Molecular Modeling -- Spring 2014
Lecture 20

Normal Mode Analysis

1. Normal modes
2. Solving for Normal Modes
3. Enzyme catalysis and ligand-induced conformational changes
4. Exercise 21

The knowledge of protein flexibility is then crucial for understanding protein function and evolution.

Manuel Rueda, Pablo Chacó and Modesto Orozco

20.1 Normal modes

20.1 Normal modes

- **modes** = harmonic (sinusoidal) motions.
- **normal** = independent of each other.

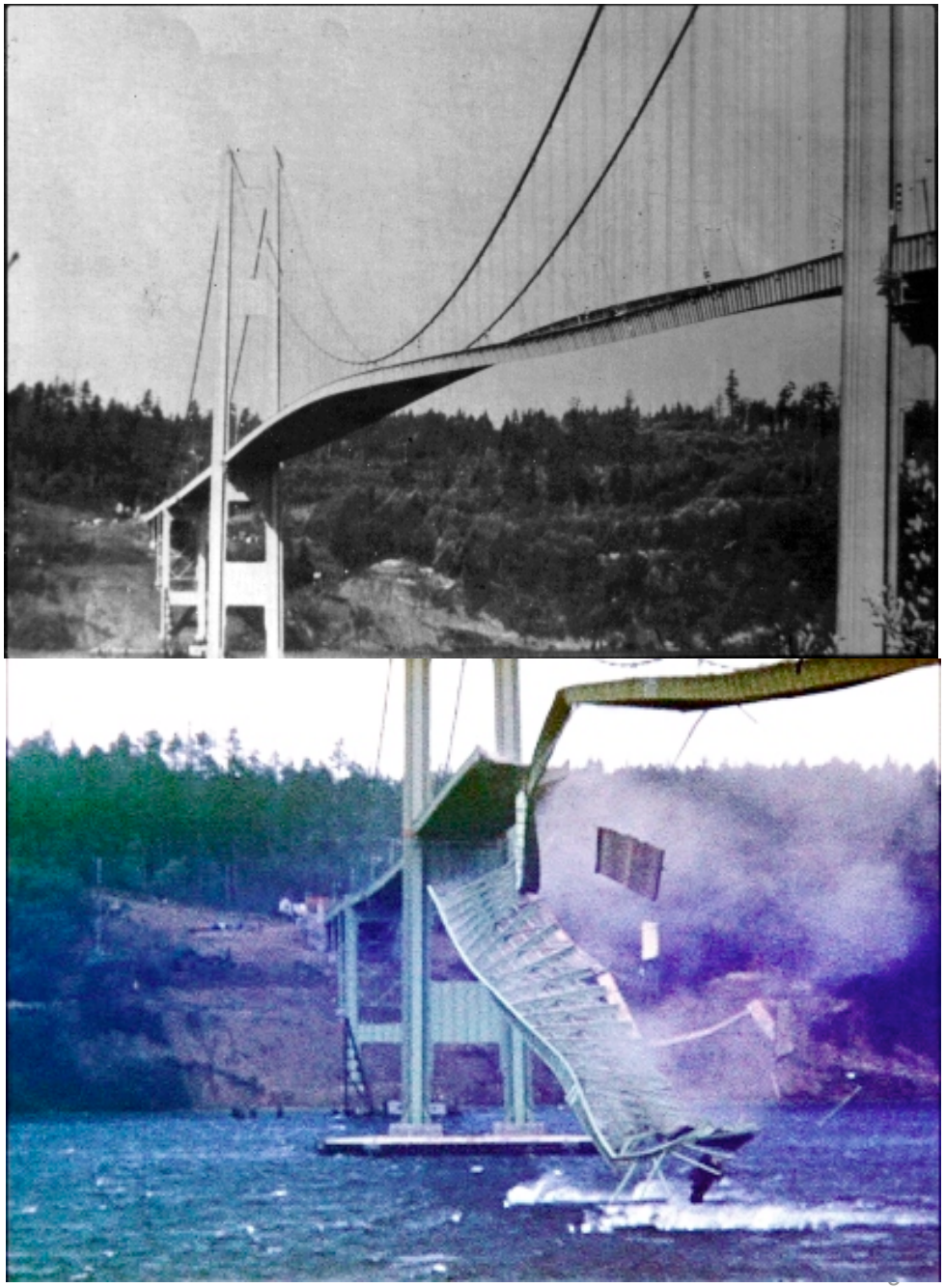
A normal mode is a motion where all parts of the system are moving sinusoidally with the same frequency and in phase.

All observed configurations of a system may be generated from its normal modes.

