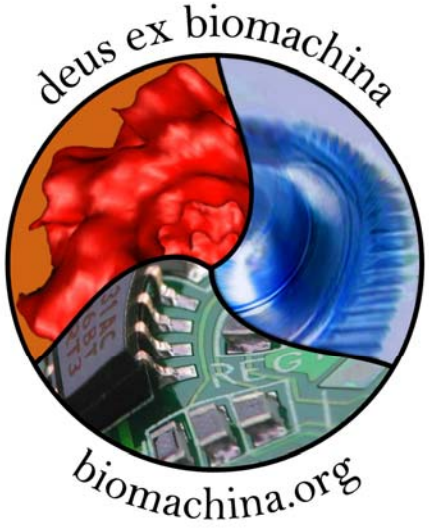


Normal Mode Analysis

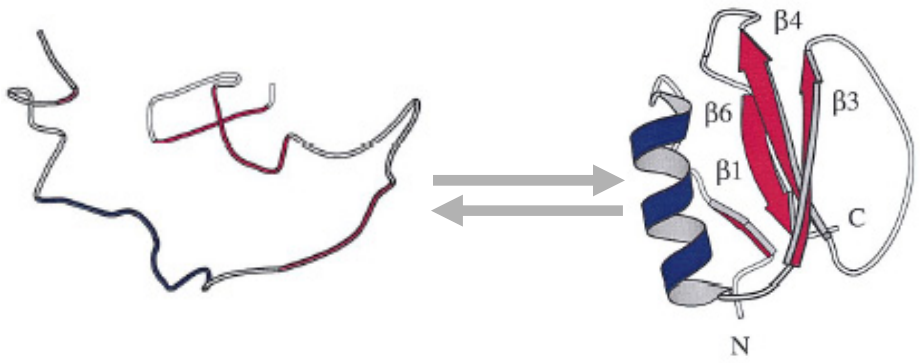
Jun Wan and Willy Wriggers
School of Health Information Sciences &
Institute of Molecular Medicine
University of Texas – Houston

Contents

- Introduction --- protein dynamics (low-frequency vibration and variables reduction)
- Normal mode analysis (NMA)
- Application and limitation
- Conclusion and future work

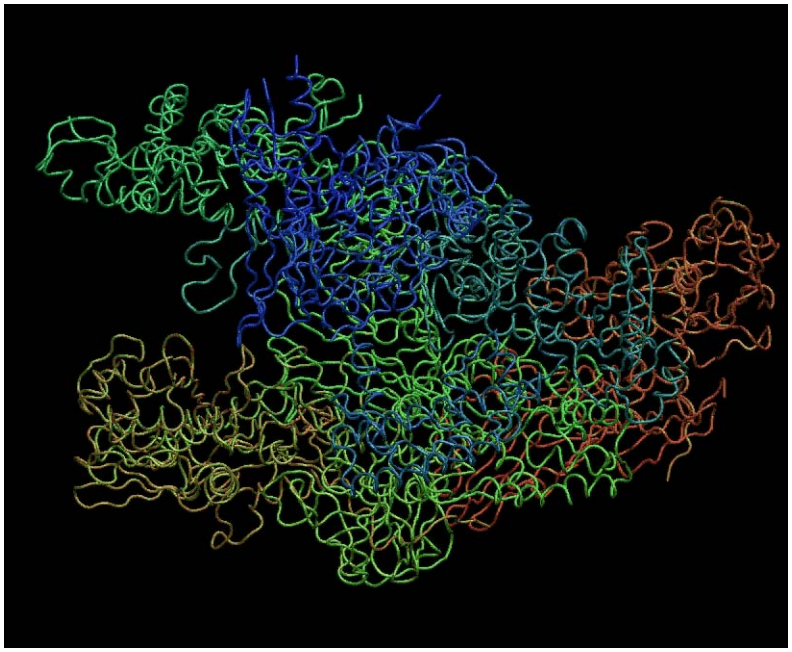
Protein Dynamics is Hierarchical

Vibration of bonds: 10^{-15} s



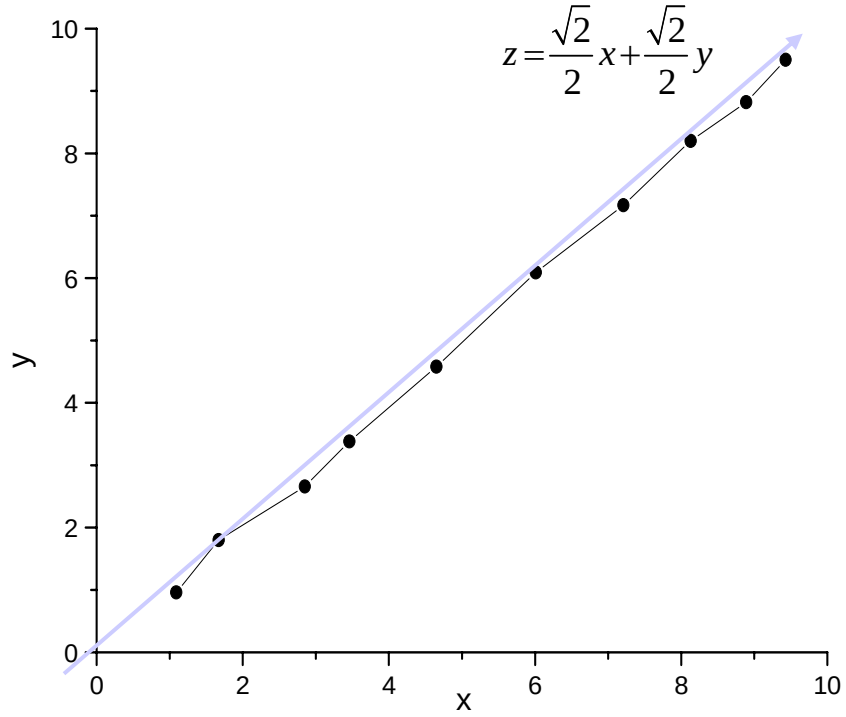
Protein folding/unfolding

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



Large-scale functional motions

Collective Coordinates and Dimensionality Reduction



Find a Model

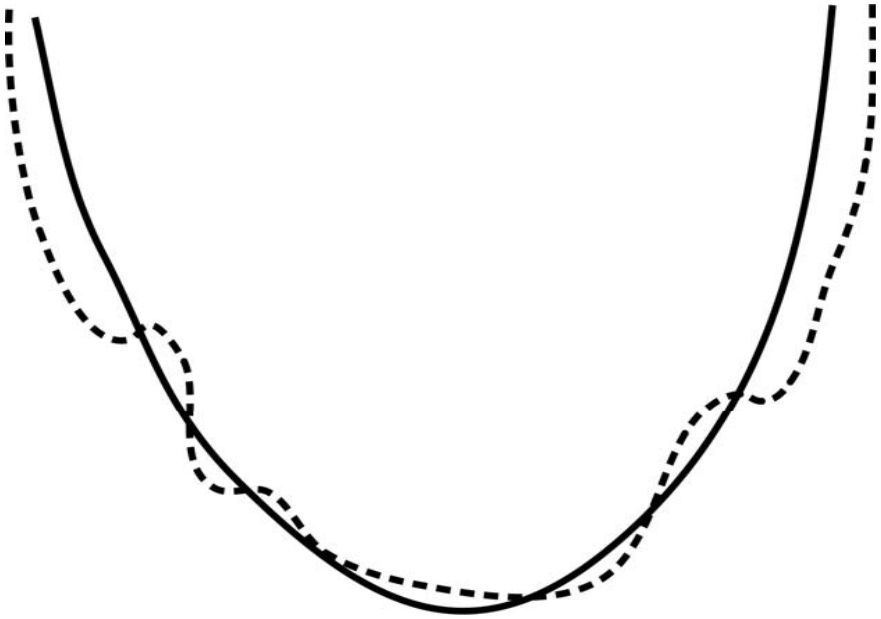
- To investigate low-frequency movement (vibration)
- To reduce the number of degrees of freedom

Normal Mode Analysis

- Theory of vibration
- Harmonic potential
- Close to the potential minimum

- Orthogonal normal modes
- Conformational fluctuation = a superposition of normal modes.

