



# ProDy

Protein Dynamics & Sequence Analysis

# Overview & Applications

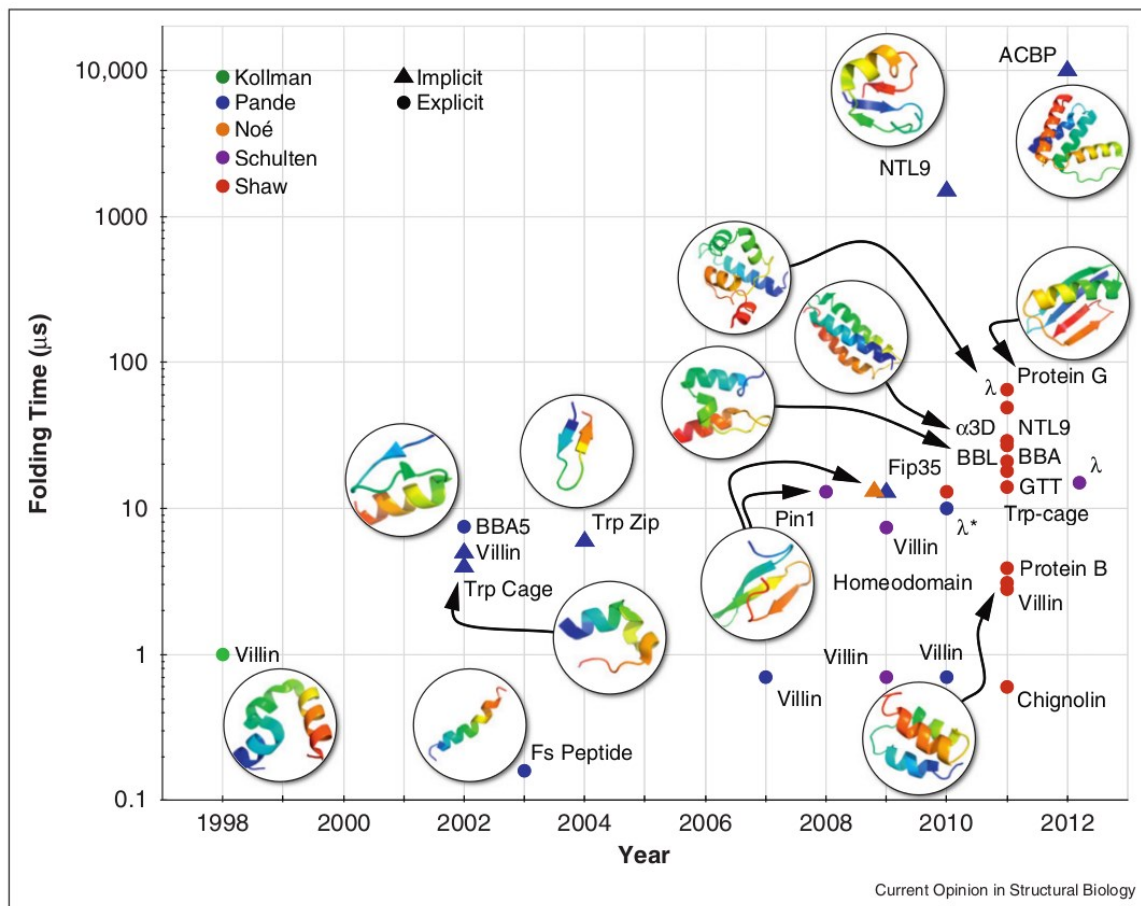
T. Lezon

Hands-on Workshop in Computational Biophysics

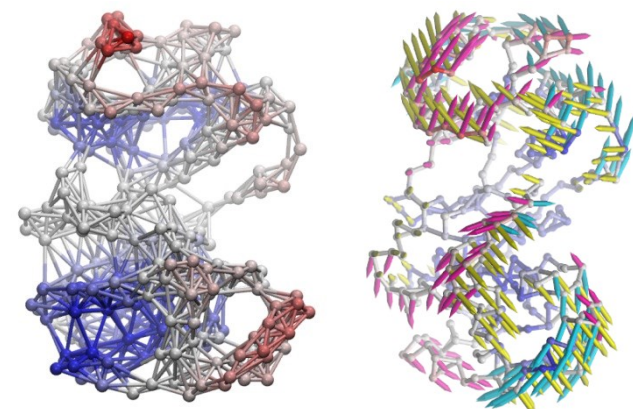
Pittsburgh Supercomputing Center

04 June, 2015

# Simulations still take time



Lane et al. 2013



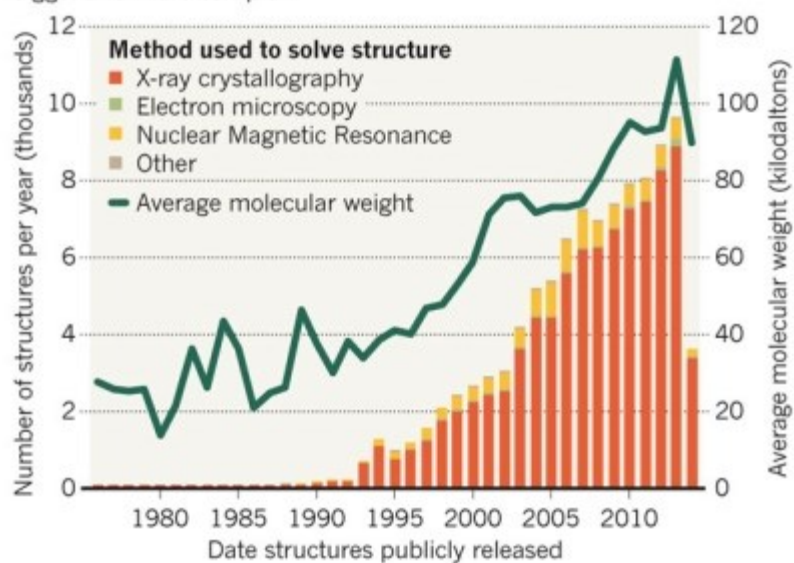
Bakan et al. Bioinformatics 2011.

Coarse-grained Elastic  
Network Models are fast

# The structural data explosion

## ONE HUNDRED THOUSAND PROTEIN STRUCTURES

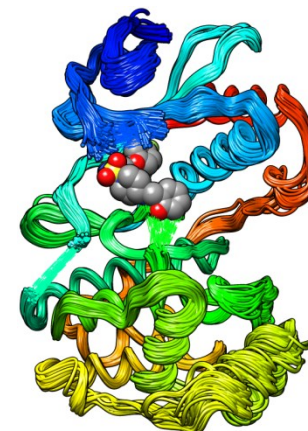
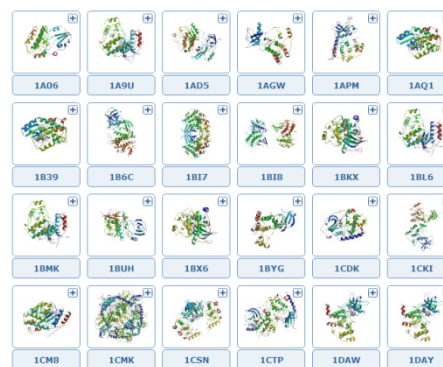
Biomolecular structures stored in the Protein Data Bank are getting bigger and more complex.



Nature, 15 May 2014.

Multiple structures for a single sequence

RCSB **PDB**  
PROTEIN DATA BANK



Dynamics may be inferred from structural data.

