ProDy

Protein Dynamics & Sequence Analysis

Overview & Applications

T. Lezon

Hands-on Workshop in Computational Biophysics

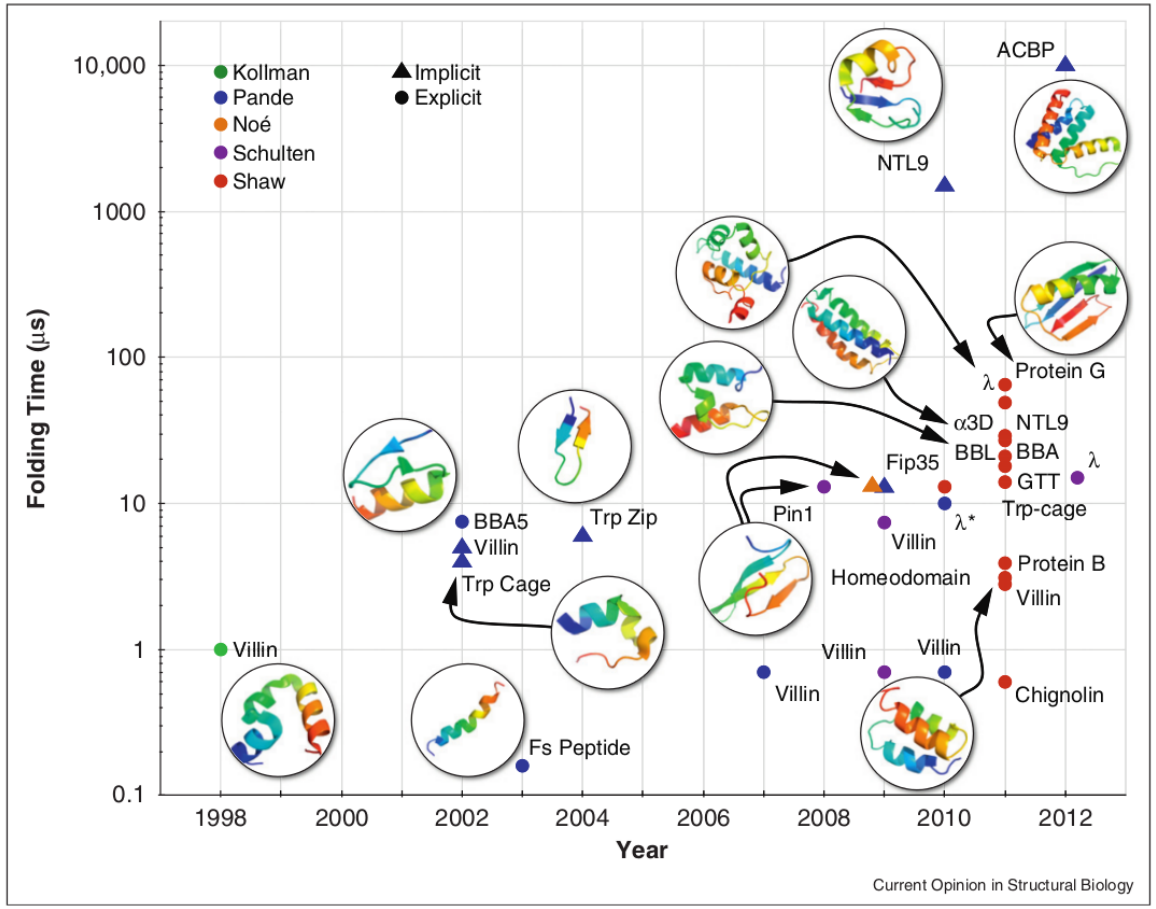
Pittsburgh Supercomputing Center

22 May, 2014

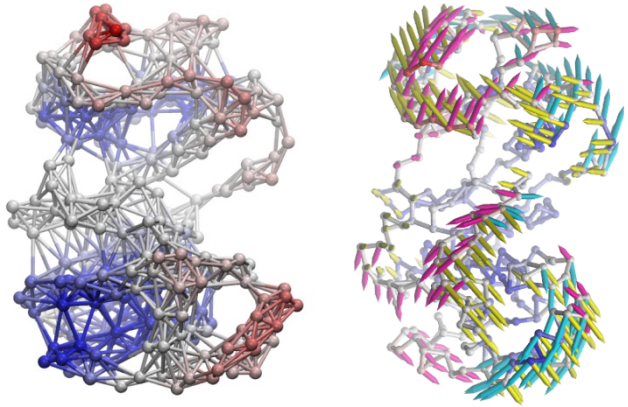
ProDy
Protein Dynamics & Sequence Analysis

MMBioS

Simulations still take time



Lane et al. 2013



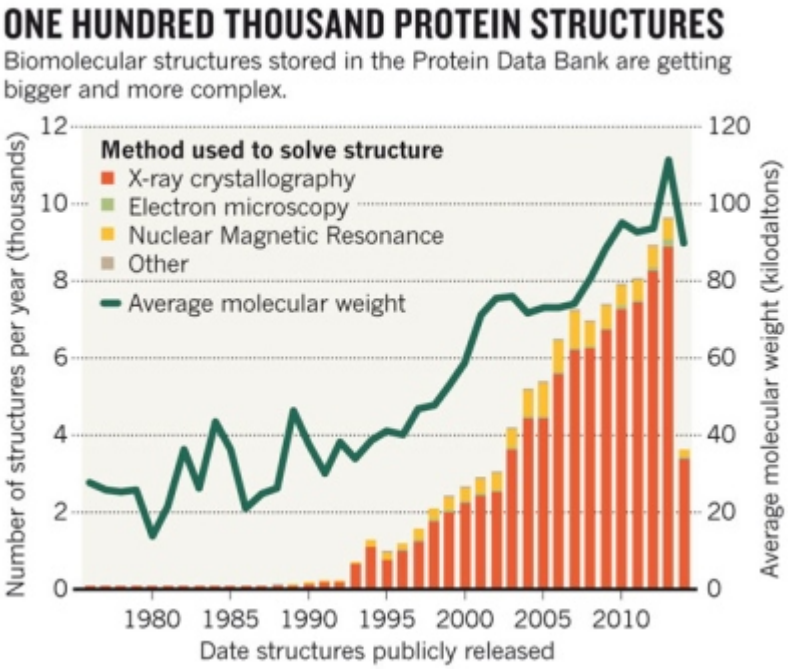
Bakan et al. Bioinformatics 2011.

Coarse-grained Elastic Network Models are fast

ProDy
Protein Dynamics & Sequence Analysis

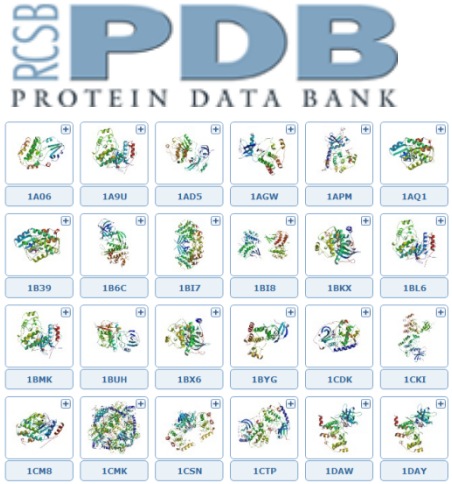
MMBioS

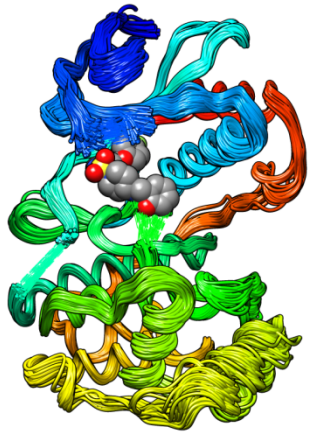
The structural data explosion



Nature, 15 May 2014.

Multiple structures for a single sequence





Dynamics may be inferred from structural data.