Harmonic Approximation

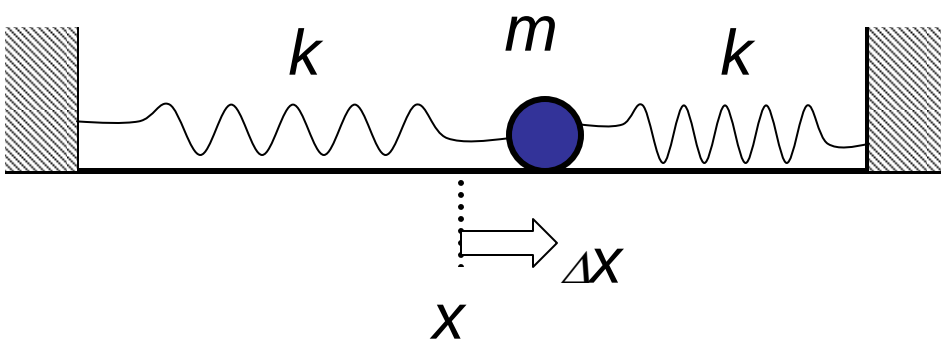


Approximation:

Potential energy => harmonic

$$E \approx E_0 + \sum_i \frac{\partial E}{\partial x_i} \Delta x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 E}{\partial x_i \partial x_j} \Delta x_i \Delta x_j$$

Harmonic Oscillator



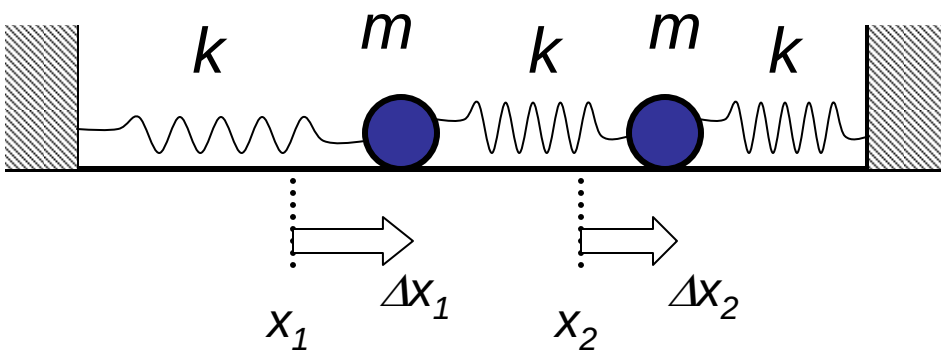
$$m\frac{d^2\Delta x}{dt^2} = -k\Delta x - k\Delta x$$

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

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

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

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

Eigenvalue and Eigenvector 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.

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

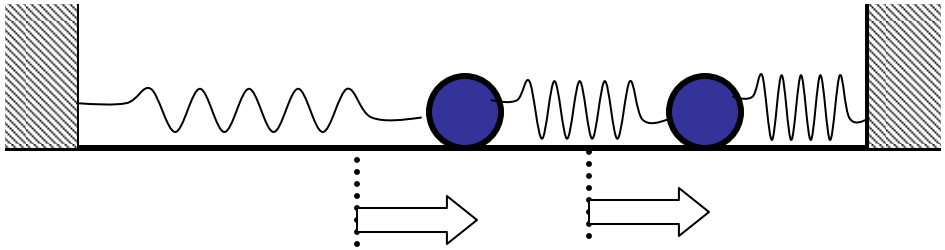
$$U^tU=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

Matrix Diagonalization

Conformational Fluctuation

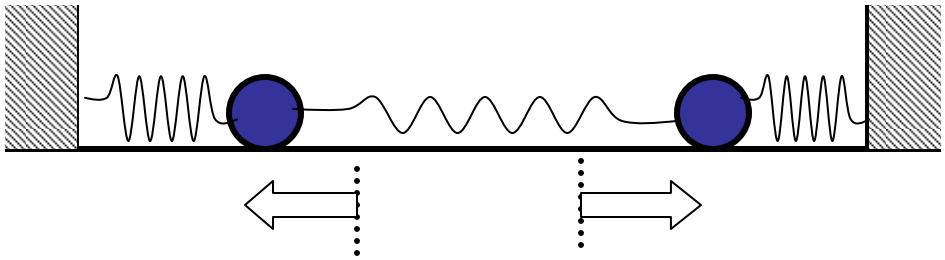
…is given by a superposition of normal modes:

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



Lower frequency mode

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



Higher frequency mode

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

Two-Atomic Molecule

Spring constant

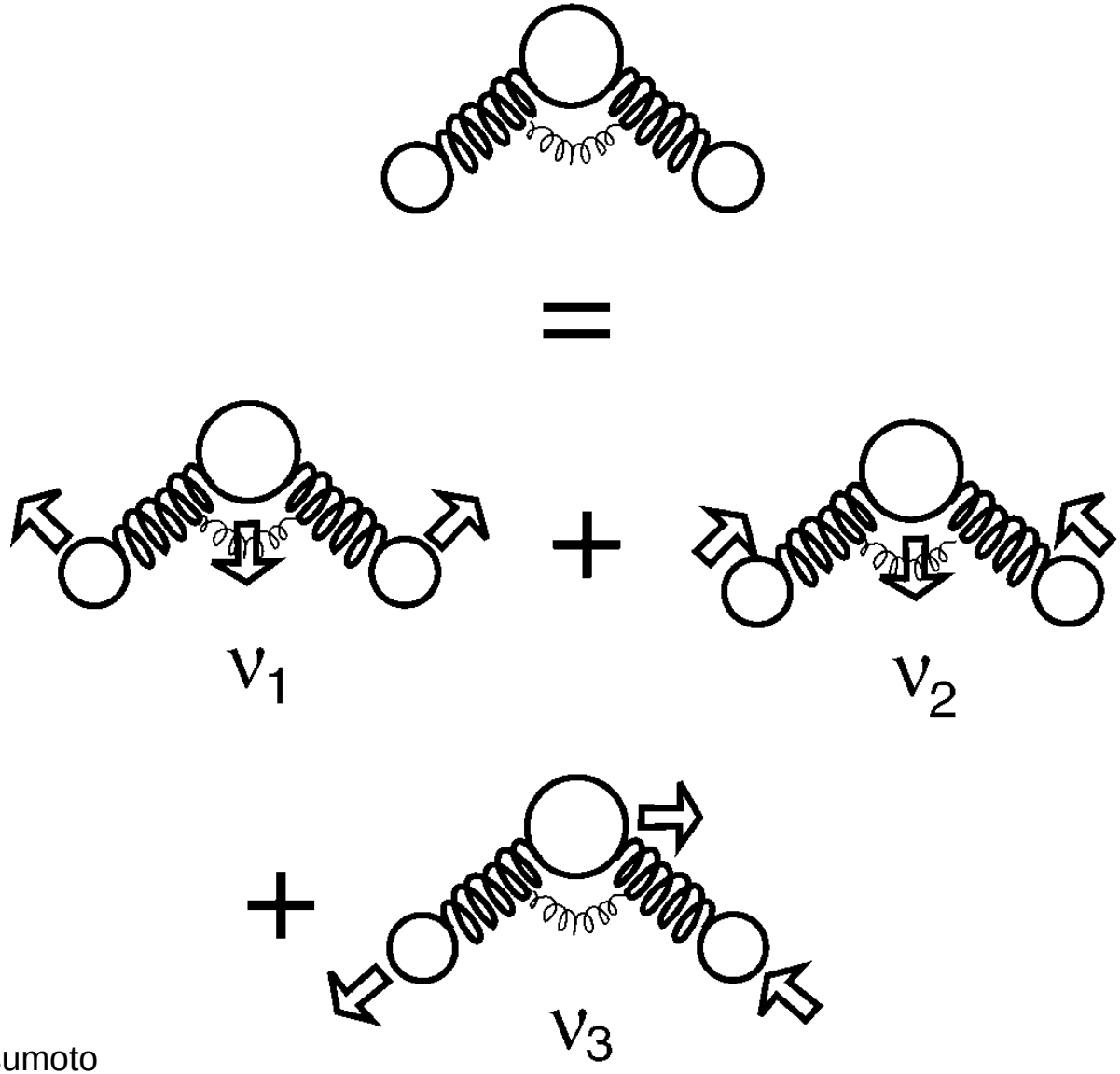
k

=

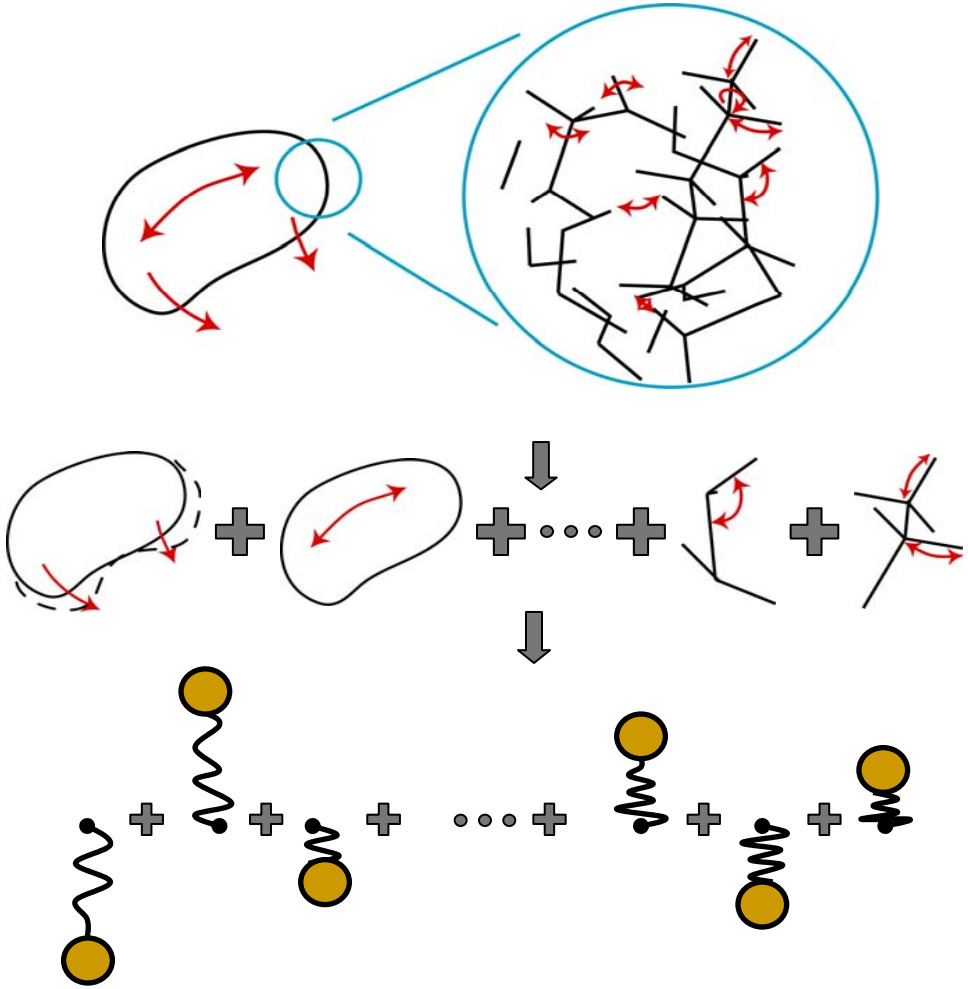
Frequency

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

Three-Atomic Molecule



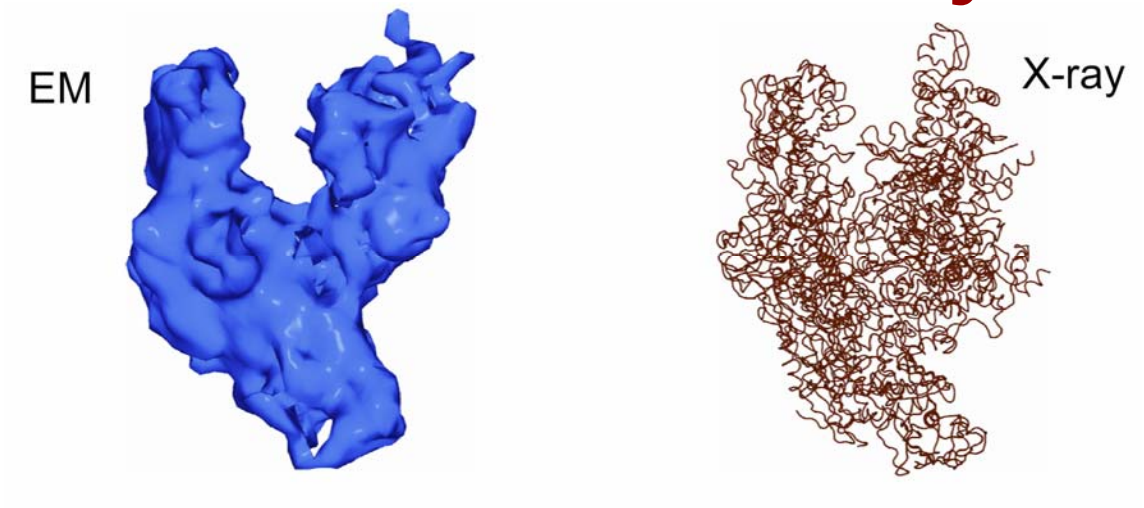
Multi-Atom Molecule



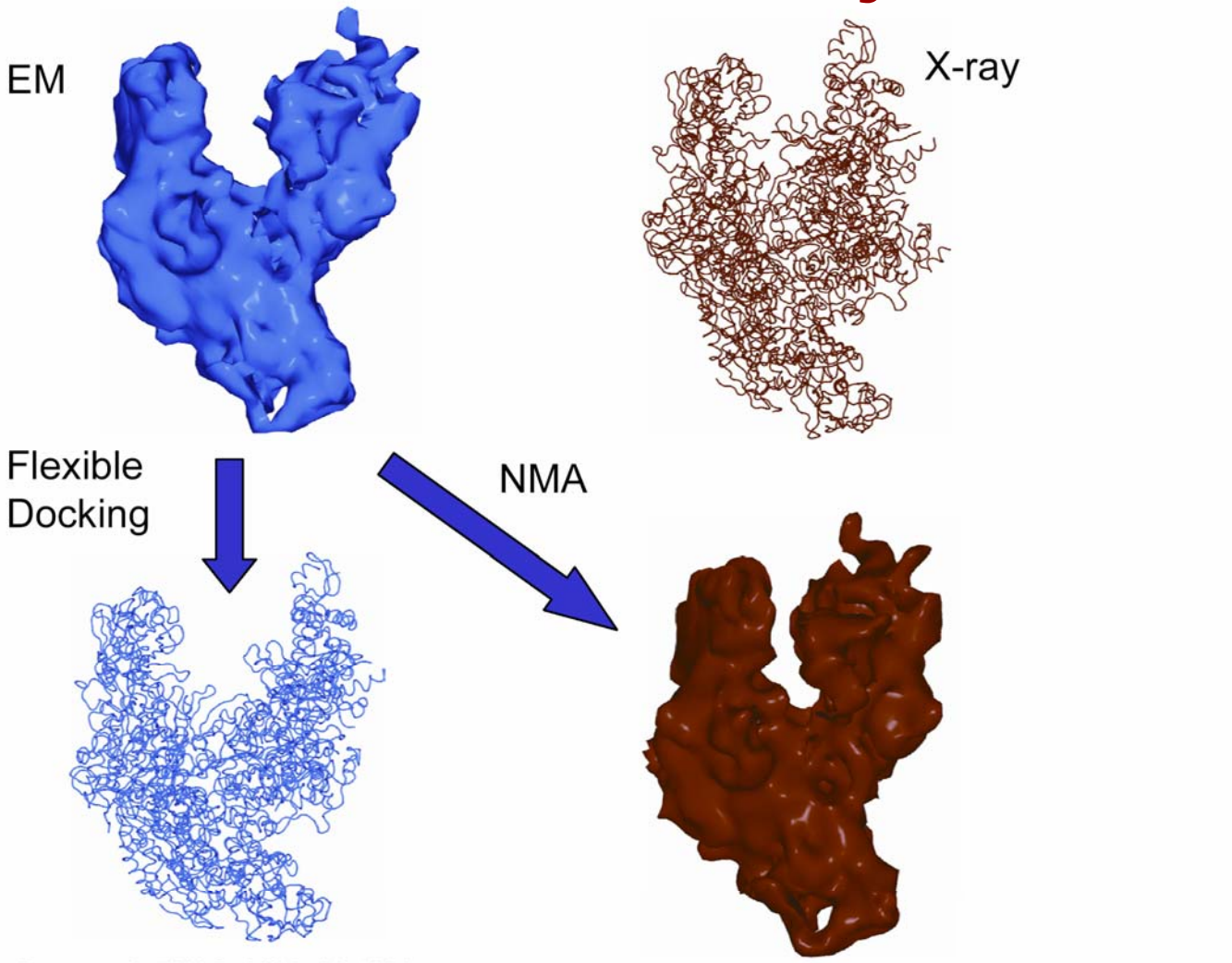
Low frequencies ⟷ High frequencies

Large collective motions — Localized motions

Success Story



Success Story



Darst et al, PNAS, 2002, 99:4296

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

Success Story

- One mode can represent 70-90% of functionally relevant motion.
- For many observed movements, the first 12 normal modes contain the relevant degrees of freedom

NMA using Molecular Mechanics

Full atomic representation and MM interactions require:

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

Computational Challenges