# Exploiting the PDB since 2010

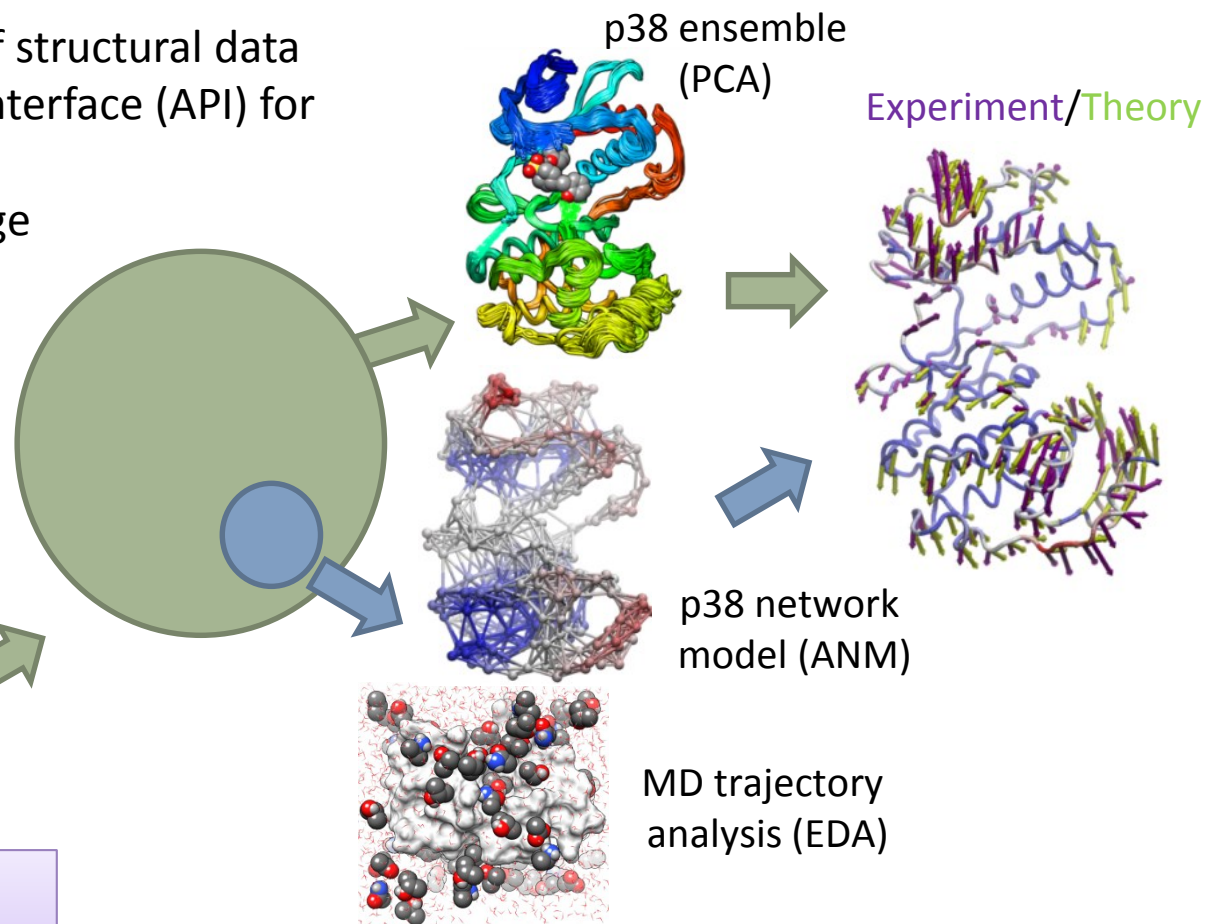
- High-throughput analysis of structural data
- Application Programming Interface (API) for development of tools
- Suitable for interactive usage

User inputs a sequence

**Usage example**

```
>1A9U:A|PDBID|CHAIN
GSSHHHHHHSSGLVPRGSHMSQ
ERPTFYRQELNKTIEWVPERYQ
NLSPVGSAYGSVCAAFDTKTG
.....
```

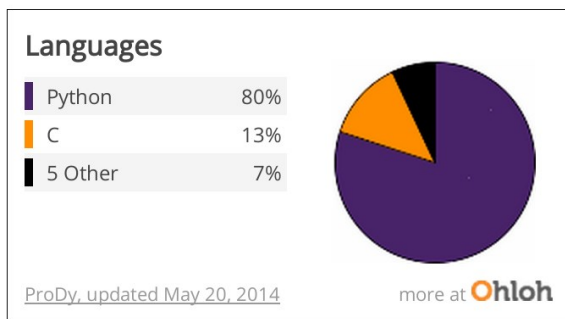
**ProDy identifies, retrieves, aligns, and analyzes (PCA) structures matching input sequence**



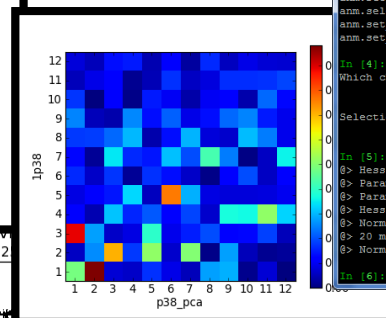
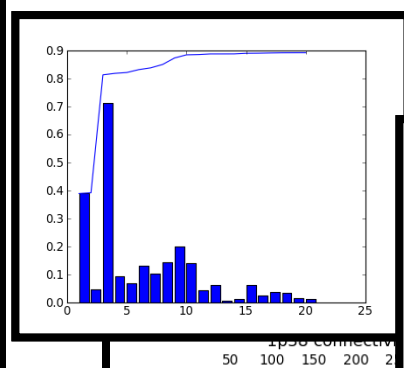
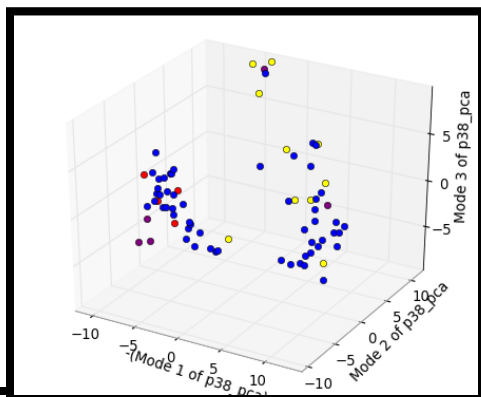
User can