ProDy
Protein Dynamics & Sequence Analysis

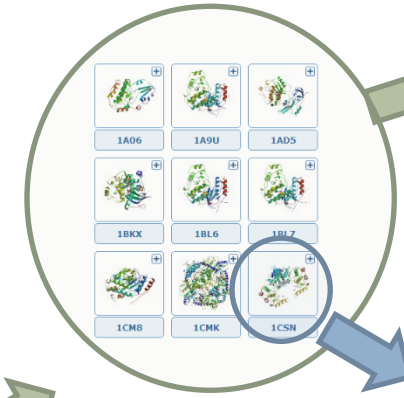
MMBioS

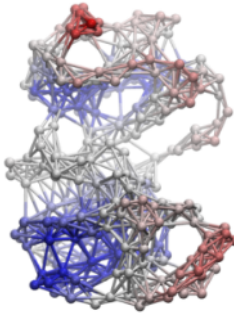
Exploiting the PDB since 2010

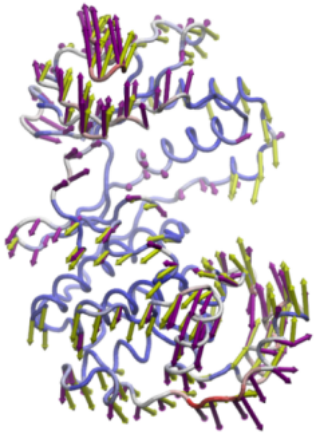
- High-throughput analysis of structural data
- Application Programming Interface (API) for development of tools
- Suitable for interactive usage

p38 ensemble (PCA)

Experiment/Theory





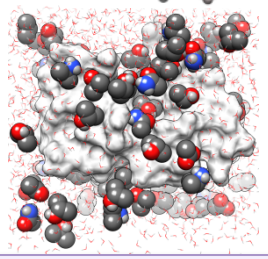


p38 network model (ANM)

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

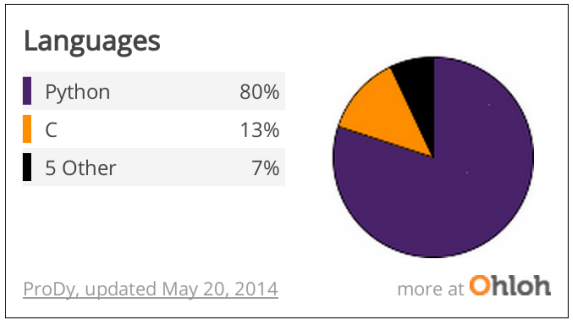
User can
- Compare experimental and theoretical models
- Sample conformations along normal modes

Bakan, Meireles & Bahar. Bioinformatics

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

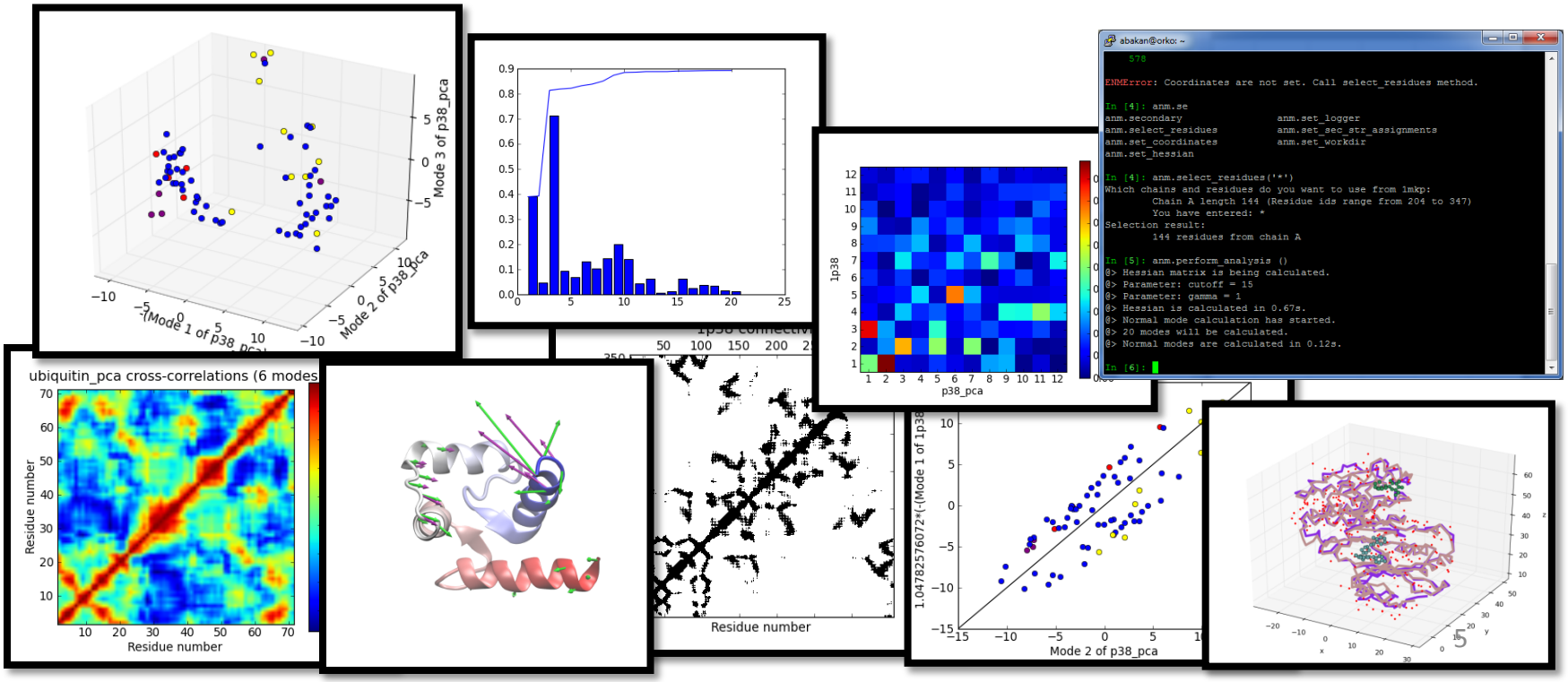
An Interactive Tool



matplotlib

NumPy

IP[y]: IPython
Interactive Computing



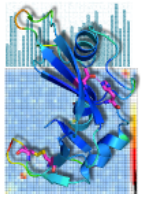
ProDy
Protein Dynamics & Sequence Analysis

MMBioS

An evolving suite of tools

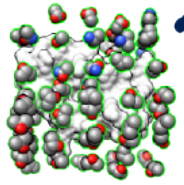
ProDy
Protein Dynamics & Sequence Analysis

Principal Component Analysis
Elastic Network Models
Normal Mode Analysis
Trajectory Analysis



Evol
Bridging Evolution & Dynamics

Multiple Sequence Alignment
Correlated Mutation Analysis
Structural Evolution



DruGUI
Druggability Suite for VMD

Computational Drug Discovery
Binding Site Prediction
Affinity Estimation

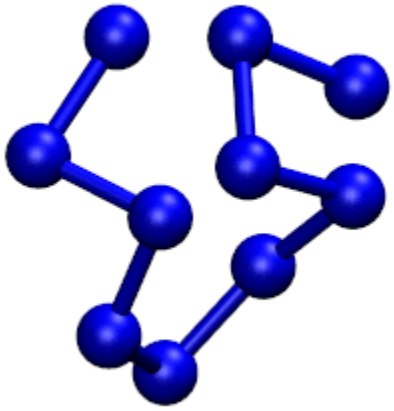
NMWiz
Normal Mode Wizard for VMD

Call ProDy from VMD
Normal Mode Visualization

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

Elastic Network Model

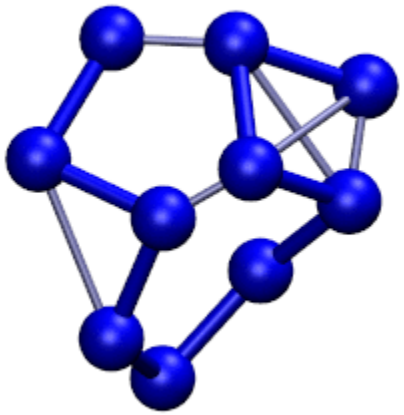


- Useful for finding global equilibrium motions of proteins
- Employs harmonic potential about native state
- Coarse-grained (Cα-only description)
- Residue pairs are connected via springs
- Normal modes are found analytically

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

Elastic Network Model

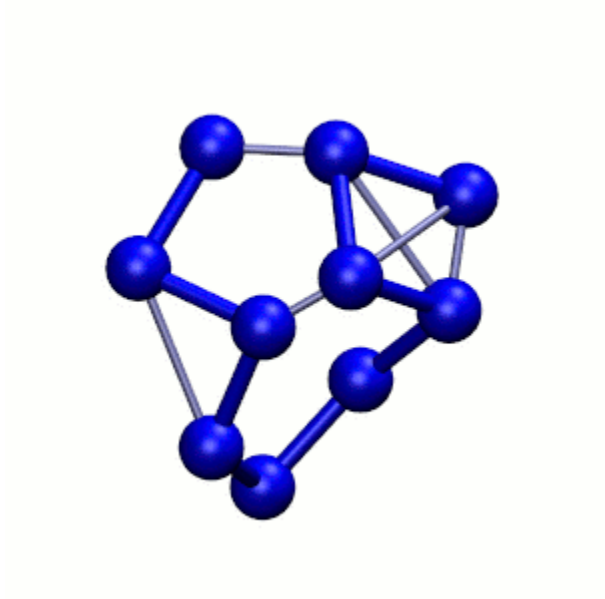


- Useful for finding global equilibrium motions of proteins
- Employs harmonic potential about native state
- Coarse-grained (Cα-only description)
- Residue pairs are connected via springs
- Normal modes are found analytically