NMA requires:

- minimization
- diagonalization of the Hessian matrix

Problems for large systems:

- expensive, cumbersome (MM)
- memory requirements

Memory-Efficient Diagonalization

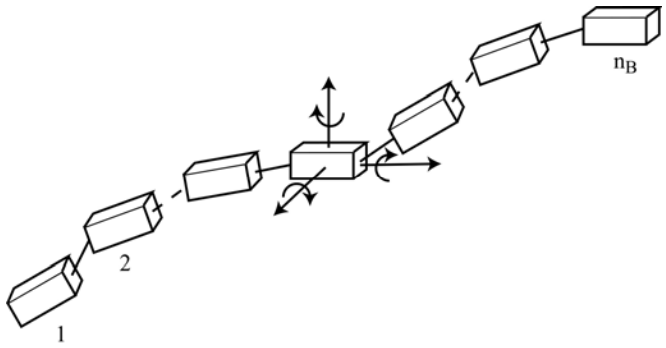
DIMB => Diagonalization in mixed basis

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

Group theory => Use symmetrical properties of viruses

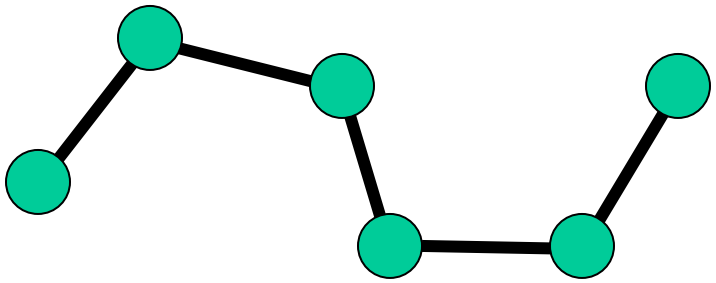
(Roux & Karplus, **1988**, *Biophys. J*, **53**, 297; Simonson & Perahia, **1992**, *Biophys. J*., **61**, 410; van Vlijmen & Karplus, **2001**, *J.Chem. Phys*, **115**, 691)

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.

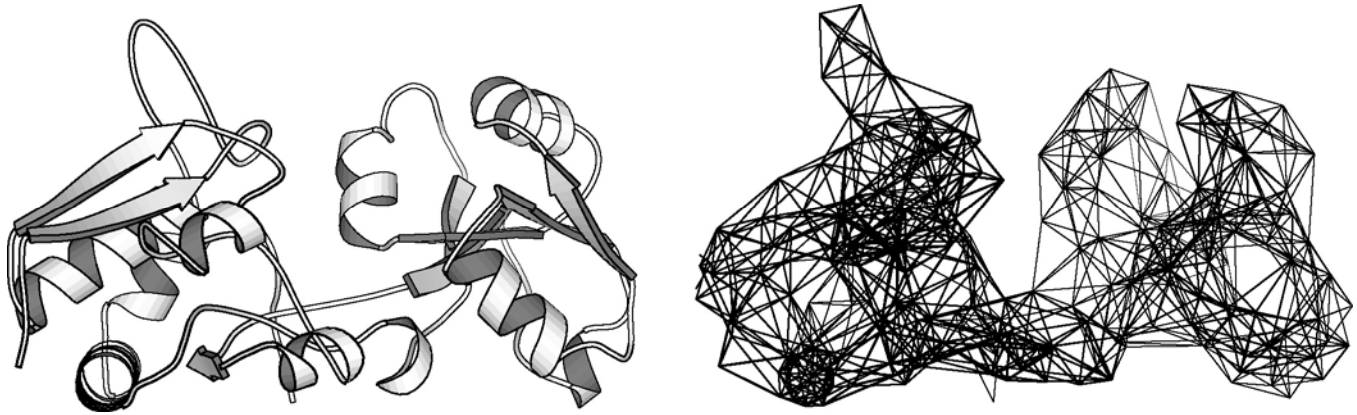
Number of variables: ~1/10

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"):

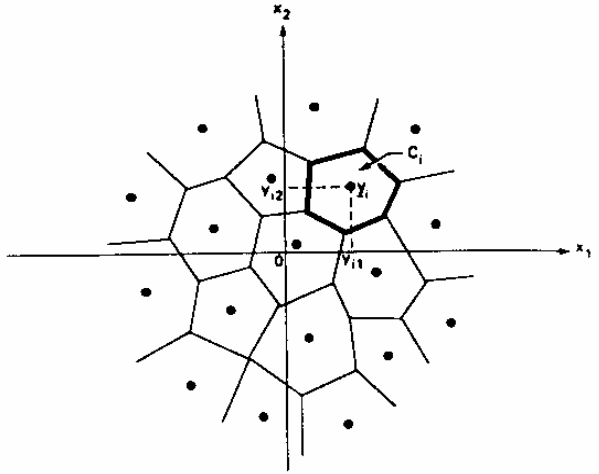
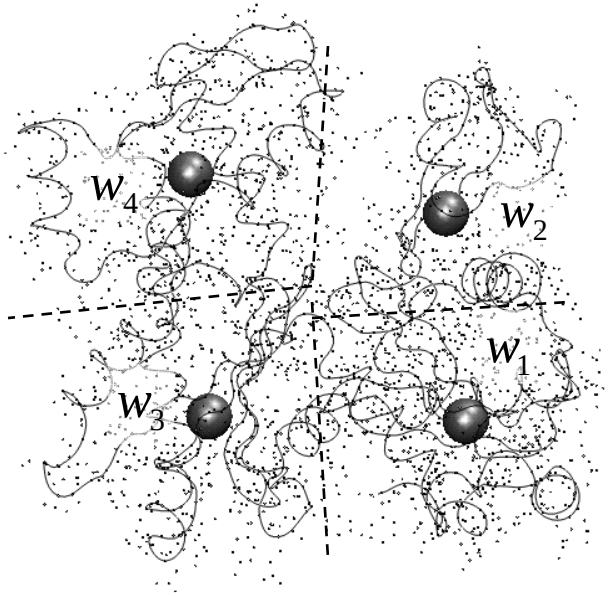


Fig. 3. Partitioning of two-dimensional space (N = 2) into L = 18 cells. All input vectors in cell C_i will be quantized as the code vector y_i. The shapes of the various cells can be very different.