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

# An Interactive Tool



**IP[y]:** IPython  
Interactive Computing



```

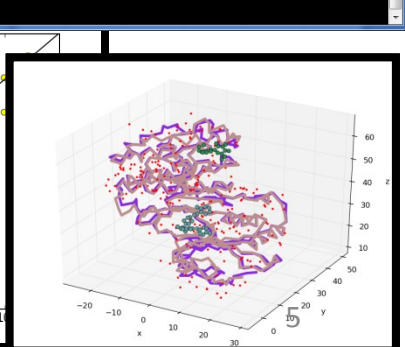
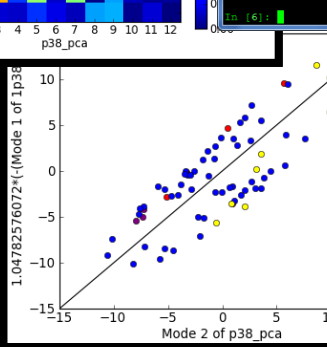
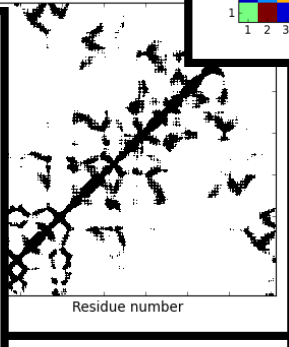
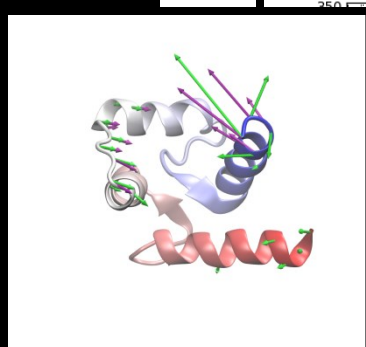
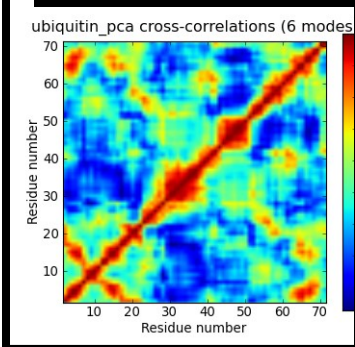
abakan@orko: ~
578
ENMError: Coordinates are not set. Call select_residues method.

In [4]: anm.se
anm.secondary          anm.set_logger
anm.select_residues    anm.set_sec_str_assignments
anm.set_coordinates    anm.set_workdir
anm.set_hessian

In [4]: anm.select_residues('*')
Which chains and residues do you want to use from lmkp:
Chain A length 144 (Residue ids range from 204 to 347)
You have entered: *
Selection result:
144 residues from chain A

In [5]: anm.perform_analysis ()
@> Hessian matrix is being calculated.
@> Parameter: cutoff = 15
@> Parameter: gamma = 1
@> Hessian is calculated in 0.67s.
@> Normal mode calculation has started.
@> 20 modes will be calculated.
@> Normal modes are calculated in 0.12s.

In [6]:
    
```



# An evolving suite of tools



Principal Component Analysis  
Elastic Network Models  
Normal Mode Analysis  
Trajectory Analysis



Multiple Sequence Alignment  
Correlated Mutation Analysis  
Structural Evolution



Computational Drug Discovery  
Binding Site Prediction  
Affinity Estimation



Call ProDy from VMD  
Normal Mode Visualization

# Tutorials: ProDy & Structure Analysis

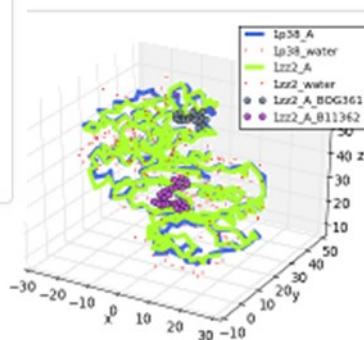


ProDy

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

Tutorial

Manual



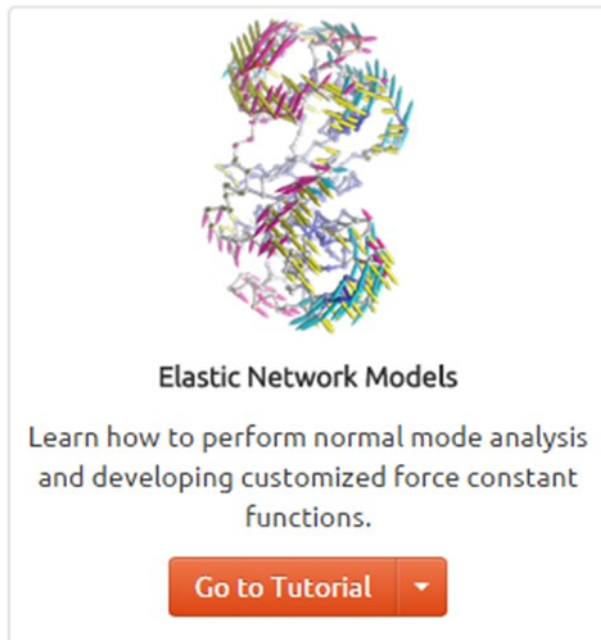
Structure Analysis


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

Go to Tutorial

- Obtaining PDB Files
- BLAST Searching the PDB
- Constructing Biomolecules from Transformations
- Aligning and Comparing Structures
- Identifying Intermolecular Contacts

# Tutorial: Elastic Network Models





**Elastic Network Models**

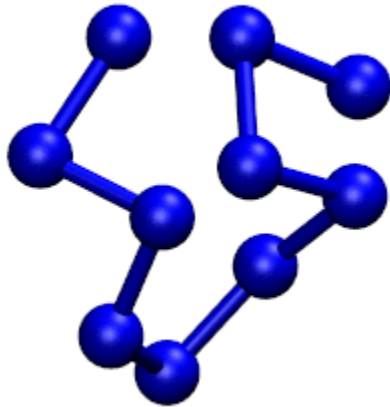
Learn how to perform normal mode analysis and developing customized force constant functions.

[Go to Tutorial](#) ▾

- Gaussian Network Model (GNM)
- Anisotropic Network Model (ANM)
- Normal Mode Algebra
- Deformation Analysis
- Customizing ENMs

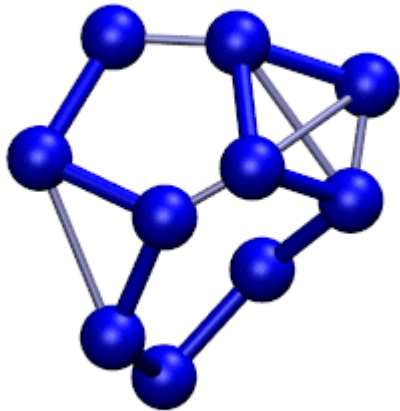


# Elastic Network Model



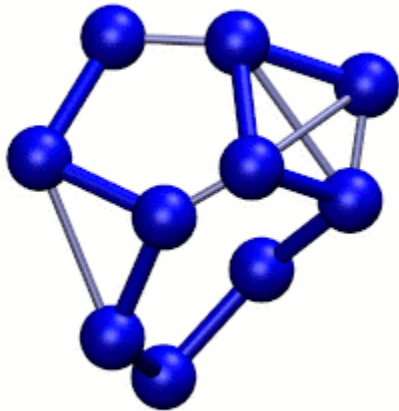
- Useful for finding global equilibrium motions of proteins
- Employs harmonic potential about native state
- Coarse-grained (C $\alpha$ -only description)
- Residue pairs are connected via springs
- Normal modes are found analytically

# Elastic Network Model



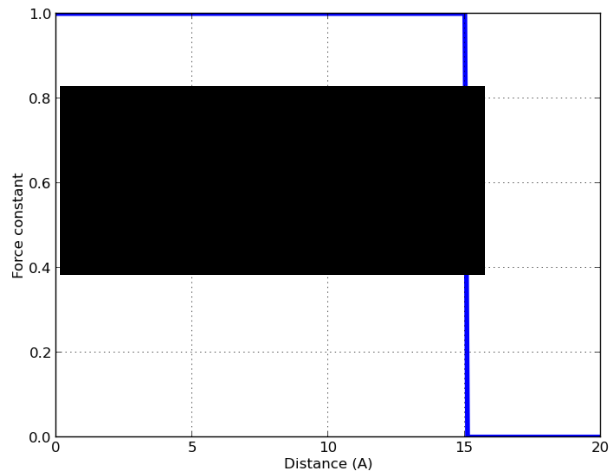
- Useful for finding global equilibrium motions of proteins
- Employs harmonic potential about native state
- Coarse-grained (C $\alpha$ -only description)
- Residue pairs are connected via springs
- Normal modes are found analytically

