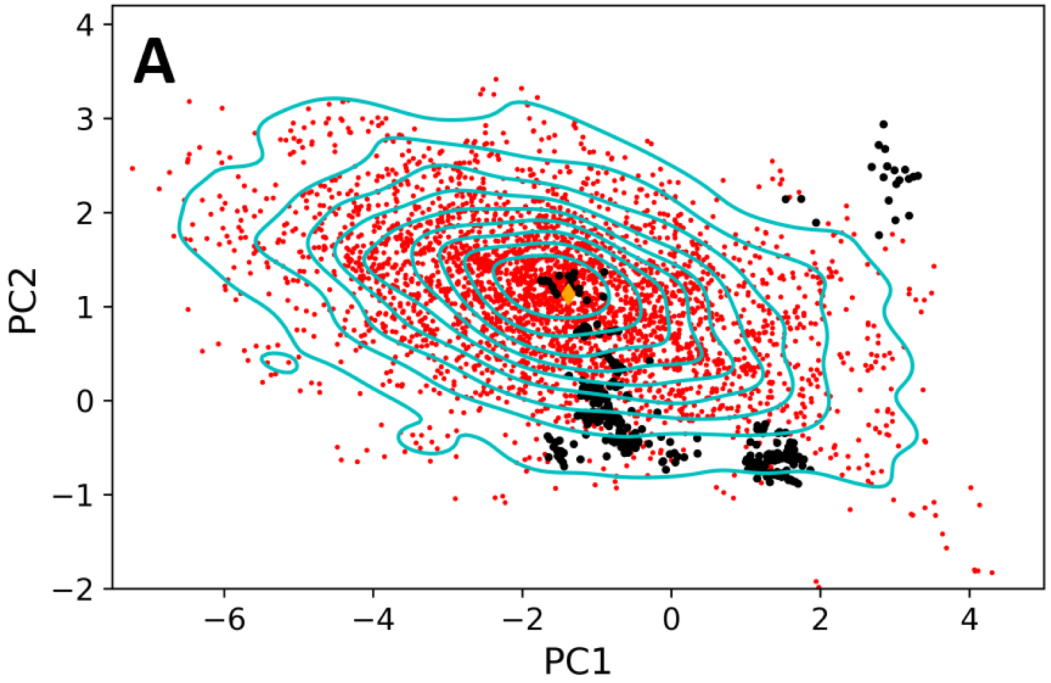
ClustENMD implementation in ProDy





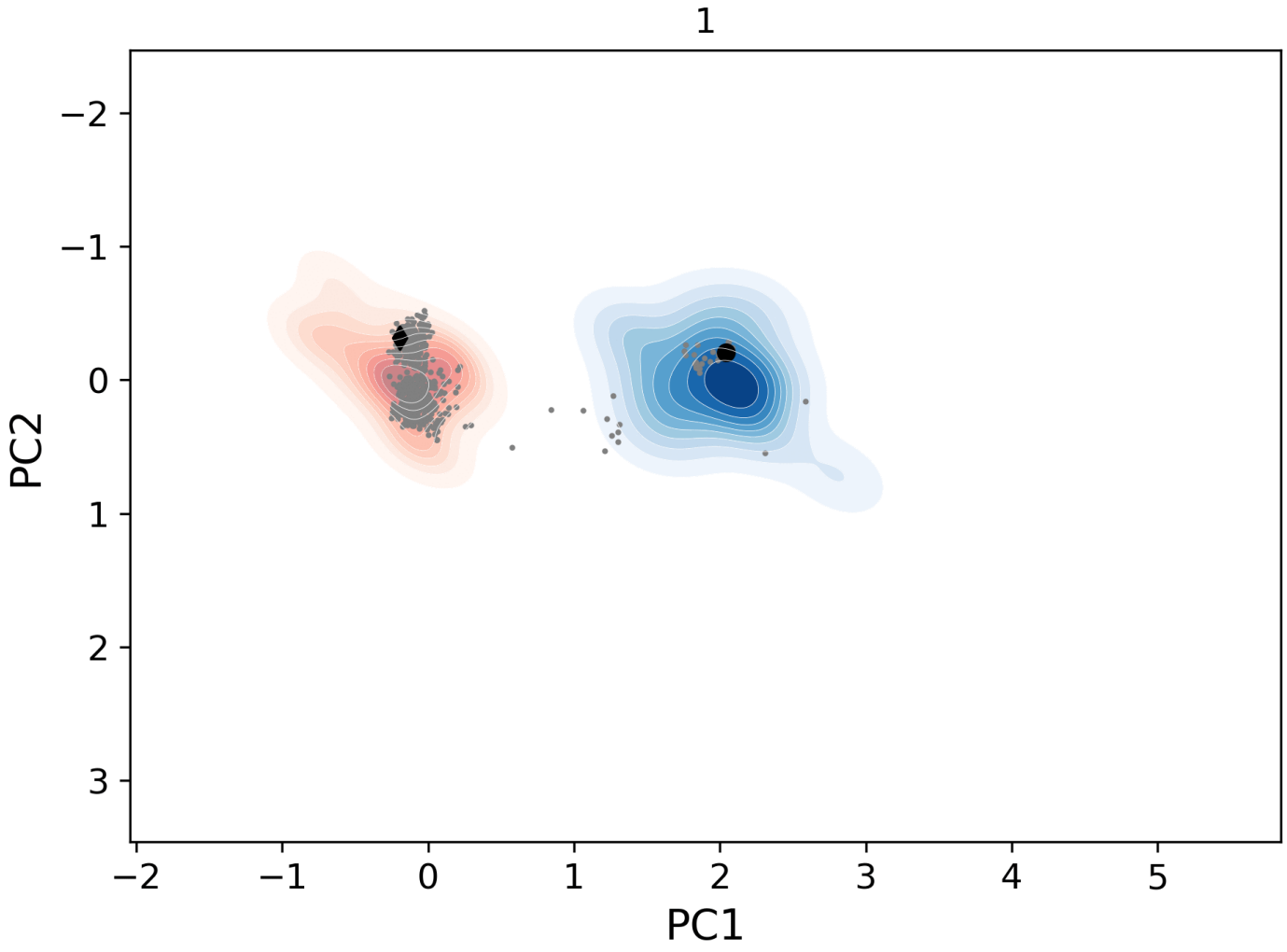
Kaynak B, Zhang S, Bahar I, Doruker P, Bioinformatics (accepted)

HIV-1 reverse transcriptase



Red dots: ClustENMD conformers
Black dots: experimental structures

ClustENMD for HIV-1 protease



Kaynak B, Zhang S, Bahar I, Doruker P, Bioinformatics (accepted)

Thank you!