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

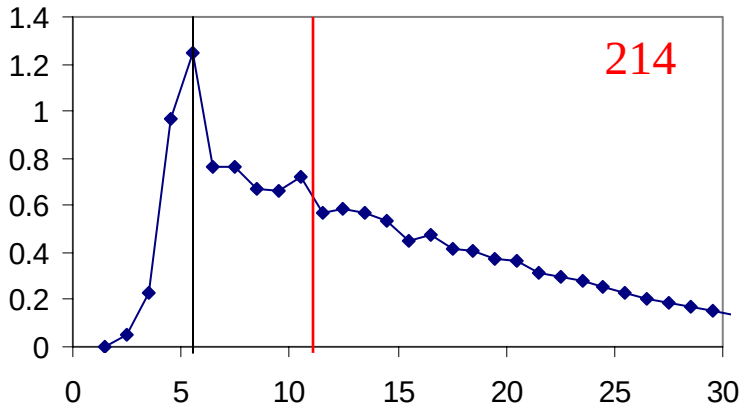
1 codebook vector ≈ 1 residue

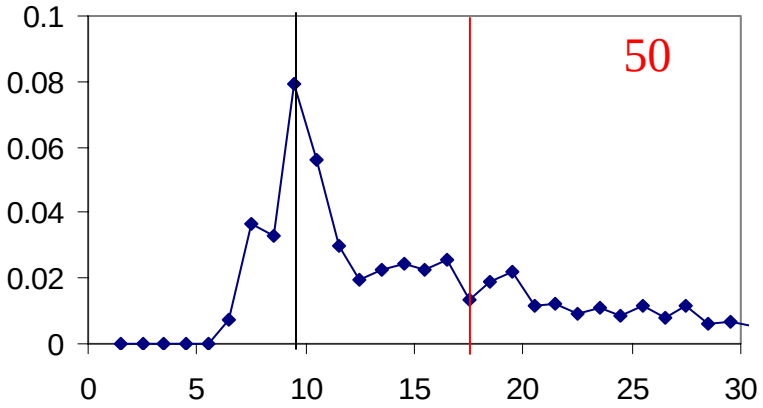
⇨10-12 Å cut-off OK

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

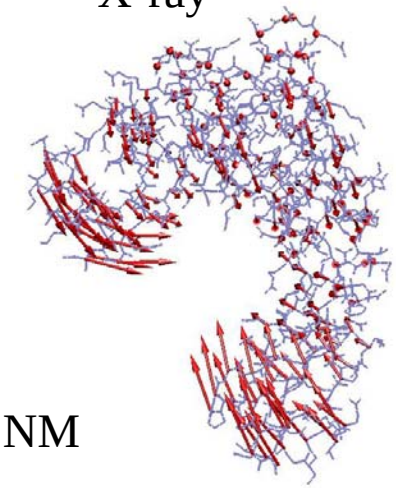
Projection onto atomic normal modes
≈ 1 for the first few modes

Low frequency NM are similar to atomic NM

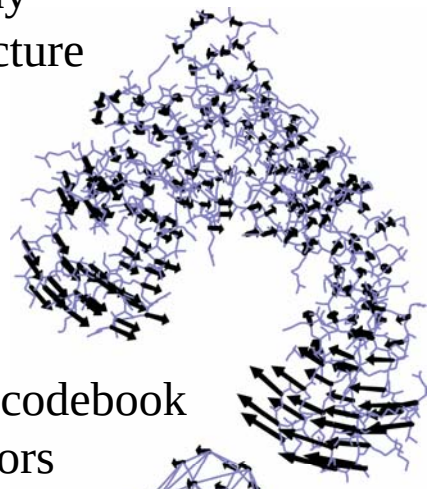
Models can reproduce functional rearrangements even at 30Å resolution