# Elastic Network Model

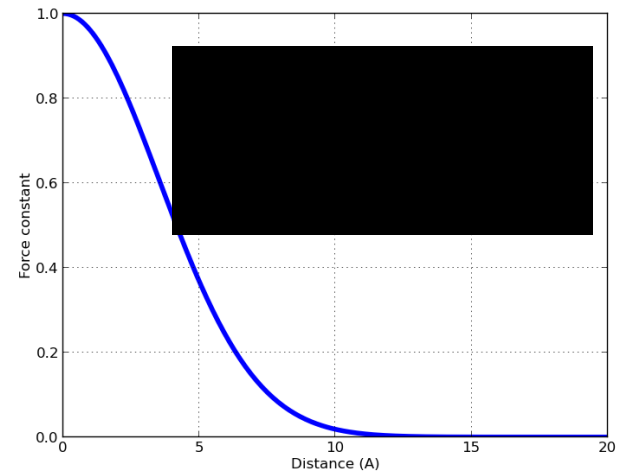


- Useful for finding global equilibrium motions of proteins
- Employs harmonic potential about native state
- Coarse-grained (C $\alpha$ -only description)
- Residue pairs are connected via springs
- Normal modes are found analytically

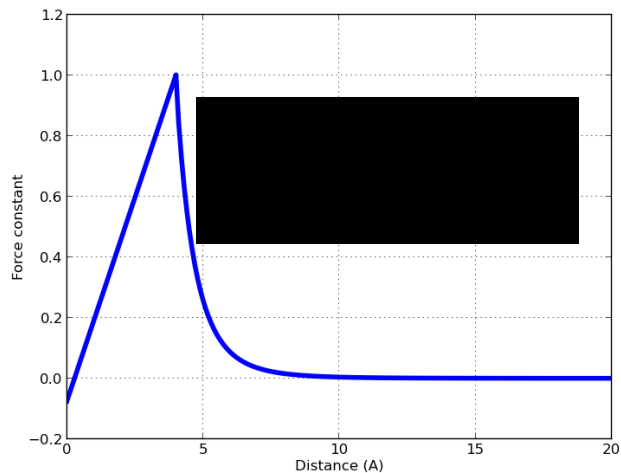
# Flexible force constants



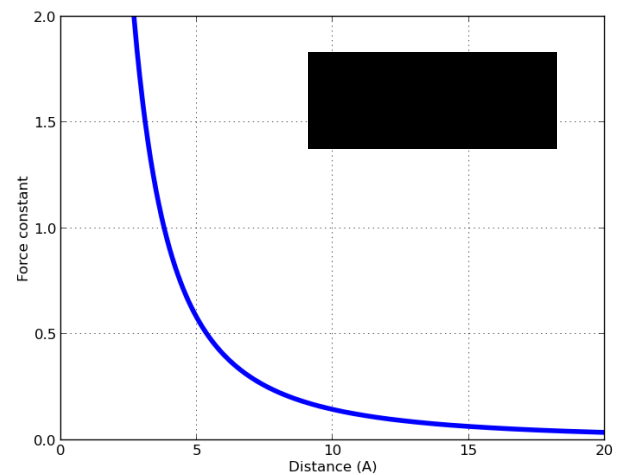
Tirion, PRL 77 (1996).



Hinsen et al. Proteins 33 (1998).



Hinsen et al. Chem Phys 261 (2000).



Yang et al. PNAS 106 (2009).

# Optimizing force constants

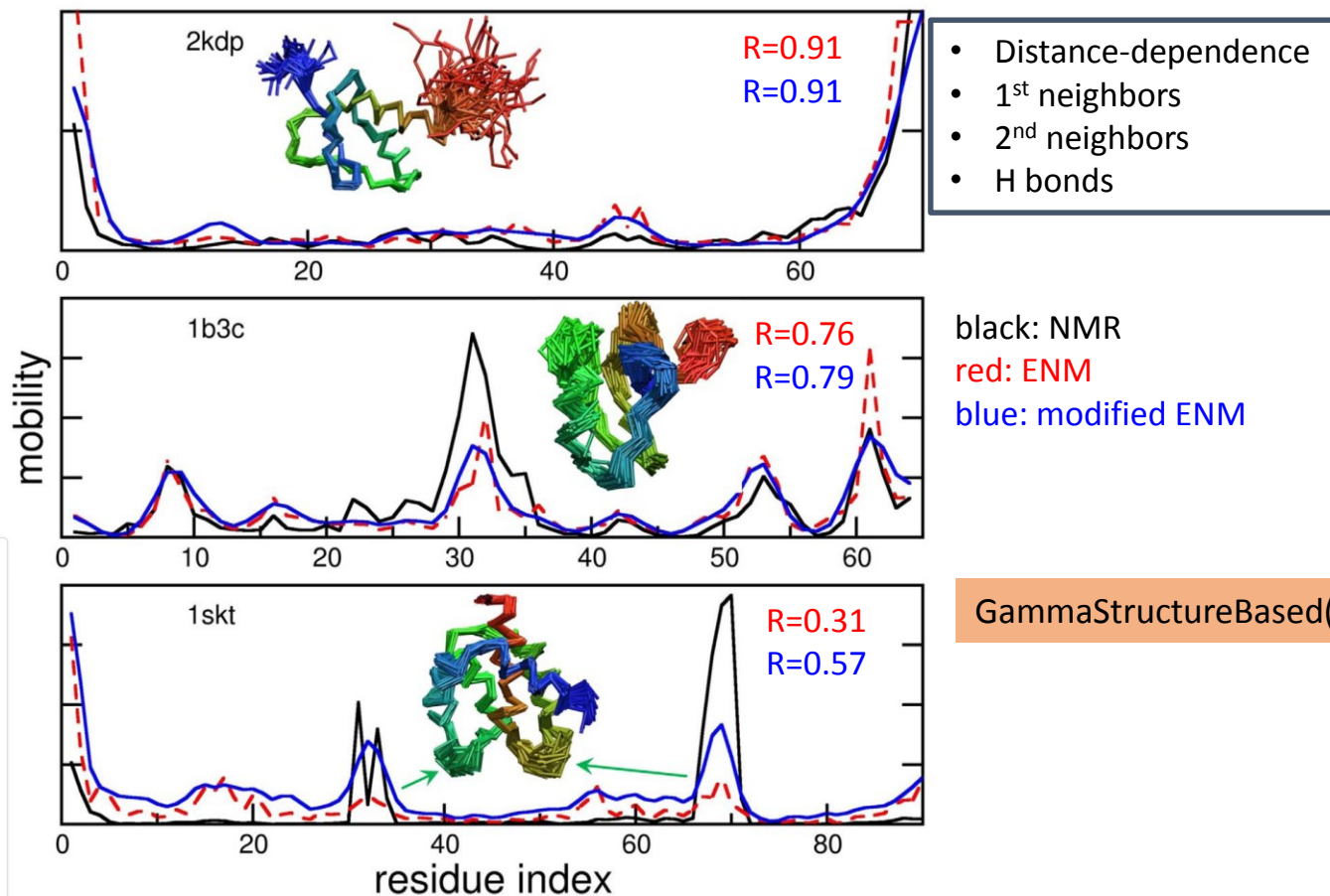
- Download NMR structures from PDB
- Calculate residue MSFs for each protein
- Assign ENM topology
- Optimize force constants to reproduce structural dynamics
- Search for trends in force constant values with structure

fetchPDB()

calcMSF()

buildHessian()