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

Elastic Network Model



- Useful for finding global equilibrium motions of proteins
- Employs harmonic potential about native state
- Coarse-grained (Cα-only description)
- Residue pairs are connected via springs
- Normal modes are found analytically

Approximating the protein free energy landscape

Rough and funnel-shaped energy landscape that is function of atomic coordinates

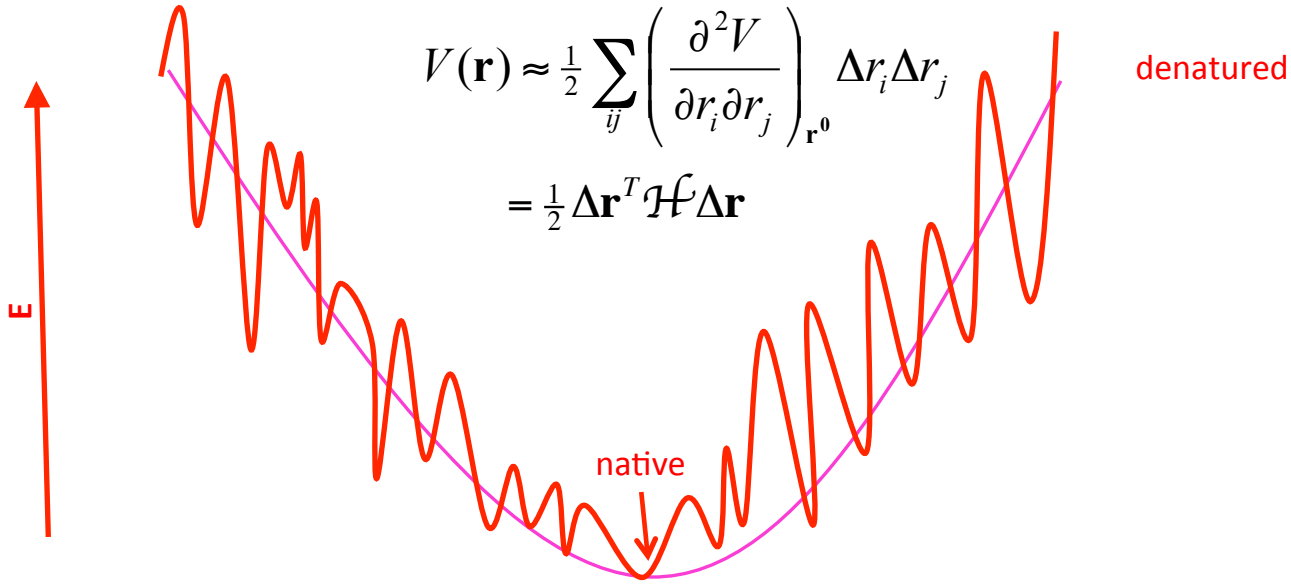
$$\mathbf{r} = (x_1, y_1, z_1, \ldots, x_N, y_N, z_N)^T$$

$$\Delta\mathbf{r} = \mathbf{r} - \mathbf{r}^0$$

$$V(\mathbf{r}) = V(\mathbf{r}^0) + \sum_i \left(\frac{\partial V}{\partial r_i}\right)_{\mathbf{r}^0} \Delta r_i + \tfrac{1}{2} \sum_{ij} \left(\frac{\partial^2 V}{\partial r_i \partial r_j}\right)_{\mathbf{r}^0} \Delta r_i \Delta r_j + \tfrac{1}{6} \sum_{ijk} \left(\frac{\partial^3 V}{\partial r_i \partial r_j \partial r_k}\right)_{\mathbf{r}^0} \Delta r_i \Delta r_j \Delta r_k + \ldots$$

$$V(\mathbf{r}) \approx \tfrac{1}{2} \sum_{ij} \left(\frac{\partial^2 V}{\partial r_i \partial r_j}\right)_{\mathbf{r}^0} \Delta r_i \Delta r_j$$

$$= \tfrac{1}{2} \Delta\mathbf{r}^T \mathcal{H} \Delta\mathbf{r}$$



ProDy
Protein Dynamics & Sequence Analysis

MMBioS

Equations of Motion

$$\mathbf{M}\Delta\ddot{\mathbf{r}} = -\mathcal{H}\Delta\mathbf{r}$$

Newton

$$\begin{cases} \mathbf{q} = \mathbf{M}^{1/2}\Delta\mathbf{r} \\ \mathbf{H} = \mathbf{M}^{-1/2}\mathcal{H}\mathbf{M}^{-1/2} \end{cases}$$

Mass-weighted coordinates

Mass-weighted Hessian

$$\ddot{\mathbf{q}} = -\mathbf{M}^{-1/2}\mathcal{H}\mathbf{M}^{-1/2}\mathbf{q}$$
$$= -\mathbf{H}\mathbf{q}$$

$$\begin{cases} \mathbf{q}(t) = \mathbf{q}e^{-i\omega t} \\ \ddot{\mathbf{q}}(t) = -\omega^2\mathbf{q}(t) \end{cases}$$

Oscillatory solution

$$\mathbf{H} = \mathbf{V}\Lambda\tilde{\mathbf{V}}$$
$$\Lambda_{ij} = \omega_i^2\delta_{ij}$$

Goldstein, Ch. 10

Anisotropic Network Model

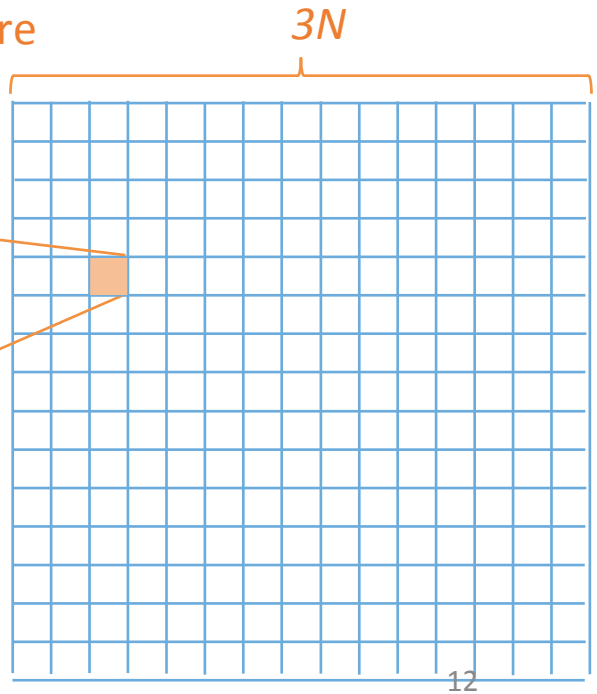
$$V(\mathbf{r}) = \frac{\gamma}{2}\sum_{i=1}^{N}\sum_{j>i}\left(\left|\mathbf{r}_{ij}\right| - \left|\mathbf{r}_{ij}^{0}\right|\right)^{2}\Theta\left(R_c - \left|\mathbf{r}_{ij}^{0}\right|\right)$$

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

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



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

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

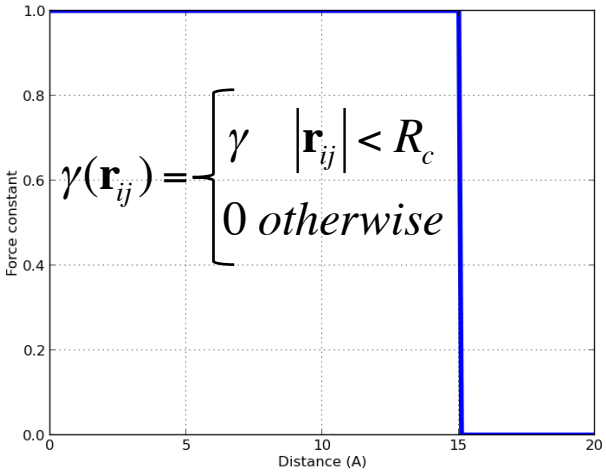
ENMs are flexible!