Open-Close

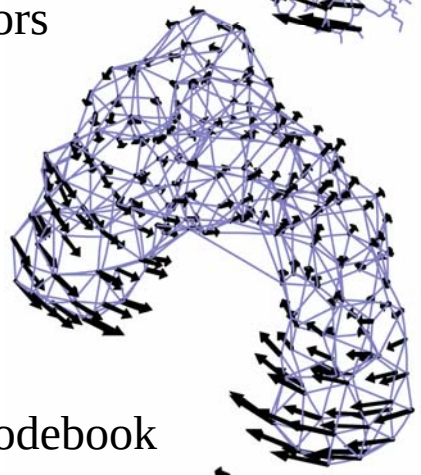
X-ray



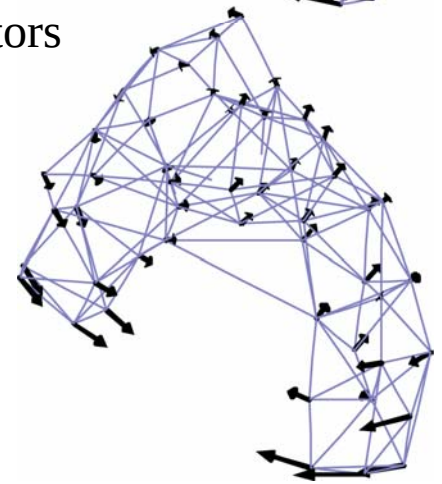
X-ray structure



214 codebook vectors

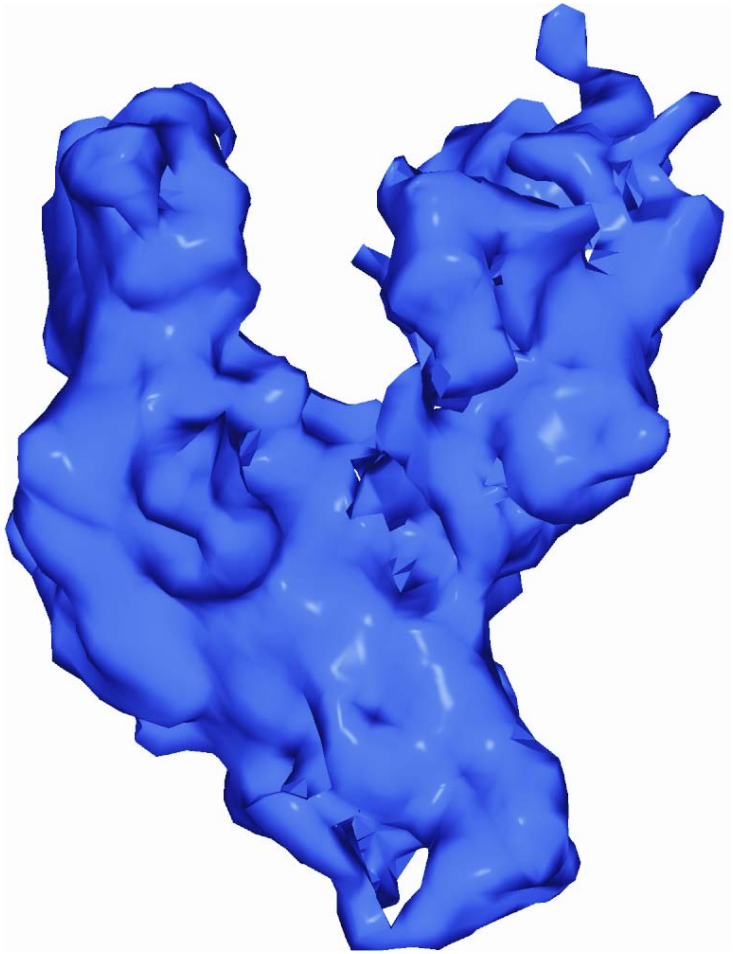


50 codebook vectors



Application to EM Data

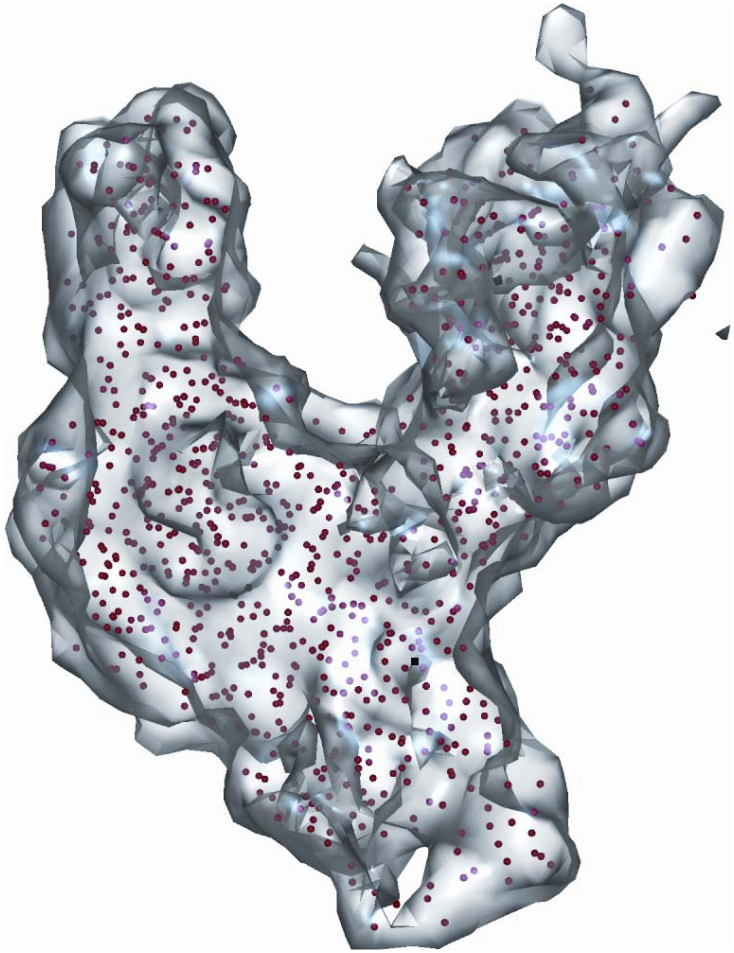
RNA Polymerase, S. Darst et al.

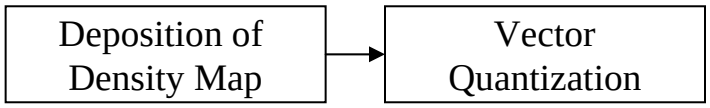


Deposition of Density Map →

Application to EM Data

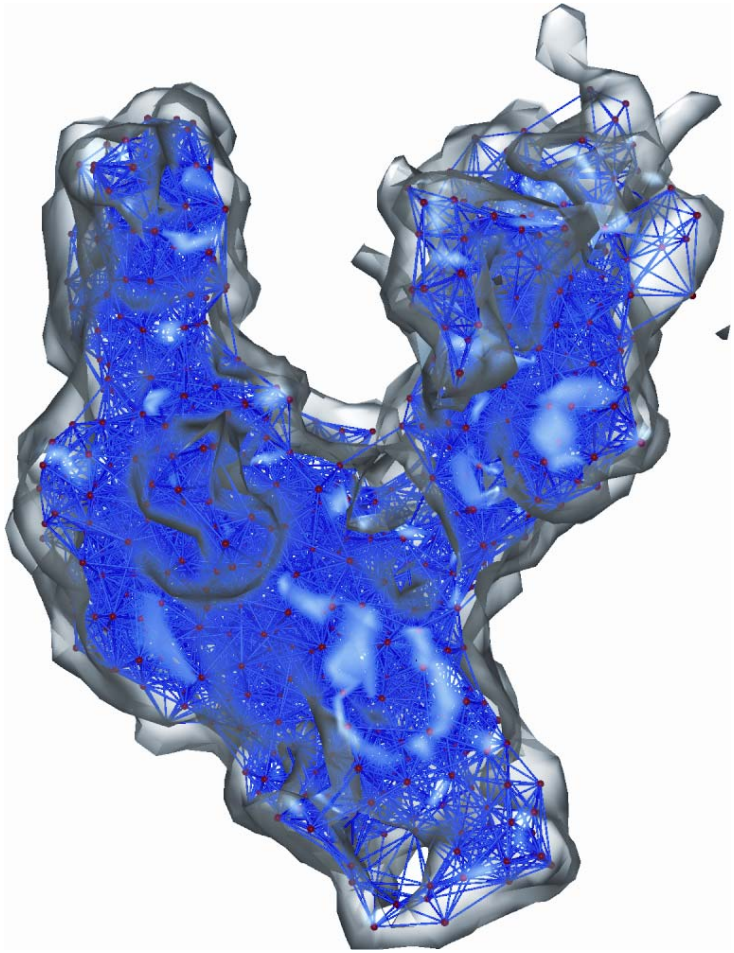
RNA Polymerase, S. Darst et al.

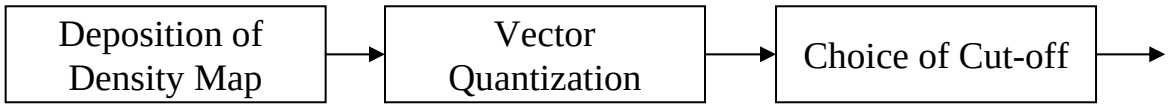




Application to EM Data

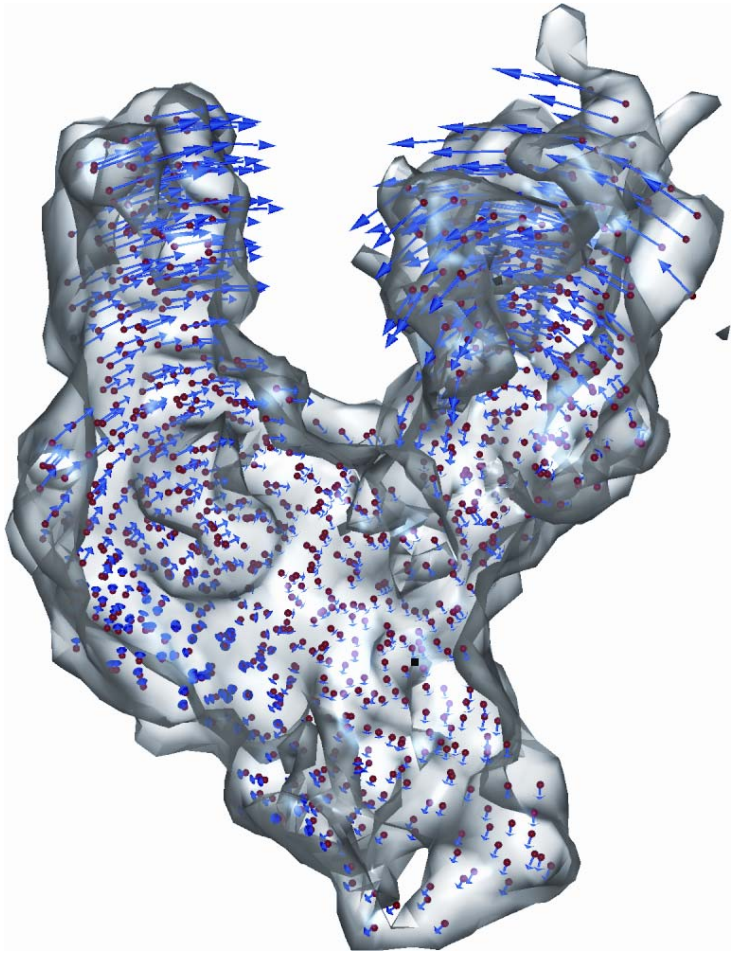
RNA Polymerase, S. Darst et al.