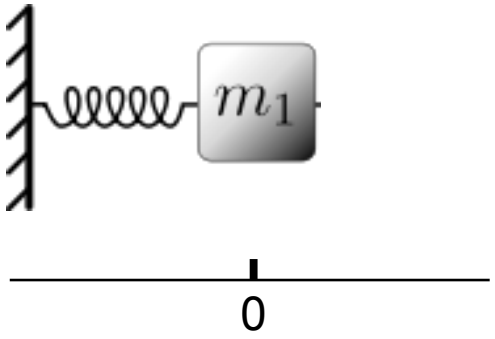
Each normal mode has a characteristic frequency, its eigenvalue.

Normal modes

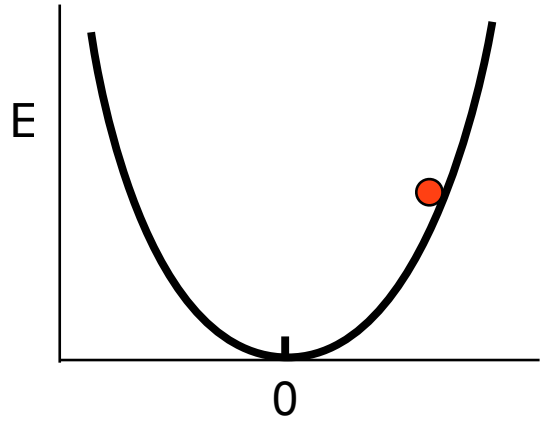
The Tacoma Narrows bridge, opened in 1940, had a normal mode that resonated with a 40 mph wind, until it broke.



Simple harmonic motion



$$\mathbf{F} = -k\mathbf{x},$$



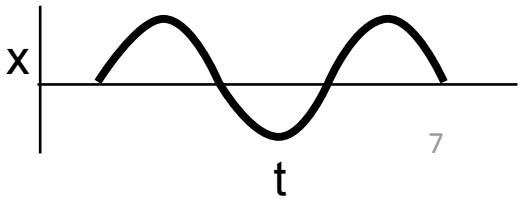
$y=1/2\ x^2$

Imagine a Hooke's Law spring, or a parabolic energy function.
The magnitude of a restoring force is proportional to displacement.
Solving the differential equation

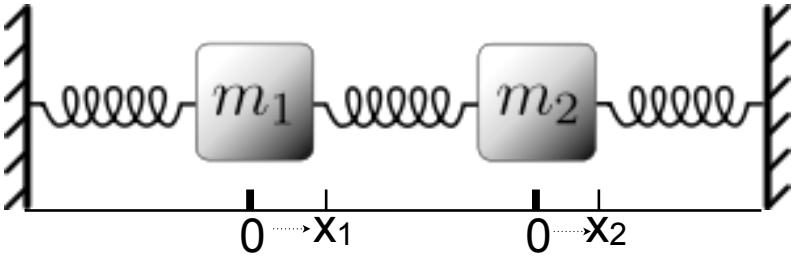
$$F_{net} = m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -kx,$$

gives a harmonic equation for x at time t,

$$x(t) = c_1\cos(\omega t) + c_2\sin(\omega t) = A\cos(\omega t - \varphi),$$



Normal modes for coupled oscillators



If we denote acceleration (the second derivative of $x(t)$ with respect to time) as , the equations of motion are:

$$m\ddot{x}_1 = -kx_1 + k(x_2 - x_1) \qquad m\ddot{x}_2 = -kx_2 + k(x_1 - x_2)$$

We assume oscillatory motion, so:

$$x_1(t) = A_1 e^{i\omega t} \qquad x_2(t) = A_2 e^{i\omega t}$$

Substituting these into the equations of motion gives us:

$$-\omega^2 m A_1 e^{i\omega t} = -2kA_1 e^{i\omega t} + kA_2 e^{i\omega t}$$
$$-\omega^2 m A_2 e^{i\omega t} = kA_1 e^{i\omega t} - 2kA_2 e^{i\omega t}$$

http://en.wikipedia.org/wiki/Normal_mode

Since the exponential factor is common to all terms, we omit it and simplify:

$$(\omega^2 m - 2k)A_1 + kA_2 = 0 \qquad kA_1 + (\omega^2 m - 2k)A_2 = 0$$

And in matrix representation:

$$\begin{bmatrix} \omega^2 m - 2k & k \\ k & \omega^2 m - 2k \end{bmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = 0$$

For this equation to have a non-trivial solution, the matrix on the left must be singular i.e. must not be invertible, such that one cannot multiply both sides of the equation by the inverse, leaving the right matrix equal to zero. It follows that the determinant of the matrix must be equal to 0, so:

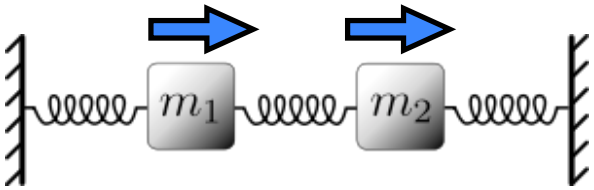
$$(\omega^2 m - 2k)^2 - k^2 = 0$$

http://en.wikipedia.org/wiki/Normal_mode

We have two solutions:

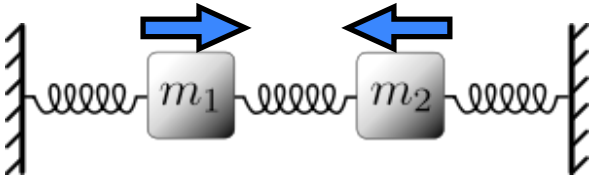
$$\omega_1 = \sqrt{\frac{k}{m}}, \qquad \omega_2 = \sqrt{\frac{3k}{m}}.$$

If we substitute ω_1 into the matrix and solve for (A_1, A_2), we get (1, 1). If we substitute ω_2, we get (1, −1). (These vectors are eigenvectors, and the frequencies ω are eigenvalues.)
The first normal mode is: (1, 1)



ω_1, low frequency mode

Which corresponds to both masses moving in the same direction at the same time.
The second normal mode is: (1, −1)



ω_2, high frequency mode

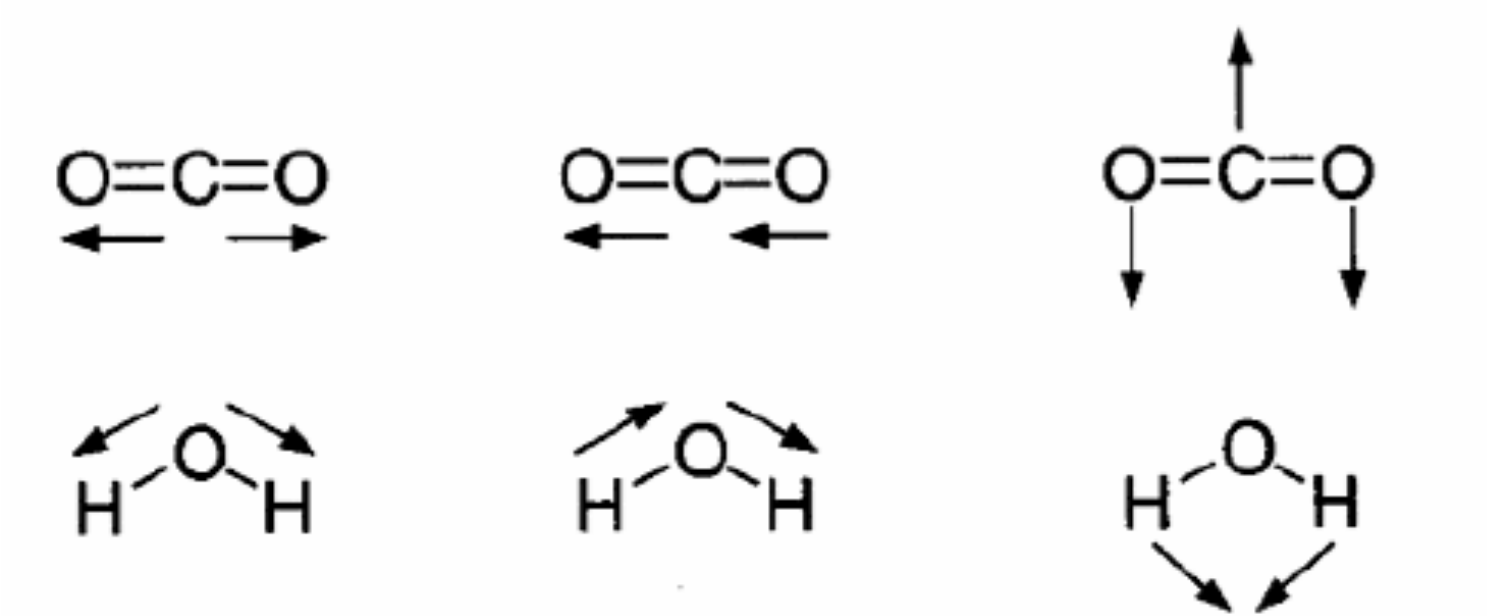
This corresponds to the masses moving in the opposite directions, while the center of mass remains stationary.

http://en.wikipedia.org/wiki/Normal_mode

Characteristics of normal mode motion

- Each normal mode acts like a simple harmonic oscillator.
- A normal mode is concerted motion of many atoms.
- Center of the mass doesn't move.
- All atom pass through their equilibrium positions at the same time.
- Normal modes are orthogonal to each other; they resonate independently.
- Directly related to vibrational spectroscopy.