- Selection of the ENM force constants
  - Cutoff distance
  - Functional form
- Changing the potential directly
- Adding rigidity
- Including external elements, or looking at only a subsystem
- Altering network topology

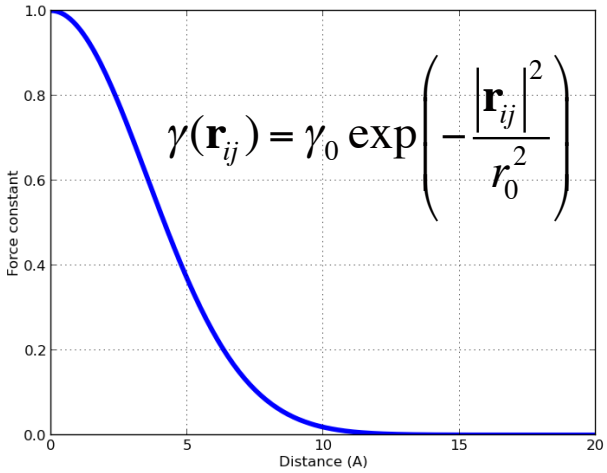
ProDy
Protein Dynamics & Sequence Analysis

MMBioS

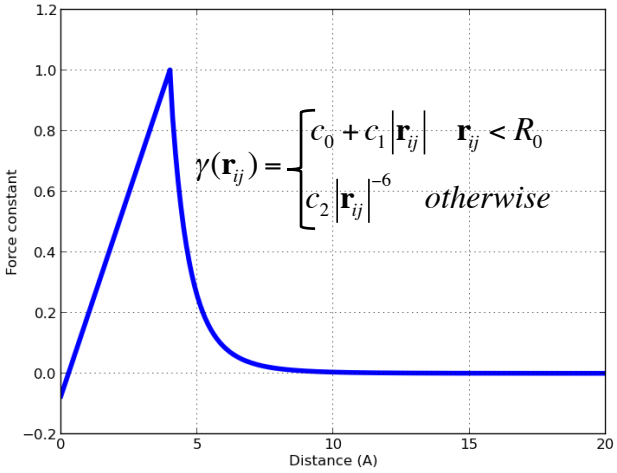
Flexible force constants



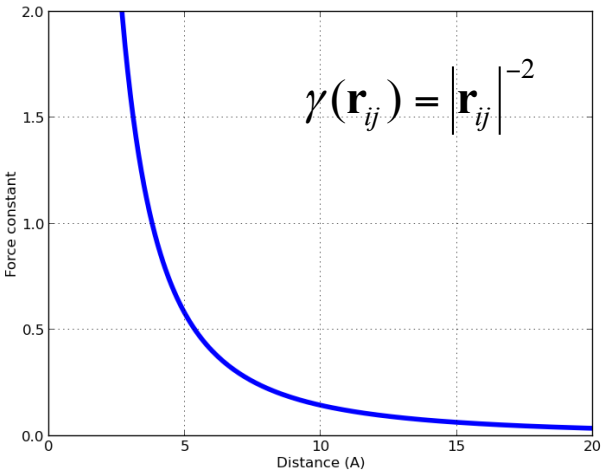
Tirion, PRL 77 (1996).



Hinsen et al. Proteins 33 (1998).



Hinsen et al. Chem Phys 261 (2000).



Yang et al. PNAS 106 (2009).

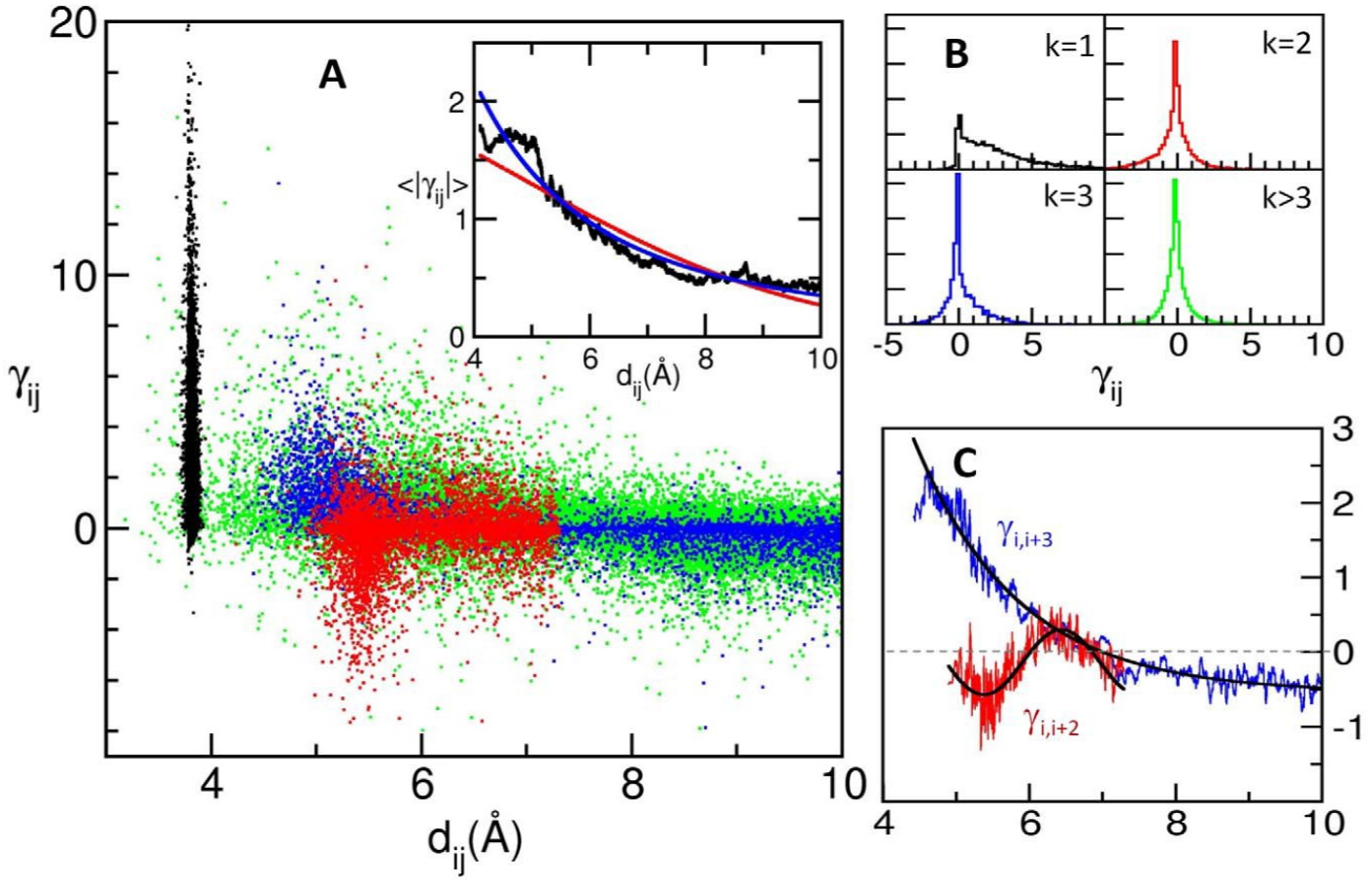
ProDy
Protein Dynamics & Sequence Analysis

MMBioS

Optimizing force constants

- Download NMR structures from PDB — fetchPDB()
- Calculate residue MSFs for each protein — calcMSF()
- Assign ENM topology — buildHessian()
- Optimize force constants to reproduce structural dynamics
- Search for trends in force constant values with structure

