ProDy

Protein Dynamics & Sequence Analysis

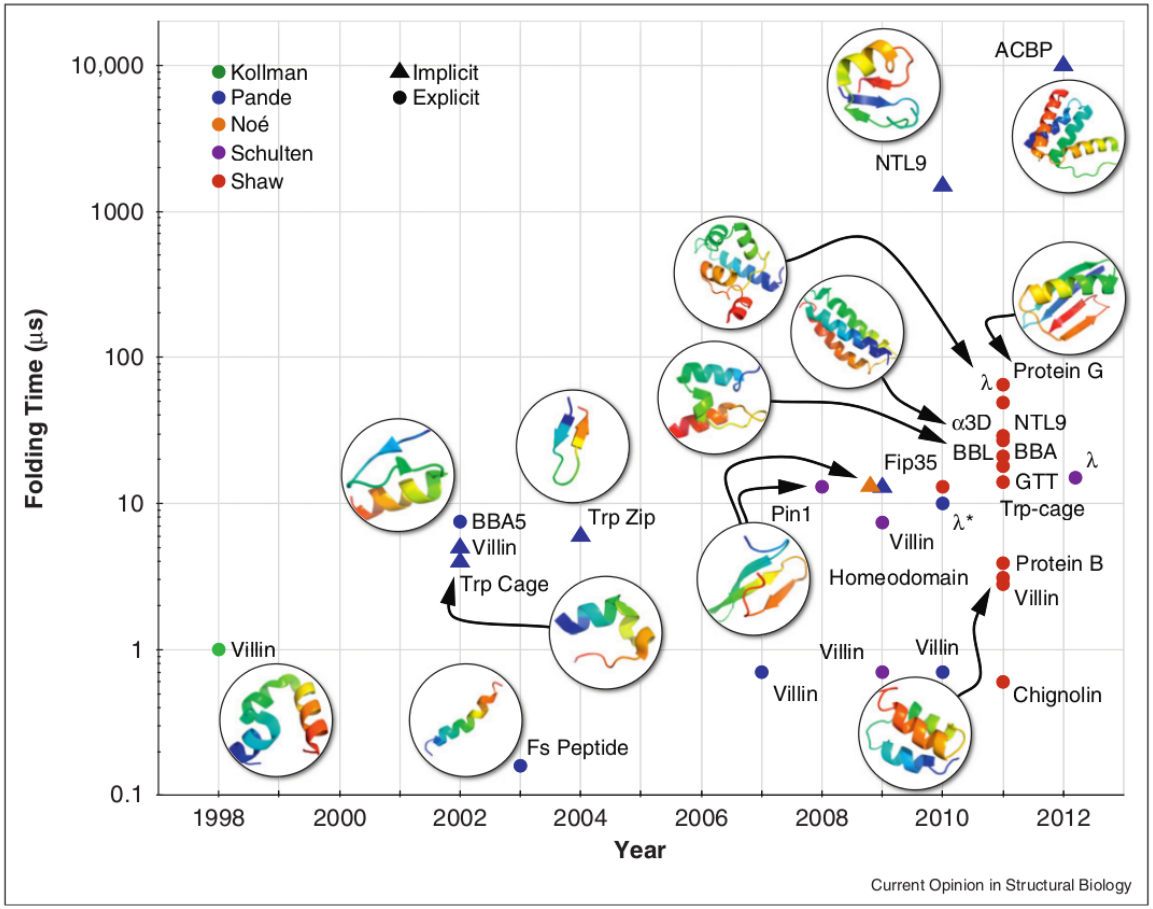
Overview & Applications

Hands-on Workshop in Computational Biophysics

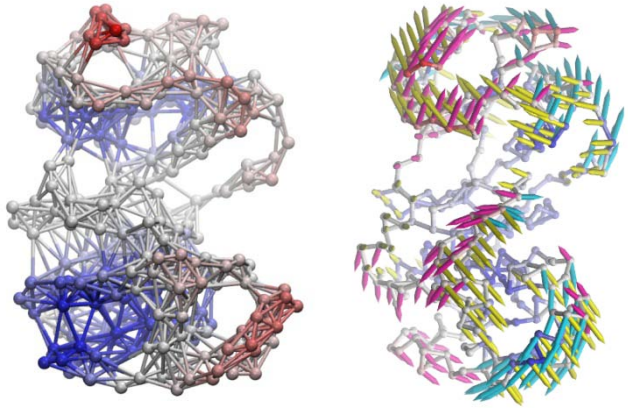
Pittsburgh Supercomputing Center

June 9, 2016

Full atomic simulations are computationally expensive



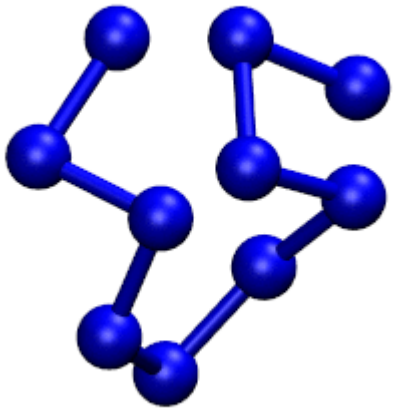
Lane et al. 2013



Bakan et al. Bioinformatics 2011.

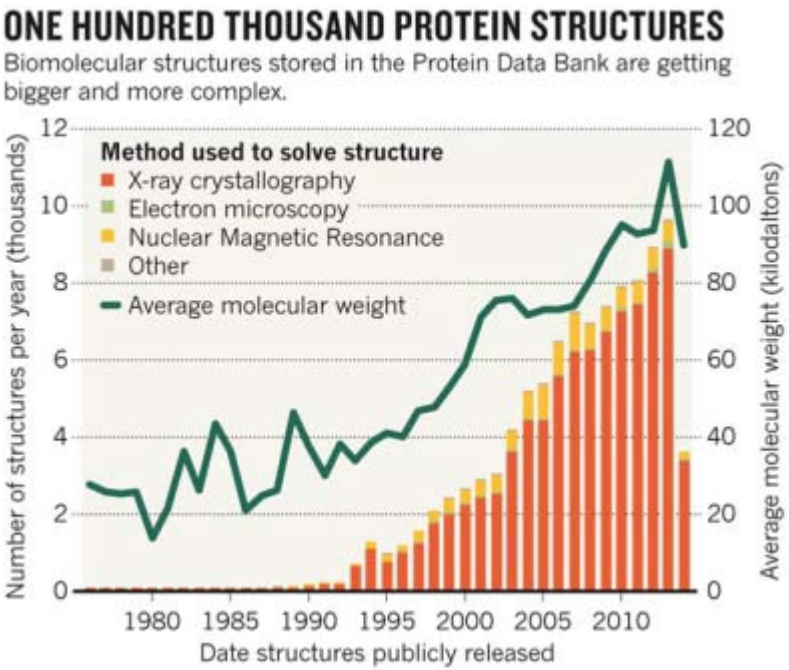
Coarse-grained Elastic Network Models are fast

Elastic Network Model



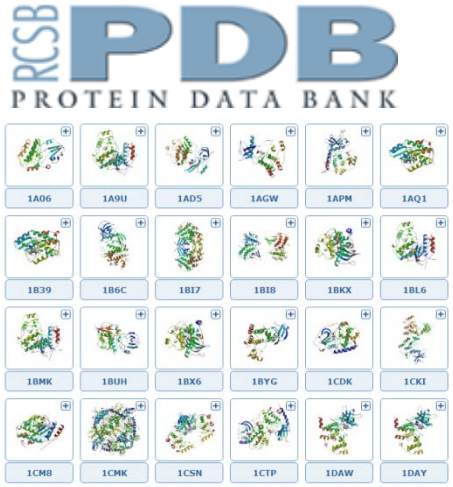
- Useful for predicting global motions of proteins
- Coarse-grained description (Cα-only usually)
- Residue pairs are connected via elastic springs with unified force constants
- You obtain a unique analytical solution for the spectrum of motion for each system – this is not a simulation

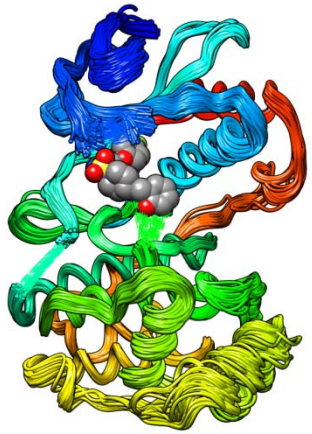
Growth in structural data



Nature, 15 May 2014.

Multiple structures for a single sequence





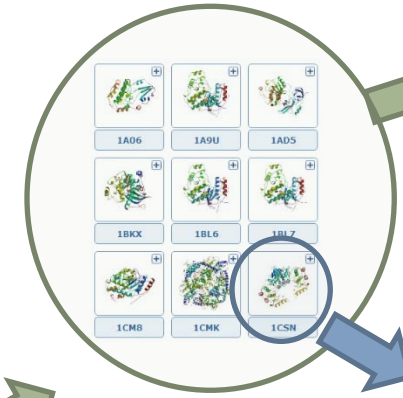
Dynamics may be inferred from structural data.