Normal modes of small molecules



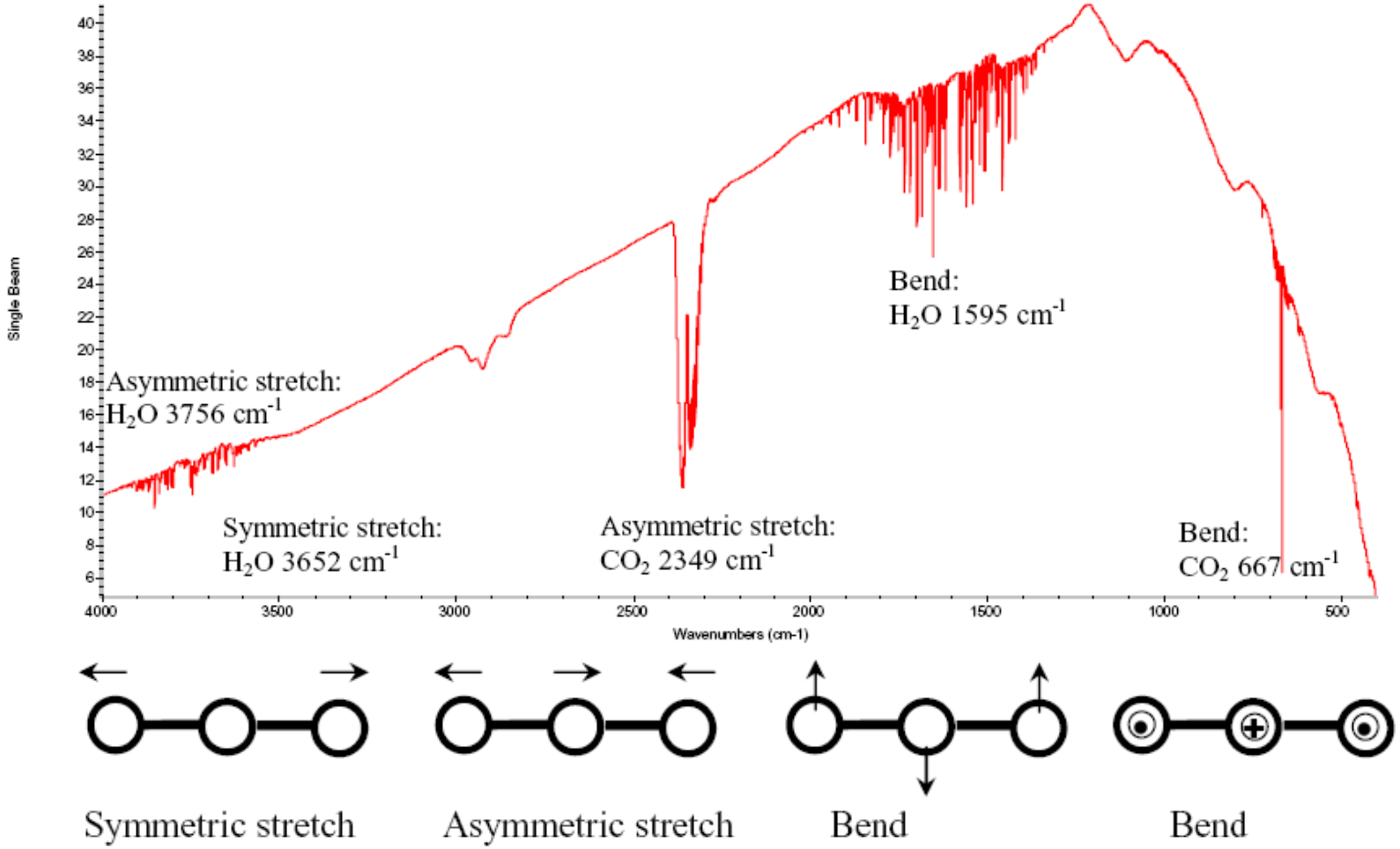
Symmetric stretch

Asymmetric stretch

Bending

Infrared spectrum of air

Normal modes resonate with light at the frequency of oscillation.



Wave number

Instead of frequency in spectroscopy in general wave numbers are used.
Wave numbers are defined by $k = 1/\lambda$ and are measured In units of cm^{-1}.

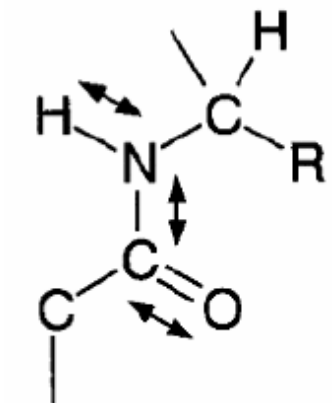
So, wave length is converted

from 500 nm to 20 000 cm^{-1}

from 1 μm to 10 000 cm^{-1}

or 1 mm to 10 cm^{-1}

Stretching modes of the protein backbone



Secondary structure defines the steric and H-bonding interactions, which change the energy of the bond. Steric clashes stretch bonds. H-bonds polarize and weaken the covalent bond.

