THE UNIVERSITY of TEXAS
HEALTH SCIENCE CENTER AT HOUSTON
SCHOOL of HEALTH INFORMATION SCIENCES

Normal Mode and Principal Component Analysis

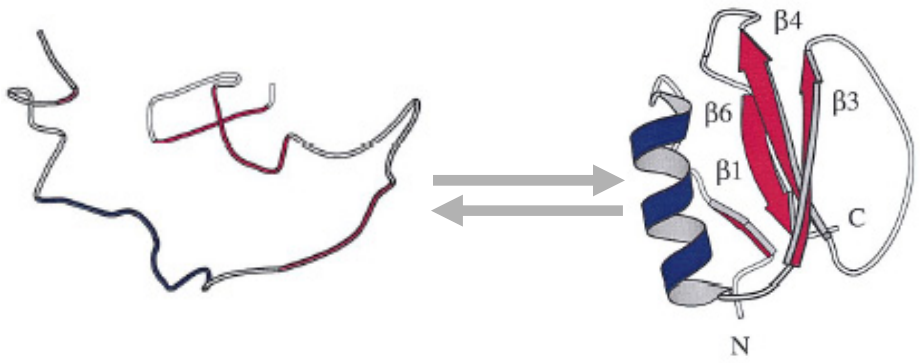
For students of HI 6327 “Biomolecular Modeling”

Willy Wriggers, Ph.D.
School of Health Information Sciences

http://biomachina.org/courses/modeling/07.html

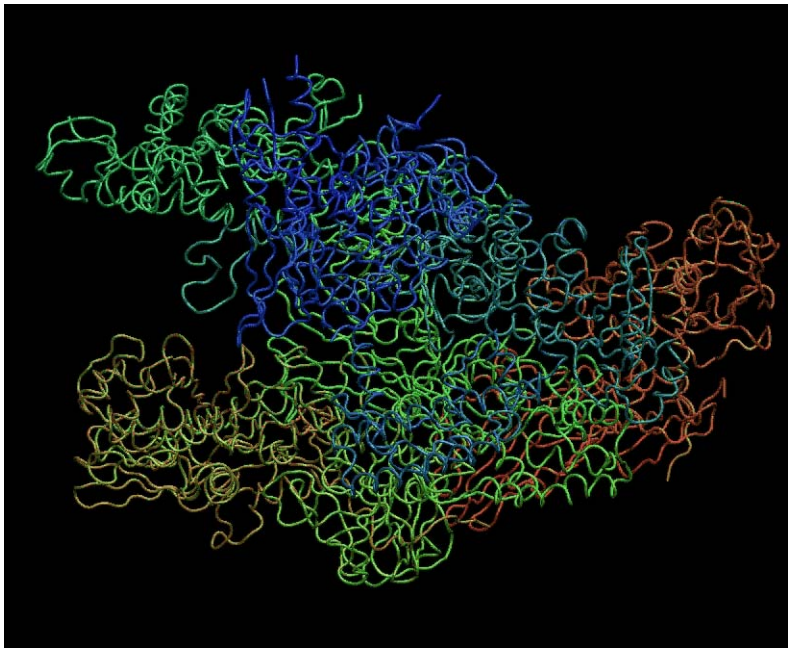
Protein Dynamics is Hierarchical

Vibration of bonds: 10^{-15} s



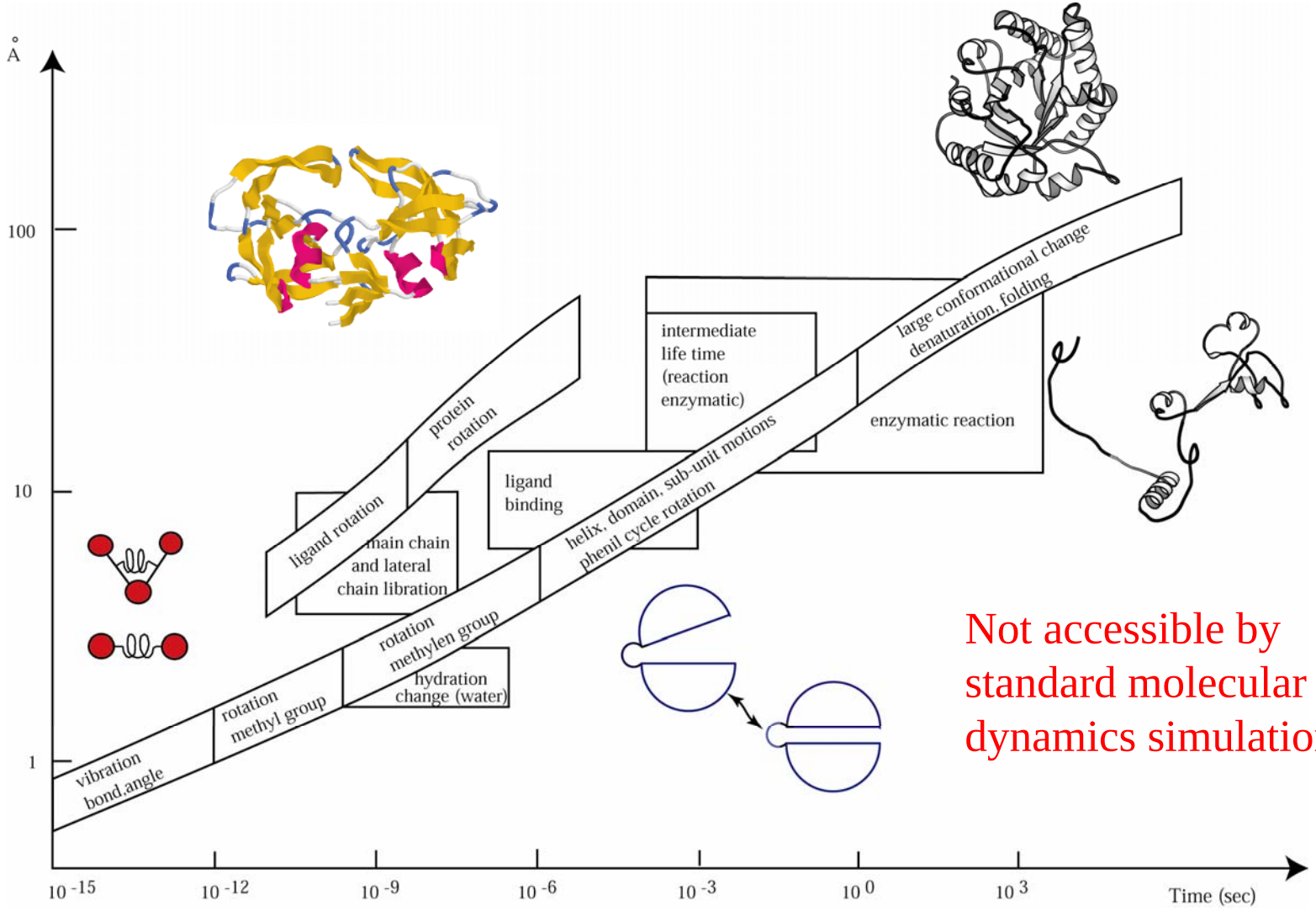
Protein folding/unfolding

10^{-6} s, 10^{-3} s, s and even longer



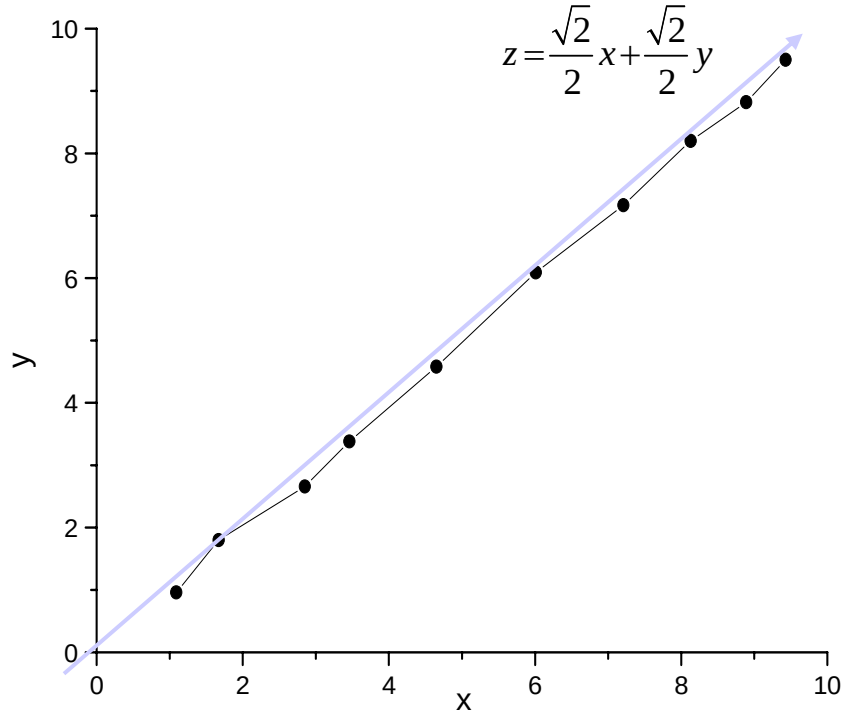
Large-scale functional motions

Dynamics of Biological Systems



Not accessible by standard molecular dynamics simulation

Collective Coordinates and Dimensionality Reduction



Collective Coordinates

- **Diagonalize Hessian matrix**

$$C = U \Lambda U^T$$

- **Principal Component Analysis from MD**

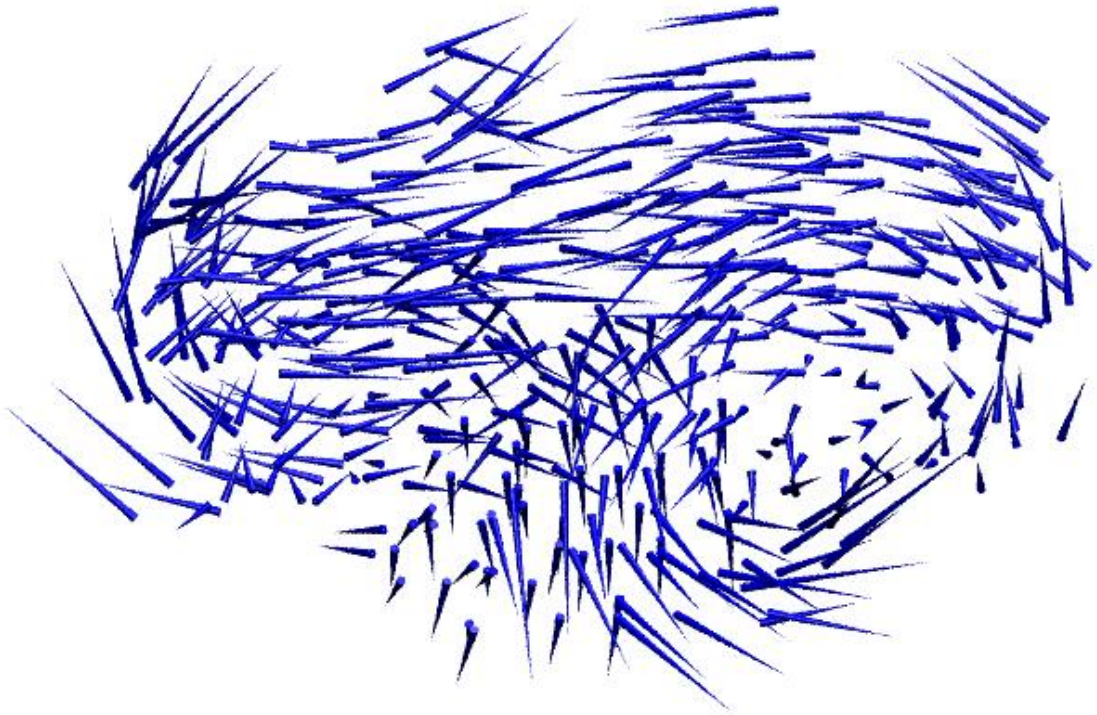
$$C_{ij} = \left\langle \left(x_i - \left\langle x_i \right\rangle \right) \left(x_j - \left\langle x_j \right\rangle \right) \right\rangle$$

- **Normal Mode Analysis**

$$C_{ij} = \partial^2 V \Big/ \partial x_i \, \partial x_j$$

Functional motions of a protein may be represented by only a few low-frequency modes.

Principal Component Analysis



Background

A mathematical technique, used to find patterns in high-dimensional datasets, such as protein structures.

Allows us to find relationships/patterns, which would be invisible from a pure visual examination.

Can be applied to MD simulation trajectories to detect the global, correlated motions of the system (the principal components).

Why are the PCs important?

Amadei *et al.* argue that we can separate the configurational space into 2 sub-spaces:

1. The Essential subspace: correlated motions comprising only a few of the degrees of freedom available to the protein = *FUNCTIONALLY IMPORTANT*

2. The "Irrelevant" subspace: independent, Gaussian fluctuations, which are constrained and of no/little functional relevance – act locally

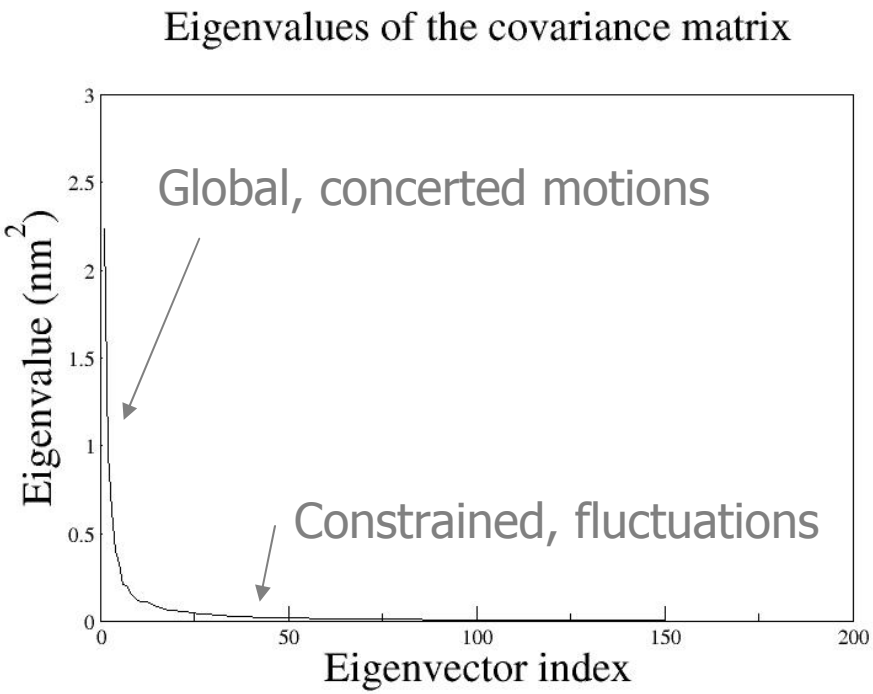
Overview

MD trajectory

Construct & diagonalize covariance matrix

Eigenvectors, ranked by eigenvalue

Analyse eigenvectors:
eg. Visualisation of motions
eg. Calculate subspace overlap



Eigenvalues and Eigenvectors

Matrix algebra

Online introduction, e.g.

http://www.sosmath.com/matrix/matrix.html

Theory (1)

Example: a 500 frame trajectory of a 300 residue protein.

BUILDING THE COVARIANCE MATRIX FROM YOUR TRAJECTORY:

Populate the 900 x 900 matrix (x, y & z component of each Cα atom):

$$\mathrm{Cov}(A,B) = \frac{\sum_{i=1}^{n} (A_i - \overline{A}) * (B_i - \overline{B})}{n}$$

$\overline{A}$ = time-averaged position

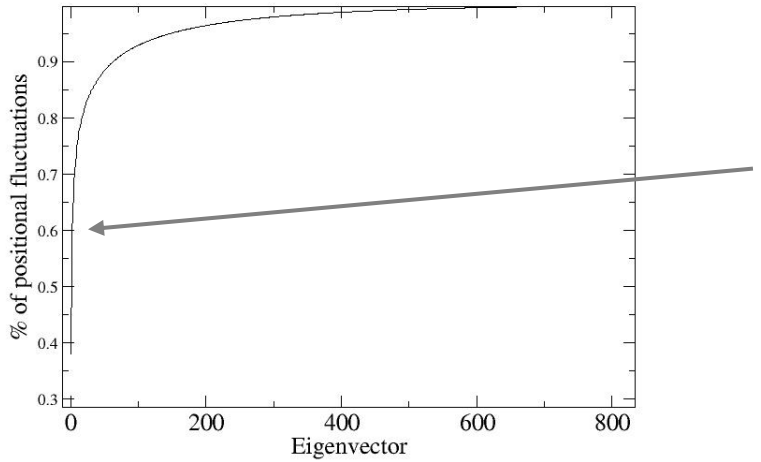
positive: both degrees of freedom move in same direction
negative: degrees of freedom move in opposite directions
zero: degrees of freedom are independent of each other

Theory (2)

The covariance matrix is then diagonalized – the columns of the transformation matrix become the eigenvectors, each associated with an eigenvalue.

Eigenvectors are then sorted by eigenvalue – the highest eigenvalues represent the most significant relationship between the dimensions: these are the principal components.

Eigenvectors represent a correlated displacement of groups of atoms through space
Eigenvalues represent the magnitude of this displacement (nm^2)

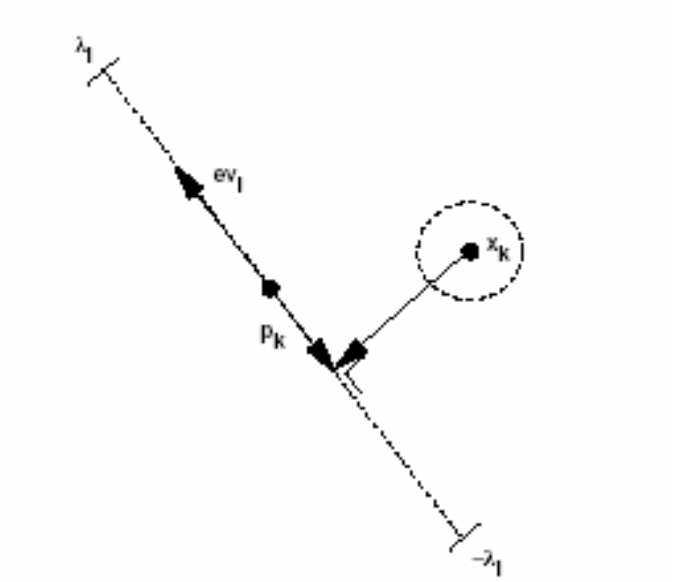


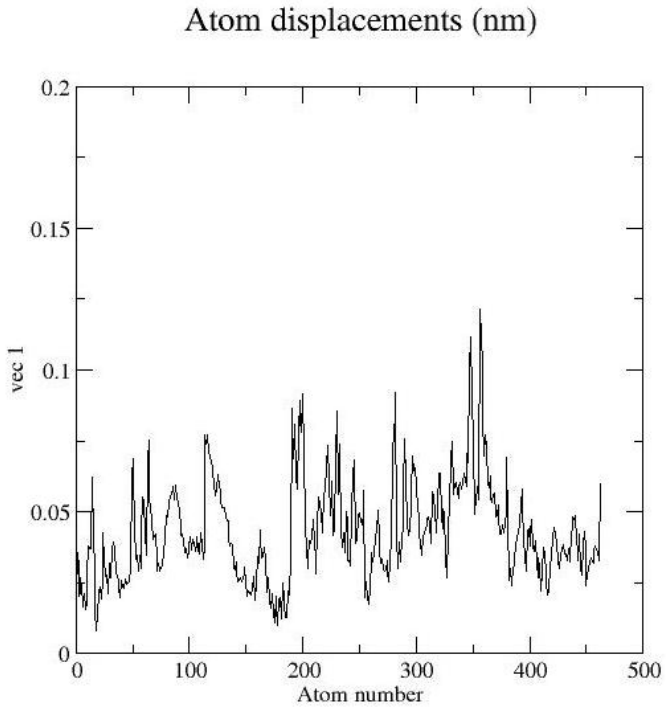
First 2 eigenvectors account for 60% of total positional fluctuations

Visualizing PCs (1)

The motion described by an eigenvector can be visualized by projecting the trajectory onto the eigenvector and taking the 2 extreme projections and interpolating between them to create an animation.