ProDy
Protein Dynamics & Sequence Analysis

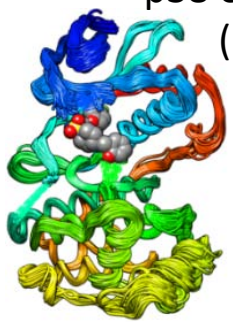
MMBioS

Leveraging the PDB since 2010

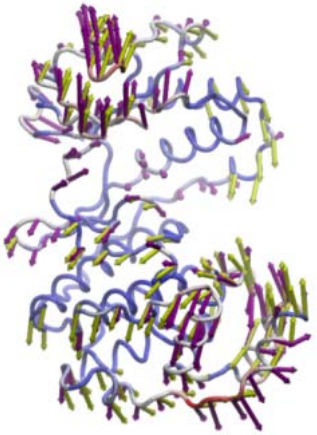
- High-throughput analysis of structural data
- Application Programming Interface (API) for development of tools
- Suitable for interactive usage

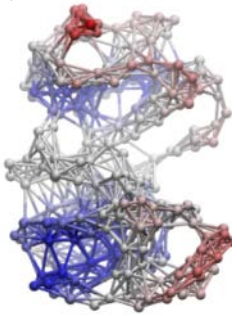


p38 ensemble (PCA)



Experiment/Theory



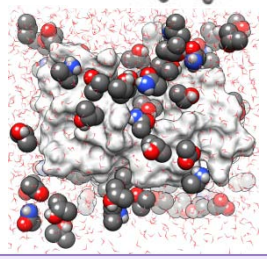


p38 network model (ANM)

User inputs a sequence

Usage example

```
>1A9U:A|PDBID|CHAIN
GSSHHHHHHSSGLVPRGSHMSQ
ERPTFYRQELNKTIWEVPERYQ
NLSPVGSGAYGSVCAAFDTKTG
......
```



MD trajectory analysis (EDA)

ProDy *identifies*, *retrieves*, *aligns*, and *analyzes (PCA)* structures matching input sequence

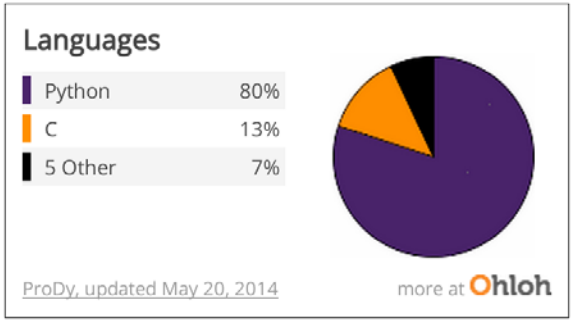
User can
- Compare experimental and theoretical models
- Sample conformations along normal modes

Bakan, Meireles & Bahar. Bioinformatics 2011.

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

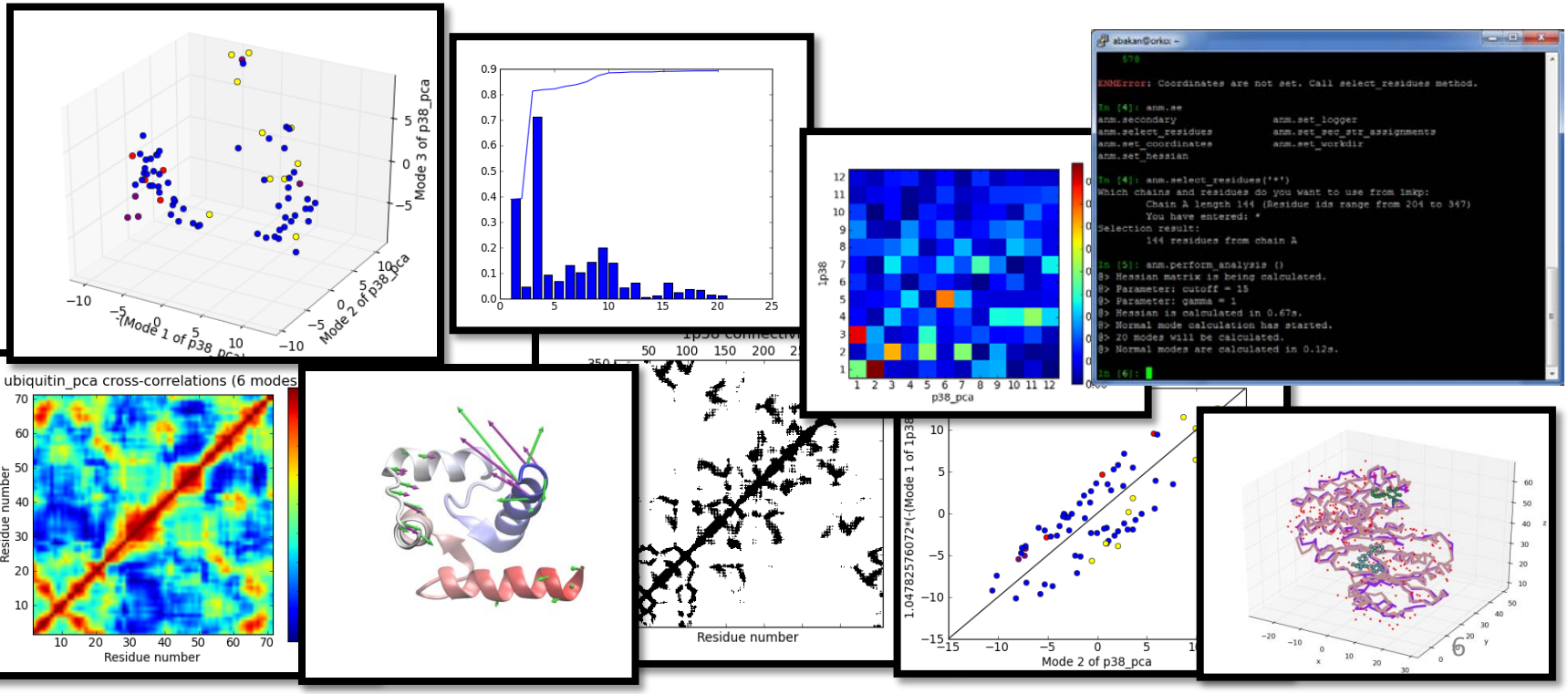
An Interactive Tool



matplotlib

NumPy

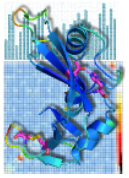
IP[y]: IPython
Interactive Computing



Suite of tools

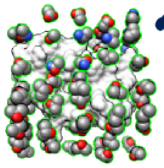
ProDy
Protein Dynamics & Sequence Analysis

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



Evol
Bridging Evolution & Dynamics

Multiple Sequence Alignment
Sequence conservation
Correlated Mutation



DruGUI
Druggability Suite for VMD

Computational Drug Discovery
Binding Site Prediction
Affinity Estimation

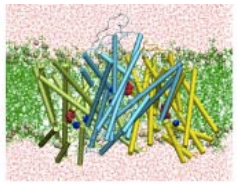
NMWiz
Normal Mode Wizard for VMD

A VMD plugin
Visualization of collective motions
Animations/movies

ProDy
Protein Dynamics & Sequence Analysis

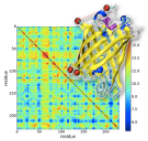
MMBioS

Suite of tools



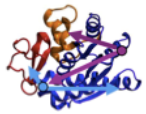
membrANM
Membrane Anisotropic Network Model

Modeling coupled protein-lipid dynamics
Useful for membrane proteins



MechStiff
Mechanical Stiffness Calculations

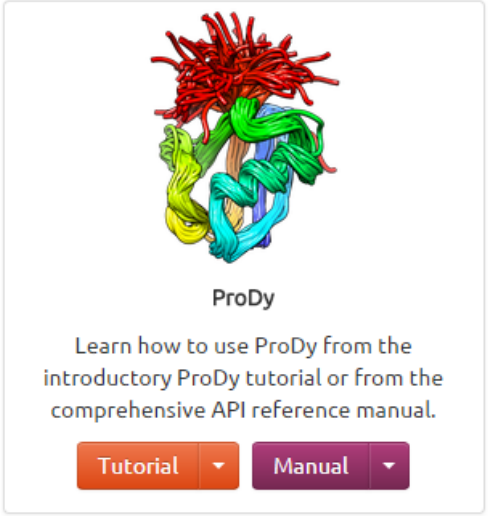
Response to external forces
Identification of mechanical stiffness

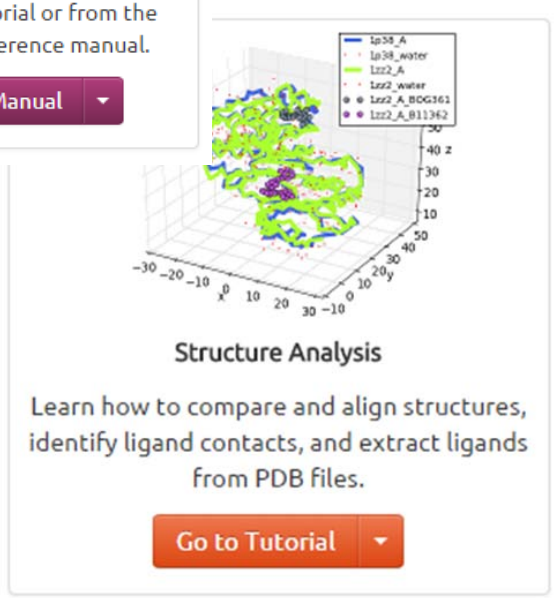


coMD
Collective Molecular Dynamics

ENM guided MD simulations
Efficient sampling of energy landscape

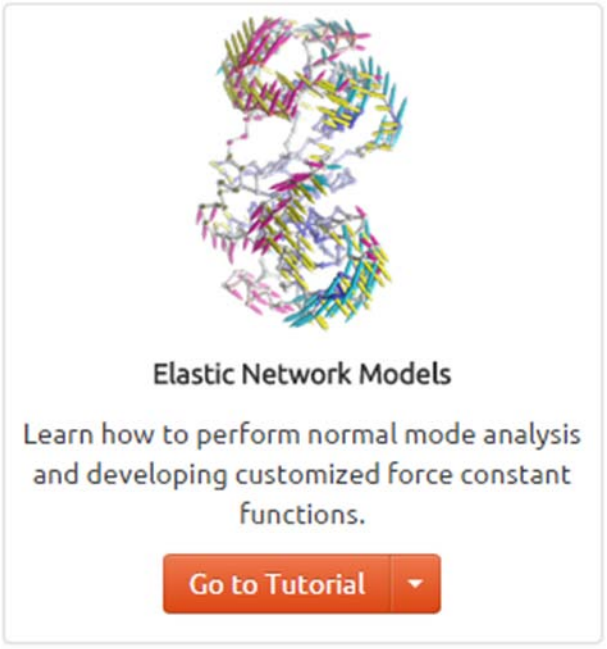
Tutorials: ProDy & Structure Analysis





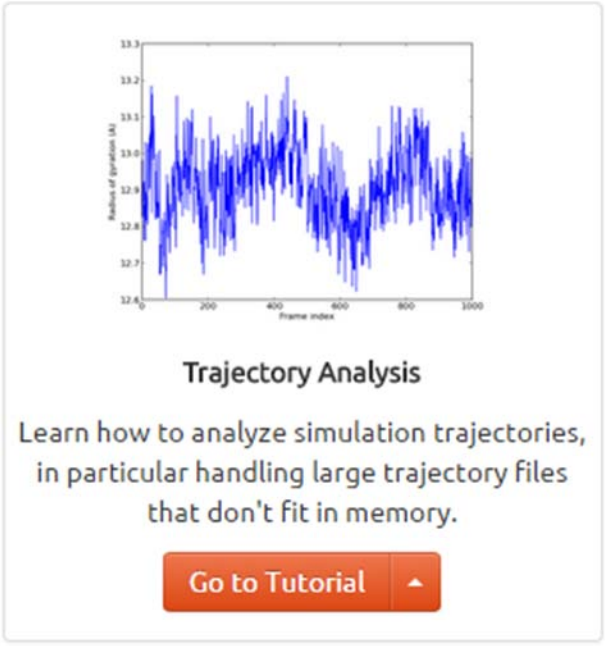
- Retrieving PDB Files
- BLAST Searching the PDB
- Constructing Biomolecular Assemblies
- Determining functional motions
- Aligning and Comparing Structures
- Identifying Intermolecular Contacts