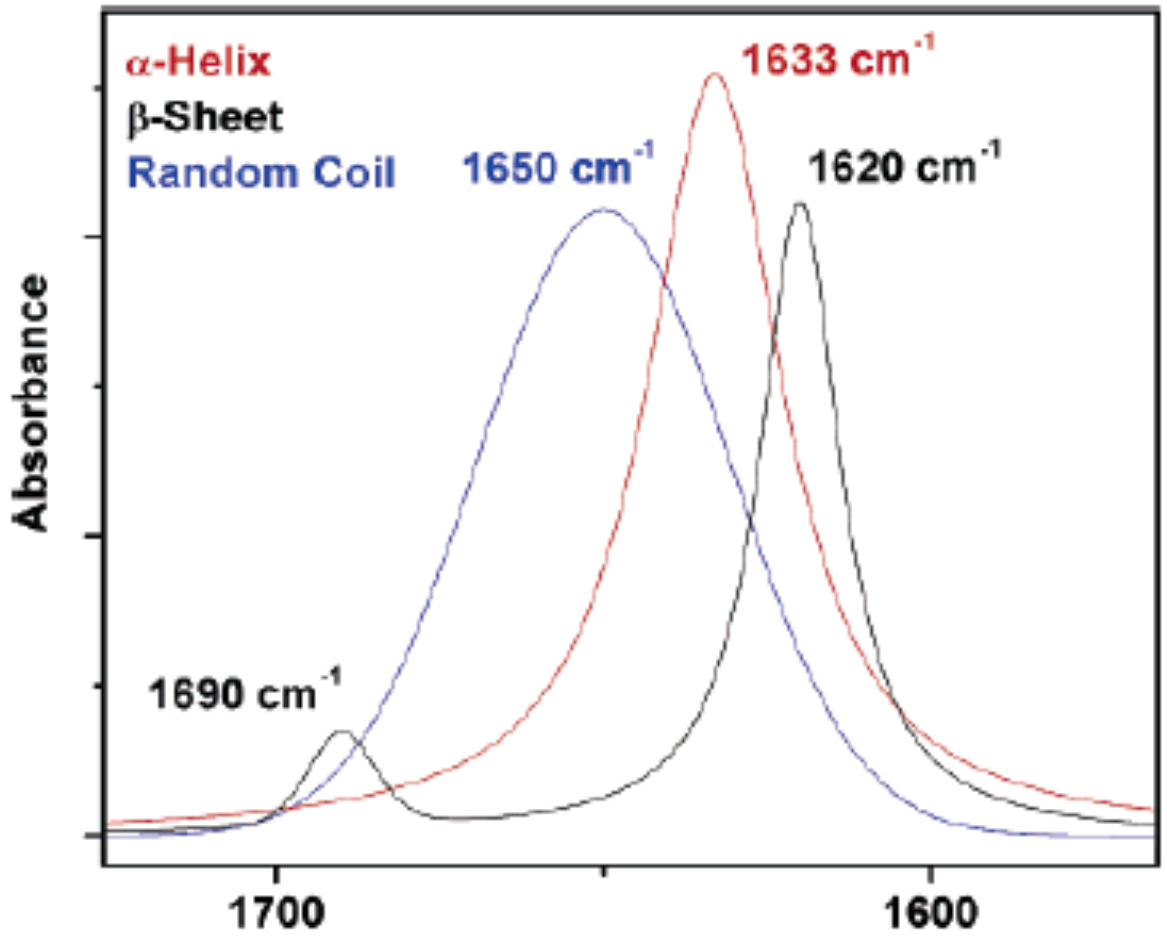
Table 8-5

Characteristics of principal infrared absorption bands of the peptide group

Vibration	$\partial\underset{\sim}{\mu}/\partial R$	Hydrogen-bonded forms: α Helix Frequency (cm^{-1})	α Helix Dichroism	Hydrogen-bonded forms: β Sheet Frequency (cm^{-1})	β Sheet Dichroism	Non-hydrogen-bonded Frequency (cm^{-1})
N—H stretch	←N—H→ ↔	3,290–3,300	∥	3,280–3,300	⊥	~3,400
Amide I (C═O stretch)	←C═O→ ↔	1,650–1,660	∥	1,630	⊥	1,680–1,700
Amide II	↑ H ←C—N→ ↙ ↓	1,540–1,550	⊥	1,520–1,525	∥	<1,520?