# Fine-tuning force constants




**Elastic Network Models**

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

[Go to Tutorial](#)

# Tutorial: Ensemble Analysis



**Ensemble Analysis**

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

[Go to Tutorial](#)

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

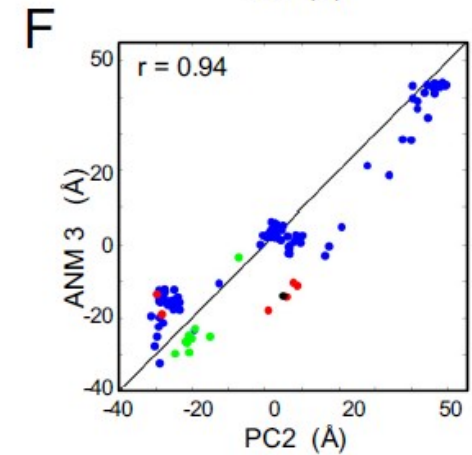
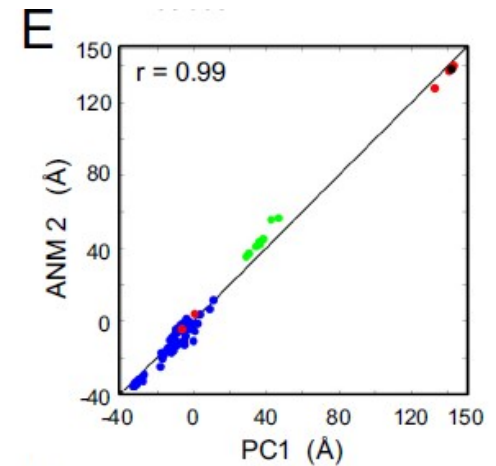
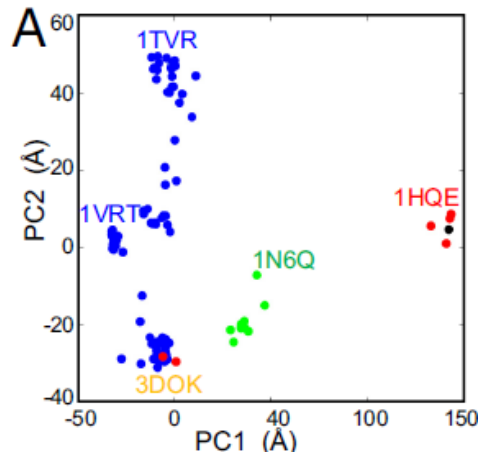
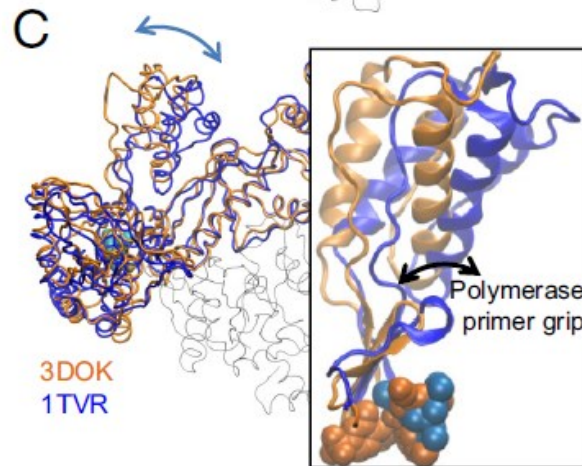
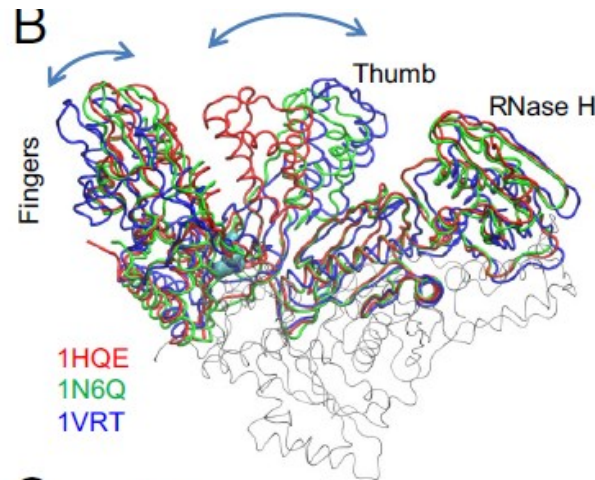
# Example: Comparing PCA and ENM

Structures of HIV1-RT

Unbound

Inhibitor bound

DNA bound





# Example: Comparing PCA and ENM

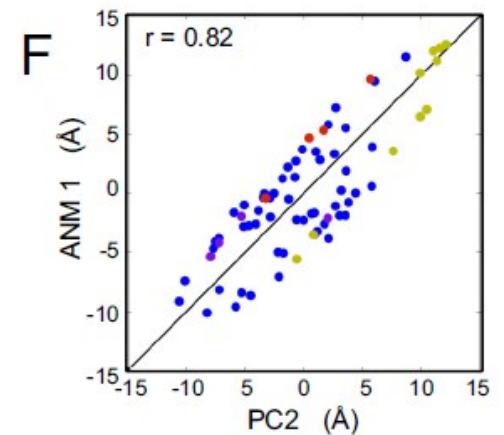
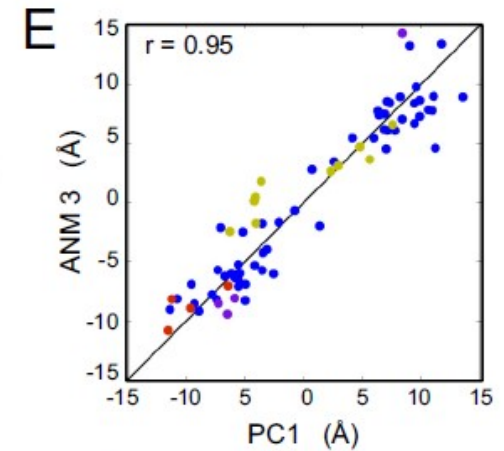
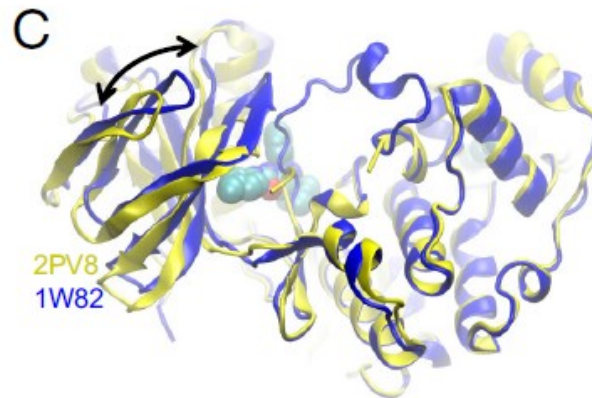
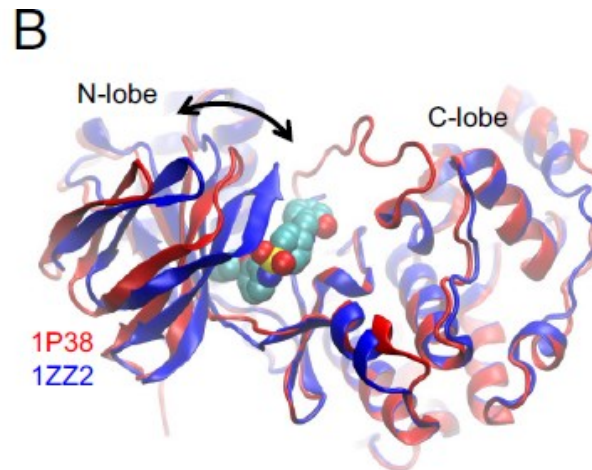
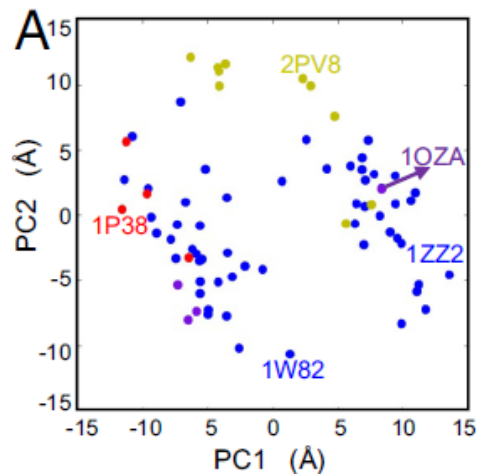
Structures of p38 MAPK