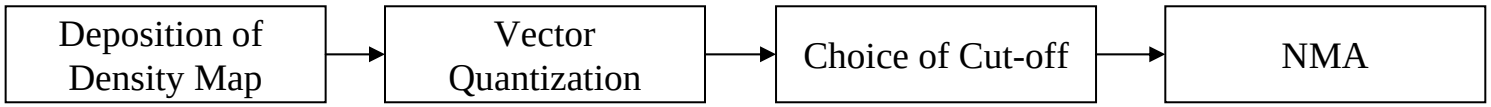




Application to EM Data

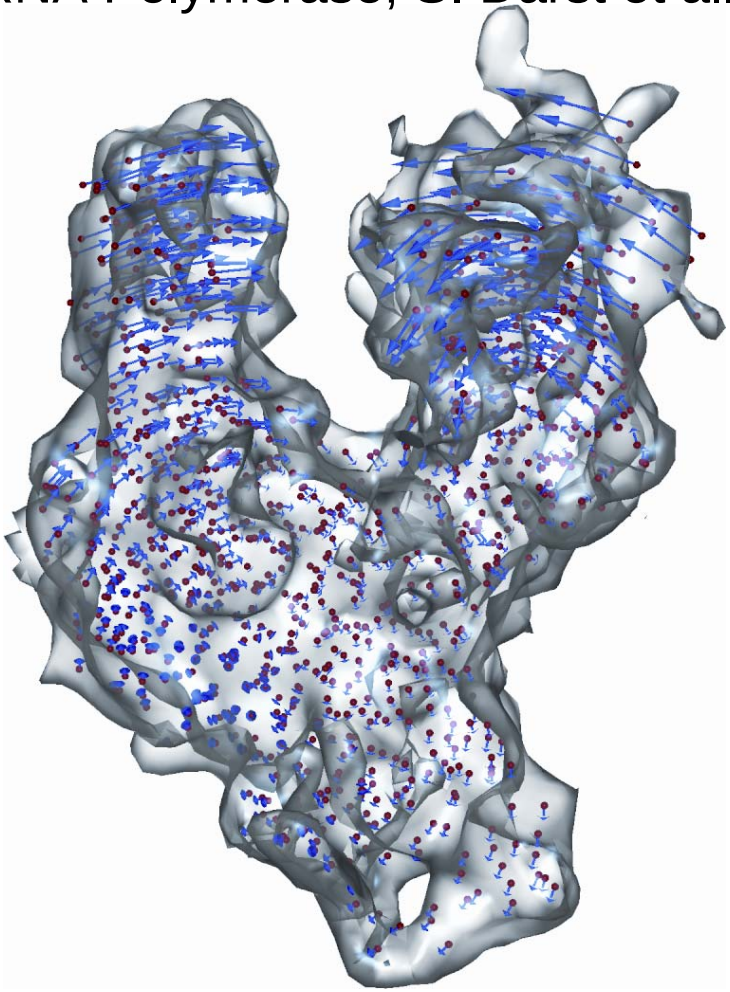
RNA Polymerase, S. Darst et al.





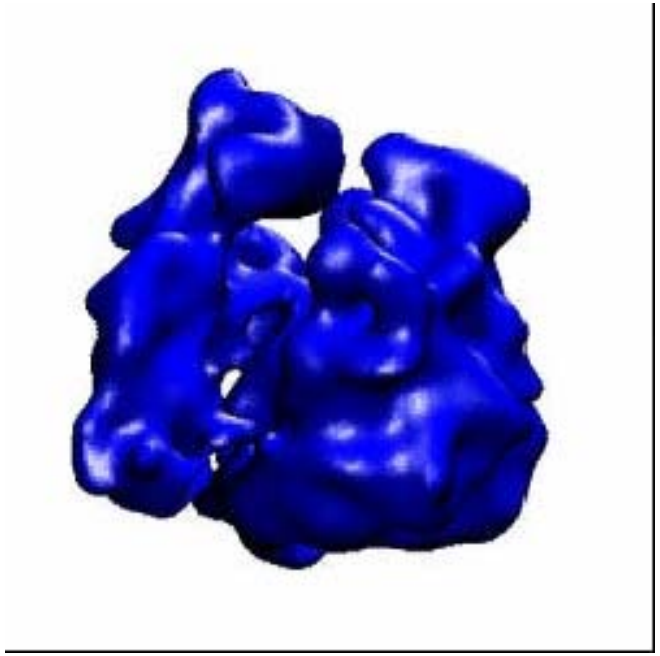
Application to EM Data

RNA Polymerase, S. Darst et al.

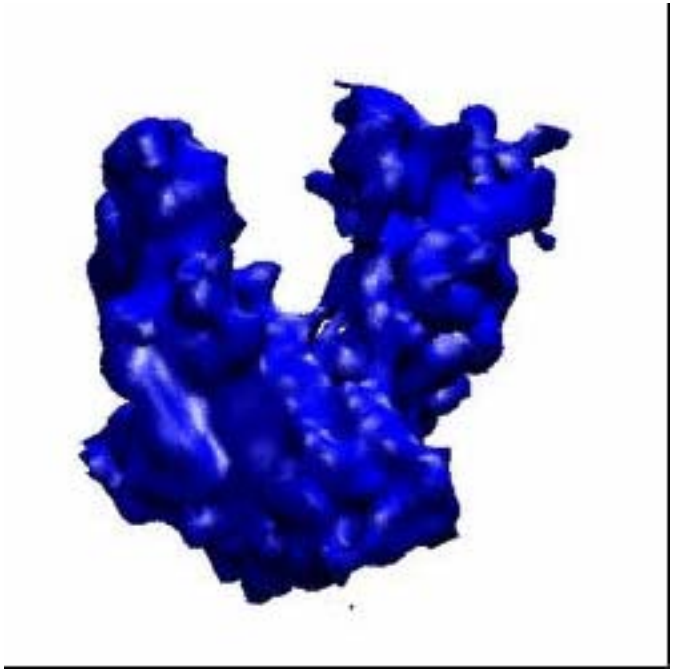


Examples

Ribosome



RNA Polymerase

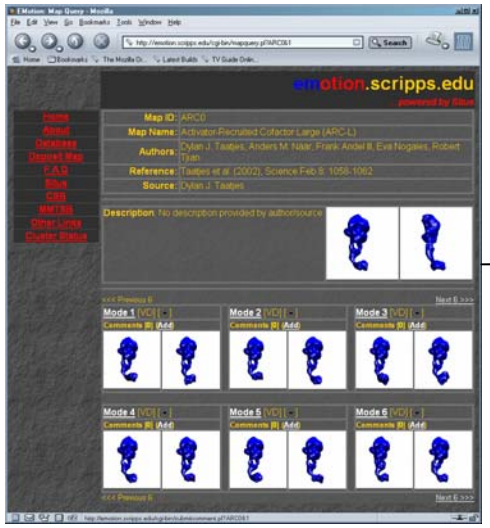


What are the Limitations of NMA (I)?