Flexible force constants

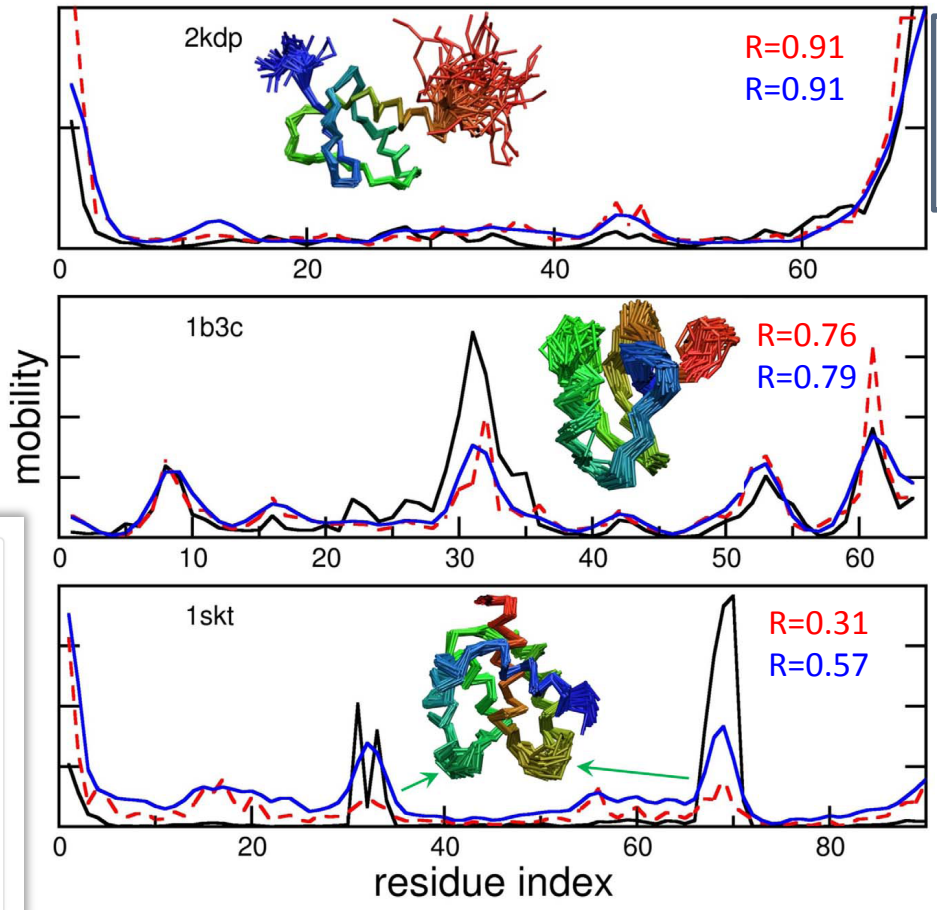


Lezon & Bahar. PLoS Comp Biol 6 (2010).

ProDy
Protein Dynamics & Sequence Analysis

MMBioS

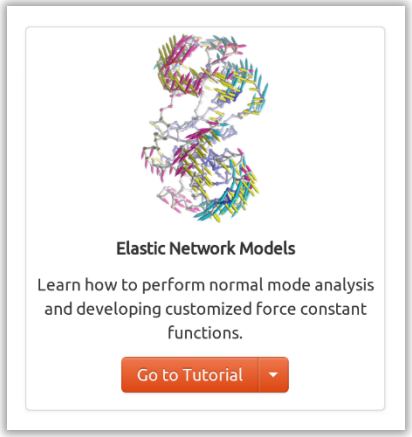
Fine-tuning force constants



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

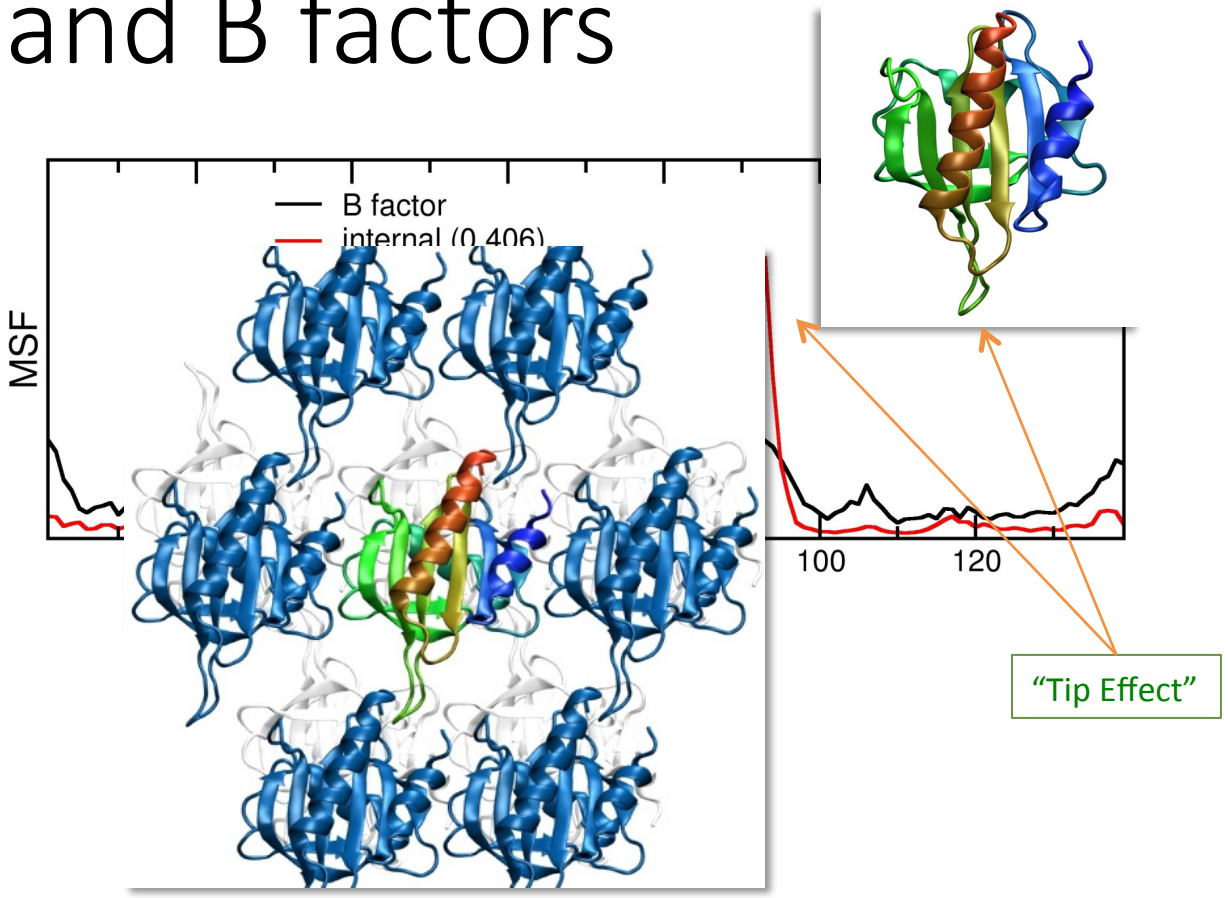
black: NMR
red: ENM
blue: modified ENM

GammaStructureBased()



Learn more at prody.csb.pitt.edu

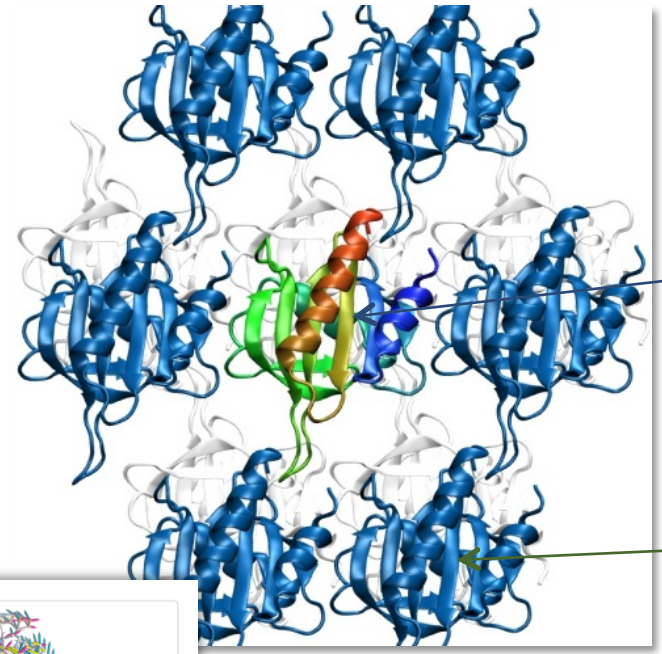
Improving correlation between ENM and B factors

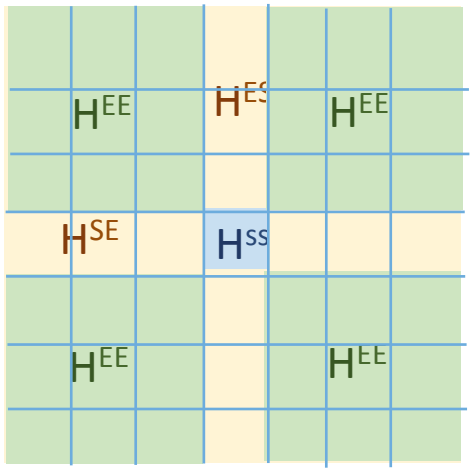


Possible solution: Add crystal contacts

Lezon. Proteins 80 (2012).