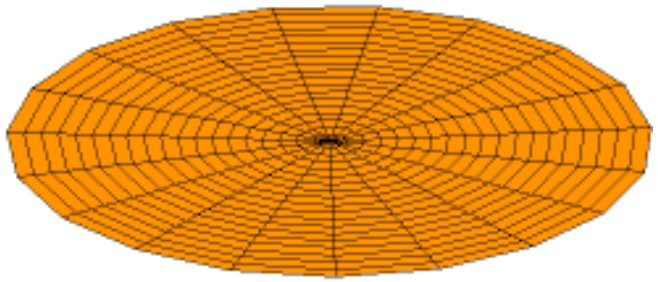
SOURCE: Adapted from J. A. Schellman and C. Schellman, in *The Proteins*, 2d ed., vol. 2, ed. H. Neurath (New York: Academic Press, 1962), p. 1.

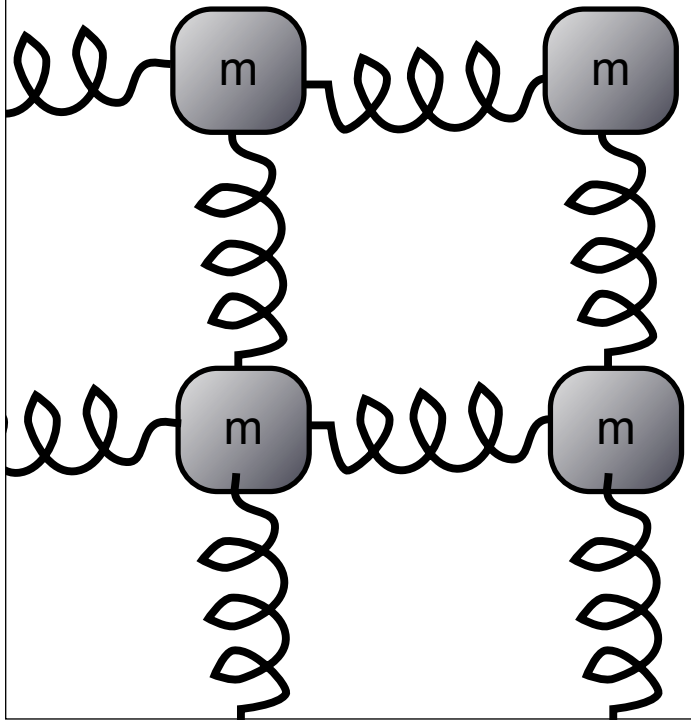
Secondary Structure and Amide I frequency



20.2 Solving for Normal modes

Normal modes for a network of springs





Normal mode analysis assumes that all pairwise forces behave like Hooke's Law springs for short displacements. In other words, all "atoms" (or points on a drum, or on a bridge, or...) are sitting in a "harmonic well", a parabola.

All points move is synchrony, in phase.

Normal modes are eigenvectors of the Hessian matrix.

- NM are fluctuations around the ground state.
- NM atom motions are approximated as straight lines.
- Anywhere on that line, the sum of all forces on an atom drives it to another point *on the same line*.
- Force in one dimension may be expressed as the second derivative of the energy times the displacement in that direction.
- Forces in multiple directions i may be expressed as the sum of the second partial derivatives for all pairs of interactions (i,j), times the displacement in direction j.

Hessian matrix

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

If there were only two atoms, the force vector on the first atom x_1 would be

$$F_1 = \nabla_1 E = \frac{\partial E}{\partial x_1} = \frac{\partial^2 E}{\partial x_1\,\partial x_1}\Delta x_1 + \frac{\partial^2 E}{\partial x_1\,\partial x_2}\Delta x_2$$

http://en.wikipedia.org/wiki/Hessian_matrix

Hessian matrix

Two atom case, both atoms

$$F_1 = \nabla_1 E = \frac{\partial E}{\partial x_1} = \frac{\partial^2 E}{\partial x_1 \partial x_1}\Delta x_1 + \frac{\partial^2 E}{\partial x_1 \partial x_2}\Delta x_2$$

$$F_2 = \nabla_2 E = \frac{\partial E}{\partial x_2} = \frac{\partial^2 E}{\partial x_2 \partial x_2}\Delta x_2 + \frac{\partial^2 E}{\partial x_2 \partial x_1}\Delta x_1$$

Note, $\frac{\partial^2 E}{\partial x_1 \partial x_2} = \frac{\partial^2 E}{\partial x_2 \partial x_1}$