Unbound

Inhibitor bound

Glucose bound

Peptide bound



# Tutorial: Trajectory Analysis

- Fast processing of long trajectories
- Enables comparison of MD trajectories and structural data or ENM results



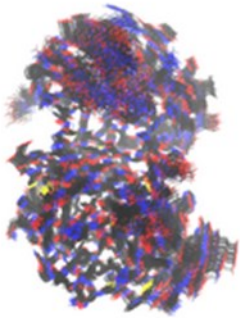
The plot shows the radius of gyration (Rg) in Angstroms (Å) on the y-axis, ranging from 12.6 to 13.3. The x-axis is labeled 'Frame index' and ranges from 0 to 1000. The data is represented by a blue line that fluctuates significantly, with peaks around 13.2 Å and troughs around 12.7 Å.

**Trajectory Analysis**

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

[Go to Tutorial](#) 

# Tutorial: Conformational Sampling



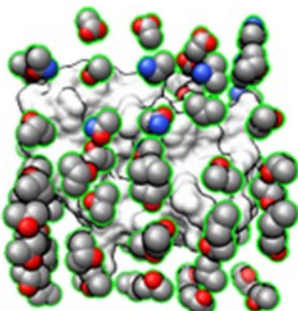
**Conformational Sampling**

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

[Go to Tutorial](#) ▶

- Sample structures along normal modes
- Refine structures using NAMD

# Tutorial: Druggability



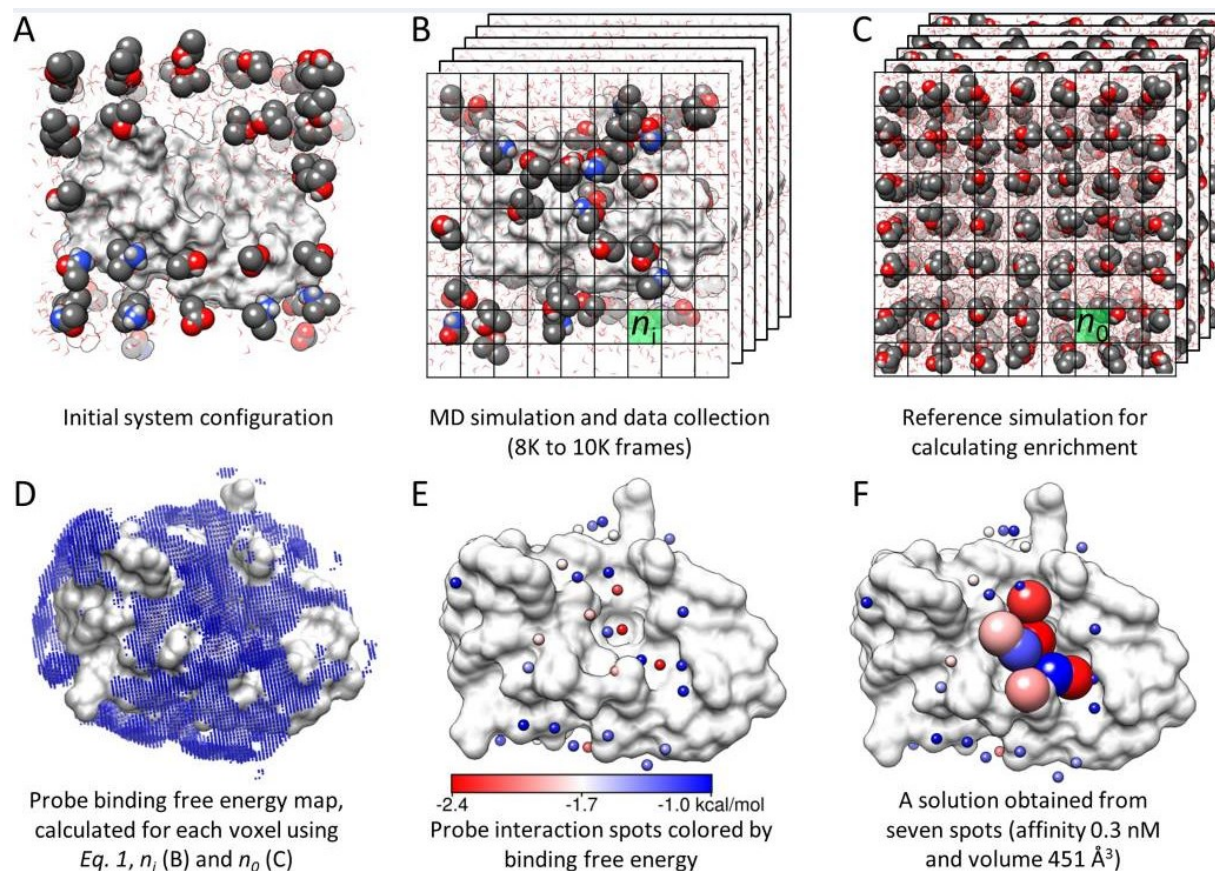
**Drugability Suite**

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

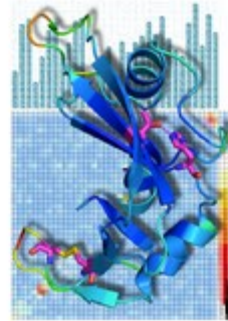
[Go to Tutorial](#)