Tutorial: Elastic Network Models



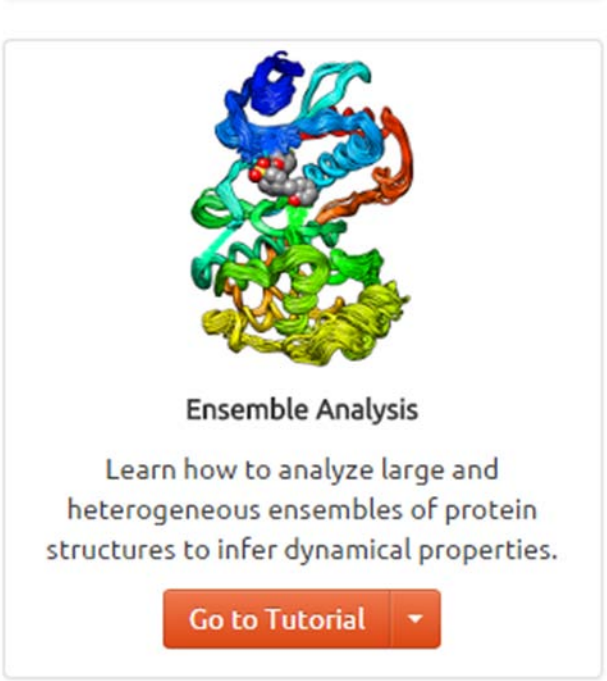
- Gaussian Network Model (GNM)
- Anisotropic Network Model (ANM)
- Normal Mode Analysis

Tutorial: Trajectory Analysis



- Fast processing of long trajectories
- Enables comparison of MD trajectories and ENM predictions

Tutorial: Ensemble Analysis



- NMR Models
- Homologous Proteins
- Multiple X-ray Structures
- Multimeric Proteins

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

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

Average position vector $\langle R_i \rangle$ of atom i

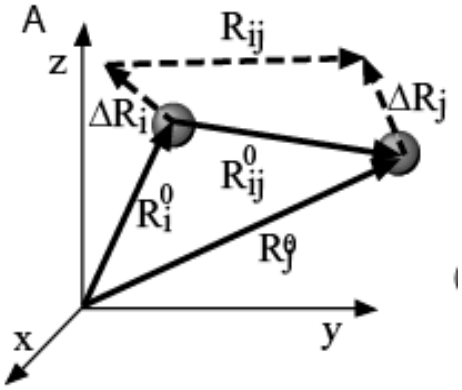
$\{\mathbf{R}_1(t_1), \mathbf{R}_1(t_2), \mathbf{R}_1(t_3), \ldots\ldots \mathbf{R}_1(t_m)\}$ for atom/residue 1

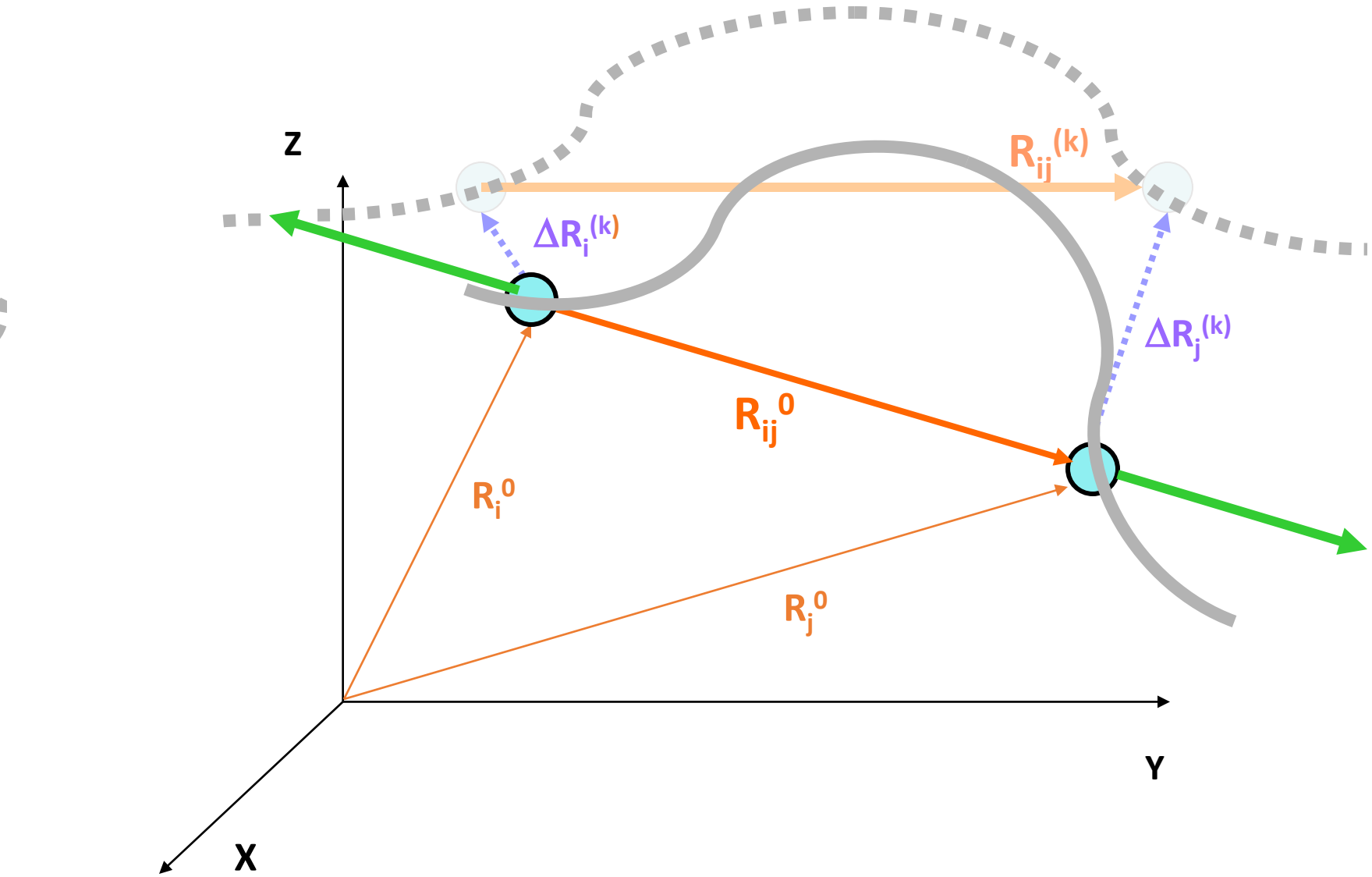
Average position vector for atom *i* over all trajectory

$\langle \mathbf{R}_1 \rangle = (1/k) \sum_k \mathbf{R}_1(t_k)$, where the summation is $k = 1, m$

Instantaneous fluctuation vector

$$\Delta\mathbf{R}_1(t_3) = \mathbf{R}_1(t_3) - \langle \mathbf{R}_1 \rangle$$




Z
$R_{ij}^{(k)}$
$\Delta R_i^{(k)}$
$\Delta R_j^{(k)}$
R_{ij}^{0}
R_i^{0}
R_j^{0}
Y
X

RMSD

with respect to starting structure R(0)

Instantaneous deviation for atom i

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

Average deviation over all atoms, at a given time,

$$\mathbf{RMSD}(t_k) = (1/N) \left[\sum_i (\Delta \mathbf{R}_i(t_k) \,.\, \Delta \mathbf{R}_i(t_k))\right]^{1/2} \quad \text{where } i = 1, N$$

Cross-correlations between fluctuations

Example: cross-correlations between fluctuation vectors of residues i and j (average over *m* snapshots/conformations)

$$\langle \Delta\mathbf{R}_i \,.\, \Delta\mathbf{R}_j \rangle = \sum_k [\Delta\mathbf{R}_i(t_k) \,.\, \Delta\mathbf{R}_j(t_k)] \,/\, m$$

For i = j, this reduces to mean-square fluctuation

$$\langle (\Delta\mathbf{R}_i)^2 \rangle$$

Covariance matrix (NxN)

$\mathbf{C} =$

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

$= \Delta R \, \Delta R^T$

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.