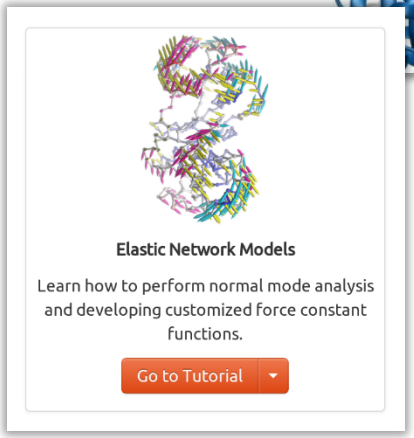
Explicit membrane models





"system"

"environment"



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

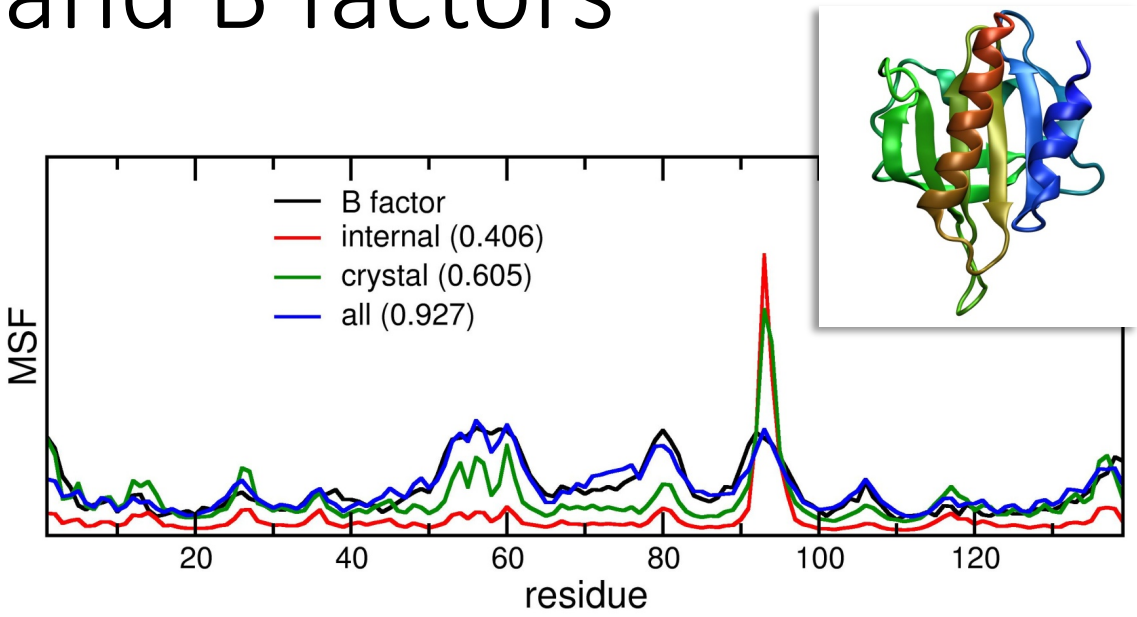
$$V_{eff}(\mathbf{s}) = \tfrac{1}{2}\Delta\mathbf{s}^T\left(\mathbf{H}^{ss\,\prime}\right)\Delta\mathbf{s}$$

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

reduceModel()

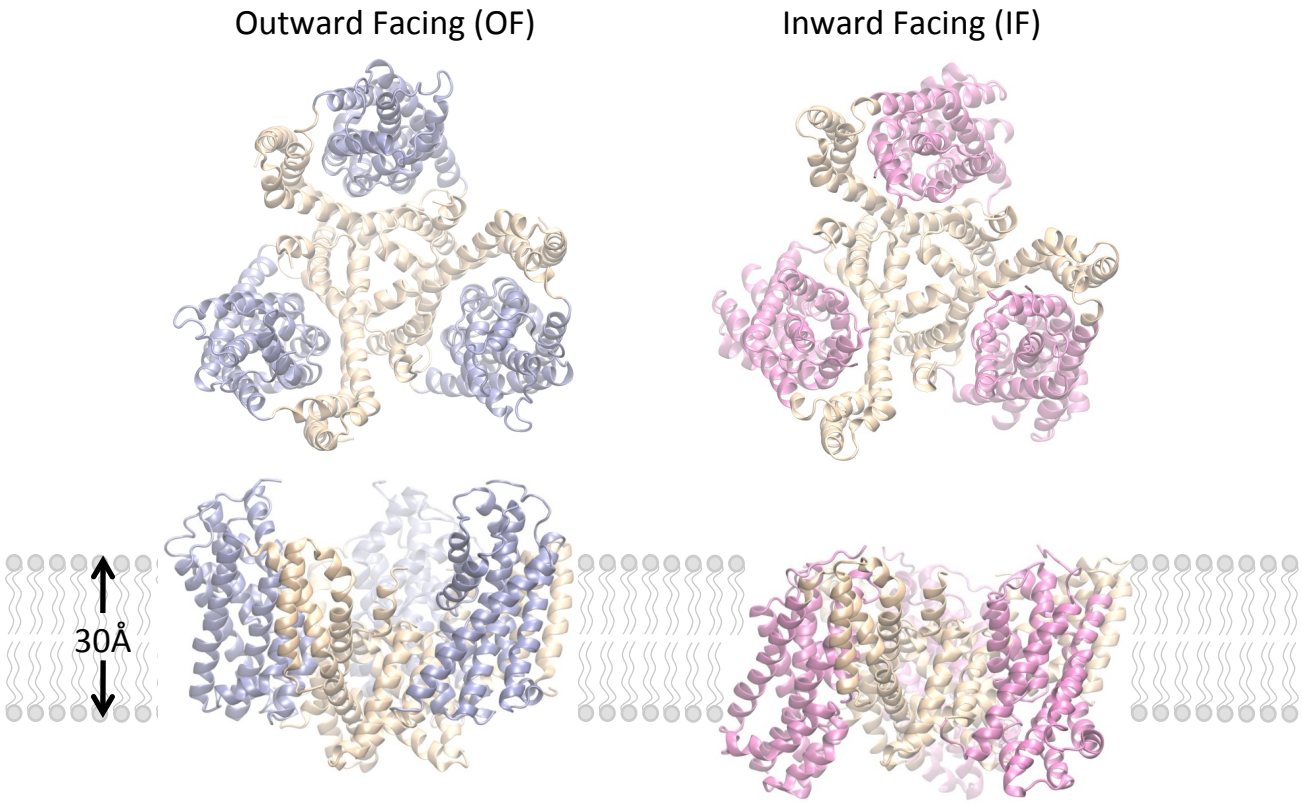
Learn more at prody.csb.pitt.edu

Improving correlation between ENM and B factors



Crystal contacts help, but do not eliminate tip effect.
What about lattice vibrations?

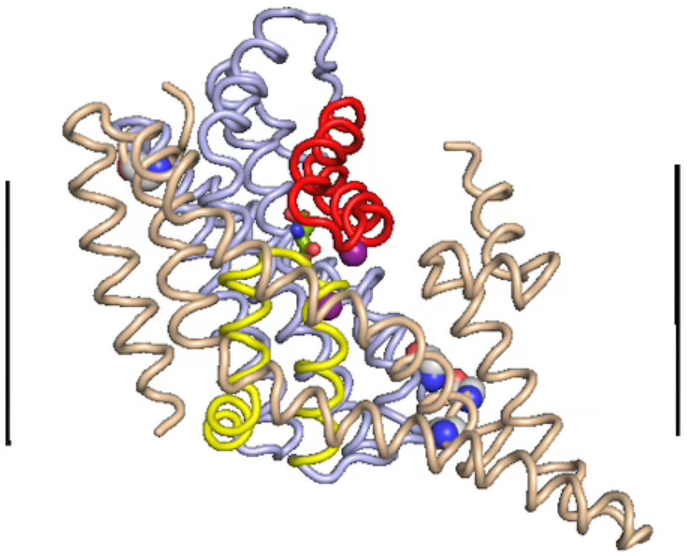
Global transitions



Yernool *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).

Rotations-Translations of Blocks

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

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

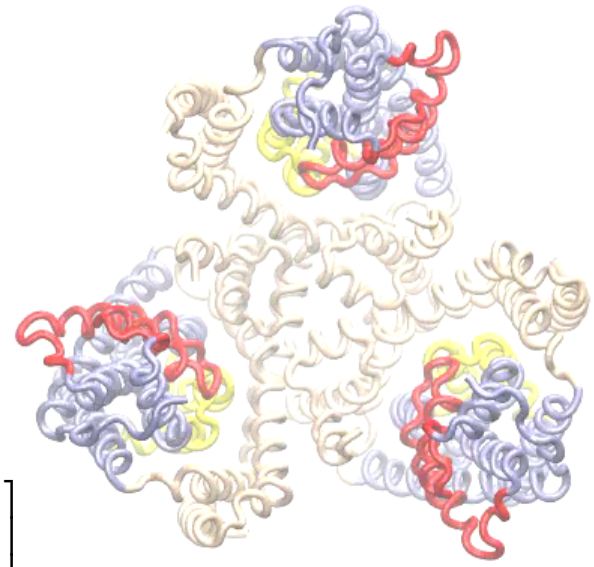
ANM predicts large radial motions of the trimer. Can we invent a better model?