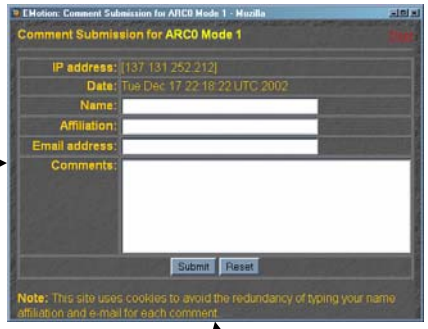
- We do not know *a priori* which is the relevant mode, but the first 12 low-frequency modes are probable candidates.
- The amplitude of the motion is unknown.
- NMA requires additional standards for parameterization, i.e. a screening against complementary experimental data to select the relevant modes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

Solution 1: Annotation of Modes

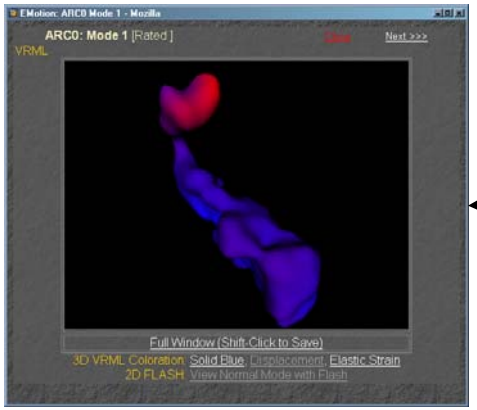
(Essam Metwally: emotion.biomachina.org)



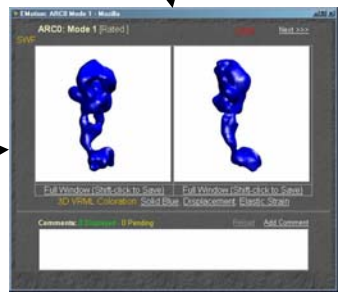
Database Query Results



Annotation/
Comments



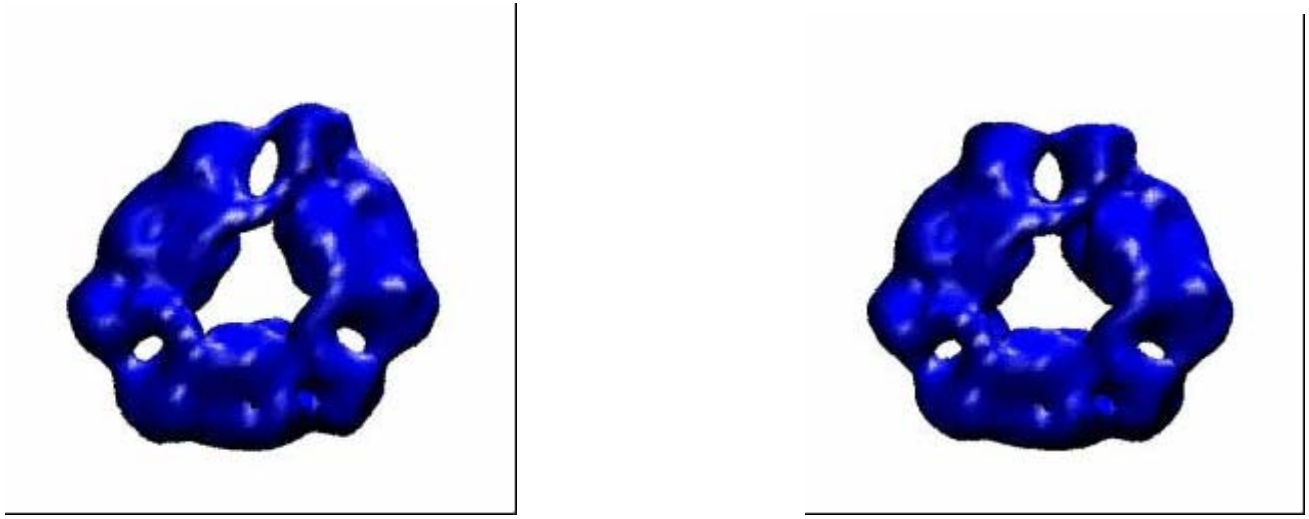
Mode View:
3D Movie
(VRML)



Mode View:
2D Movie &
Comments

What are the Limitations of NMA (II)?

- Normal modes may break the symmetry of structures due to forced orthogonalization:



Global representation → Local description of dynamic feature

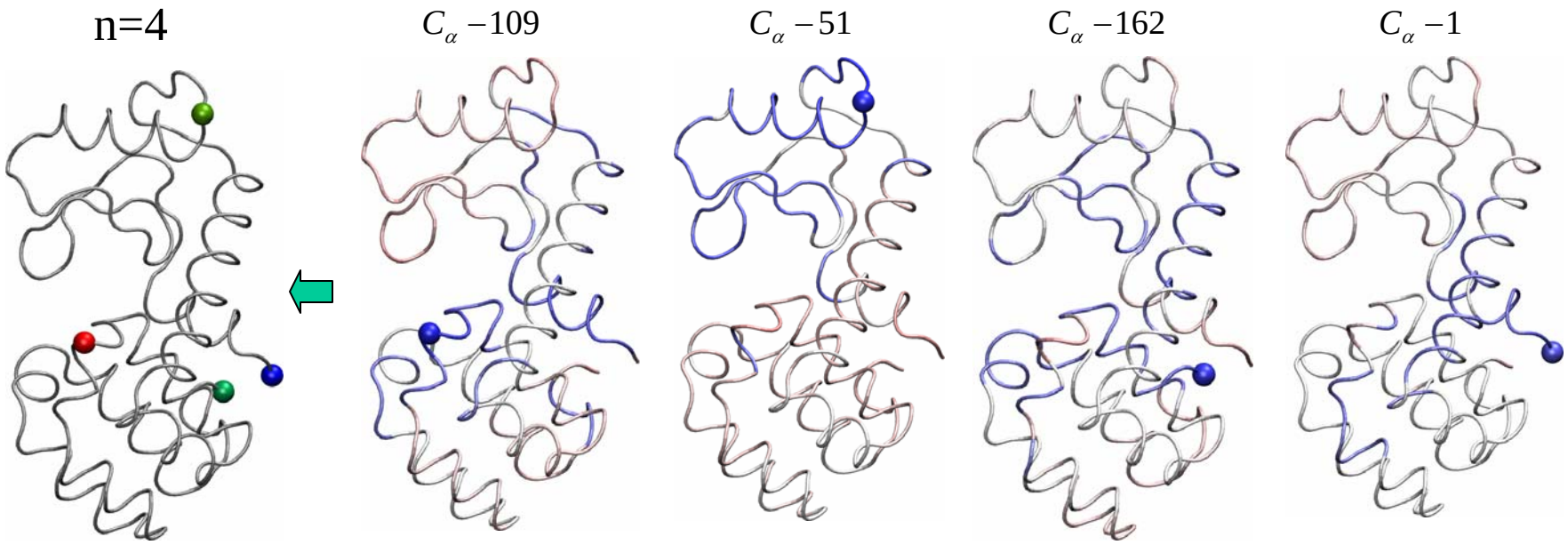
Local Feature Analysis

(Zhiyong Zhang)

Global modes replaced by 3N LFA highly correlated output functions

Sparsification: n “seed atoms” + their neighboring correlated regions

T4 Lysozyme:



Conclusion

- Normal mode analysis is an alternative method to study dynamics of molecules.
- Normal mode analysis does not require trajectory, working with single structure.
- Conformational fluctuation is given by a superposition of normal modes.
- We are using normal mode analysis to refine small-angle X-ray scattering profiles.

Acknowledgement

Dr. Zhiyong Zhang
Dr. Pablo Chacon
Dr. Florence Tama
Dr. Atsushi Matsumoto at Wright-Rieman Laboratory, Rutgers University