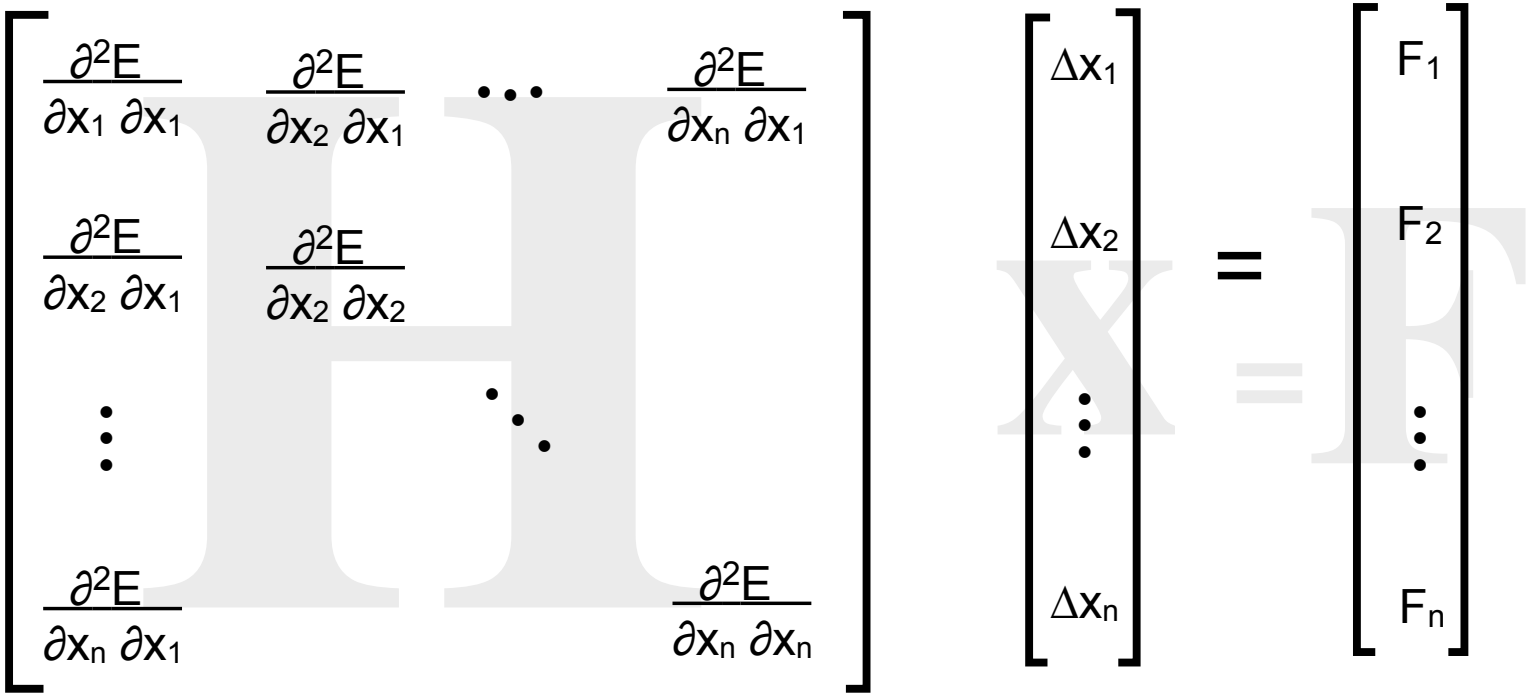
In matrix notation,

$$\begin{bmatrix} \frac{\partial^2 E}{\partial x_1 \partial x_1} & \frac{\partial^2 E}{\partial x_2 \partial x_1} \\ \frac{\partial^2 E}{\partial x_2 \partial x_1} & \frac{\partial^2 E}{\partial x_2 \partial x_2} \end{bmatrix} \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}$$

http://en.wikipedia.org/wiki/Hessian_matrix

Hessian matrix

Many atom case



So the force vectors **F** as a function of the vector of displacement vectors **x** is

$$\mathbf{F = H \cdot x}$$

Normal modes are eigenvectors of the Hessian matrix.

$$\mathbf{H} \cdot \mathbf{x} = \mathbf{F}$$

Normal modes are the special case where the forces line up with the displacements. In other words,

$$\mathbf{H} \cdot \mathbf{x} = \lambda \mathbf{x}$$

where λ is a scaler. This means **x** is an eigenvector by definition. To solve for x, first we combine terms.

$$(\mathbf{H} - \lambda \mathbf{I}) \cdot \mathbf{x} = 0$$

where **I** is the identity matrix.

Since we are looking for non-trivial solutions, the matrix (H-λI) *cannot be invertible*, therefore it must be *singular*, therefore the *determinant is zero*.

Solving the eigenvector problem

$$\det(\mathbf{H}-\lambda\mathbf{I}) = 0$$

Offered without proof: There are n solutions to the equation above, where n is the rank of H. Each solution has a vector **x** and a scaler value λ.
λ is the eigenvalue associate with each eigenvector **x**.