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

# Exploring binding with probe molecules



# Tutorial: Evol

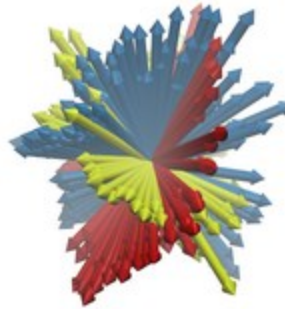


## Evol

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

[Go to Tutorial](#) ▼

# Tutorial: Normal Mode Wizard

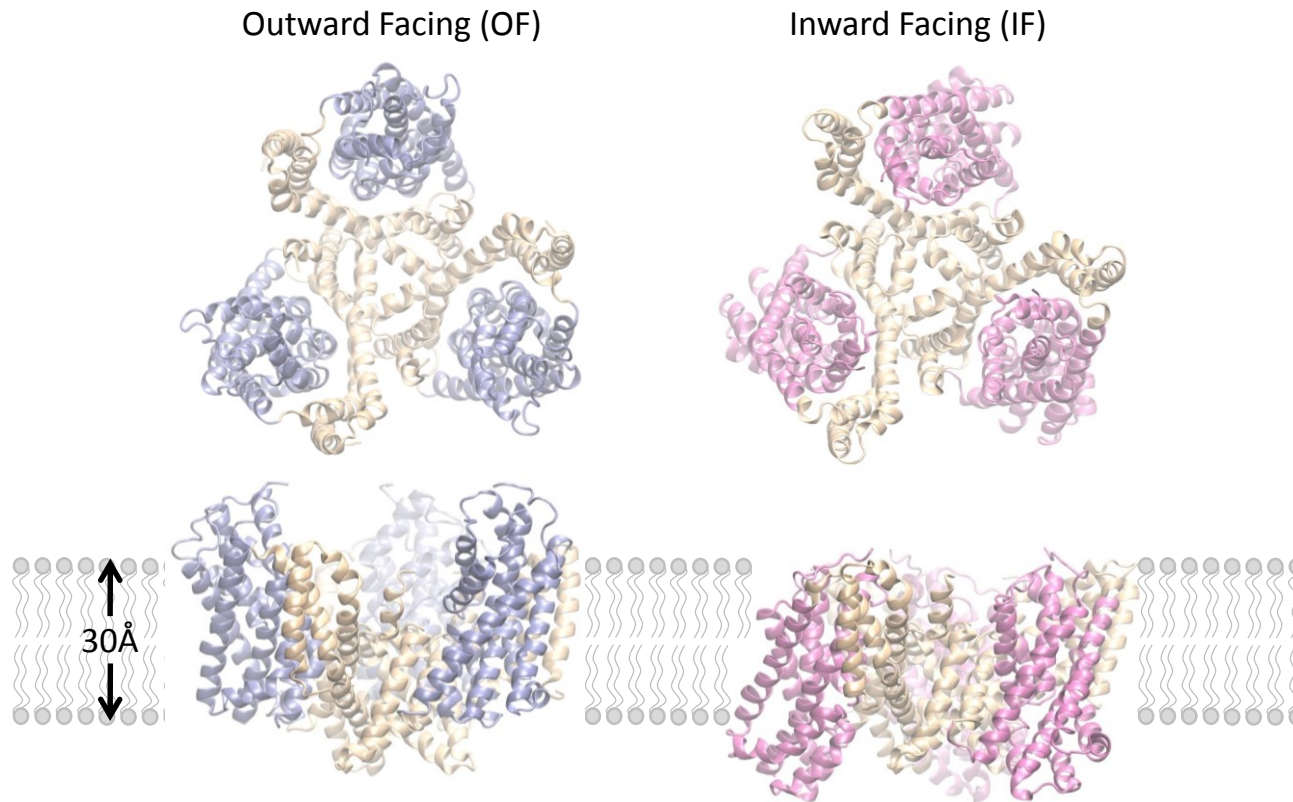


## Normal Mode Wizard

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

[Go to Tutorial](#) ▼

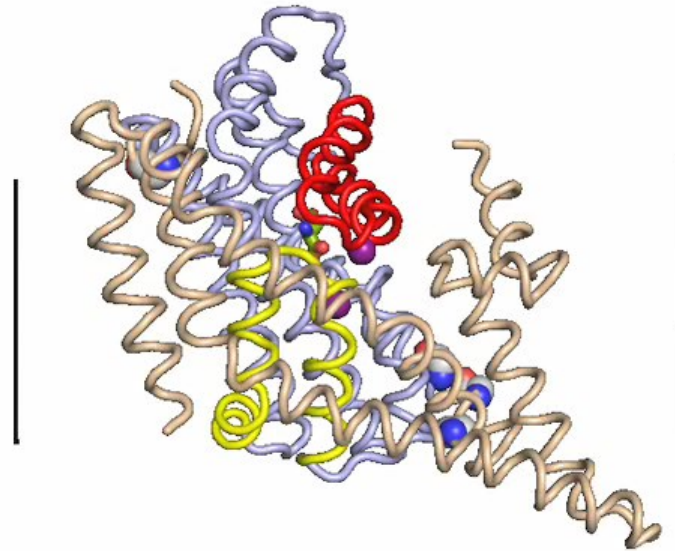
# Global transitions



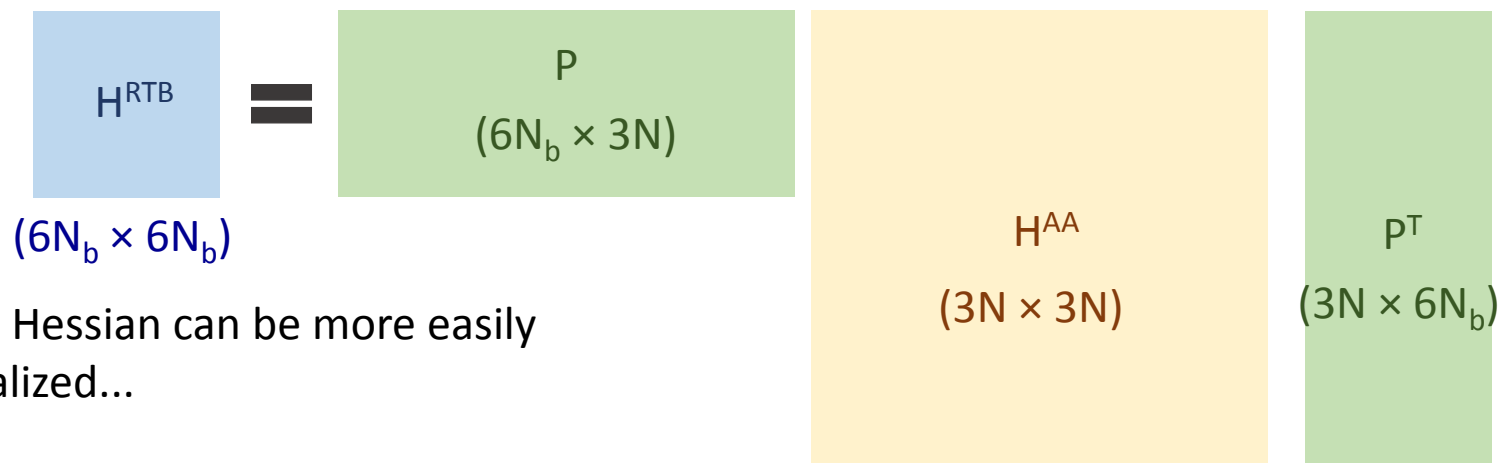


# Global transitions

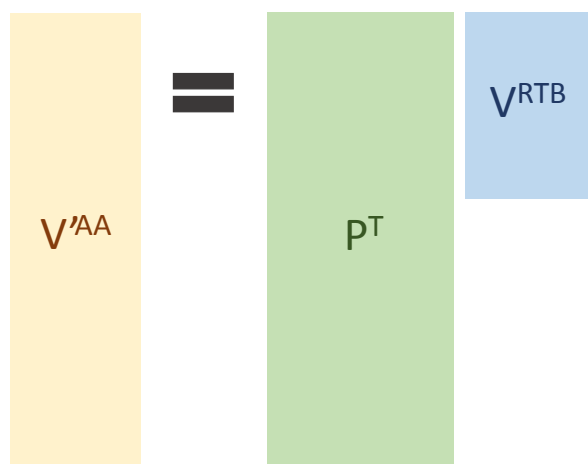
Single subunit showing the  
transport domain moving  
across the membrane



# Rotations-Translations of Blocks



Smaller Hessian can be more easily diagonalized...



- 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).