Cross-correlations between Components of fluctuation vectors

Example: cross-correlations between the X-component of R_i and Y component of R_j

$$\langle \Delta X_i \, \Delta Y_j \rangle = \Sigma_k \, [\Delta X_i(t_k) \, \Delta Y_j(t_k)] \, / \, m$$

To be organized in a 3x3 matrix as

$\langle \Delta X_i \Delta X_j \rangle$	$\langle \Delta X_i \Delta Y_j \rangle$	$\langle \Delta X_i \Delta Z_j \rangle$
$\langle \Delta Y_i \Delta X_j \rangle$	$\langle \Delta Y_i \Delta Y_j \rangle$	$\langle \Delta Y_i \Delta Z_j \rangle$
$\langle \Delta Z_i \Delta X_j \rangle$	$\langle \Delta Z_i \Delta Y_j \rangle$	$\langle \Delta Z_i \Delta Z_j \rangle$

Covariance matrix (3Nx3N)

C_{3N} =

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

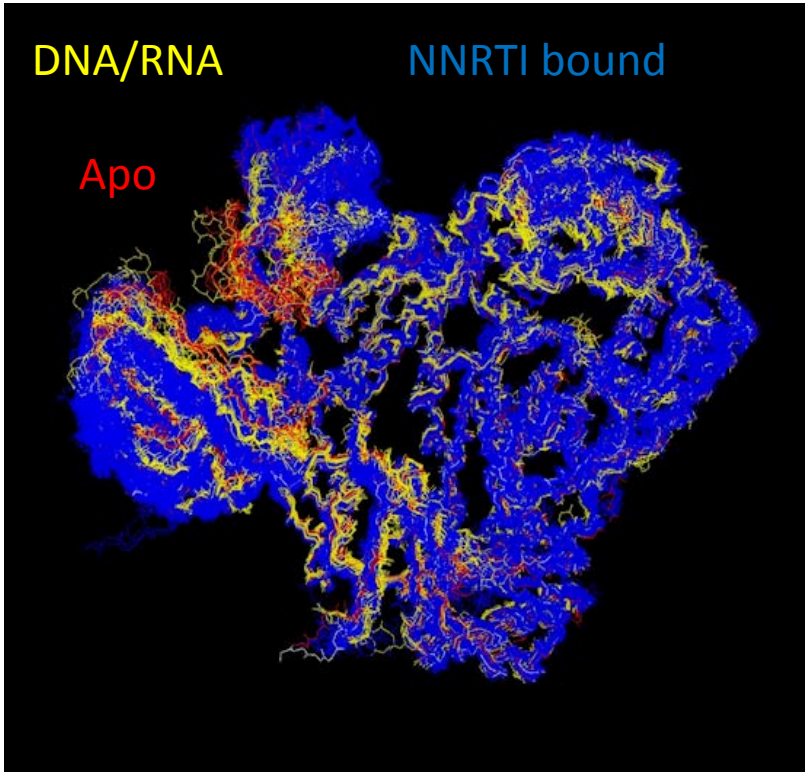
3N x 3N

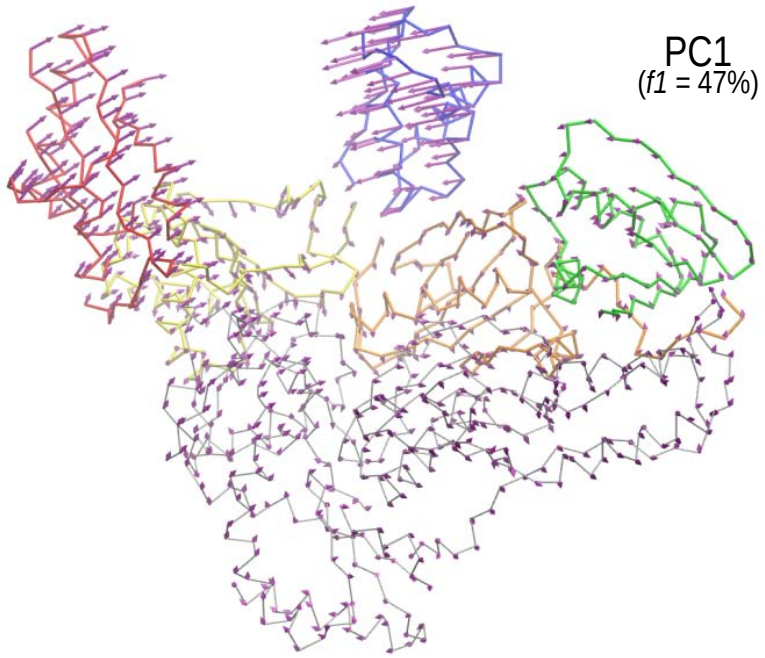
$\langle \Delta X_1 \Delta X_2 \rangle$	$\langle \Delta X_1 \Delta Y_2 \rangle$	$\langle \Delta X_1 \Delta Z_2 \rangle$
$\langle \Delta Y_1 \Delta X_2 \rangle$	$\langle \Delta Y_1 \Delta Y_2 \rangle$	$\langle \Delta Y_1 \Delta Z_2 \rangle$
$\langle \Delta Z_1 \Delta X_2 \rangle$	$\langle \Delta Z_1 \Delta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

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

Principal Component Analysis (PCA)

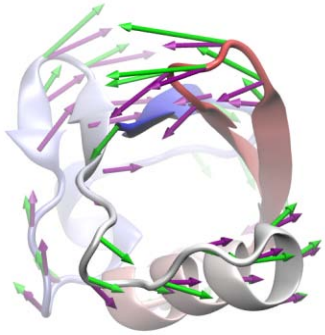
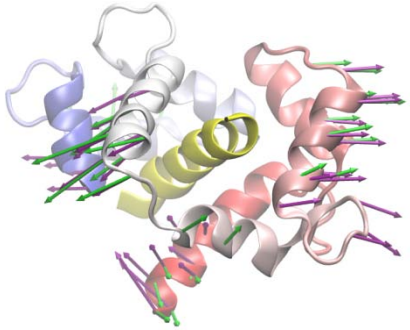




$$\mathbf{C}^{(ij)} = \begin{bmatrix} \langle \Delta x_i \Delta x_j \rangle & \langle \Delta x_i \Delta y_j \rangle & \langle \Delta x_i \Delta z_j \rangle \\ \langle \Delta y_i \Delta x_j \rangle & \langle \Delta y_i \Delta y_j \rangle & \langle \Delta y_i \Delta z_j \rangle \\ \langle \Delta z_i \Delta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

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

Global motions inferred from theory and experiments

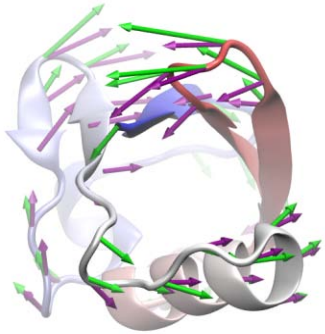
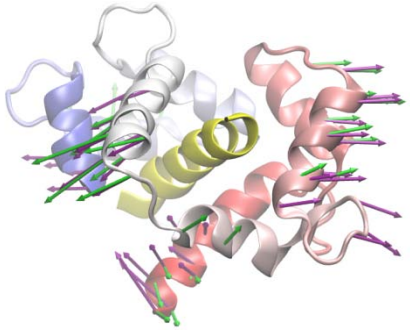




→ PCA of the ensemble of resolved structures
→ ANM analysis of a single structure from the ensemble

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

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

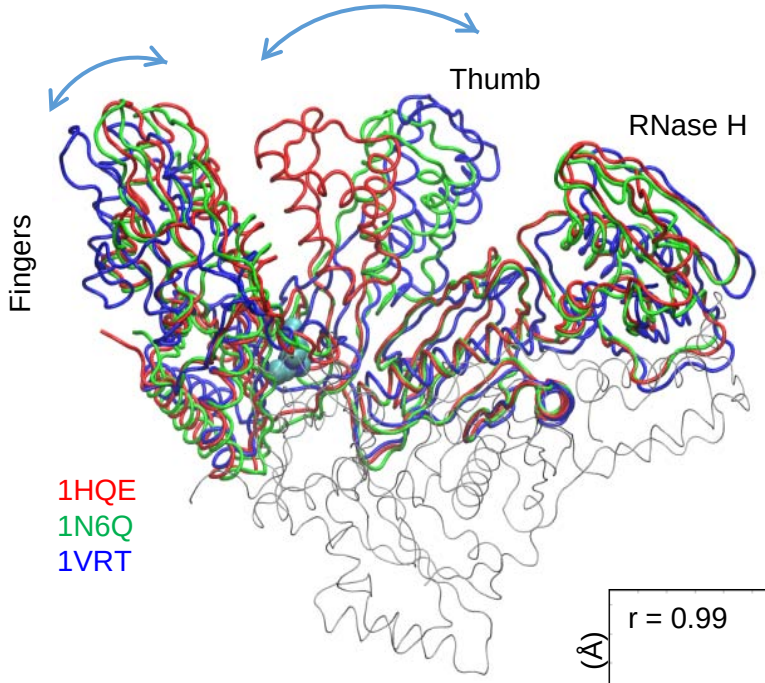
Reference:

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

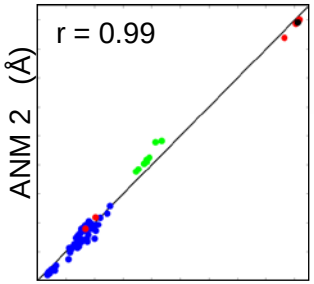
ProDy
Protein Dynamics & Sequence Analysis

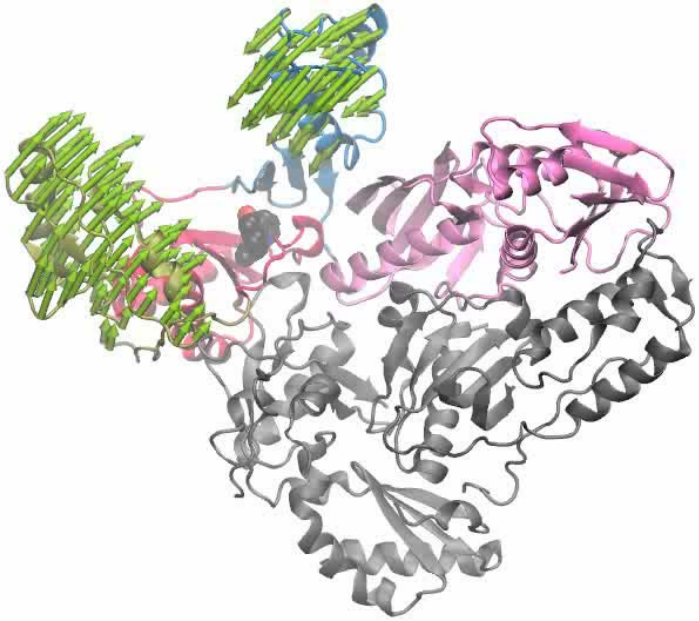
MMBioS

Soft modes enable functional movements



Experiments





Theory

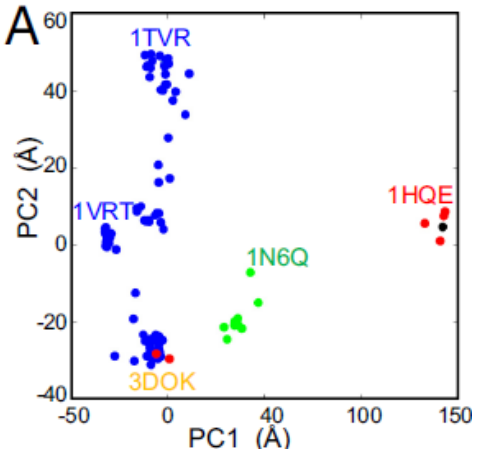
http://www.youtube.com/watch?v=1OUzdzm68YY