Projection of atom from a trajectory onto eigenvector

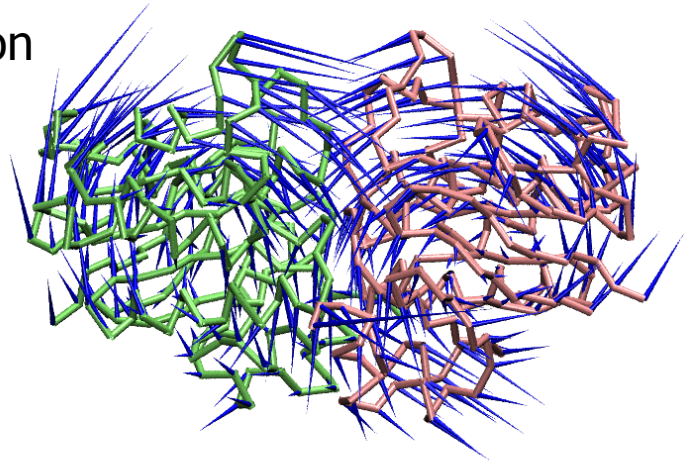




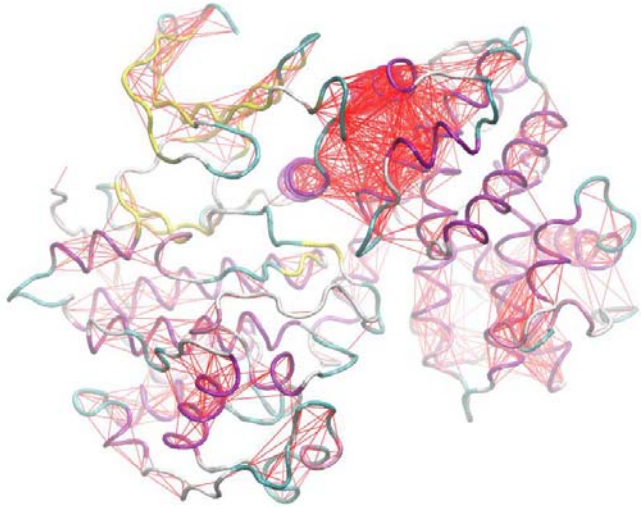
Visualizing PCs (2)

Porcupine plots can be used to display the motion described by an eigenvector in a static image.

A cone extending from the C-alpha position shows the direction of the atom along the eigenvector.



Covariance plots are a tool to visualize atoms which have a high correlation coefficient from the covariance matrix



Validation

How relevant are the PCs we have calculated and visualised?

Divide simulation into 2 or more parts and compare the eigenvectors for each part, to measure subspace overlap:

Higher overlap indicates sampling of only a single energy minimum
Lower overlap indicates more complete sampling

Can also measure cosine content of eigenvectors.

Hess *et al.* showed that the first few PCs of high-dimensional random diffusion are cosines and that several protein simulation PCs resemble these cosines (see Suggested Reading).

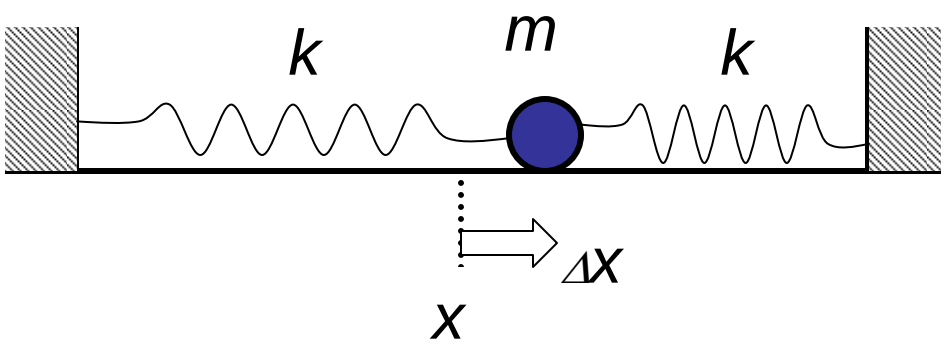
So high cosine content may mean that the fluctuations in your simulation are due to random diffusion:

Typically seen when simulation timescales are too short to reach energy barriers

Normal Mode Analysis

- An alternative method to study dynamics of molecules.
- Does not require trajectory, works with single structure.
- It is based on the theory of vibration.
- Conformational fluctuation is given by a superposition of normal modes.

Harmonic Oscillator



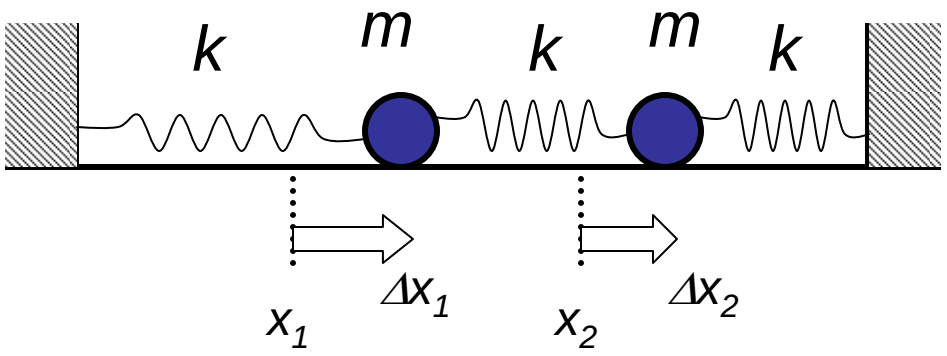
$$m\frac{d^2\Delta x}{dt^2} = -k\Delta x - k\Delta x$$

Newton / Hooke

$$\Leftrightarrow \quad m\frac{d^2\Delta x}{dt^2} + 2k\Delta x = 0$$

$$\begin{cases} \Delta x = A\cos(\omega t + \delta) \\ \omega^2 = 2k/m \end{cases}$$

Coupled Oscillators



$$\begin{cases} m\dfrac{d^2\Delta x_1}{dt^2} = -k\Delta x_1 + k(\Delta x_2 - \Delta x_1) \\ m\dfrac{d^2\Delta x_2}{dt^2} = -k(\Delta x_2 - \Delta x_1) - k\Delta x_2 \end{cases} \Leftrightarrow m\frac{d^2}{dt^2}\begin{pmatrix}\Delta x_1 \\ \Delta x_2\end{pmatrix} + \begin{pmatrix} 2k & -k \\ -k & 2k\end{pmatrix}\begin{pmatrix}\Delta x_1 \\ \Delta x_2\end{pmatrix} = 0$$

Matrix Diagonalization

$$\Leftarrow \begin{pmatrix}\Delta x_1 \\ \Delta x_2\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1\end{pmatrix}\begin{pmatrix}\Delta u_1 \\ \Delta u_2\end{pmatrix}$$

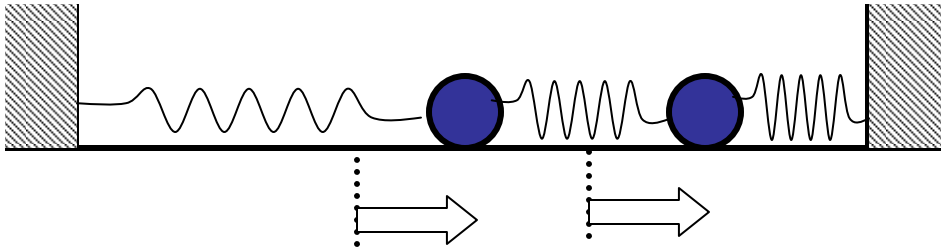
$$m\frac{d^2}{dt^2}\begin{pmatrix}\Delta u_1 \\ \Delta u_2\end{pmatrix} + \begin{pmatrix} k & 0 \\ 0 & 3k\end{pmatrix}\begin{pmatrix}\Delta u_1 \\ \Delta u_2\end{pmatrix} = 0 \Leftrightarrow \begin{cases} m\dfrac{d^2\Delta u_1}{dt^2} + k\Delta u_1 = 0 \\ m\dfrac{d^2\Delta u_2}{dt^2} + 3k\Delta u_2 = 0 \end{cases}$$

$$\begin{cases} \Delta u_1 = A_1\cos(\omega_1 t + \delta_1) \\ \Delta u_2 = A_2\cos(\omega_2 t + \delta_2) \end{cases} \qquad \begin{aligned} \omega_1^2 &= k/m \\ \omega_2^2 &= 3k/m \end{aligned}$$

Conformational Fluctuation

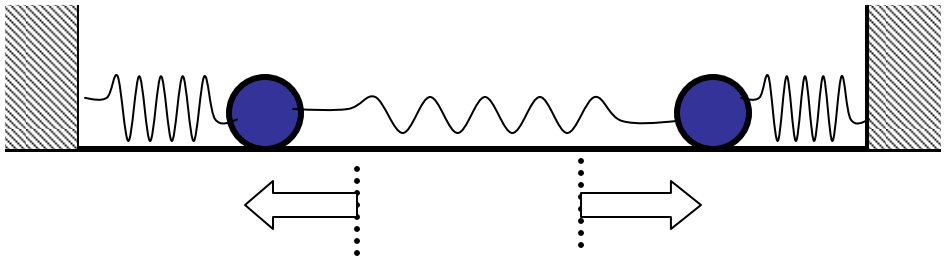
…is given by a superposition of normal modes:

$$\begin{pmatrix}\Delta x_1\\ \Delta x_2\end{pmatrix} = \frac{A_1}{\sqrt{2}}\begin{pmatrix}1\\ 1\end{pmatrix}\cos(\omega_1 t+\delta_1)+\frac{A_2}{\sqrt{2}}\begin{pmatrix}1\\ -1\end{pmatrix}\cos(\omega_2 t+\delta_2)$$



Lower frequency mode

$$\omega_1 = \sqrt{k/m}$$



Higher frequency mode

$$\omega_2 = \sqrt{3k/m}$$

Eigenvalue Problem

$$m\frac{d^2}{dt^2}\begin{pmatrix}\Delta x_1\\ \Delta x_2\end{pmatrix}+\underbrace{\begin{pmatrix}2k & -k\\ -k & 2k\end{pmatrix}}_{F}\begin{pmatrix}\Delta x_1\\ \Delta x_2\end{pmatrix}=0 \quad \Leftarrow \quad \begin{pmatrix}\Delta x_1\\ \Delta x_2\end{pmatrix}=\underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}}_{U}\begin{pmatrix}\Delta u_1\\ \Delta u_2\end{pmatrix}$$

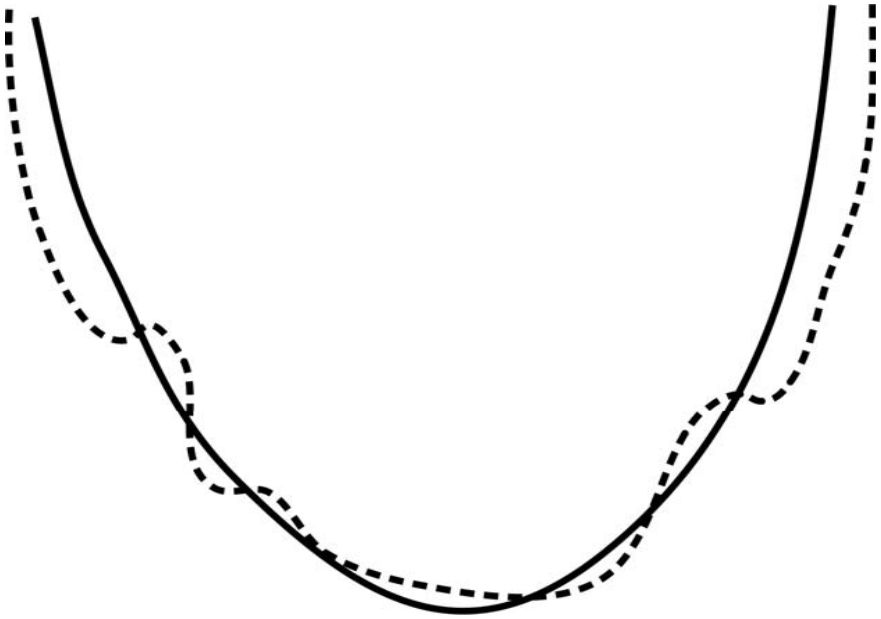
U is chosen so that it satisfies the following conditions.

$$U^t F U=\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}$$

$$U^t U=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

Matrix Diagonalization

Harmonic Approximation



Approximation:

Potential energy => harmonic

Harmonic Approximation

The potential energy must be written in the following form,

$$E = E_0 + \frac{1}{2}\sum_{i,j} F_{ij}\Delta x_i \Delta x_j$$

The form of potential energy for a molecule is usually very complicated. However, we can get such a form around a minimum-energy point, because the function can be written as follows:

$$E = E_0 + \sum \frac{\partial E}{\partial x_i}\Delta x_i + \frac{1}{2!}\sum \frac{\partial^2 E}{\partial x_i \partial x_j}\Delta x_i \Delta x_j + \frac{1}{3!}\sum \frac{\partial^3 E}{\partial x_i \partial x_j \partial x_k}\Delta x_i \Delta x_j \Delta x_k + \ldots\ldots\ldots\ldots$$

Taylor series

NMA using Molecular Mechanics

Full atomic representation and MM interactions require:

- energy minimization
- diagonalization of the 2nd derivative of the potential energy (Hessian)
- Hessian is 3N x 3N matrix (memory requirements!)

Two-Atomic Molecule

Spring constant

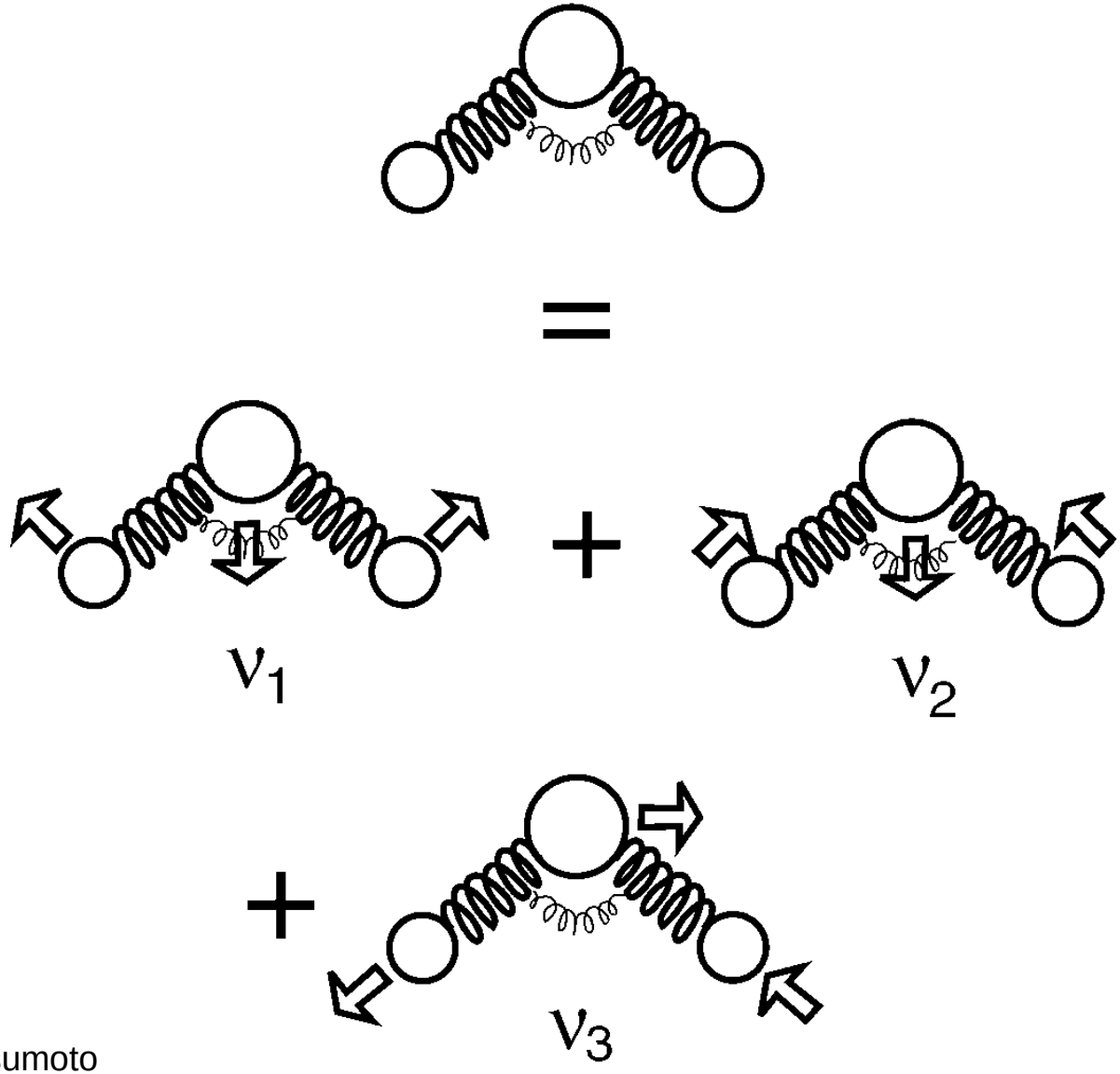
k

=

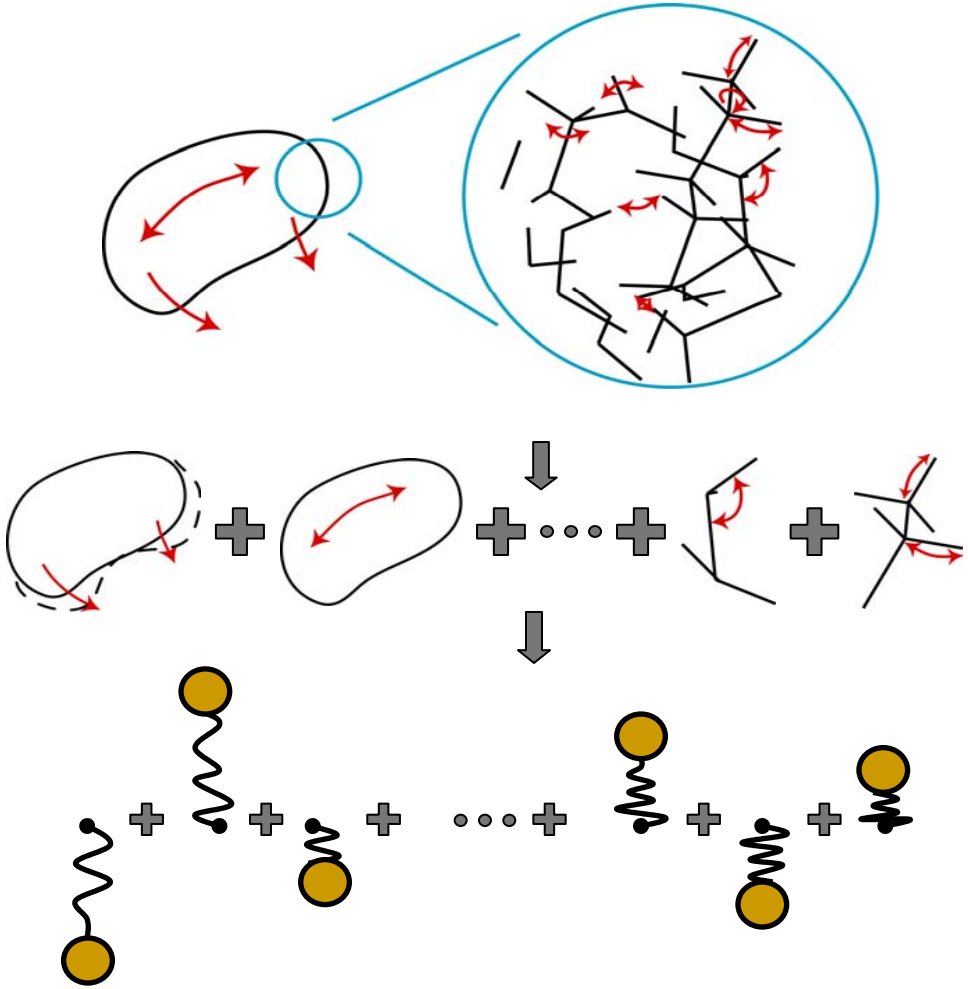
Frequency

$$\nu \propto \sqrt{k}$$

Three-Atomic Molecule



Multi-Atom Molecule



Low frequencies ⟷ High frequencies

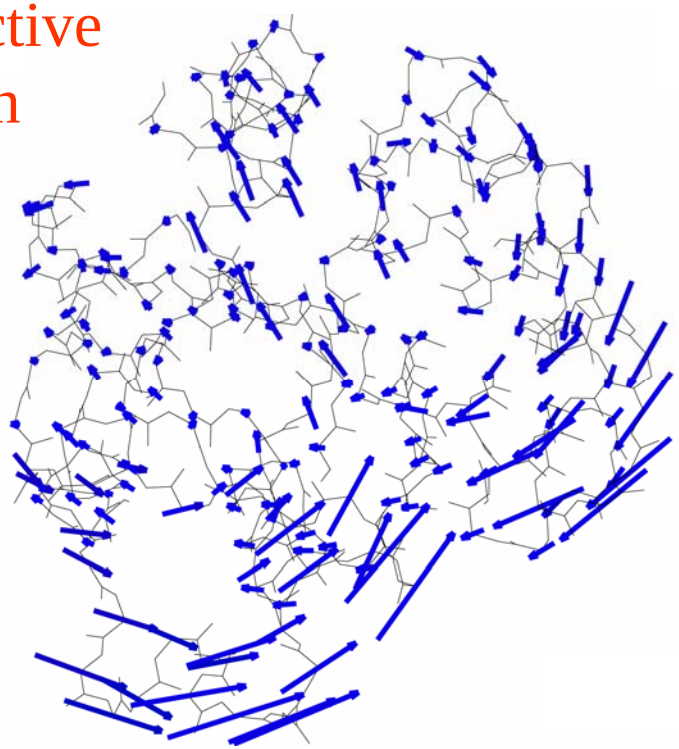
Large collective motions — Localized motions