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

# Exploring structural transitions: Glutamate transporter

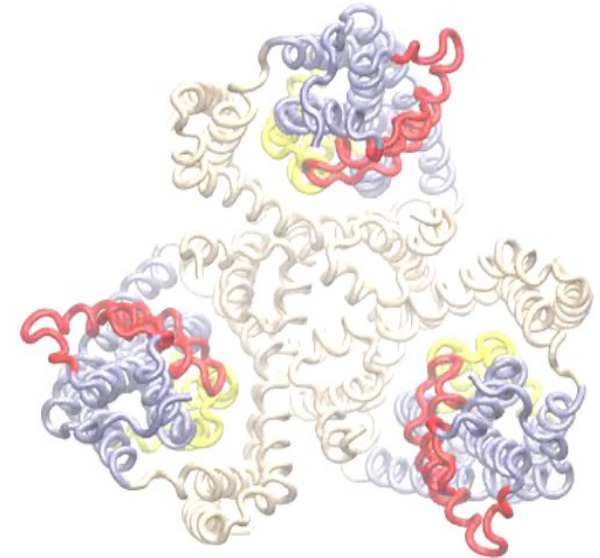
ANM predicts large radial motions of the trimer. Can we invent a better model?

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

Altered radial force constants:

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

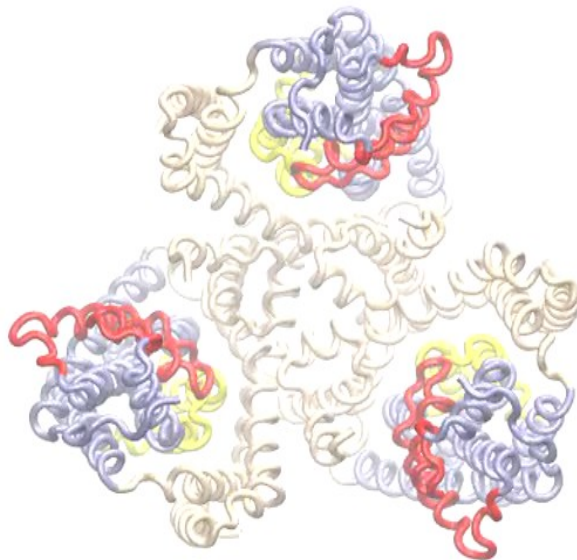
$$\mathbf{H}_{ij} = -\frac{\gamma}{(R_{ij}^0)^2} \begin{bmatrix} (x_{ij}^0)^2 & x_{ij}^0 y_{ij}^0 & cx_{ij}^0 z_{ij}^0 \\ x_{ij}^0 y_{ij}^0 & (y_{ij}^0)^2 & cy_{ij}^0 z_{ij}^0 \\ cx_{ij}^0 z_{ij}^0 & cy_{ij}^0 z_{ij}^0 & (cz_{ij}^0)^2 \end{bmatrix}$$



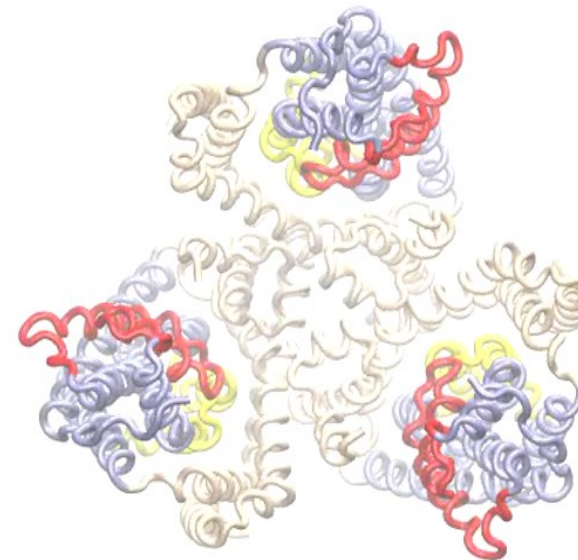
RTB.buildHessian()



## Exploring structural transitions: Glutamate transporter

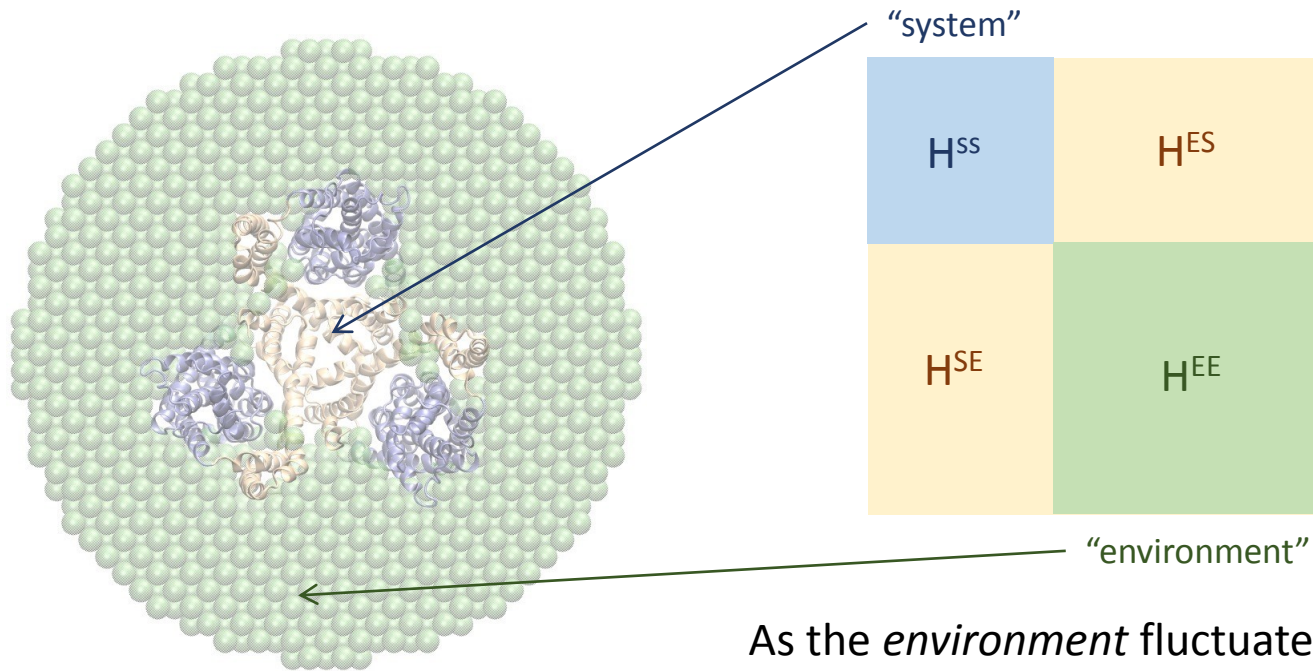


ANM: Large radial motions

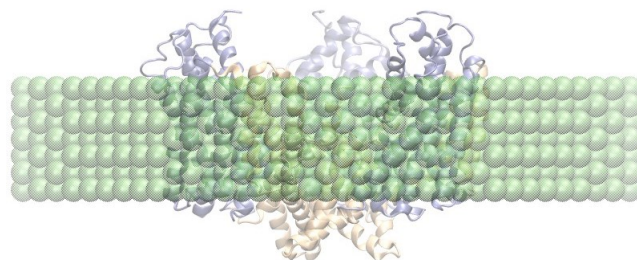


imANM

# Explicit membrane models



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



```

[Redacted]
[Redacted]
[Redacted]
reduceModel()
    
```