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

Altered radial force constants:

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

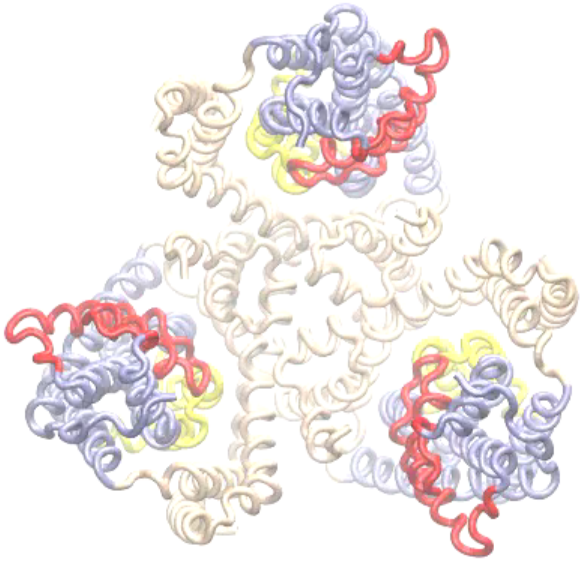
$$\mathbf{H_{ij}} = -\frac{\gamma}{\left(R_{ij}^0\right)^2}\begin{bmatrix} \left(x_{ij}^0\right)^2 & x_{ij}^0 y_{ij}^0 & cx_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & \left(y_{ij}^0\right)^2 & cy_{ij}^0 z_{ij}^0 \\ cx_{ij}^0 z_{ij}^0 & cy_{ij}^0 z_{ij}^0 & \left(cz_{ij}^0\right)^2 \end{bmatrix}$$



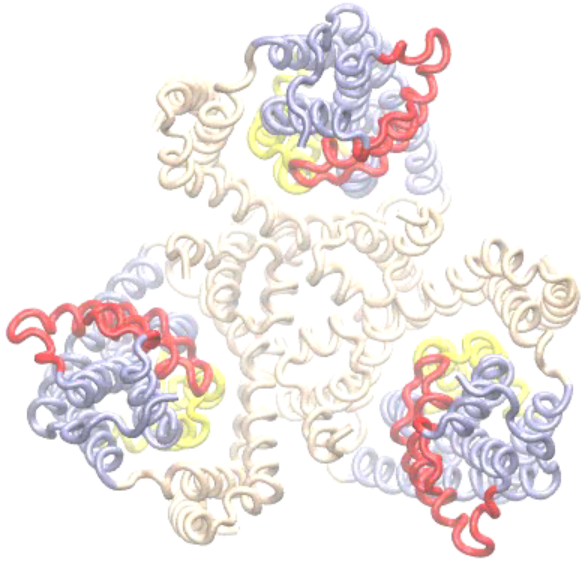
RTB.buildHessian()

Lezon & Bahar. Biophys J 102 (2012).

Exploring structural transitions: Glutamate transporter

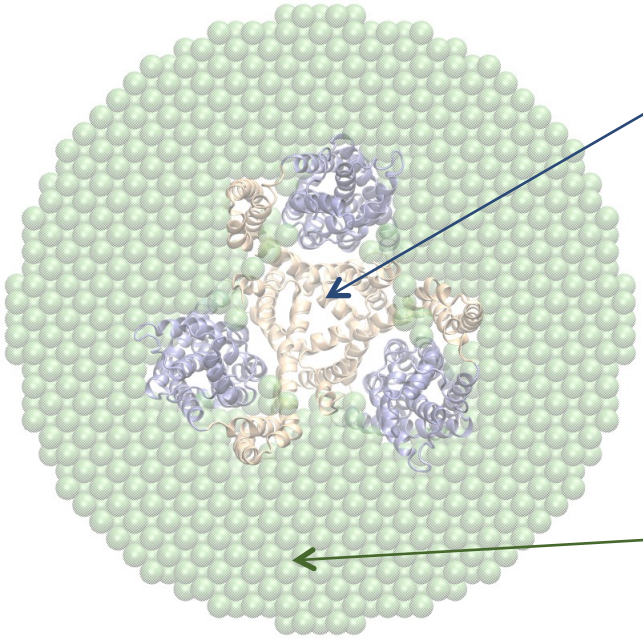


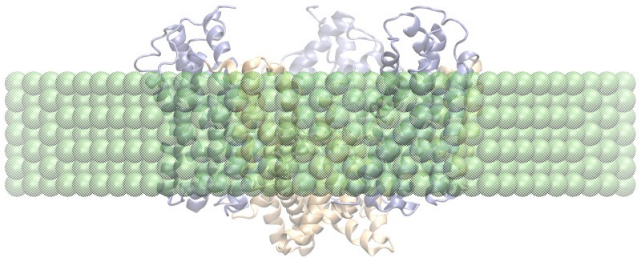
ANM: Large radial motions

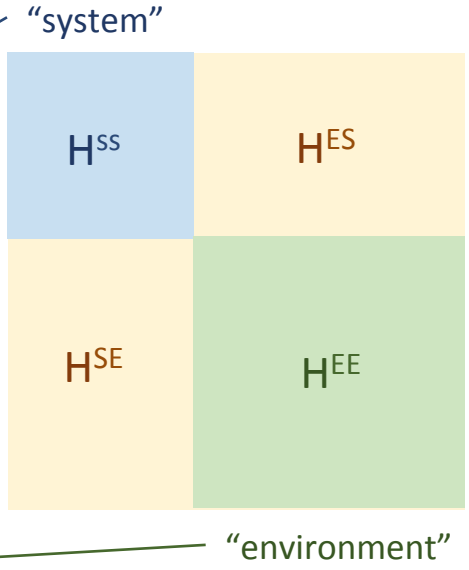


imANM

Explicit membrane models







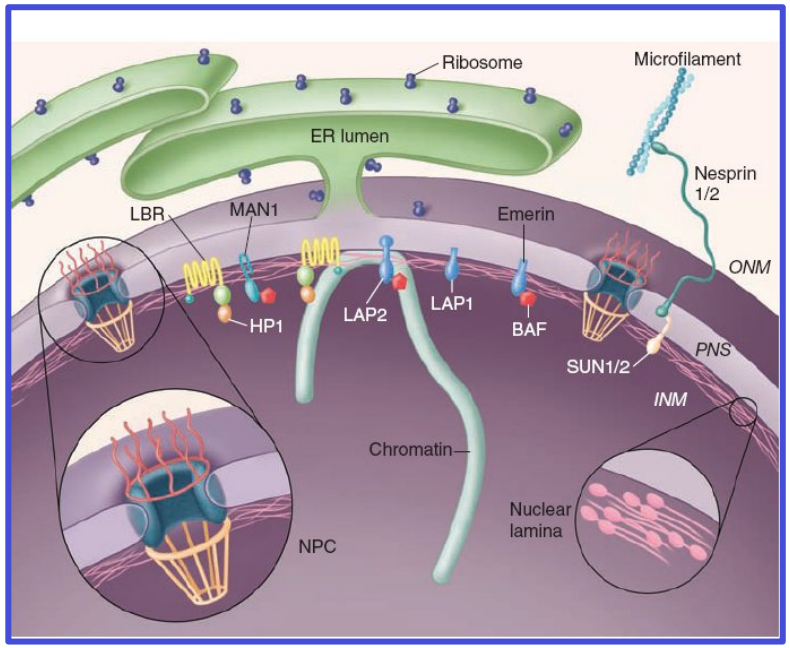
As the *environment* fluctuates randomly, the effective motion of the *system* is given by