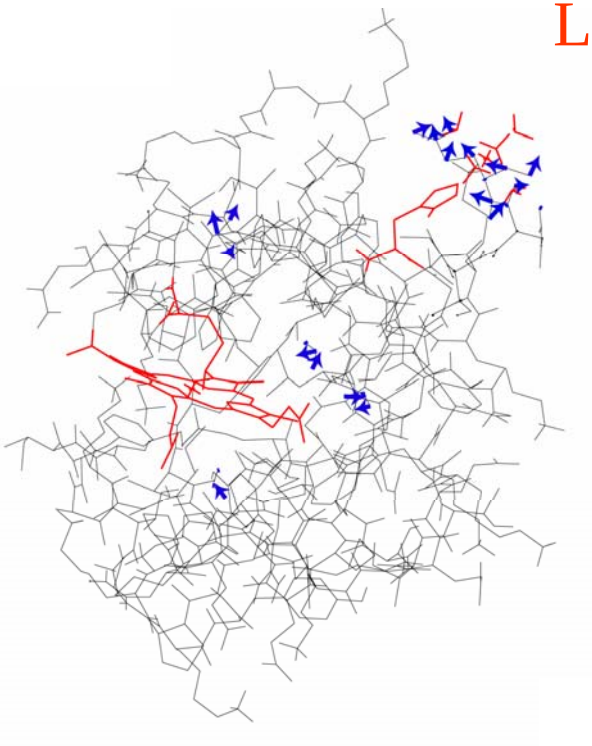
Example

Adenylate kinase:
Mode 1 (2.95 cm^{-1})

Cytochrome c:
Mode 2757 (1519 cm^{-1})

Collective motion



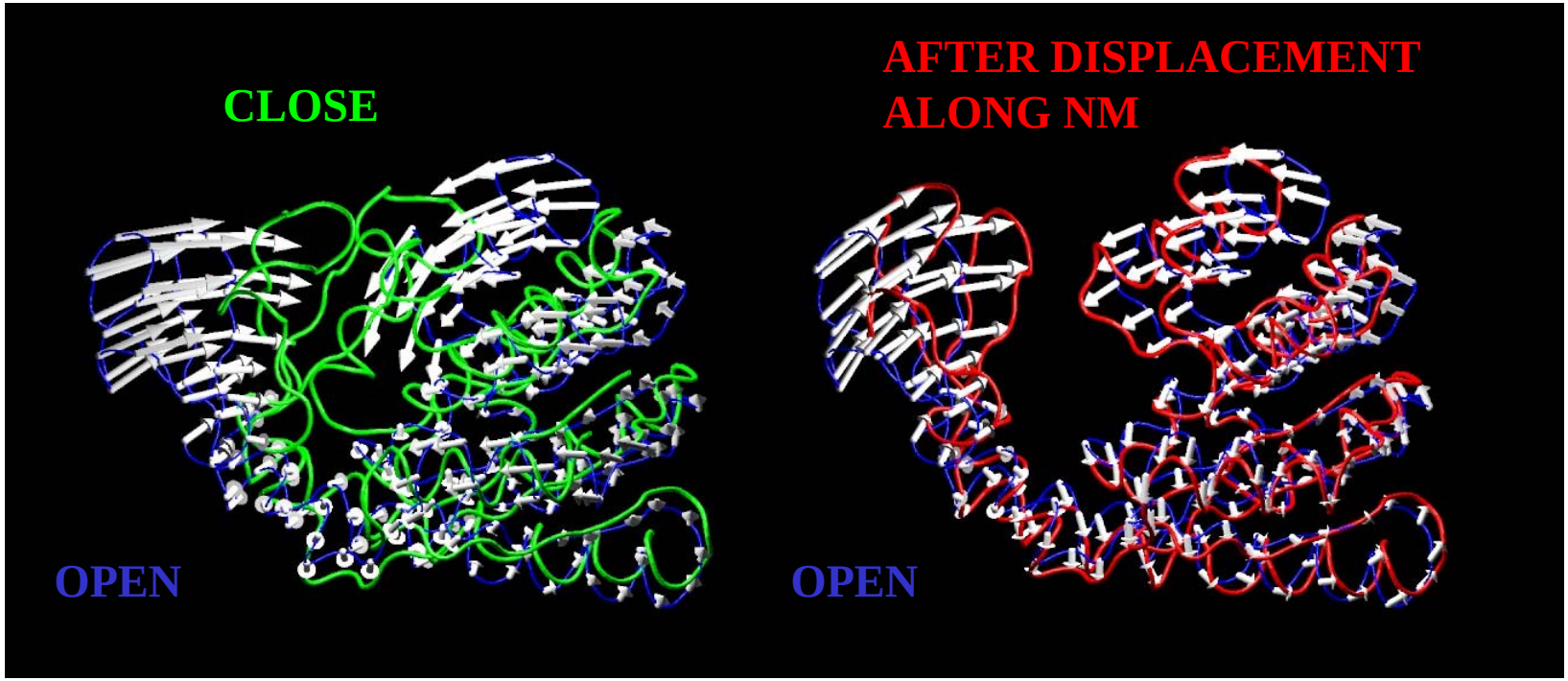


Localized motion

Success Story

Adenylate Kinase =>

large conformational change upon ligand binding



1 normal mode can represent up to 80-90 % of the overall conformational change

Computational Challenges

NMA requires:

- minimization
- diagonalization of the Hessian matrix

Problems for large systems:

- expensive, cumbersome (MM)
- memory requirements

Memory-Efficient Diagonalization

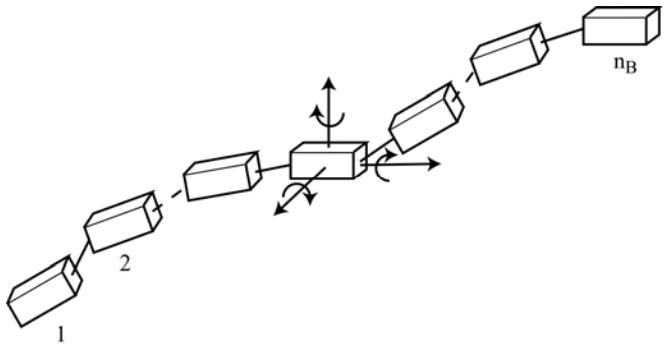
DIMB => Diagonalization in mixed basis

(Perahia & Mouawad, **1995**, *J. Comp. Chem*. **19**, 241)

Group theory => Use symmetrical properties of viruses

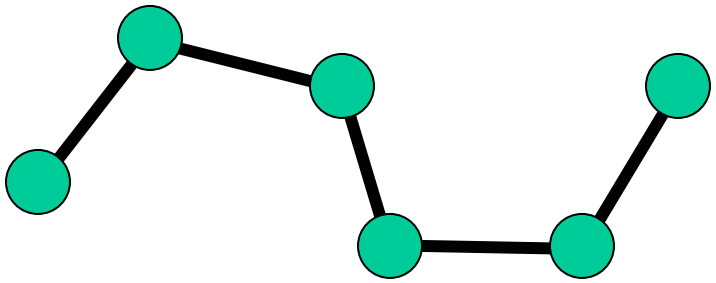
(Roux & Karplus, **1988**, *Biophys. J*, **53**, 297; Simonson & Perahia, **1992**, *Biophys. J*., **61**, 410; van Vlijmen & Karplus, **2001**, *J.Chem. Phys*, **115**, 691)

RTB => Rotation Translation Blocks method gives approximate low-frequency NM (Tama et al. **2000**, *Proteins: Struc. Funct. Genet*., **41**, 1)



- block = 1 or several residues
- rotation + translation of block => new basis
- expression of Hessian in this new basis
- diagonalization of a matrix $6n_B * 6n_B$

Reducing the Number of Variables



Cartesian coordinate space

3N-6 variables are necessary

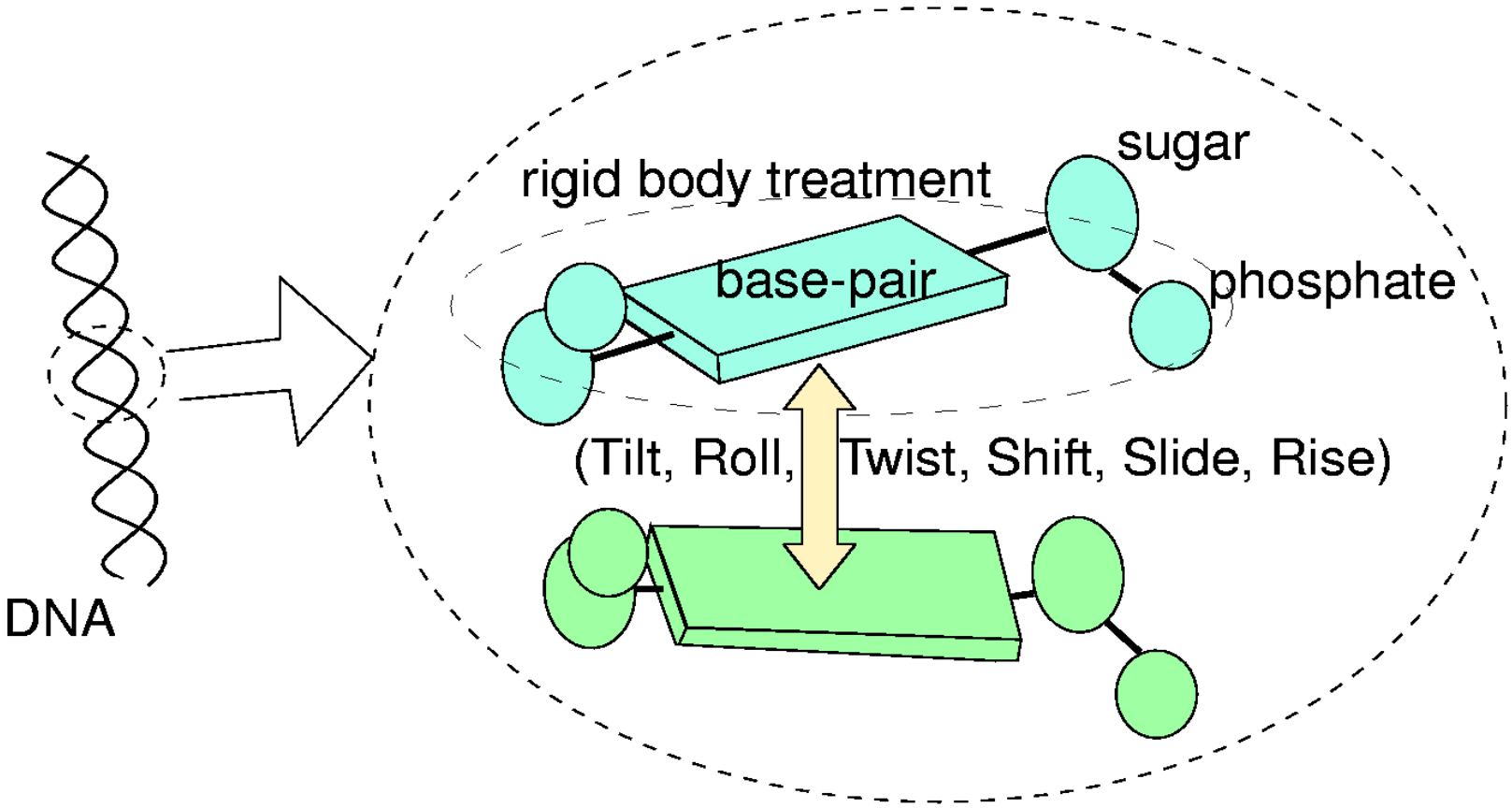
N : number of atoms

Torsion angle space

Bond angles and bond lengths are fixed, and only torsion angles are allowed to vary.

Number of variables: ~1/10

Reduced NMA of DNA



The relative position and orientation of adjacent base pairs are expressed by six "step" variables ---(Tilt, Roll, Twist, Shift, Slide, Rise). The entire structure of DNA with (n+1) base pairs is expressed by 6n variables.

Definition of Base-Pair Step Parameters

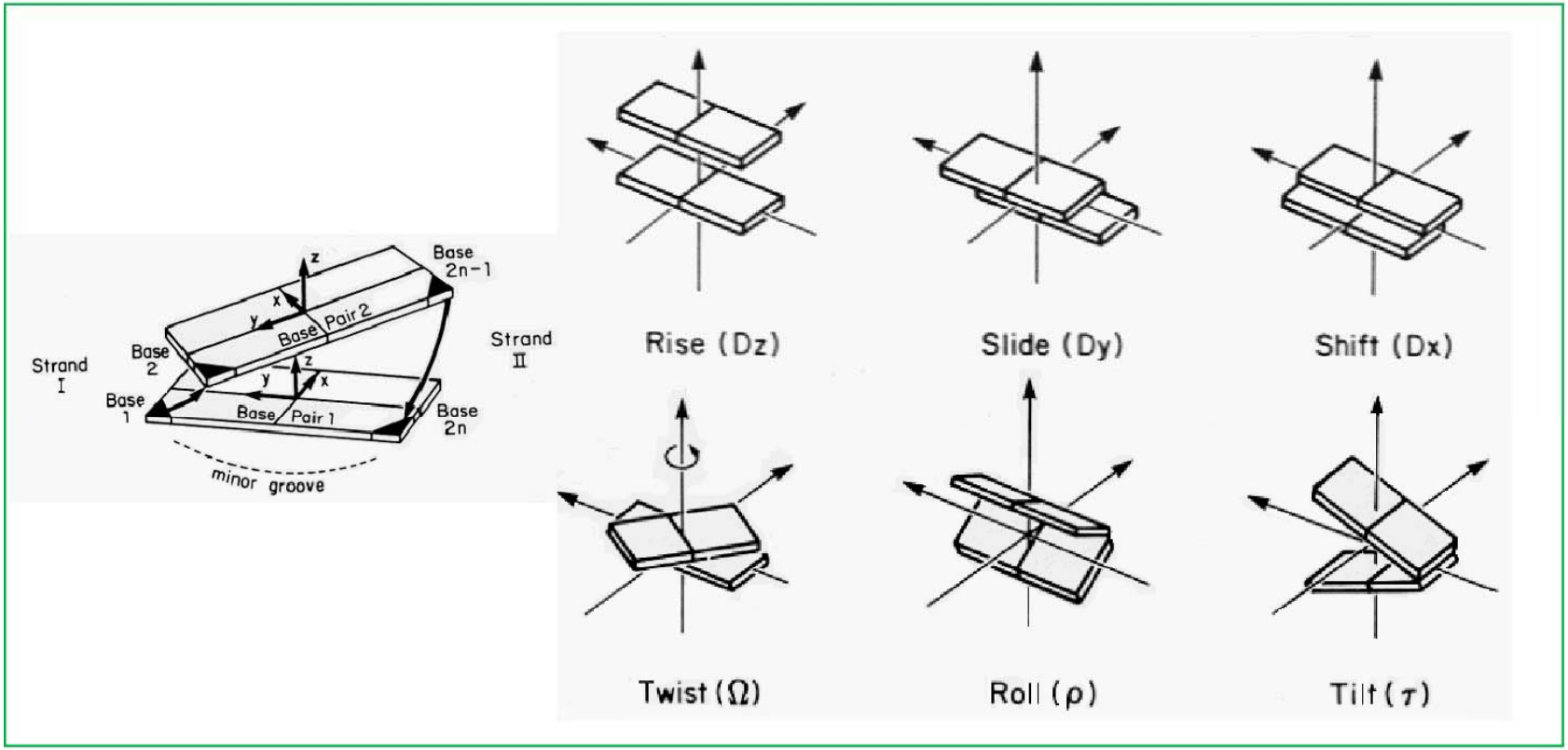
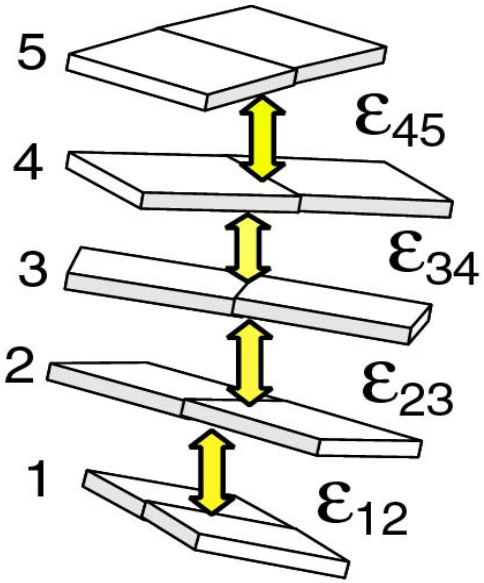


Image based on Cambridge convention of base pair parameters: Dickerson, R. E. *et al.* (1989) "Definitions and nomenclature of nucleic acid strutcutre parameters", *J. Mol. Biol.* **208**, 787-791



Energy Function

The entire conformational energy E of DNA is represented as a sum of dimer step energies.