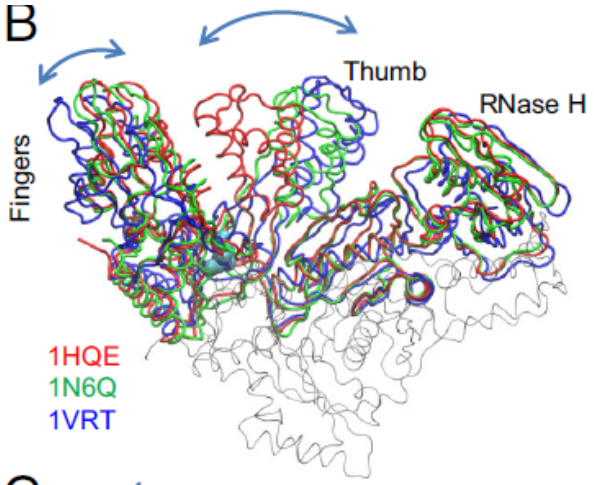
References:

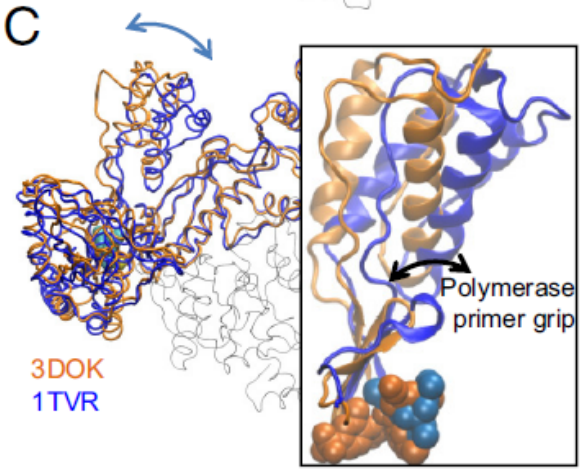
Bakan & Bahar (2009) PNAS **106**, 14349-54.

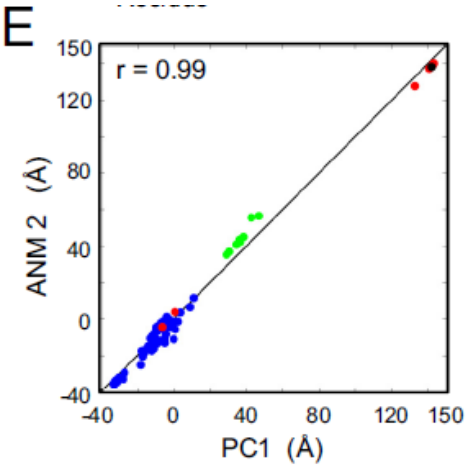
Comparing PCA and ENM

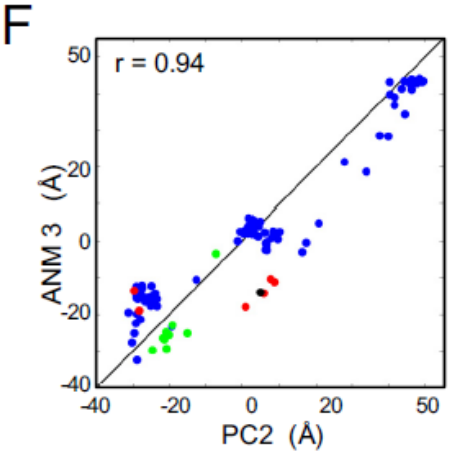
Structures of HIV-1 RT
Unbound
Inhibitor bound
DNA bound







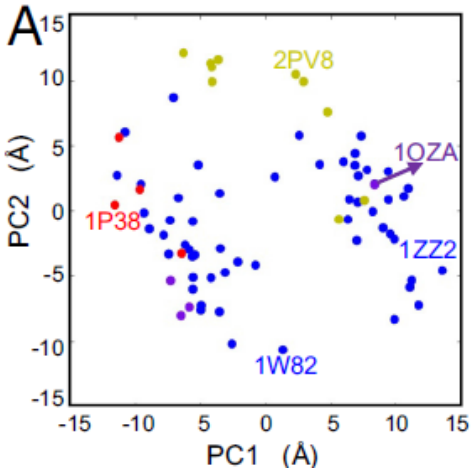


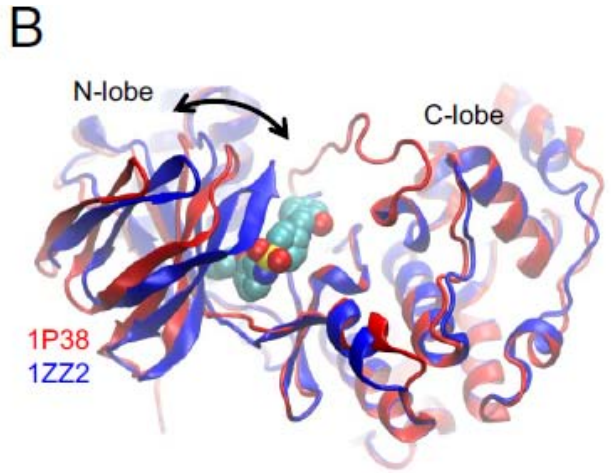


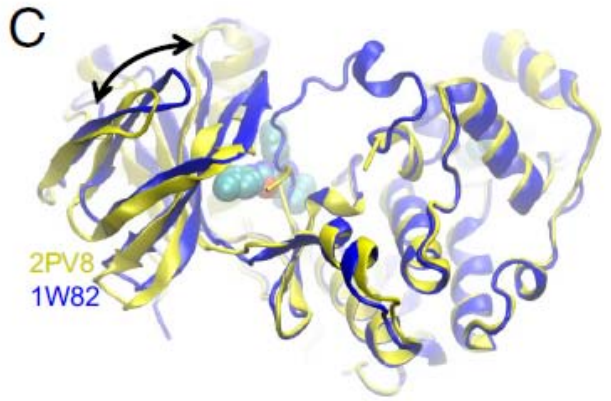
Bakan & Bahar. PNAS 106 (2009).

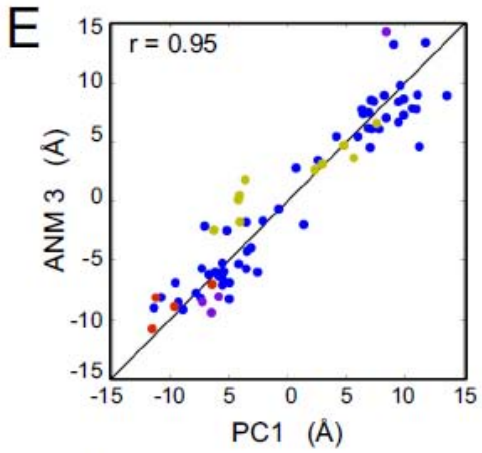
Example: Comparing PCA and ENM

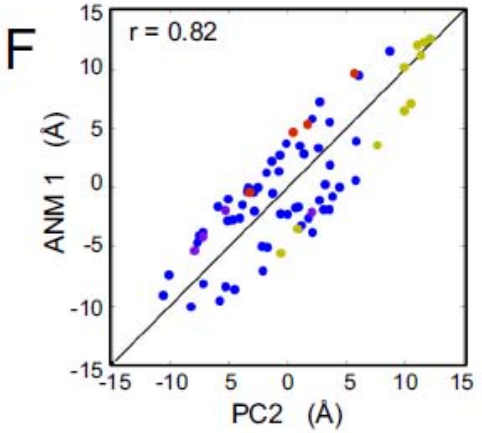
Structures of p38 MAPK
Unbound
Inhibitor bound
Glucose bound
Peptide bound





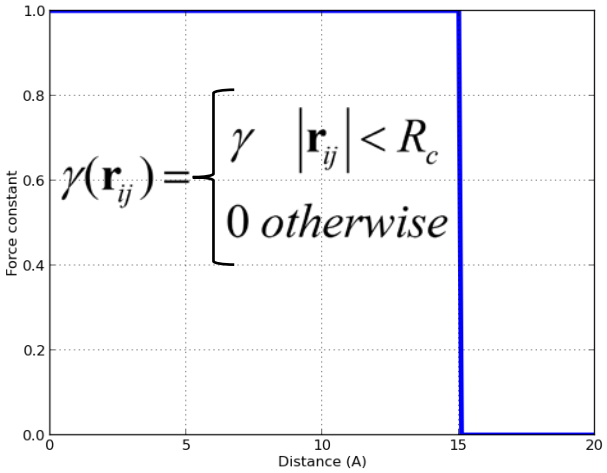




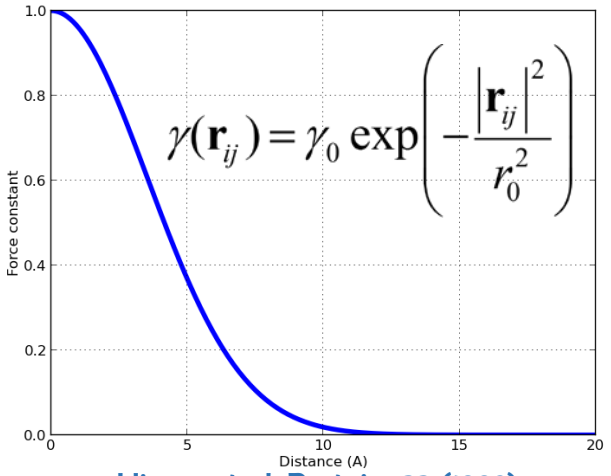


Bakan & Bahar. PNAS 106 (2009).