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

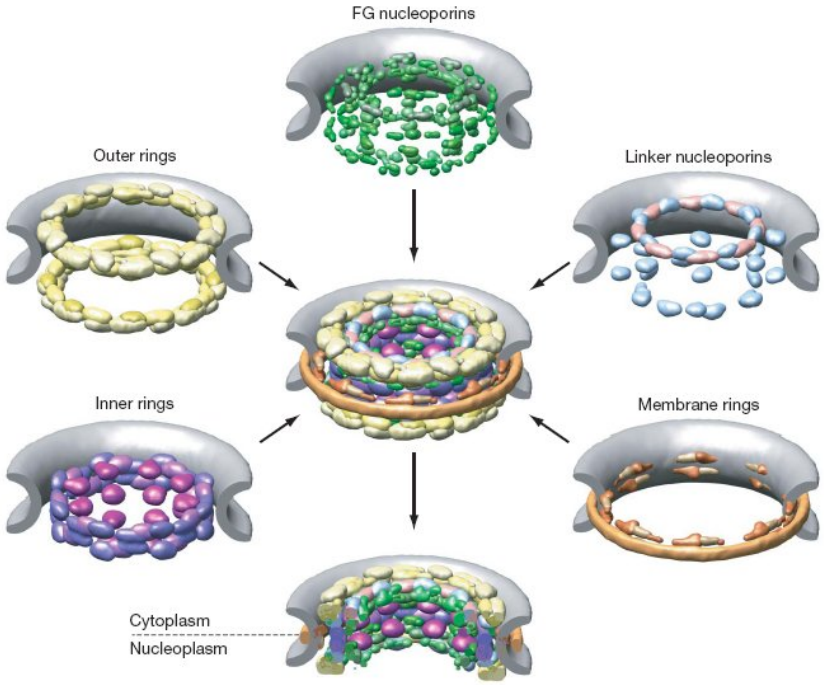
$$\mathbf{H}^{ss\,\prime} = \mathbf{H}^{ss} - \mathbf{H}^{SE}\left(\mathbf{H}^{EE}\right)^{-1}\mathbf{H}^{ES}$$

reduceModel()

Thinking bigger: NPC



Stewart *et al*. Science 318 (2007).



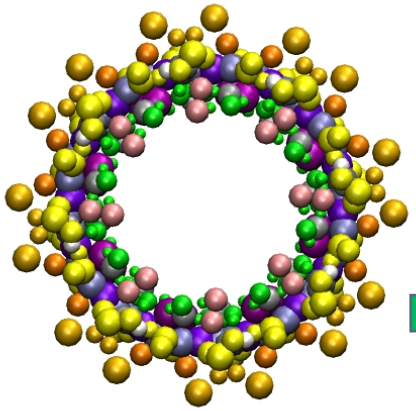
Alber *et al*. Nature 450 (2007).

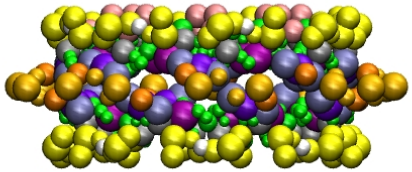
ENM of NPC

NPC Model

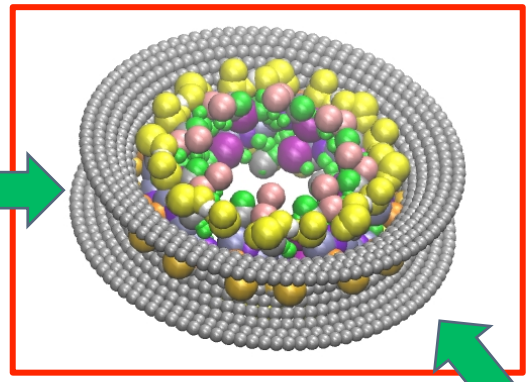
R_c = 12nm protein-protein, 8nm NE-NE, 10nm protein-NE

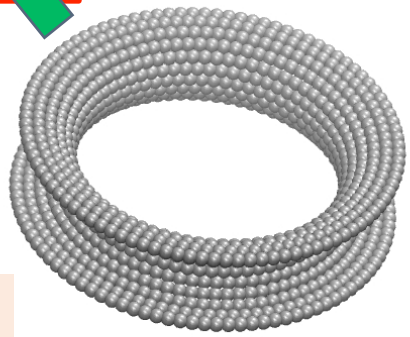
γ = 1.0 protein-protein, 10 NE-NE





456 proteins:
57 proteins/spoke
30 unique proteins





NE:
2070 discrete points

gammaVariableCutoff()
reduceModel()
parseSparseMatrix()

Lezon *et al.* PLoS Comp Biol 5 (2009).

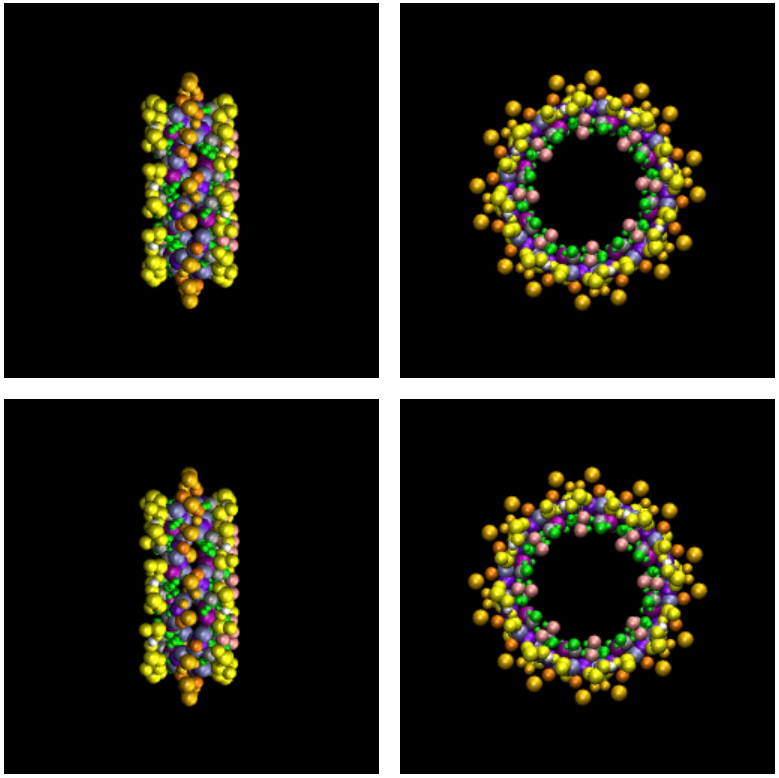
ProDy
Protein Dynamics & Sequence Analysis

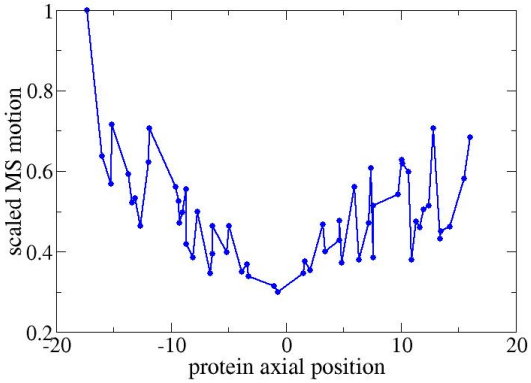
MMBioS

NPC slow modes

Mode 1

Mode 2

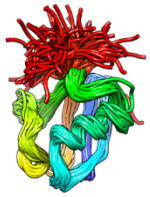




ProDy
Protein Dynamics & Sequence Analysis

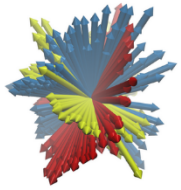
MMBioS

Tutorials



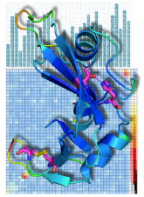
ProDy

Learn how to use ProDy from the introductory ProDy tutorial or from the comprehensive API reference manual.



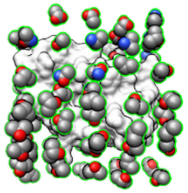
Normal Mode Wizard

Learn how to depict normal modes and generate animations of protein dynamics along them with NMWiz.



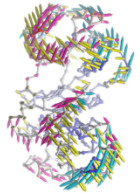
Evol

Learn how to identify conserved and coevolved residues and characterizing their dynamical properties.



Drugability Suite

Learn how to setup and analyze druggability simulations containing small organic molecules using DruGUI.

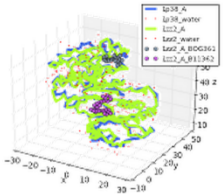


Elastic Network Models

Learn how to perform normal mode analysis and developing customized force constant functions.

Ensemble Analysis

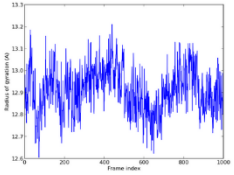
Learn how to analyze large and heterogeneous ensembles of protein structures to infer dynamical properties.



Structure Analysis

Learn how to compare and align structures, identify ligand contacts, and extract ligands from PDB files.

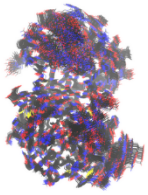
Go to Tutorial



Trajectory Analysis

Learn how to analyze simulation trajectories, in particular handling large trajectory files that don't fit in memory.

Go to Tutorial



Conformational Sampling

Learn how to generate alternate protein conformations along ANM modes and to refine them using NAMD.

Go to Tutorial