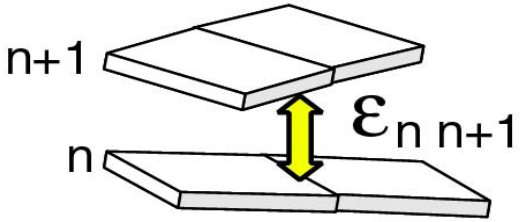
$$E = \varepsilon_{12}+\varepsilon_{23}+\varepsilon_{34}+\varepsilon_{45}$$

The dimer step energy is a function of the base-pair step parameters θ_i (=Tilt, Roll, Twist, Shift, Slide, Rise).

$$\varepsilon_{n\,n+1} = (1/2)\sum_{i}^{6}\sum_{j}^{6} f_{ij}\,(\theta_i-\theta_i^{\circ})(\theta_j-\theta_j^{\circ})$$

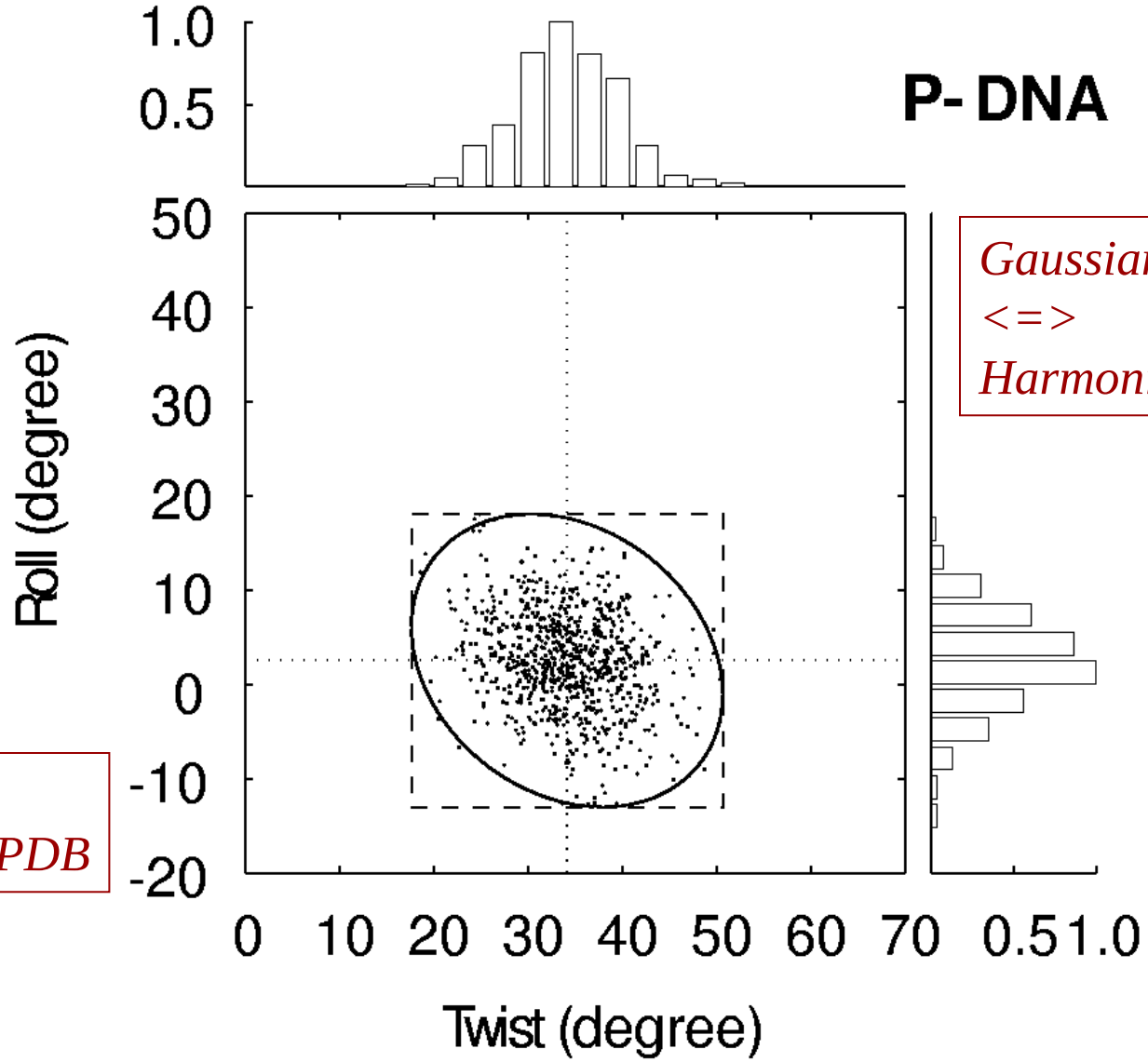
θ_i : Instantaneous value

θ_i° : Equilibrium value



f_{ij} and θ_i° are constants that depend on the type of dimer step, *e.g.*, the constants for the AA dimer are different from those for the AT step.
(Olson *et. al.*, *Proc. Natl. Acad. Sci. ,U.S.A.,* 1998, **95**, 11163-11168)

Parametrization



Gaussian Distribution
<=>
Harmonic Potential

Statistical
Mining of PDB

Results (DNA)

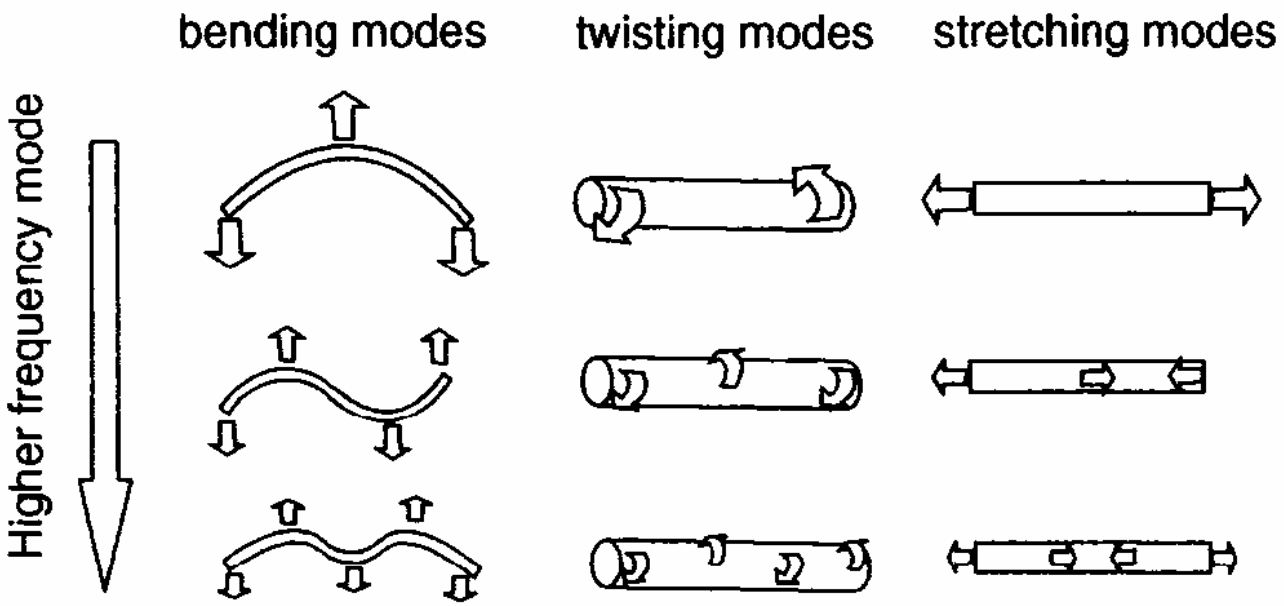


FIGURE 2 Schematic illustration of representative low frequency normal modes of an elastic rod. The arrows point in the directions of bending, twisting, and stretching motions in each mode.

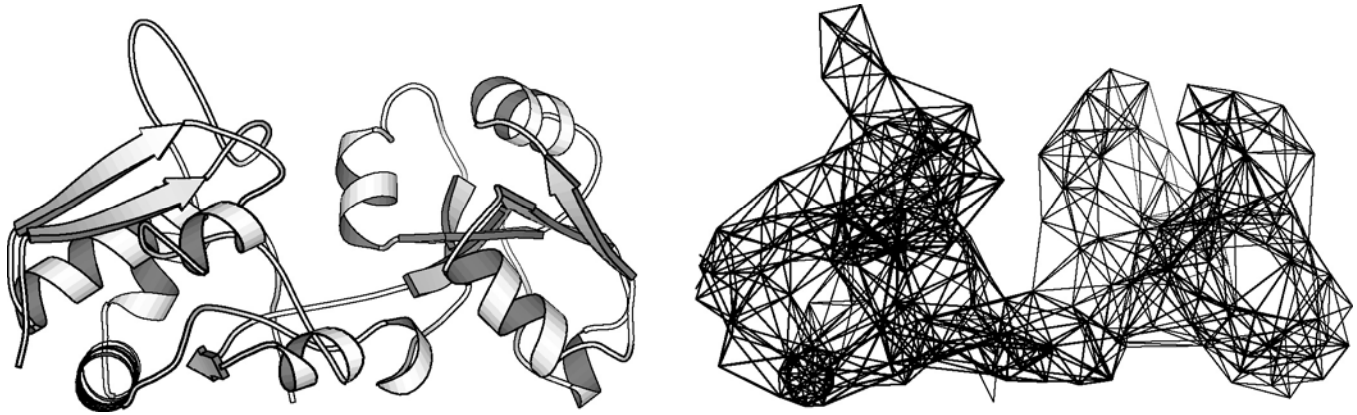
Atsushi Matsumoto and Wilma Olson, Biophys. J., 2002, 83:22-41.

Elastic Network Model

Monique M Tirion (1996) *Phys Rev Lett*. **77**, 1905-1908

Simplified force-field: no MM, already minimized

$$E(r_a, r_b) = \frac{C}{2}\left(|r_{a,b}| - |r^0_{a,b}|\right)^2 \qquad E_p = \sum_{a,\,b} E\left(r_a,\, r_b\right)$$



Possibility to reduce level of detail (up to 1 point for 40 residue)

Vector Quantization

Encode data (in $\Re^{d=3}$) using a finite set $\{w_j\}$ (j=1,…,k) of *codebook vectors*.

Delaunay triangulation divides $\Re^3$ into k *Voronoi polyhedra* ("receptive fields"):

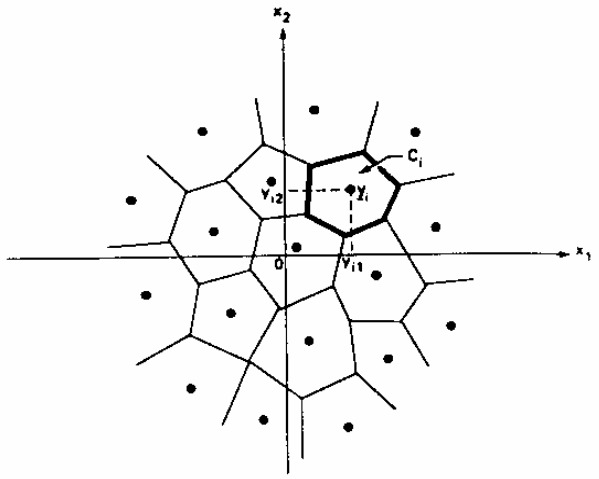
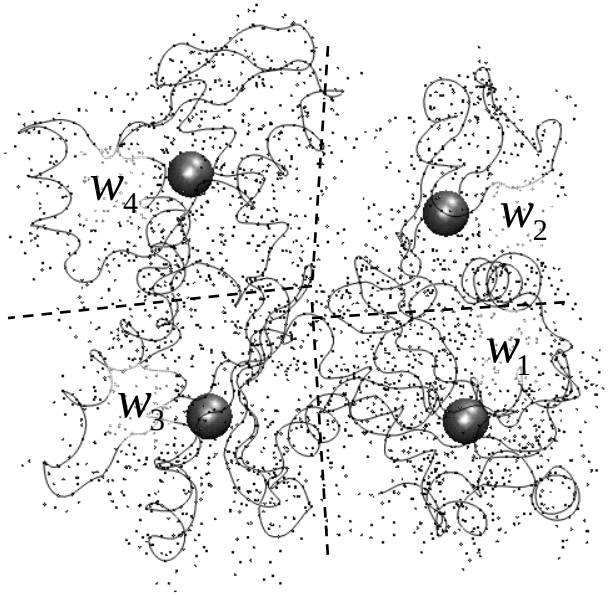


Fig. 3. Partitioning of two-dimensional space (N = 2) into L = 18 cells. All input vectors in cell C_i will be quantized as the code vector y_i. The shapes of the various cells can be very different.



Encoding Distortion Error: $$E = \sum_{i\ (\text{atoms, voxels})} \left\| v_i - w_{j(i)} \right\|^2 m_i$$

Linde, Buzo, & Gray (1980): Gradient descent finds nearest local minimum of E.

Martinetz & Schulten (1993): Global search with topology-representing neural nets.

Choice of Cut-off

Example: Adenylate kinase, 214 residues

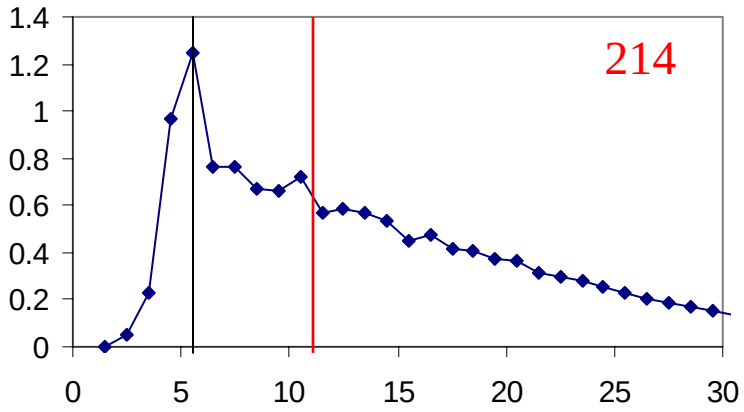
1 codebook vector ≈ 1 residue

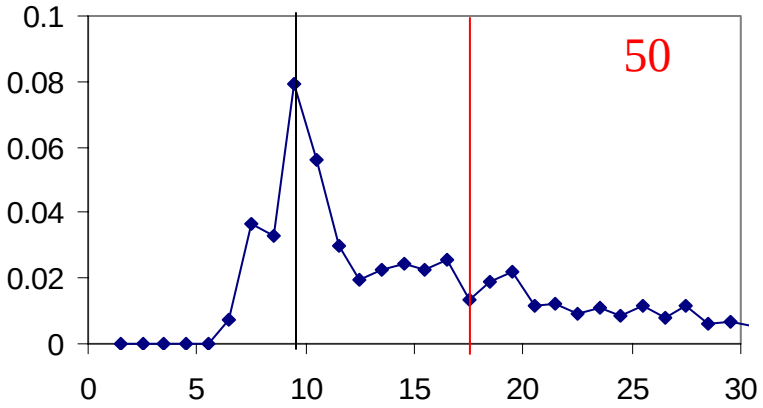
⇨10-12 Å cut-off OK

Reducing number of codebook vectors

⇨ too sparse connectivity

Inspect the pair-distance distribution of codebook vectors and increase cutoff beyond first peak.





Level of Detail not Important

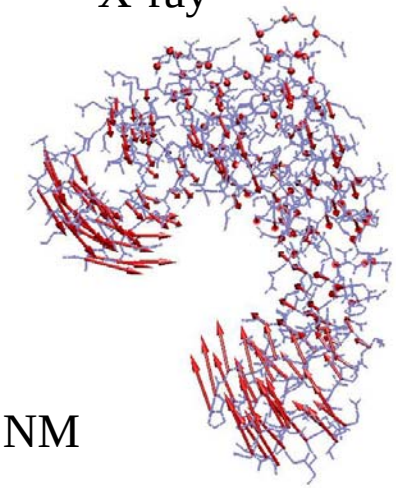
Projection onto atomic normal modes
≈ 1 for the first few modes

Low frequency NM are similar to atomic NM

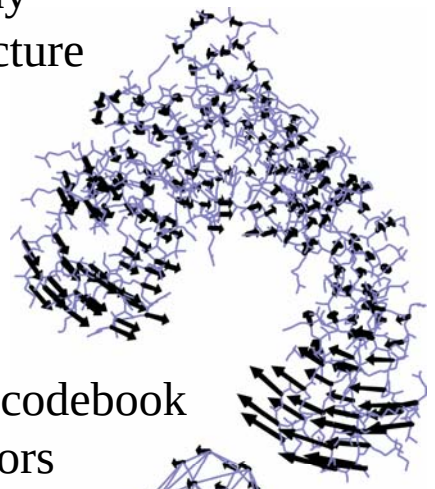
Models can reproduce functional rearrangements even at 30Å resolution