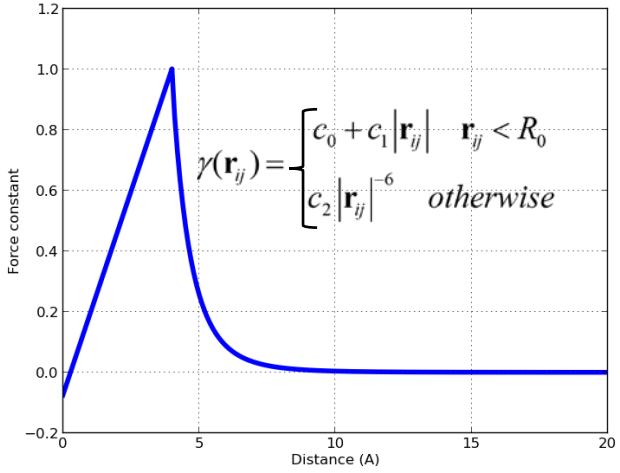
Different types of spring 'constants'



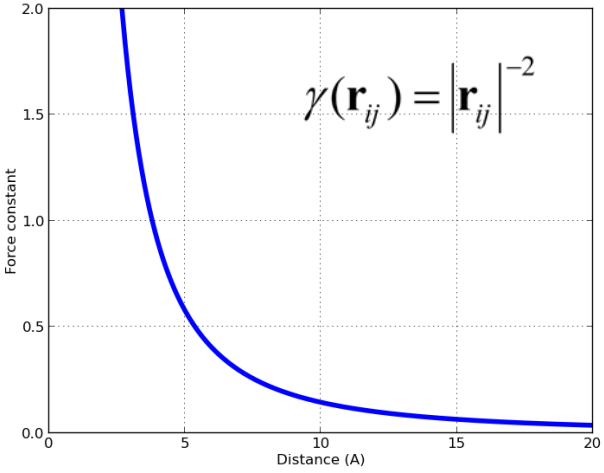
Tirion, PRL 77 (1996).



Hinsen et al. Proteins 33 (1998).

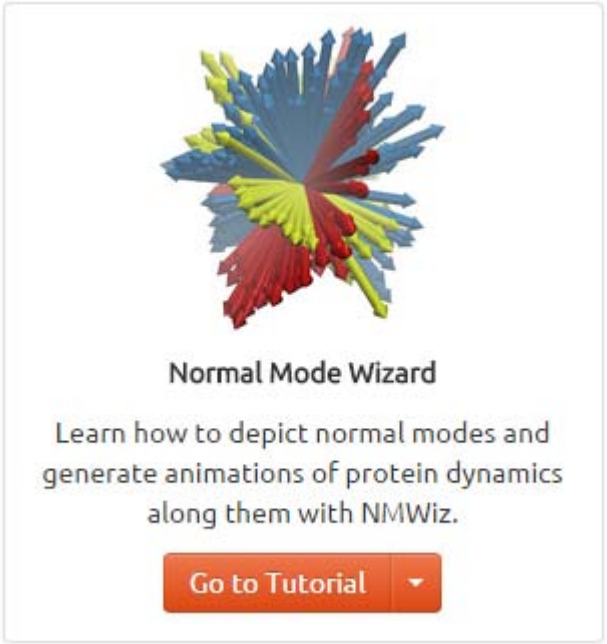


Hinsen et al. Chem Phys 261 (2000).



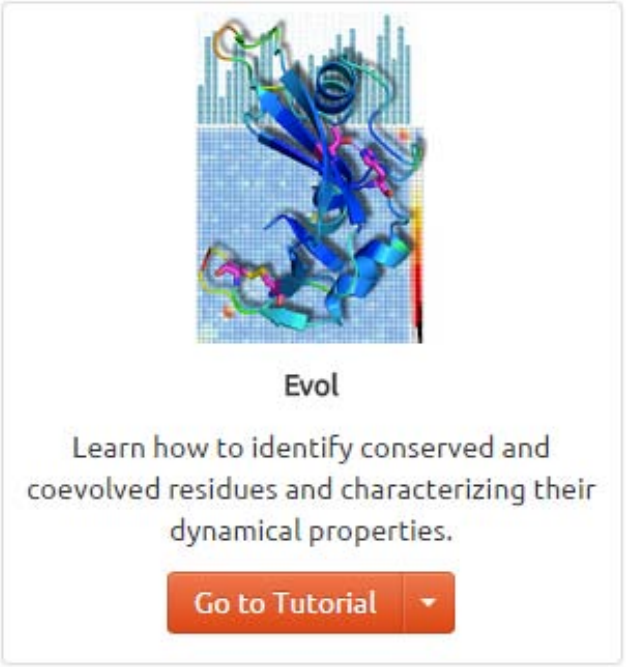
Yang et al. PNAS 106 (2009).

Tutorial: Normal Mode Wizard



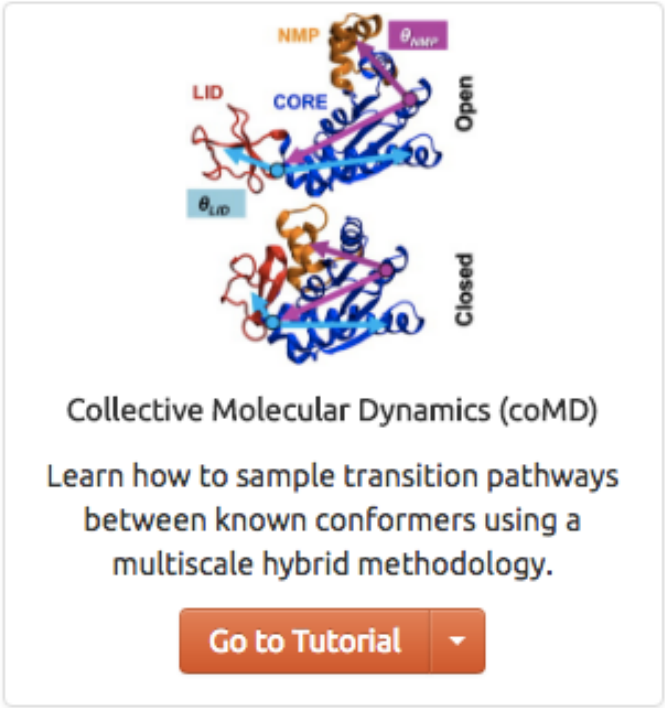
- perform ANM, GNM, and PCA/EDA calculations
- draw customizable normal mode arrows
- make animations (sample conformations)
- make interactive square-fluctuations plots
- compare two structures and draw deformation arrows

Tutorial: Evol



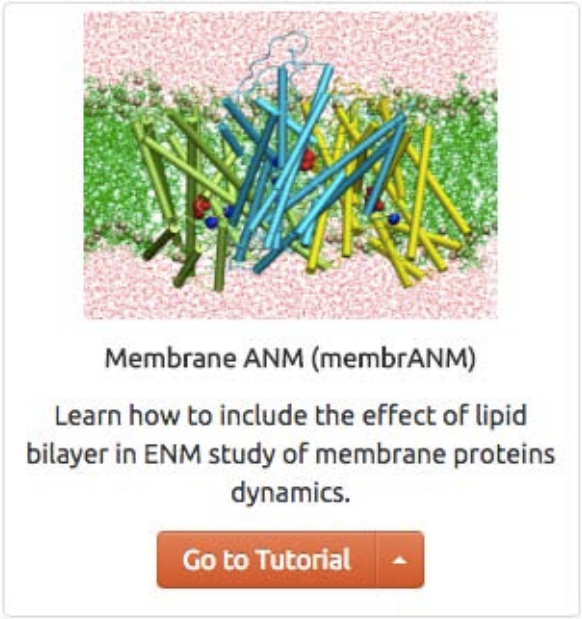
- identification of conserved and coevolving residues
- Retrieving multiple sequence alignments (MSAs) from Pfam DB
- extremely fast MSA I/O functions
- Generation of conservation profiles (1D plots) and co-evolution maps (2D plots)

Tutorial: collective Molecular Dynamics



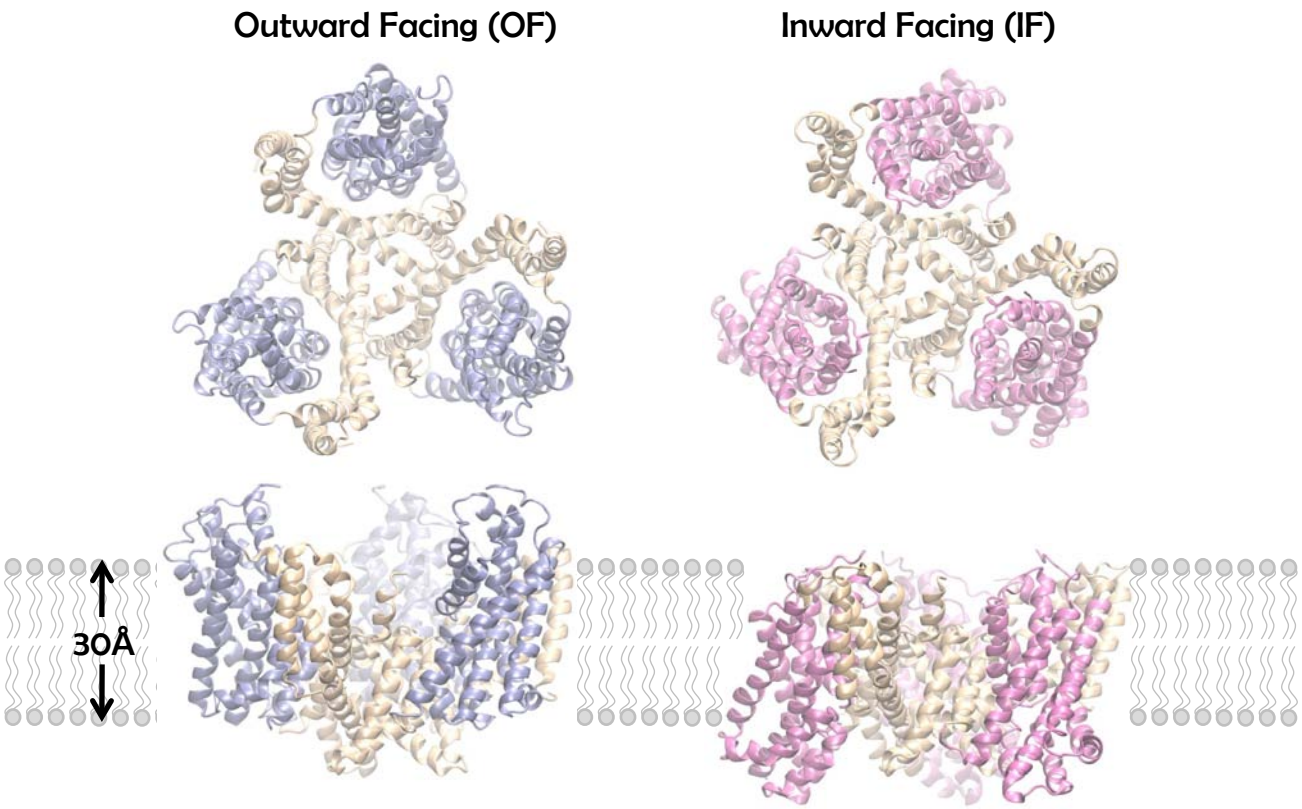
- Sampling the conformational space near native state
- Identification of substates and accessible transitions
- Generating transition paths between substates
- Obtaining information on global dynamics at atomic resolution
- Generating the conformational energy landscape for the investigated system