Open-Close

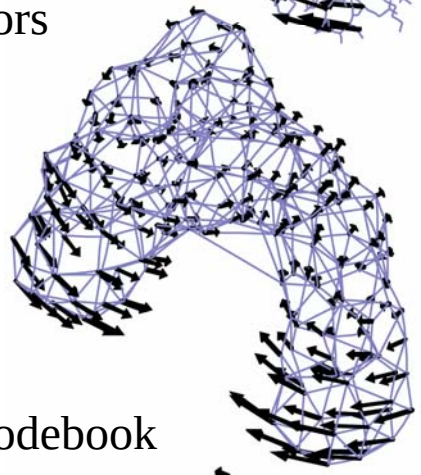
X-ray



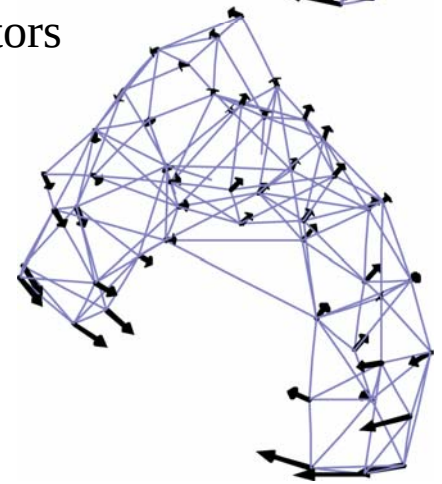
X-ray structure



214 codebook vectors

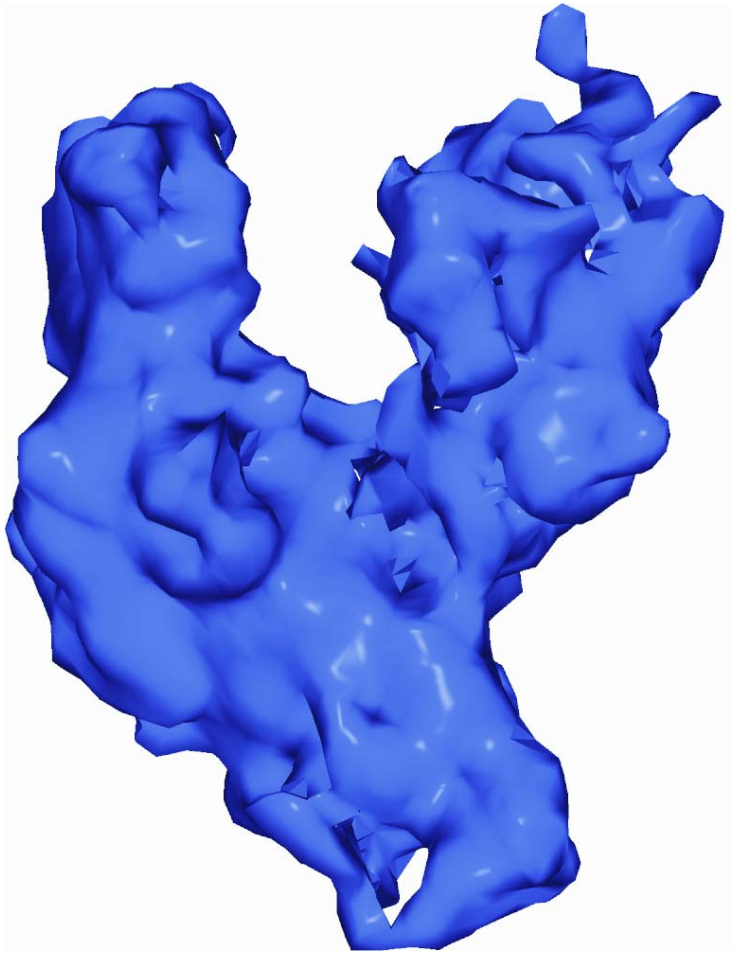


50 codebook vectors



Application to EM Data

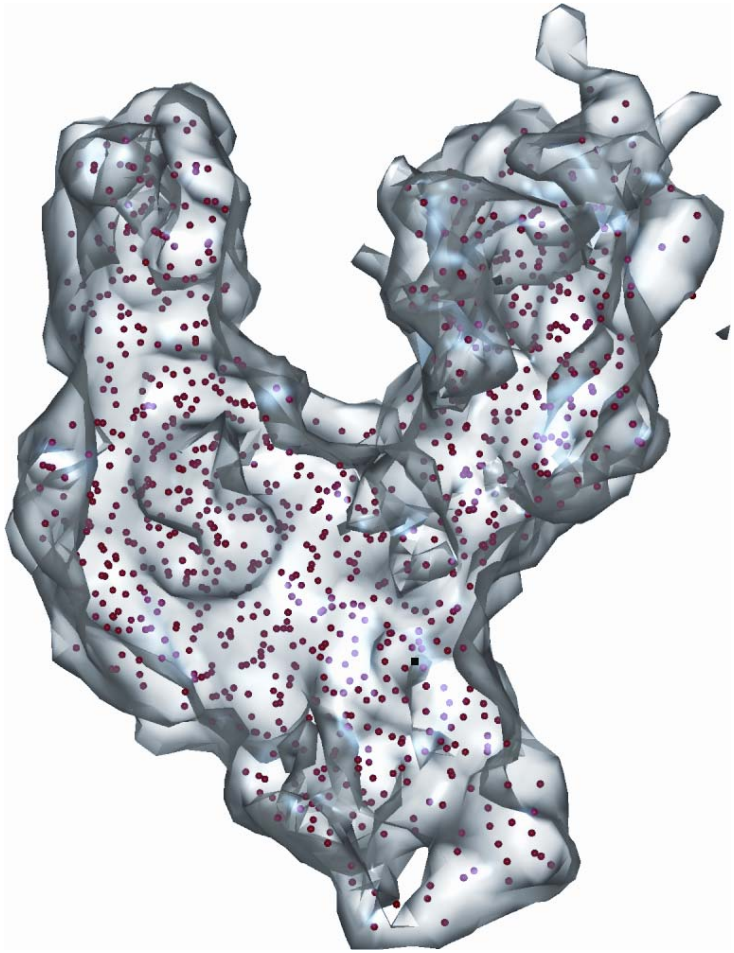
RNA Polymerase, S. Darst et al.

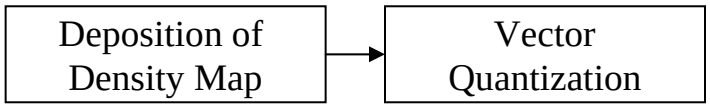


Deposition of Density Map →

Application to EM Data

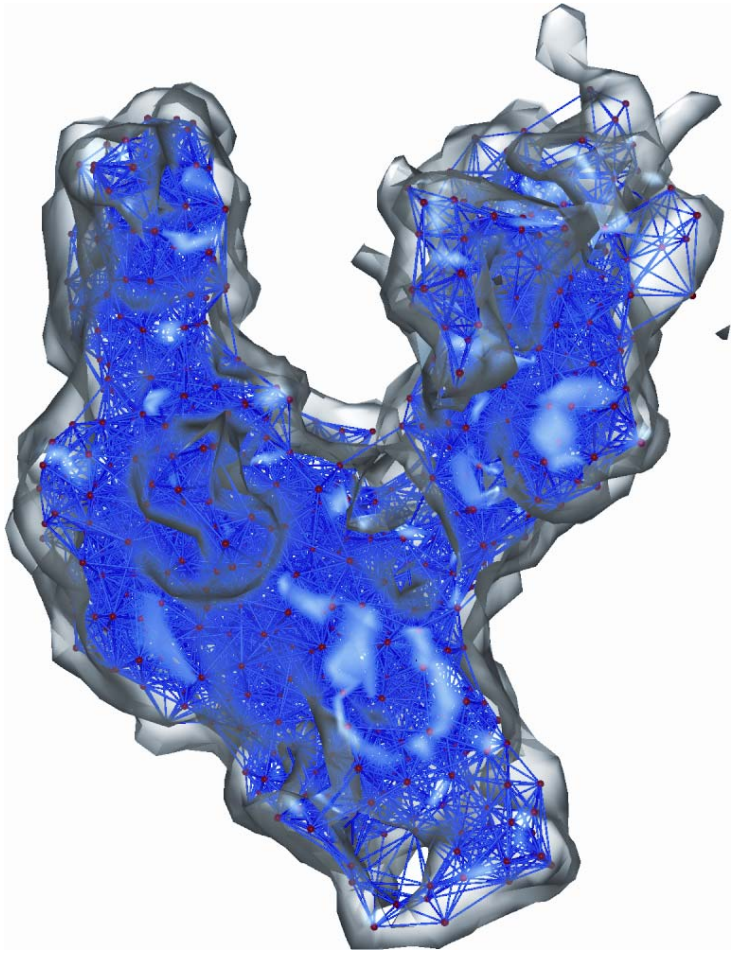
RNA Polymerase, S. Darst et al.

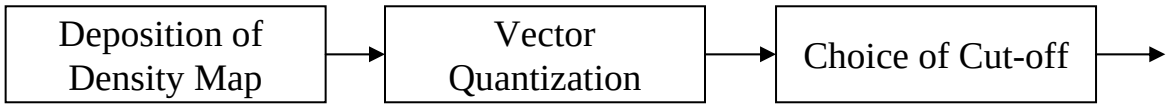




Application to EM Data

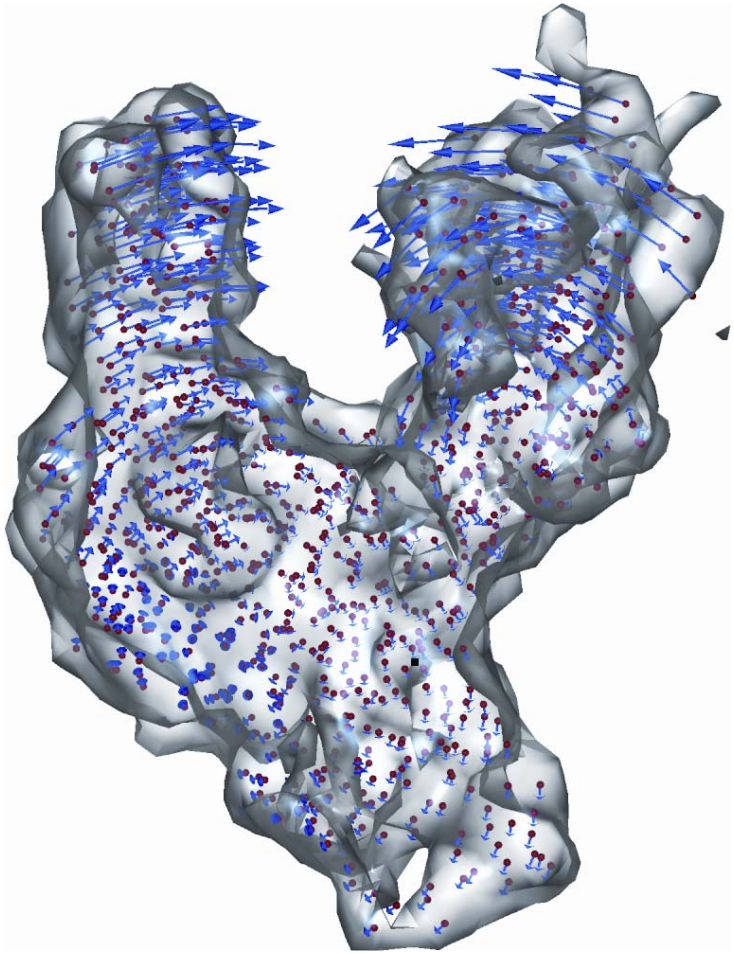
RNA Polymerase, S. Darst et al.

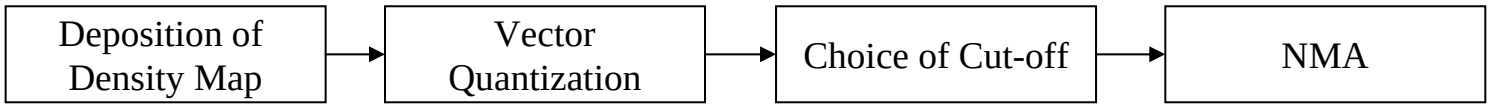




Application to EM Data

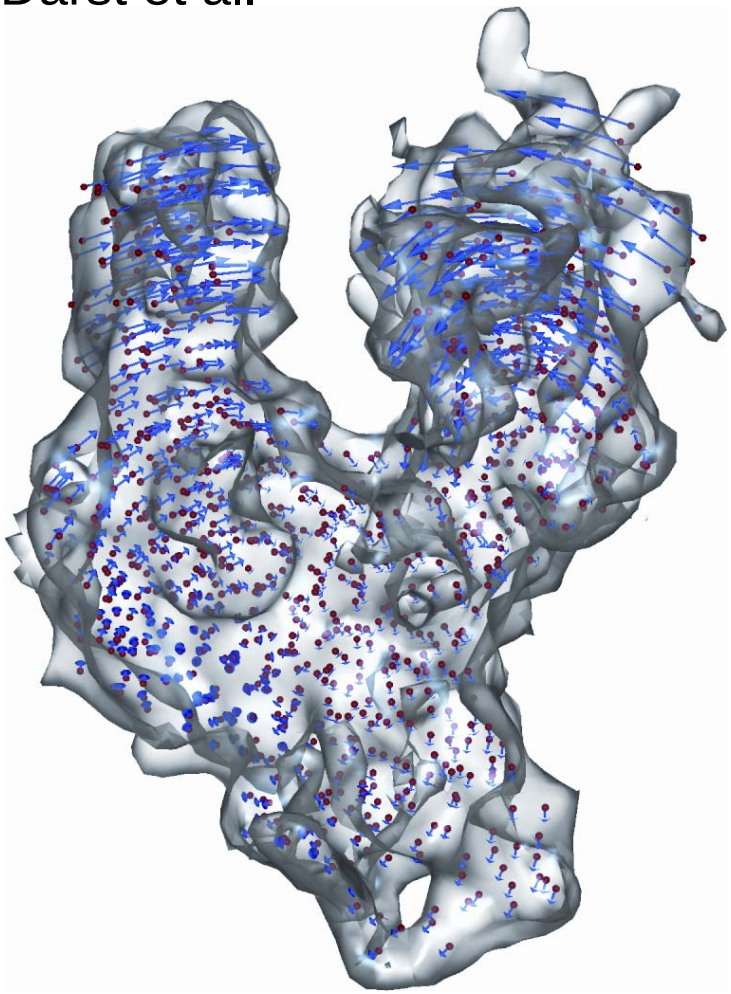
RNA Polymerase, S. Darst et al.





Application to EM Data

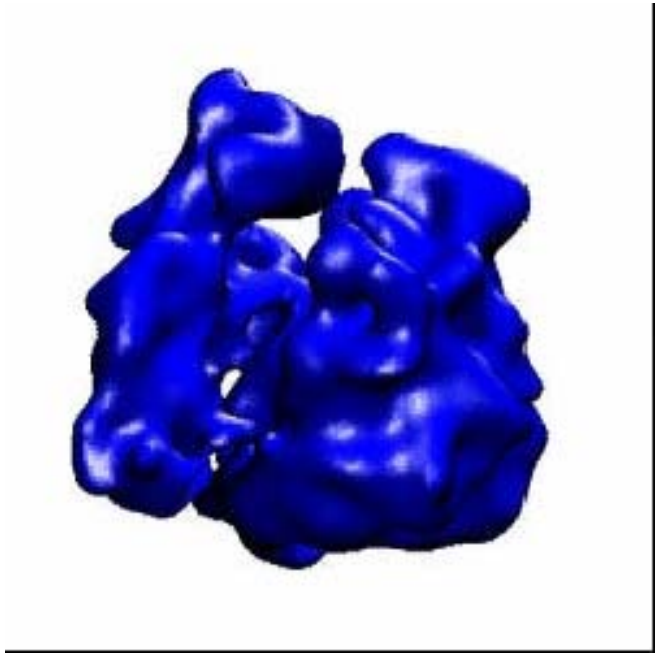
RNA Polymerase, S. Darst et al.



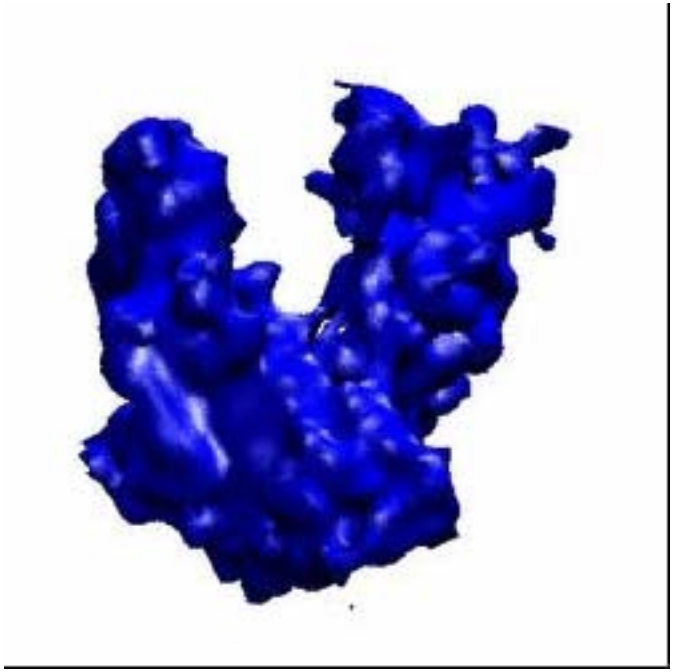
Deposition of Density Map → Vector Quantization → Choice of Cut-off → NMA → Apply Displacements

Examples

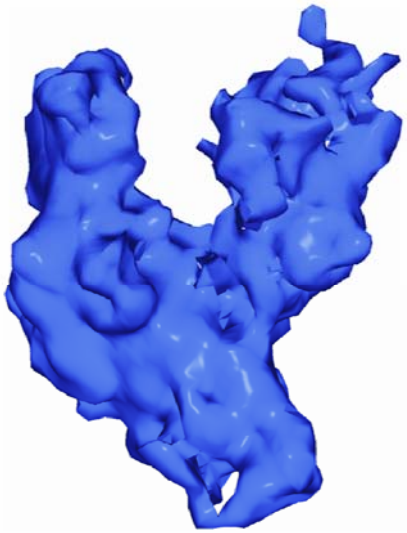
Ribosome

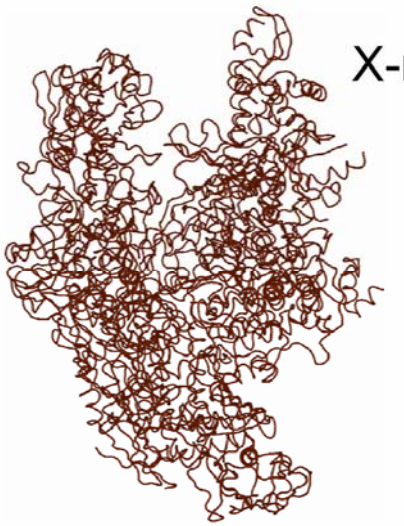


RNA Polymerase

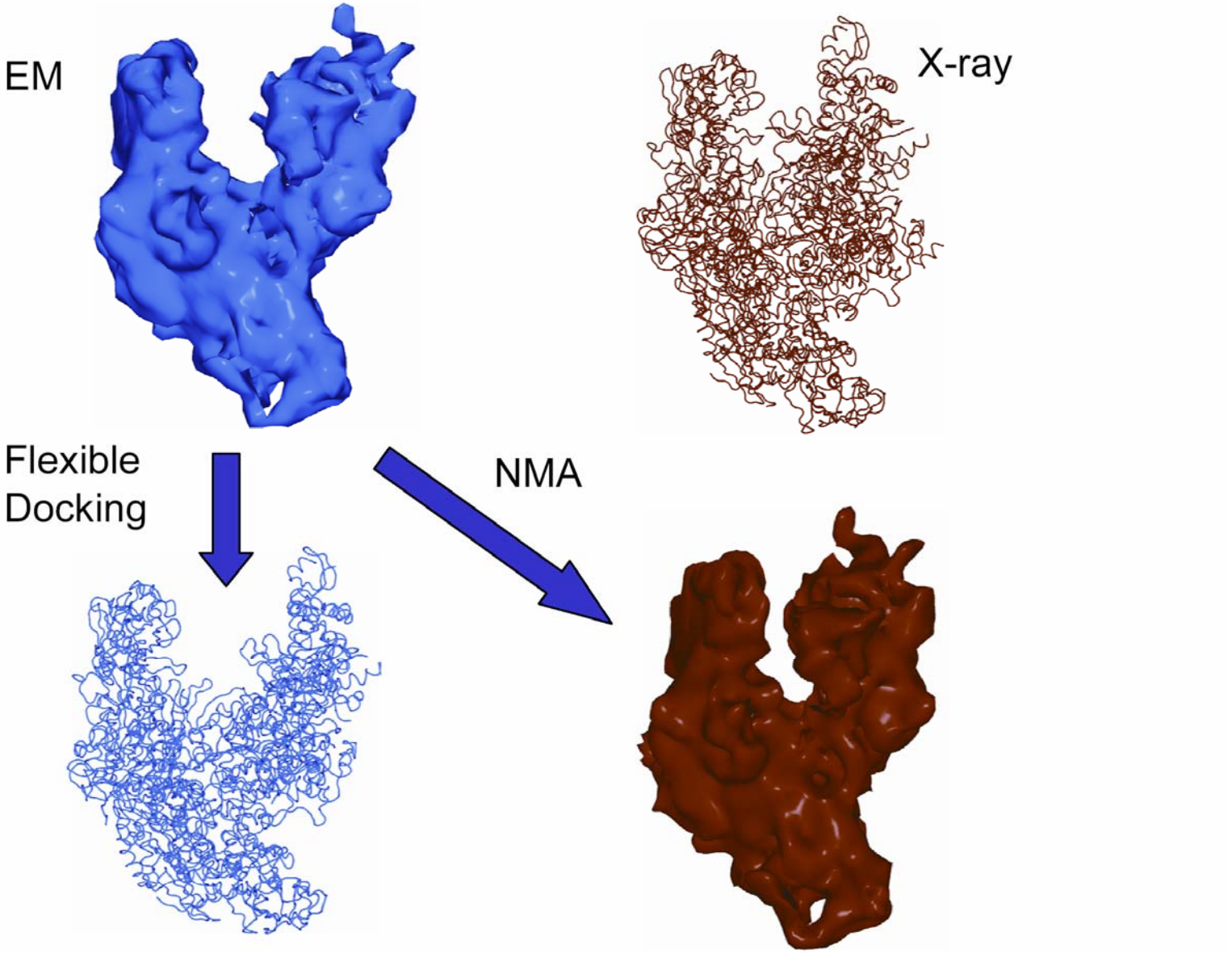


EM





X-ray

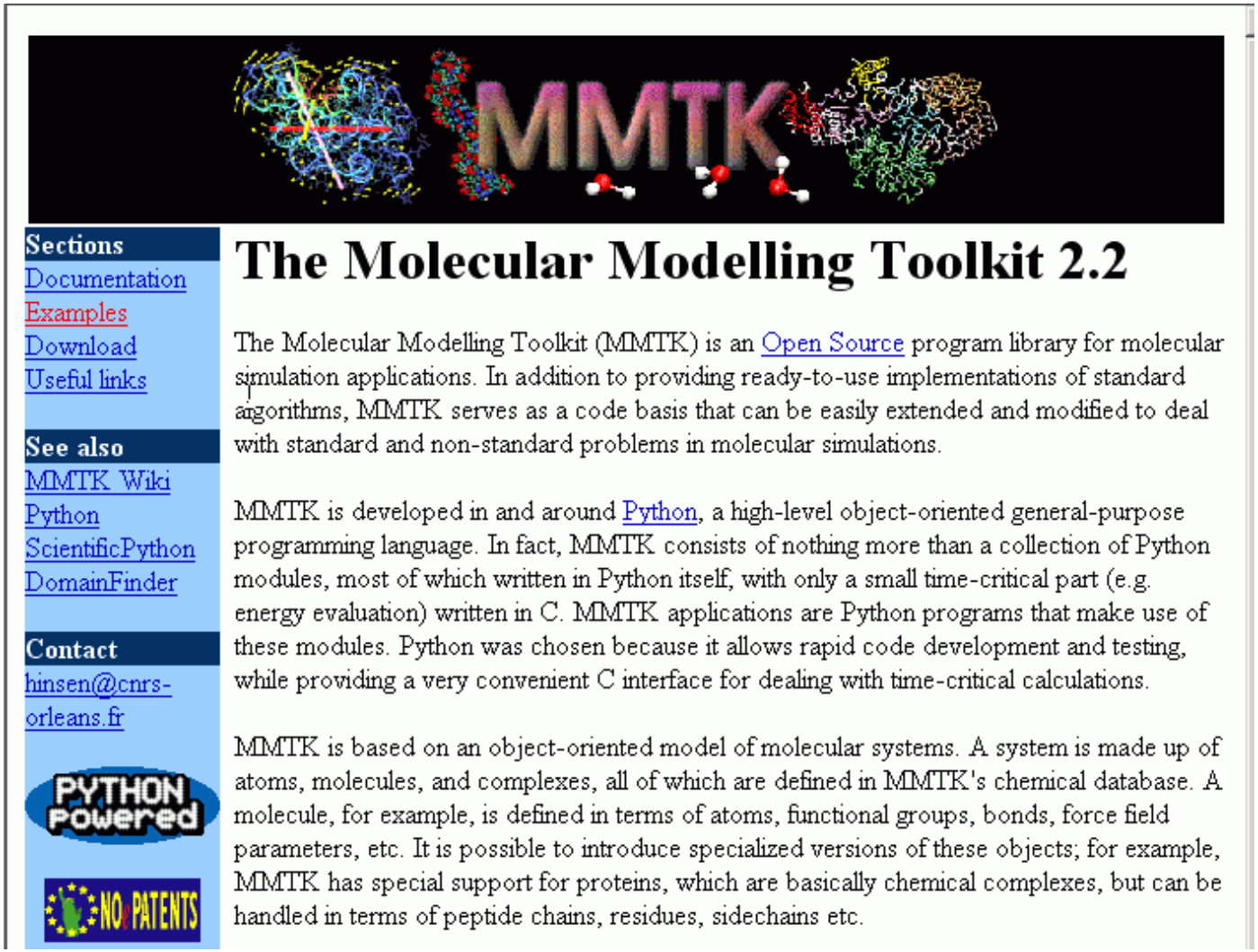


Darst et al, PNAS, 2002, 99:4296

70% overlap between the direction of the observed displacements with the direction of mode 1

Quick NMA for Atomic Structures

http://dirac.cnrs-orleans.fr/MMTK

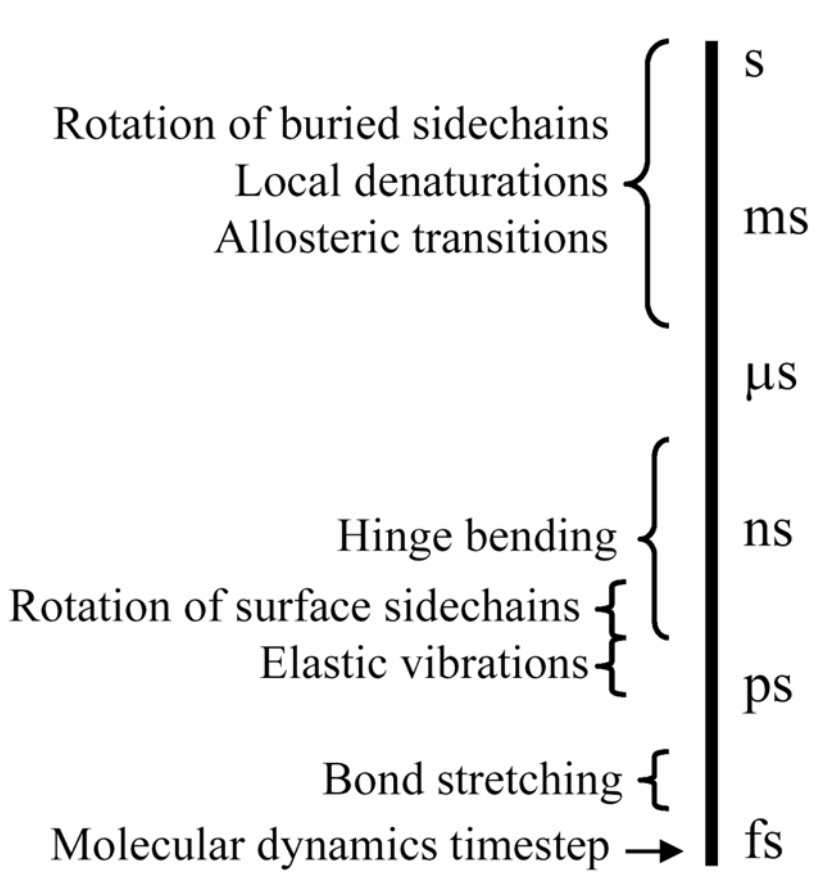


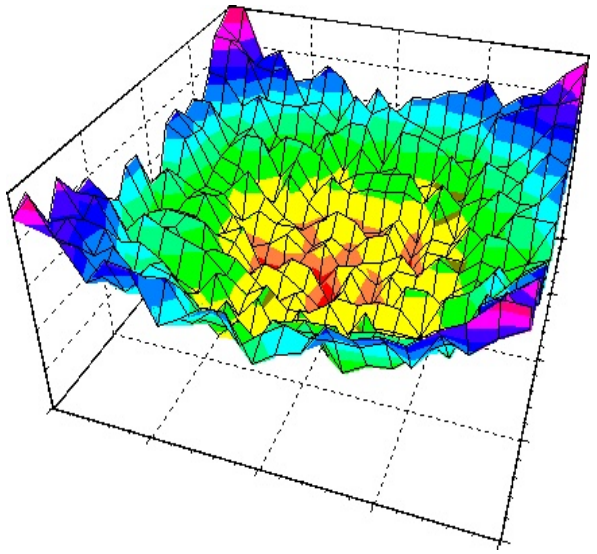
What are the Limitations of NMA?

- We do not know *a priori* which is the relevant mode, but the first 12 low-frequency modes are probable candidates.
- The amplitude of the motion is unknown.
- NMA requires additional standards for parameterization, i.e. a screening against complementary experimental data to select the relevant modes and amplitude.
- Expert user input / evaluation required
- Not based on first principles of physics (like MD).

What are the Limitations of MD?

sampling problem





Energy Surface →
Exploration by Simulation..

Solution: Enhanced Sampling MD

drive MD by collective coordinates (PCA or NMA)

First approach with PCA:

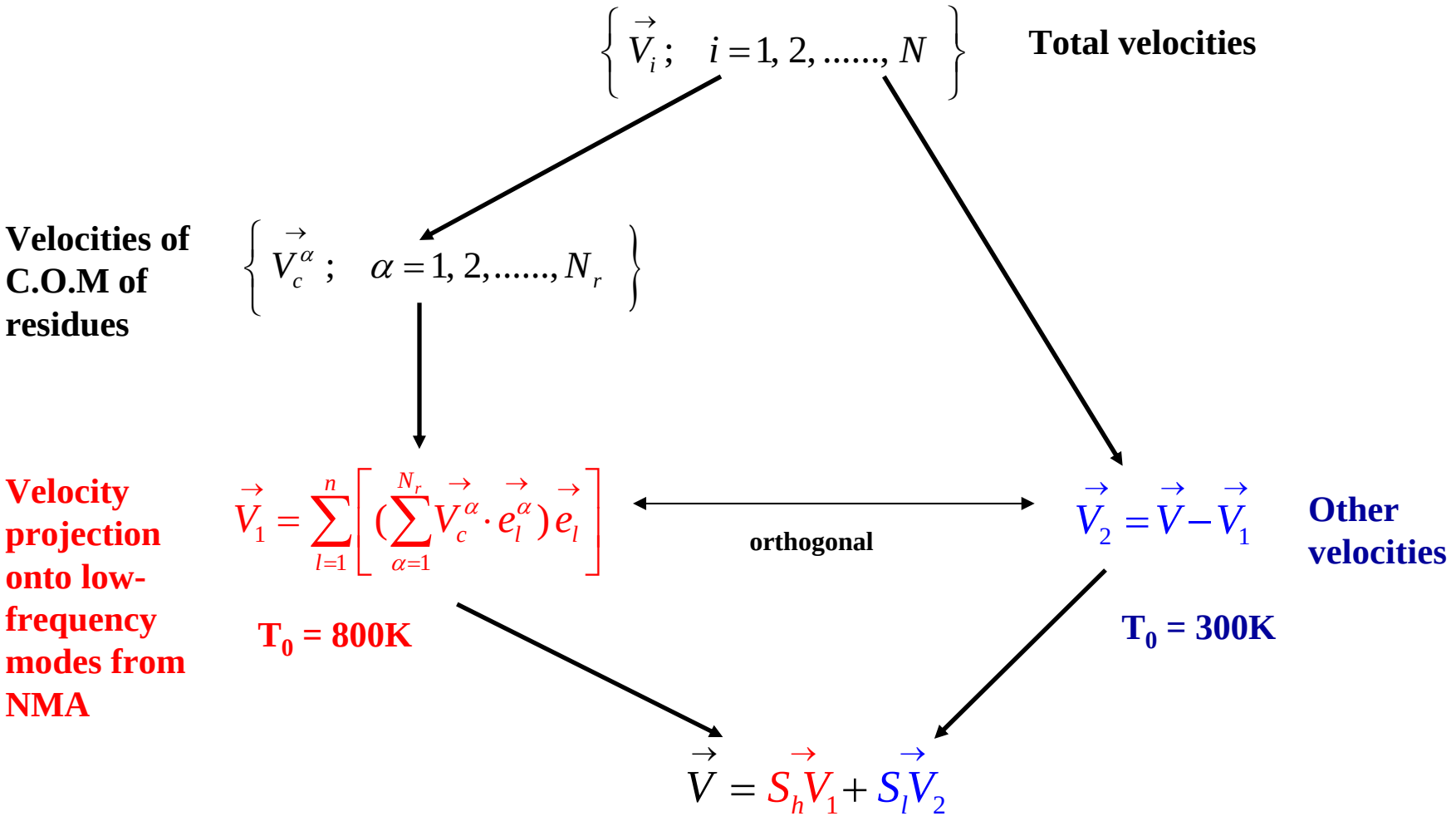
“Essential Molecular Dynamics”

Amadei, Linsen, Berendsen
“Essential Dynamics of Proteins” – Proteins (1993), 17:412-425

Use the PCs from free MD to drive a protein from one conformation to another
Used by Daidone et al. to study Cytochrome c folding with MD
Only 106 degrees of freedom out of a total 3000 were used to bias the simulation

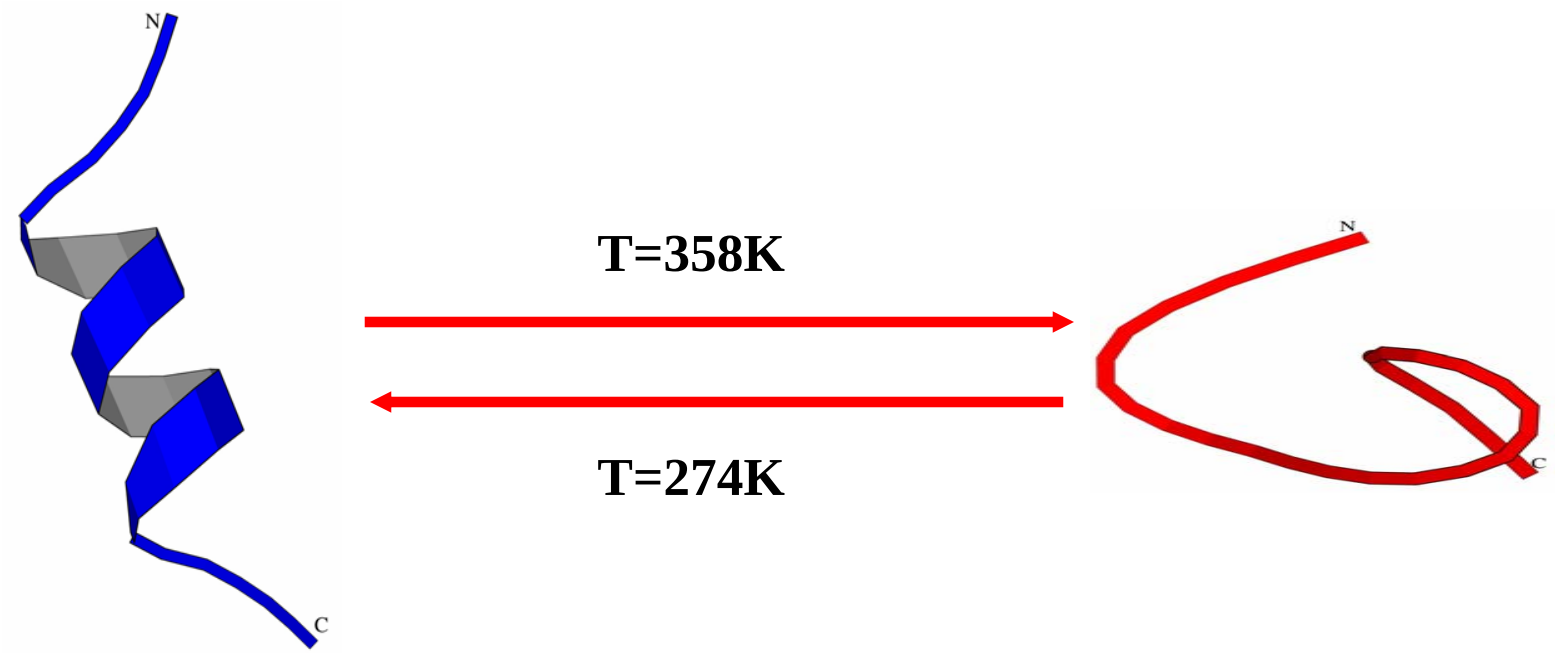
Amplified Collective Motion (ACM)

Zhang et al., Biophys J. (2003) 84:3583-93.



Folding/Unfolding of S-Peptide Analog

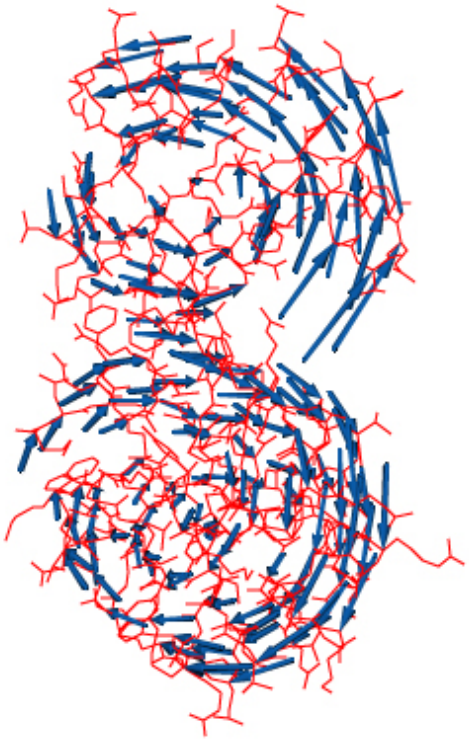
Zhang et al., Biophys J. (2003) 84:3583-93.



ACM 30-ns 3-modes @ 358K + other-DOF @ 274K
Control simulation 30-ns all-DOF 274K
implicit water model: Generalized Born

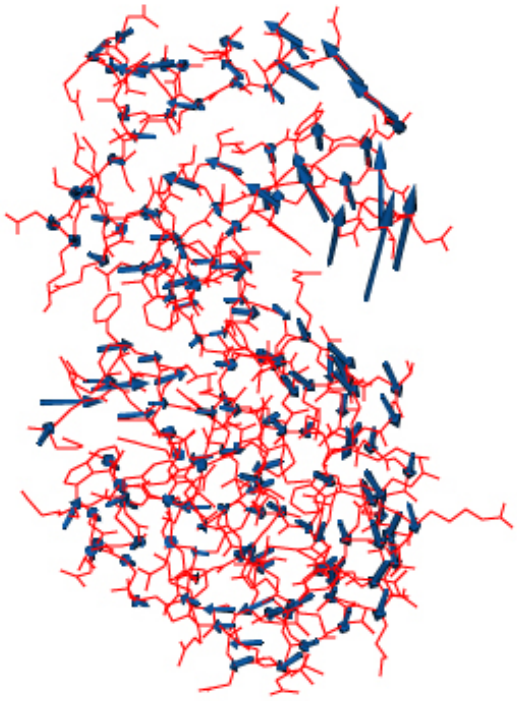
Folding/Unfolding of S-Peptide Analog

Domain Motions in Bacteriophage T4 Lysozyme (T4L)