Tutorial: membrANM



- Evaluating membrane proteins' dynamics in the presence of lipid bilayer, also modeled as an elastic network model, explicit or implicit
- Comparing protein global motions in the presence and absence of membrane
- Understanding mechanisms of protein-membrane remodeling or coupling to facilitate function

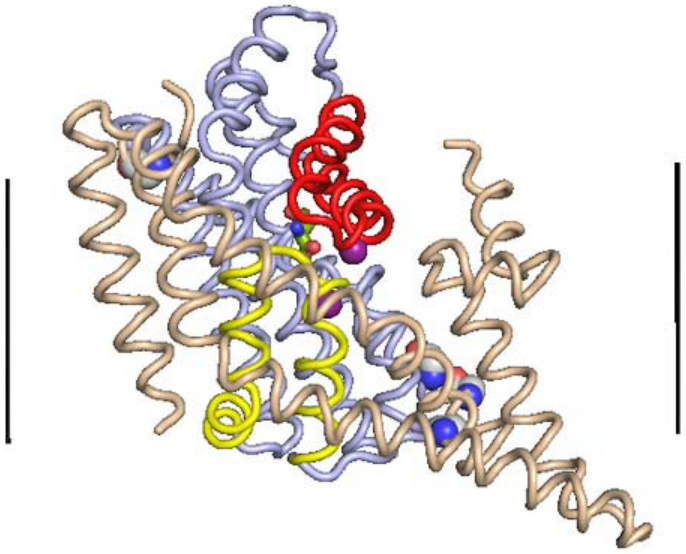
Global transitions



Vernool *et al.* Nature 431 (2004).
Reyes *et al.* Nature 462 (2009).

Global transitions

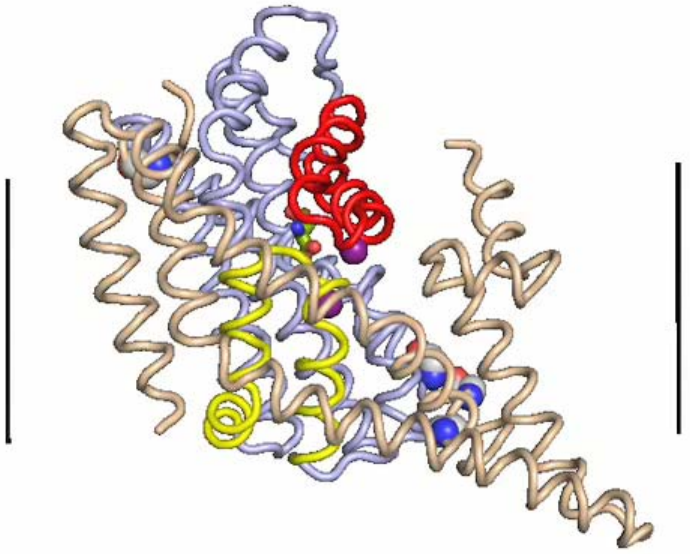
Single subunit showing the transport domain moving across the membrane



Reyes *et al.* Nature 462 (2009).

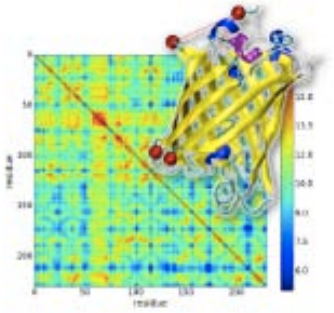
Global transitions

Single subunit showing the transport domain moving across the membrane



Reyes *et al*. Nature 462 (2009).

Tutorial: MechStiff



MechStiff

Learn how to evaluate the effective resistance of residues to deformation in the particular 3D structure

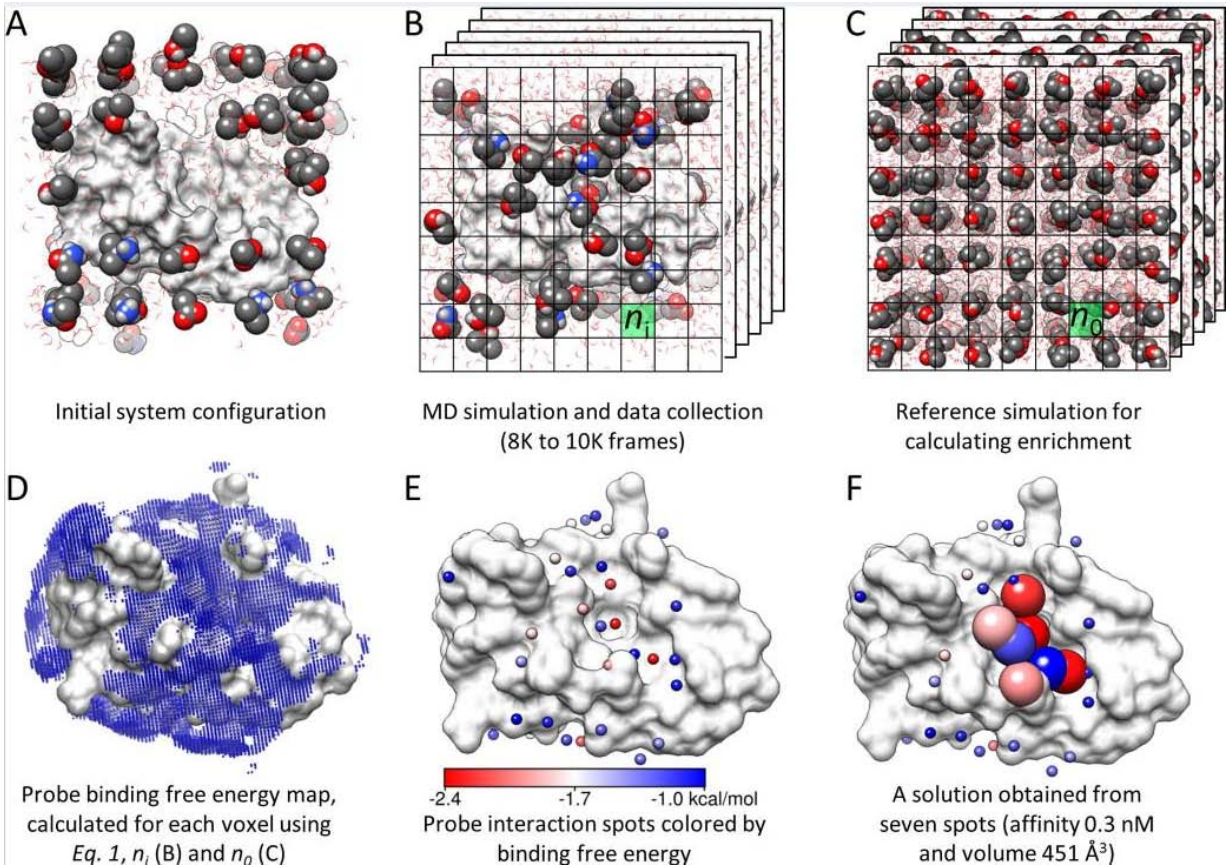
Go to Tutorial

- Identification of the anisotropic response of the structure to external perturbations
- determination of the weak/strong pairs of interactions depending on the direction of the external force and the sites that are subjected to perturbation (uniaxial tension)
- Determination of the effective spring constant observed macroscopically, to be compared with data from Molecule Force Spectroscopy (SMFS) or atomic force microscopy (AFM).
- Evaluating the contributions of each mode to deformations along selected directions

Tutorial: Druggability

- Set up NAMD simulations
- Analyze trajectories to identify binding hot spots

Exploring binding with probe molecules

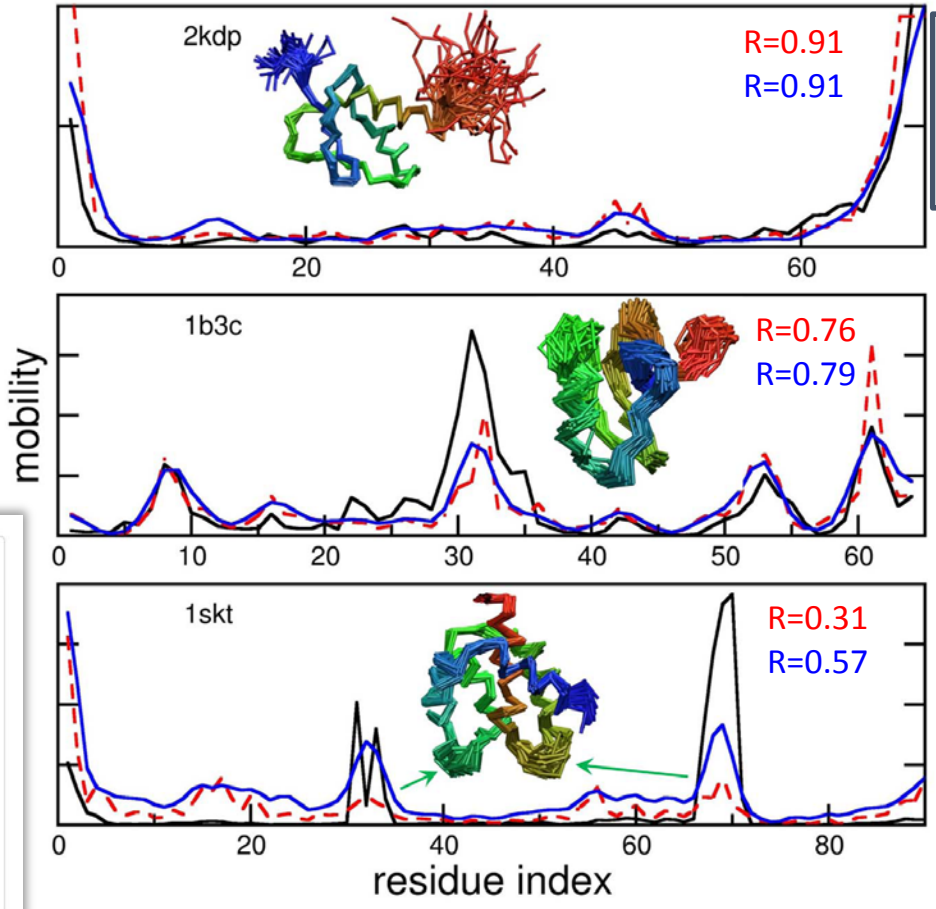


Bakan et al. J Chem Theor Comput (2012).