Closure mode

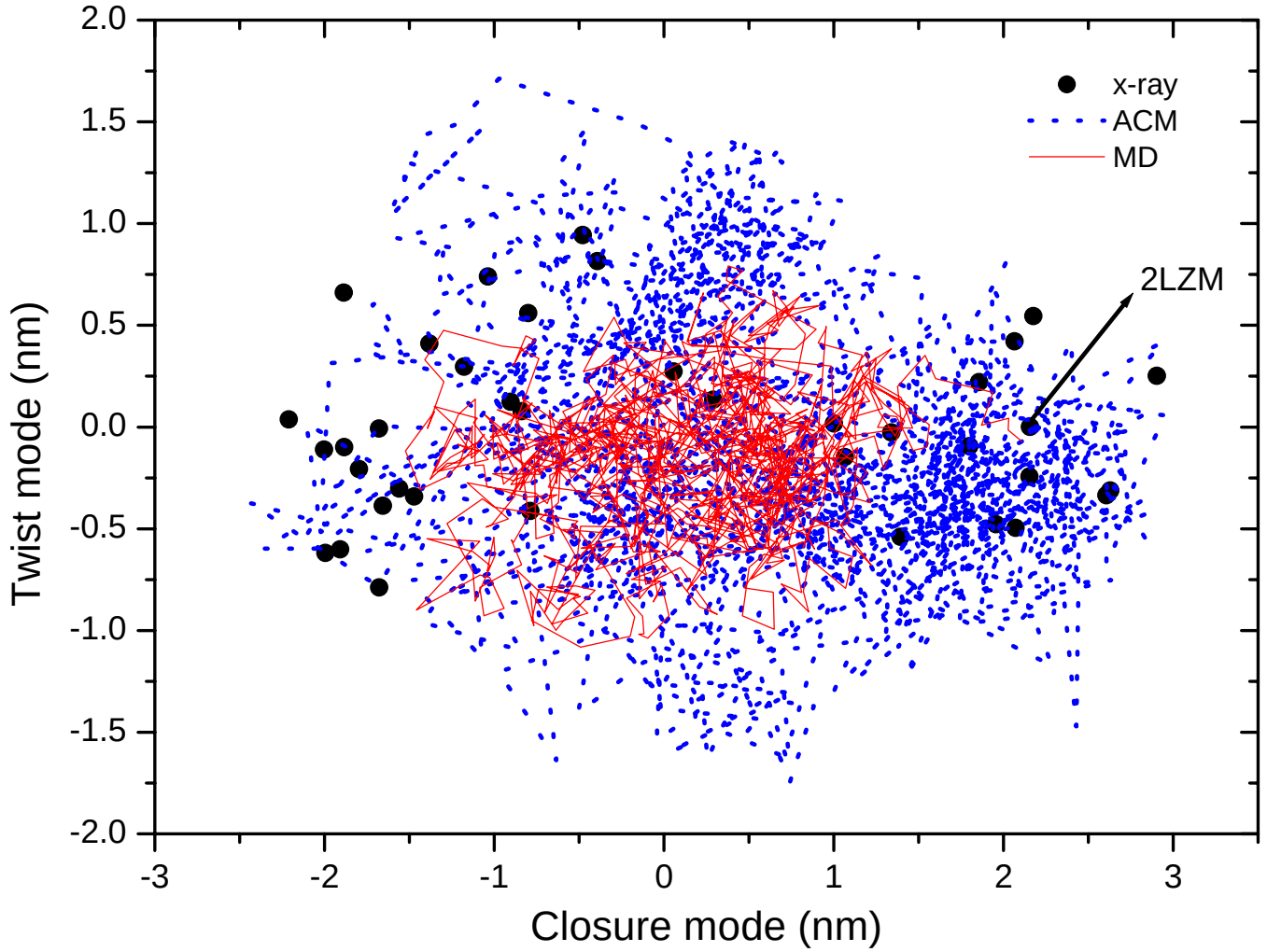
(178L vs 152L)



Twist mode

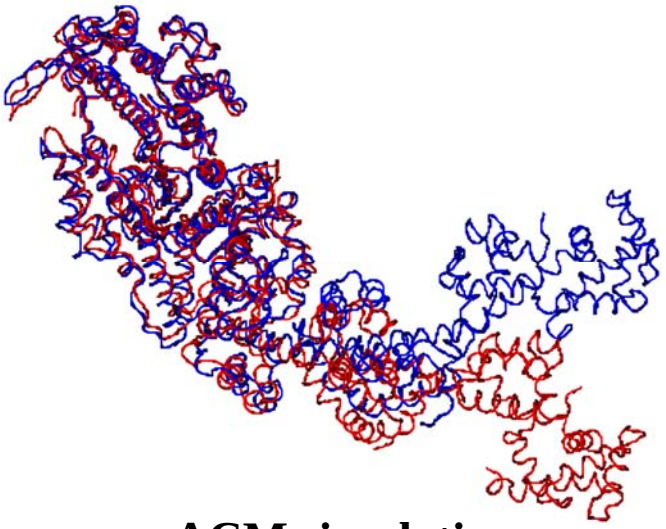
(174L vs 150L)

Projections onto the Functional Subspace

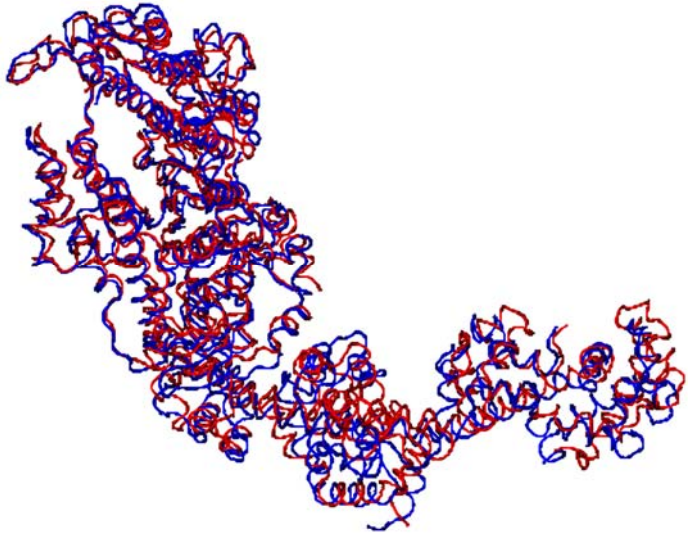


ACM: 3-ns 3-modes @ 800K other @ 300K Standard MD 3-ns all @ 300K

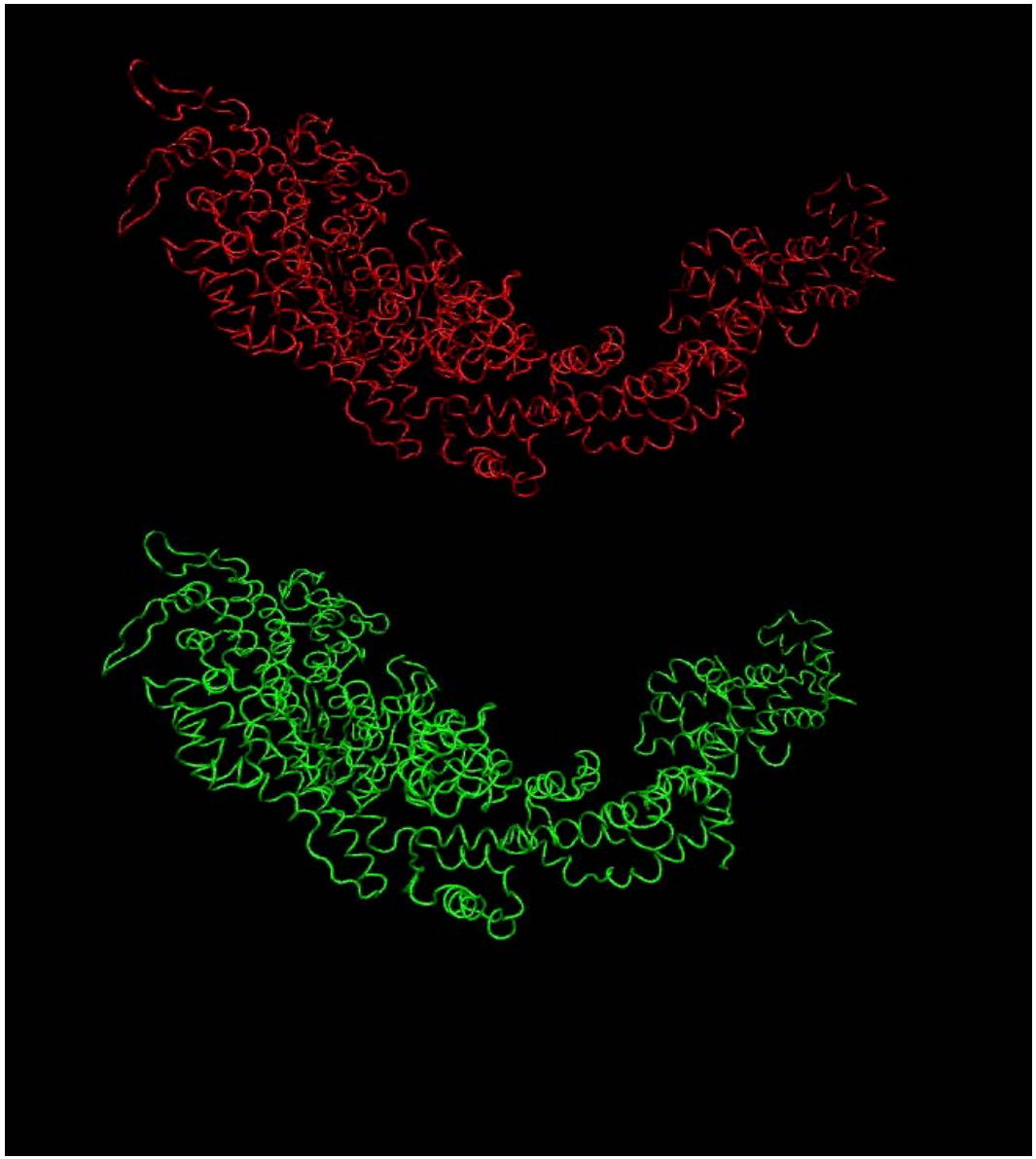
Myosin: ACM vs. MD



ACM simulation



MD simulation



Global Collective Coordinates: What are the Limitations?

- In NMA, we do not know *a priori* which is a functionally relevant mode, the first 12 low-frequency modes are probable candidates.

- In PCA, the global modes don't converge due to time limitations of the molecular dynamics simulation (sampling problem):

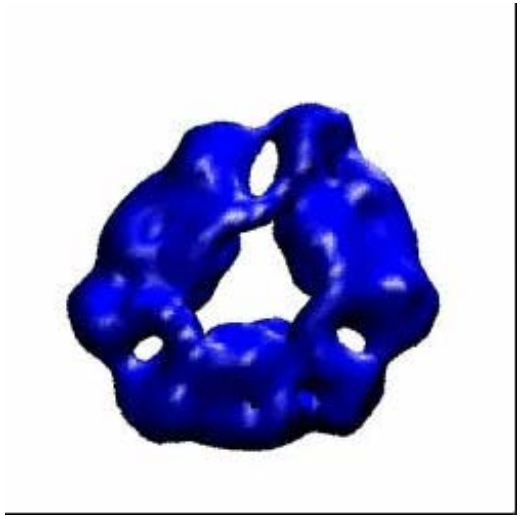
  Balsera et al. argue that PCA is not useful for identifying long-timescale protein motions and for reduced-dimension simulations.

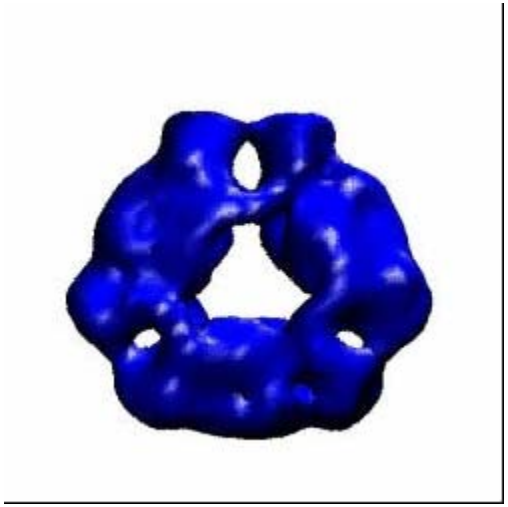
  Balsera, Wriggers, Oono, Schulten
  "Principal Component Analysis and Long Time Protein Dynamics"
  J. Phys. Chem. 100:2567-2572 (1996)

  The motions identified are only applicable to the timescale analysed.

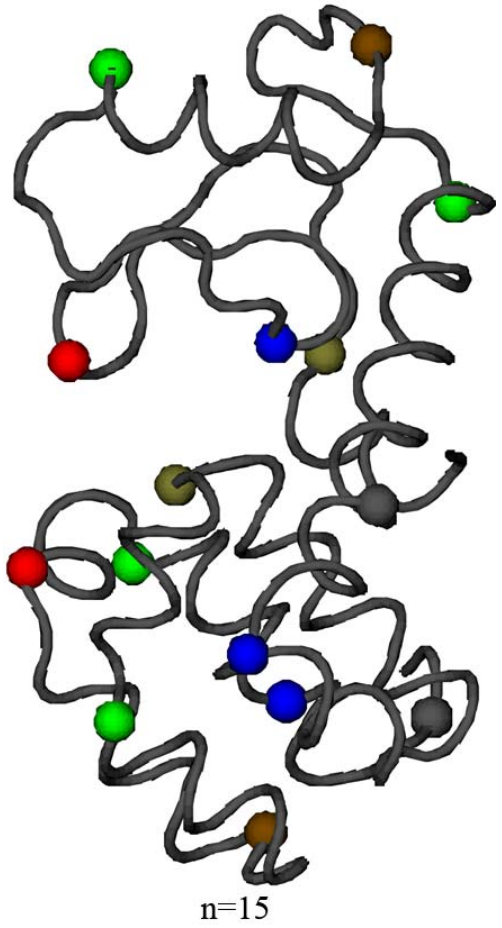
Global Collective Coordinates: What are the Limitations?

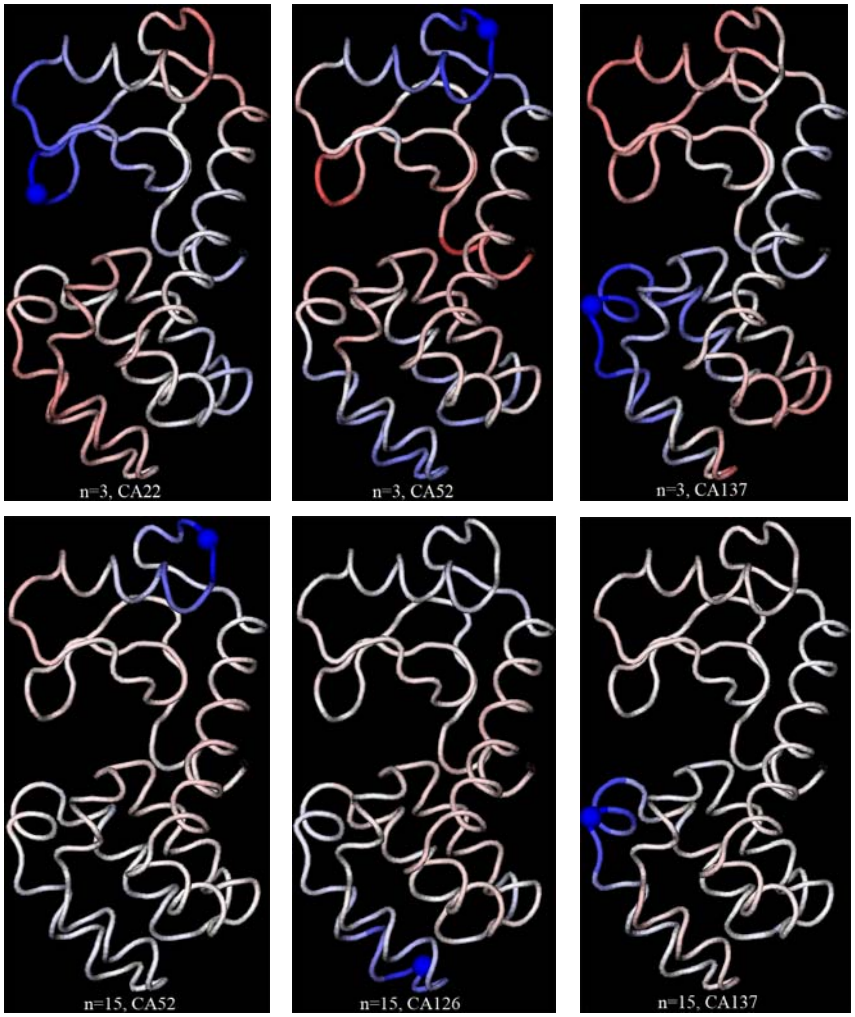
- Both PCA and NMA break the symmetry of structures due to forced orthogonalization:





Solution: Local Feature Analysis





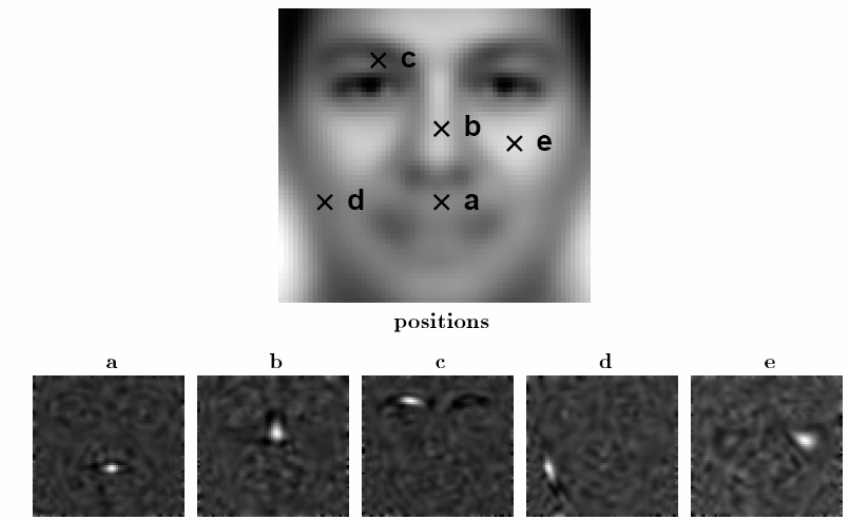
n=3

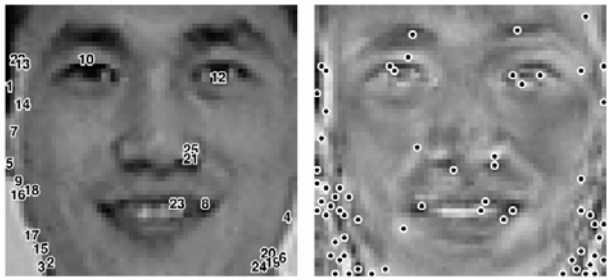
n=15

Local Feature Analysis (LFA)

Goal: an alternative statistical theory that describe dynamic features locally and that does not suffer from the sampling and orthogonalization problems.

Unlike the global eigenmodes, LFA describes objects in terms of statistically derived local features and their positions.





Is LFA applicable to protein dynamics?

From: Penev PS, Atick JJ: **Local Feature Analysis: A General Statistical Theory for Object Representation.** Network: computation in neural systems 1996, **7:**477-500.

Local Feature Analysis (LFA)

- Theory (I)

Covariance matrix from the MD simulation: $C(i,\,j) \equiv \langle \Delta x_i \Delta x_j \rangle \equiv \langle (x_i - \langle x_i \rangle)\,(x_j - \langle x_j \rangle) \rangle$

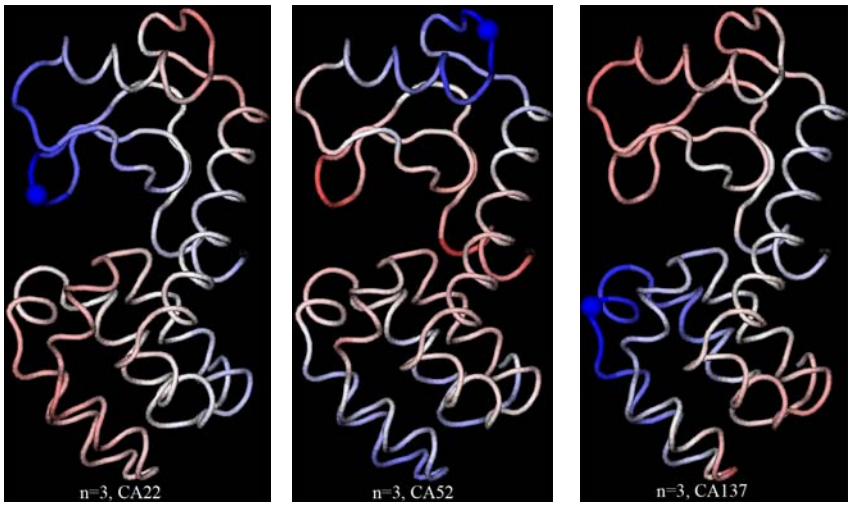
PCA: $C(i,\,j) = \sum_{r=1}^{3N} \Psi_r(i) \lambda_r \Psi_r(j)$ $\longrightarrow$ **PCA output:** $A_r = \sum_{i=1}^{3N} \Psi_r(i) \Delta x_i \equiv \sum_{i=1}^{3N} K_r(i) \Delta x_i$

General form for the LFA kernel: $K(i,\,j) = \sum_{r,\,s=1}^{n} \Psi_r(i) Q_{rs} \Psi_s(j)$ $\longrightarrow$ $K(i,\,j) = \sum_{r=1}^{n} \Psi_r(i) \frac{1}{\sqrt{\lambda_r}} \Psi_r(j)$

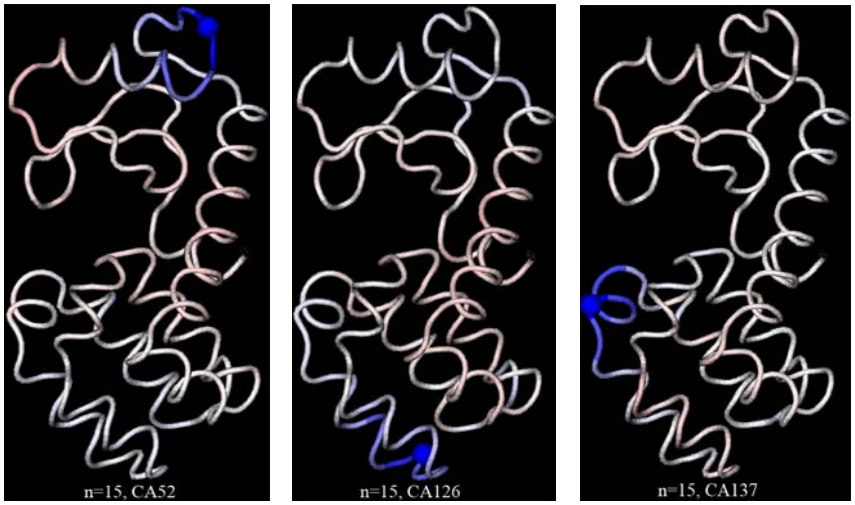
LFA output: $O(i) \equiv \sum_{j=1}^{3N} K(i,\,j) \Delta x_j$ $\longrightarrow$ $O(i) = \sum_{j=1}^{3N} \left(\sum_{r=1}^{n} \Psi_r(i) \frac{1}{\sqrt{\lambda_r}} \Psi_r(j) \right) \Delta x_j = \sum_{r=1}^{n} \frac{A_r}{\sqrt{\lambda_r}} \Psi_r(i)$

Residual correlation: $\langle O(i) O(j) \rangle = \sum_{r=1}^{n} \Psi_r(i) \Psi_r(j) \equiv P(i,\,j)$

Output Correlation



n=3



n=15

Local Feature Analysis (LFA)

- Theory (II)

We replaced the n global PCA modes with the full 3N LFA output functions. Therefore an additional dimensionality reduction step is required in the LFA output space. We approximate the entire 3N outputs with only a small subset of them that correspond to the strongest local features by taking advantage of the fact that neighboring outputs are highly correlated.

Reconstruct the outputs:

$$O^{rec}(i) = \sum_{m=1}^{|\mathcal{M}|} a_m(i) O(i_m)$$

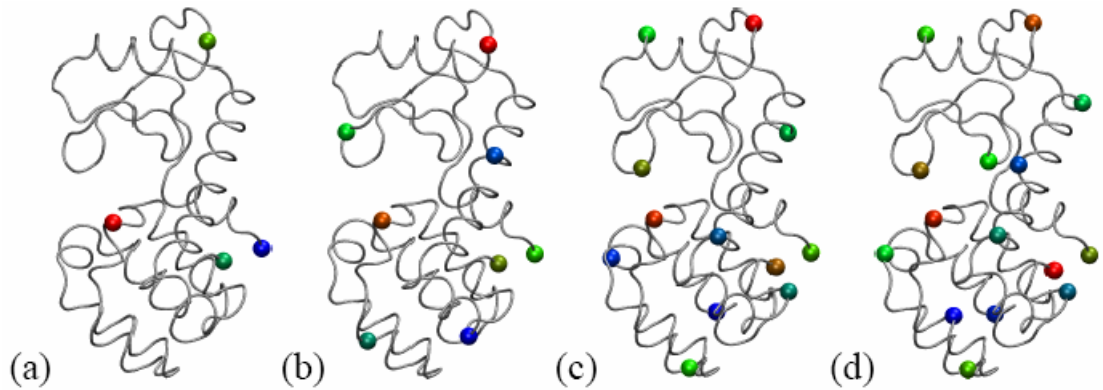
Optimal linear prediction coefficients:

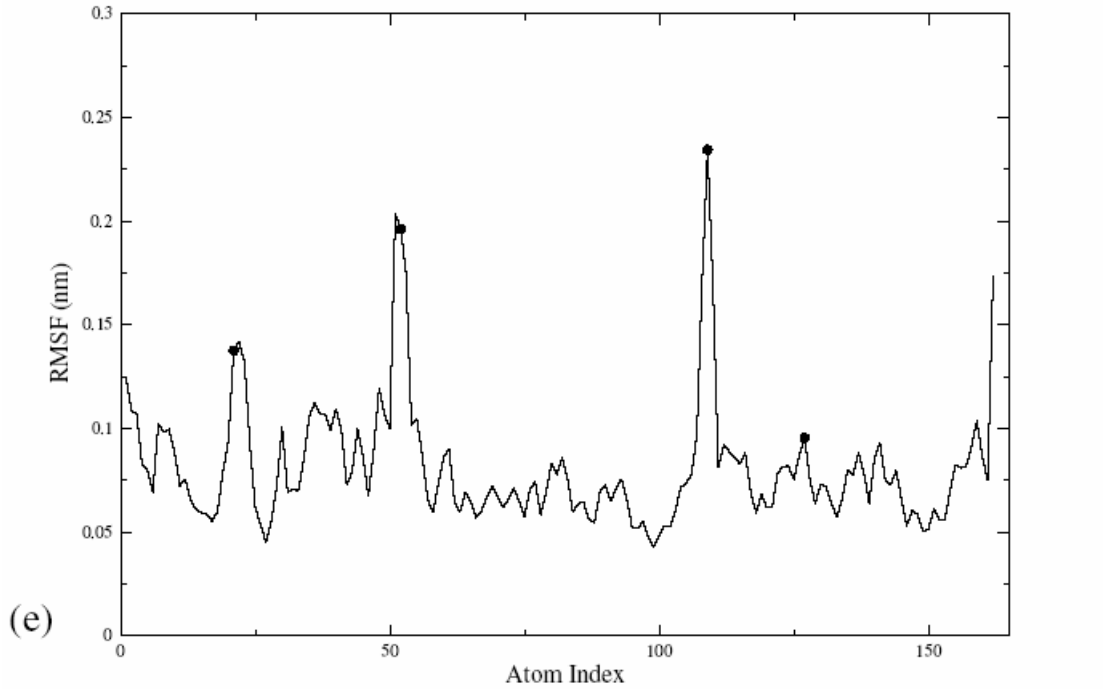
$$a_m(i) = \sum_{l=1}^{|\mathcal{M}|} P(i,\, i_l)(P'^{-1})_{lm}$$

Average reconstruction mean square error:

$$E^{rec} = \left\langle \|O^{err}(i)\|^2 \right\rangle \equiv \left\langle \|O(i) - O^{rec}(i)\|^2 \right\rangle$$

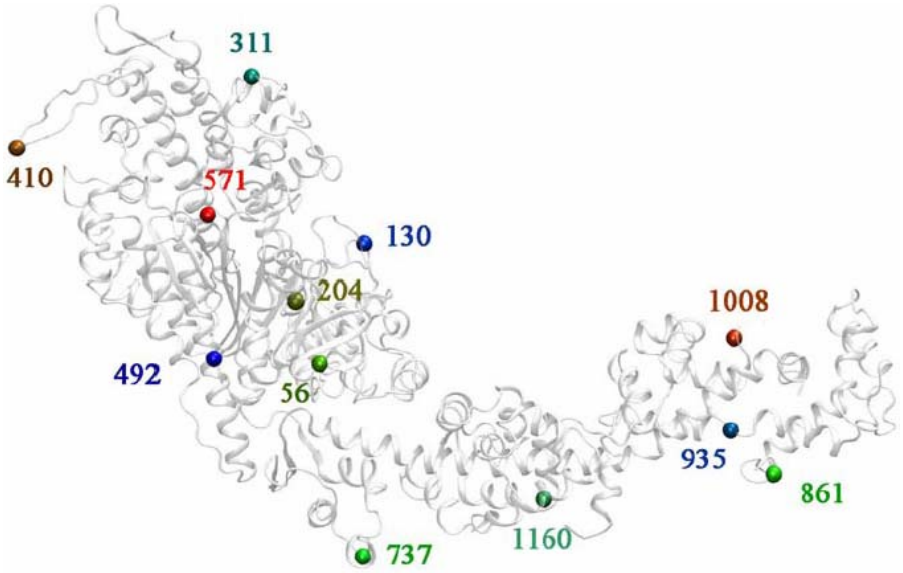
Sparsification



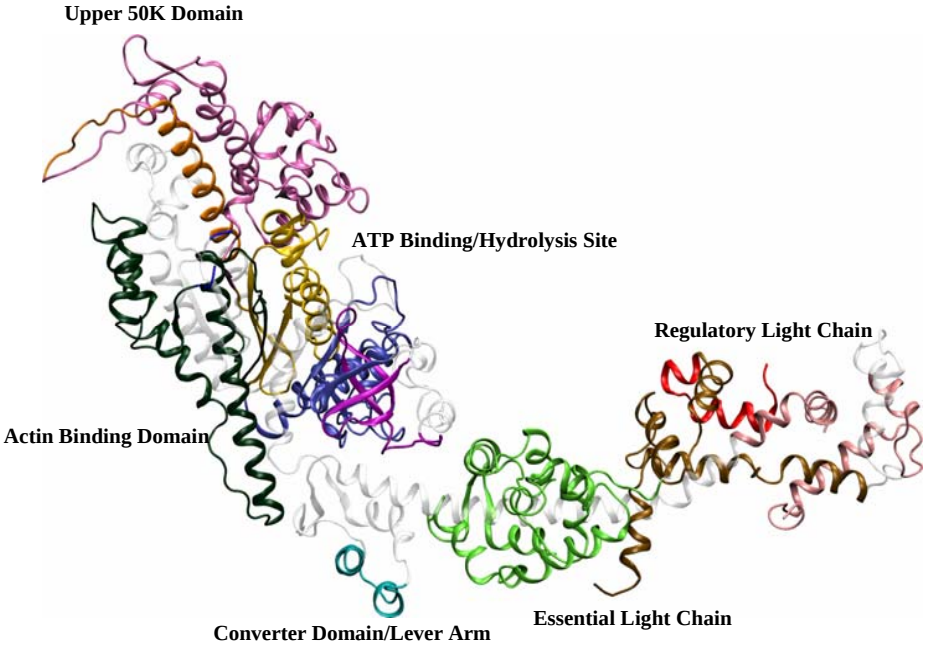


(a) The first 4 PCA modes were used to do LFA, n=4; (b) n=8, (c) n=12, and (d) n=15. (e) Root-mean-square fluctuations of C_alpha atoms in T4L.

Local Feature Analysis of Myosin



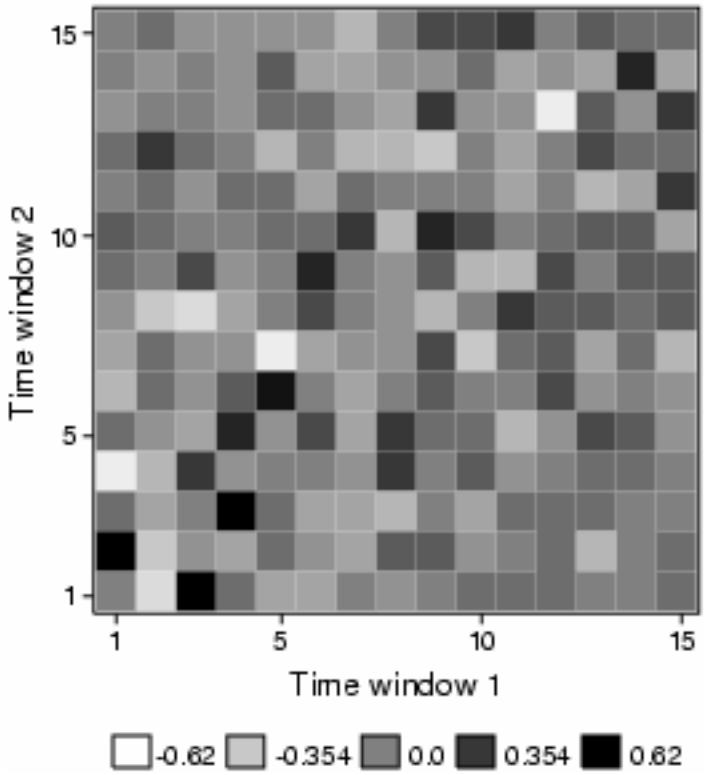
Twelve seed atoms



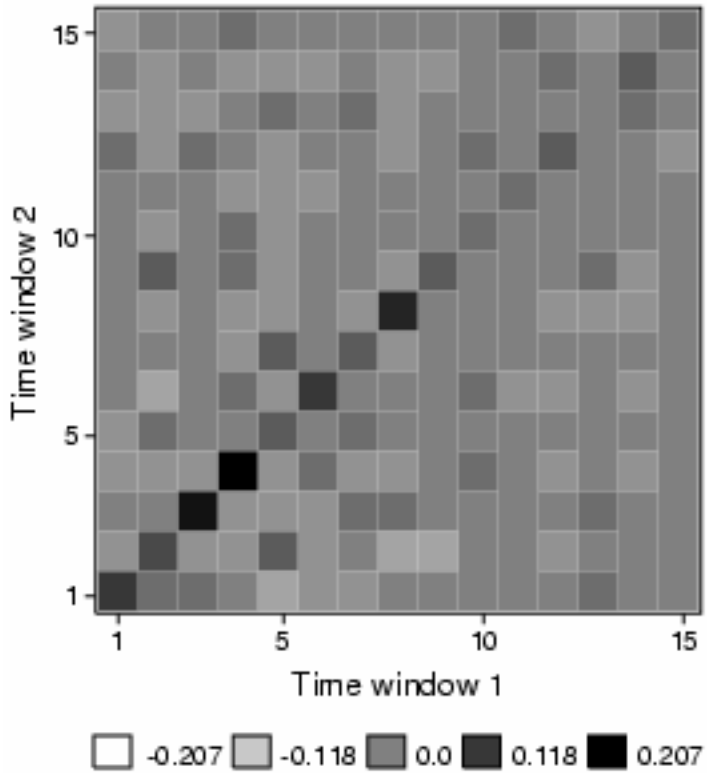
Twelve local dynamic domains

Convergence Properties

Overlap between 15 modes from first and second half of 10ns trajectory (T4 lysozyme, standard MD)

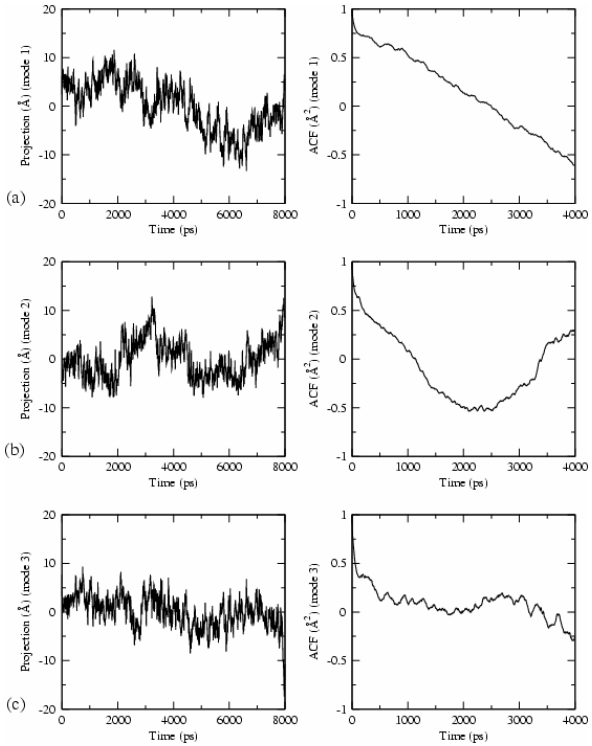


PCA

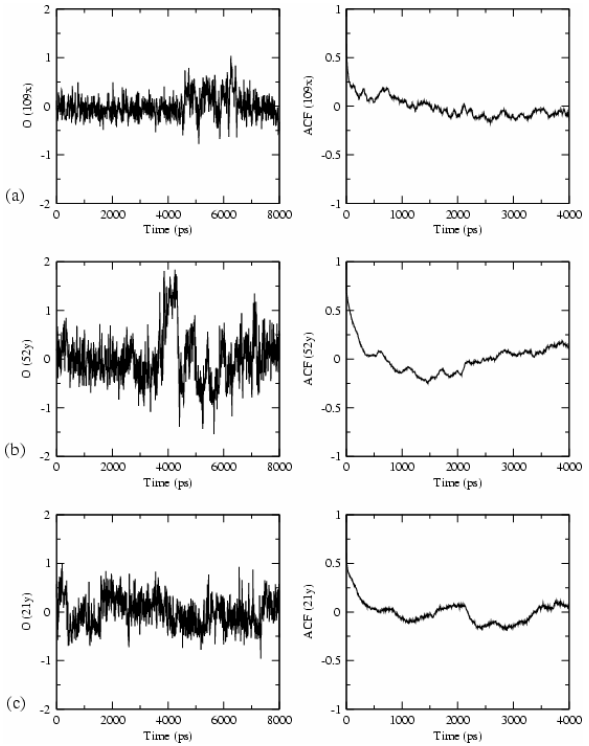


LFA

Convergence Properties



PCA autocorrelation



LFA autocorrelation

The intrinsic dynamics of local domains is more extensively sampled than that of globally coherent PCA modes.

Outlook: Predicting Functional Motion

- It appears that PCA and NMA over-estimate the coherence of global motion across large biopolymers and create artifacts due to orthogonalization.
- LFA captures local dynamic features reproducibly and is less sensitive to the MD sampling problem.
- We perform a statistical analysis that emphasizes dynamic domains that are moving independently from each other.
- LFA paper just appeared in *Proteins* (Zhang & Wriggers, 2006)

Resources and Further Reading

WWW:
http://www.sosmath.com/matrix/matrix.html
http://dirac.cnrs-orleans.fr/MMTK
http://dynamite.biop.ox.ac.uk/dynamite

Papers:
L. I. Smith “A tutorial on Principal Component Analysis” (2002) e.g. at
http://kybele.psych.cornell.edu/%7Eedelman/Psych-465-Spring-2003/PCA-tutorial.pdf
Monique M Tirion (1996) Phys Rev Lett. **77:**1905-1908
Zhang et al., Biophys J. (2003) 84:3583-93.
Chacón et al. J. Mol. Biol., 2003, 326: 485-492.
Hess, Phys. Rev. E 62:8438-8448 (2000)
Amadei, Linsen, Berendsen, Proteins (1993), 17:412-425
Balsera, Wriggers, Oono, Schulten J. Phys. Chem. 100:2567-2572 (1996)
Zhang & Wriggers, Proteins (2006), 64:391-403

Text Book:
Schlick, Chapter 8.2

Acknowledgements/Credits

Florence Tama, The Scripps Research Institute

Atsushi Matsumoto, Rutgers University, Wright-Rieman Laboratories

Sansom lab, http://indigo1.biop.ox.ac.uk/MD_workshops