A few commands in ProDy

- Download NMR structures from PDB — fetchPDB()
- Calculate residue MSFs for each protein — calcMSF()
- Determine ENM topology — buildHessian()

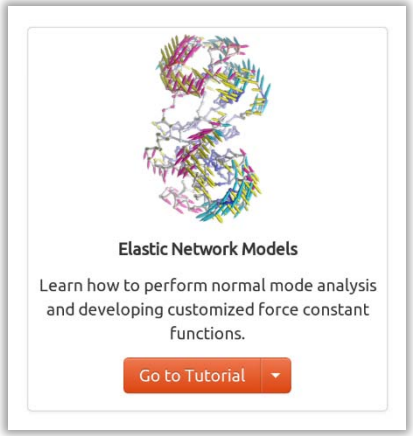
Fine-tuning force constants



- Distance-dependence
- 1st neighbors
- 2nd neighbors
- H bonds

black: NMR
red: ENM
blue: modified ENM

GammaStructureBased()



Learn more at prody.csb.pitt.edu

Rotations-Translations of Blocks

H^{RTB} = P $(6N_b \times 3N)$ H^{AA} $(3N \times 3N)$ P^T $(3N \times 6N_b)$

$(6N_b \times 6N_b)$

Smaller Hessian can be more easily diagonalized...

V'^{AA} = P^T V^{RTB}

...and modes projected back into all-residue space

H: ANM Hessian (3 rows/cols per residue)
P: Projection matrix from all-residue space to rigid block space
H^{RTB}: RTB Hessian (no internal motions of blocks)
V'^{AA}: Approximate ANM motions

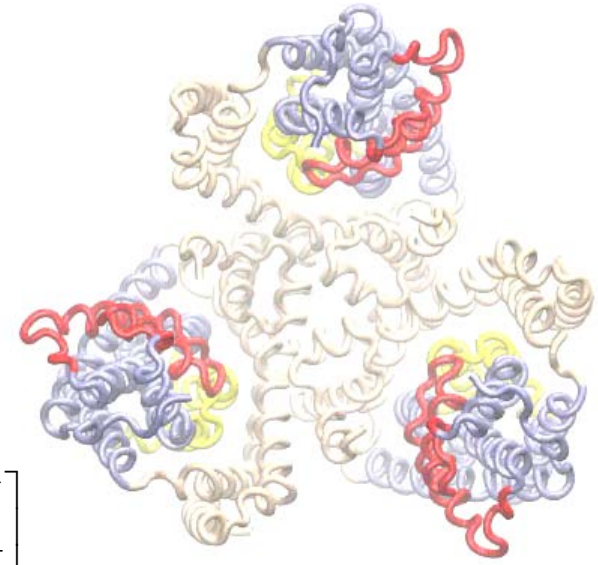
RTB.buildHessian()

Ming & Wall. PRL 95 (2005).
Zheng & Brooks. Biophys J 89 (2005).

Exploring structural transitions: Glutamate transporter

ANM predicts large radial motions of the trimer.
Can we design a better model?

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



Altered radial force constants:

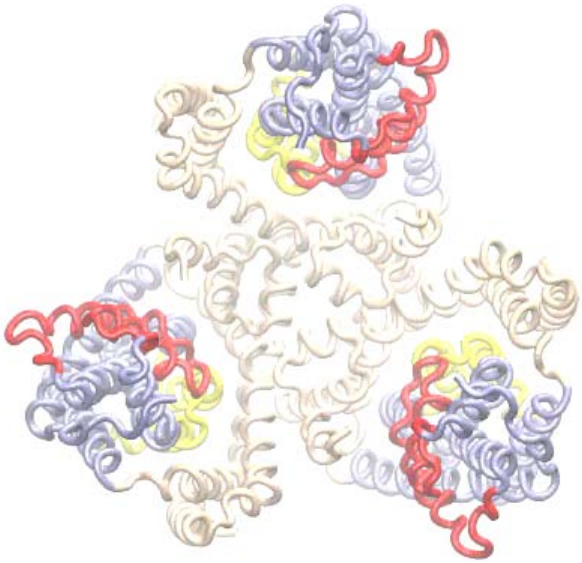
$$\mathbf{H_{ij}} = -\left(R_{ij}^0\right)^{-2}\begin{bmatrix} \left(x_{ij}^0\sqrt{\gamma_x}\right)^2 & x_{ij}^0 y_{ij}^0\sqrt{\gamma_x\gamma_y} & x_{ij}^0 z_{ij}^0\sqrt{\gamma_x\gamma_z} \\ x_{ij}^0 y_{ij}^0\sqrt{\gamma_x\gamma_y} & \left(y_{ij}^0\sqrt{\gamma_y}\right)^2 & y_{ij}^0 z_{ij}^0\sqrt{\gamma_y\gamma_z}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

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

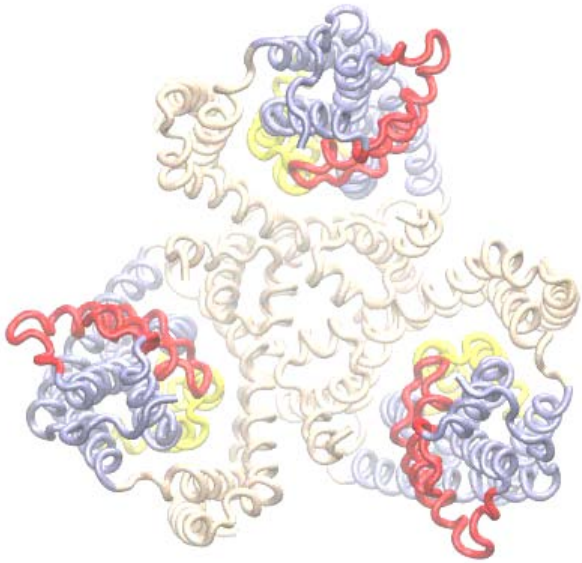
RTB.buildHessian()

Lezon & Bahar. Biophys J 102 (2012).

Exploring structural transitions: Glutamate transporter

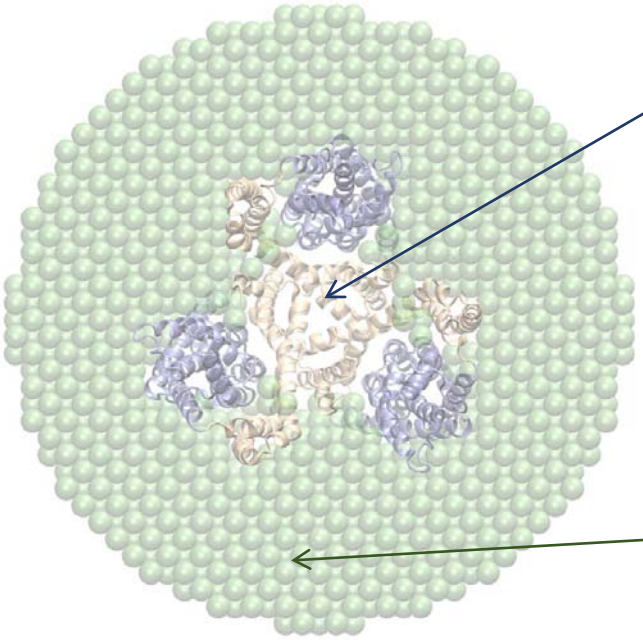


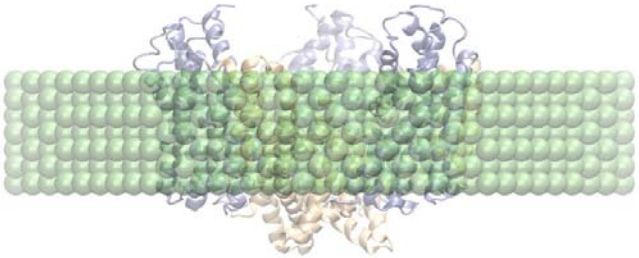
ANM: Large radial motions

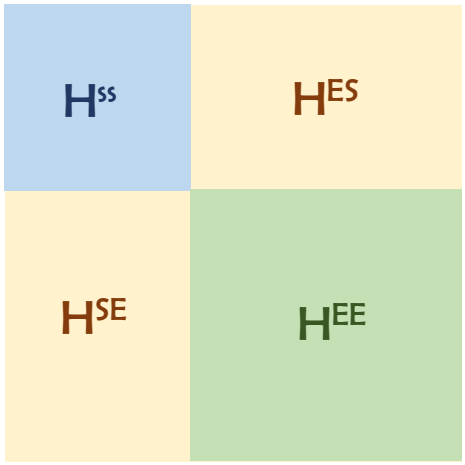


imANM

System/environment approximation







As the *environment* fluctuates randomly, the effective motion of the *system* is given by

$$V_{eff}(\mathbf{s}) = \tfrac{1}{2}\Delta\mathbf{s}^T\left(\mathbf{H}^{ss\,\prime}\right)\Delta\mathbf{s}$$

$$\mathbf{H}^{ss\,\prime} = \mathbf{H}^{ss} - \mathbf{H}^{SE}\left(\mathbf{H}^{EE}\right)^{-1}\mathbf{H}^{ES}$$

reduceModel()