Each eigenvector **x** consists of a displacement of each atom in each direction (x,y,z).

$$H \cdot x_i = \lambda_i x_i, \quad i = 1,\ldots,3N$$

where N = number of atoms.

Numerical second derivatives

A closed form for the second derivative is usually not available, so they are calculated numerically. The gradient of the energy with respect to *i* is evaluated after displacement of atom *j* in each direction (x,y,z). The difference between the gradients divided by the displacement in *j* is the second derivative.

$$H_{i,j} = \frac{\left(\frac{\delta E}{\delta x_i}\right)_{+0.5\Delta x_j} - \left(\frac{\delta E}{\delta x_i}\right)_{-0.5\Delta x_j}}{\Delta x_j}.$$

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

The difference between the slopes is (approx.) the 2nd deriv.

Properties of eigenvalues

The product of the eigenvalues is equal to the determinant of **A**

$$\det(\mathbf{A}) = \prod_{i=1}^{N_\lambda} \lambda_i^{n_i}$$

The sum of the eigenvalues is equal to the trace of **A**

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{N_\lambda} n_i \lambda_i$$

Eigenvectors are real, and mutually orthogonal

Eigenvectors provide a basis for $\mathbb{R}^N$

General procedure for normal mode analysis in a protein

- Energy minimization
- Calculate Hessian
- Divide by mass matrix (to get accelerations, because F=ma)
- "Diagonalize", solve for eigenvectors.
  - Eigenvalues: high eigenvalue mean low force constant, low frequency.
    - Low frequencies are important for proteins
  - Eigenvectors: directions of movement: visualize

Successive approximation method

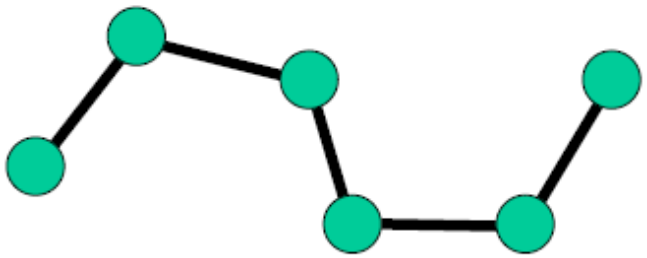
- Since for all normal modes $\mathbf{x}$, $\mathbf{H}\cdot\mathbf{x} = \lambda\mathbf{x}$:
  - Randomly initialize a vector of displacements $\mathbf{y}$
  - Multiply $\mathbf{y}$ by $\mathbf{H}$ to get $\mathbf{y'}$
  - Rescale $\mathbf{y'}$
  - Multiply $\mathbf{y'}$ by $\mathbf{H}$ to get $\mathbf{y''}$
  - Rescale $\mathbf{y''}$
  - Etc. Stop when it converges. The result is the first normal mode $\mathbf{x}$.
  - Project matrix $\mathbf{H}$ onto vector $\mathbf{x}$, reducing rank of $\mathbf{H}$ by 1.
  - Repeat process to get normal mode 2.

- Here is a tool being developed by a PhD student Osman Burak Okan, RPI Materials Engineering.

  http://midst.sabanciuniv.edu/anm/

- Other NMA tools:
  - webnm http://apps.cbu.uib.no/webnma
  - elNemo http://www.igs.cnrs-mrs.fr/elnemo/start.html

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

Level of Detail not Important

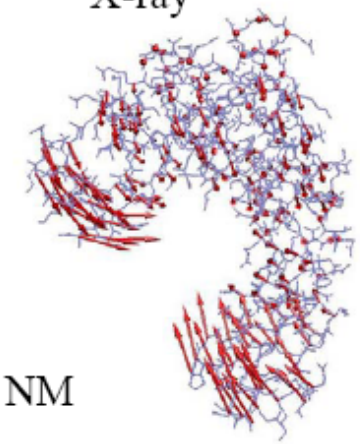
Projection onto atomic normal modes
≈ 1 for the first few modes

Low frequency NM are similar to atomic NM

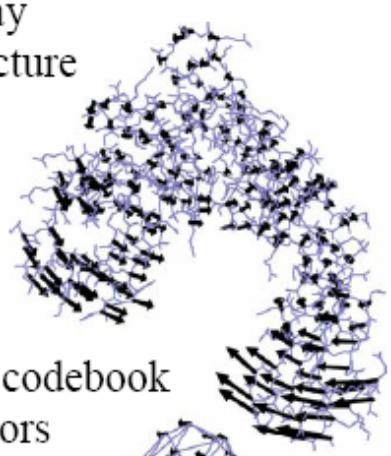
Models can reproduce functional rearrangements even at 30Å resolution

Open-Close

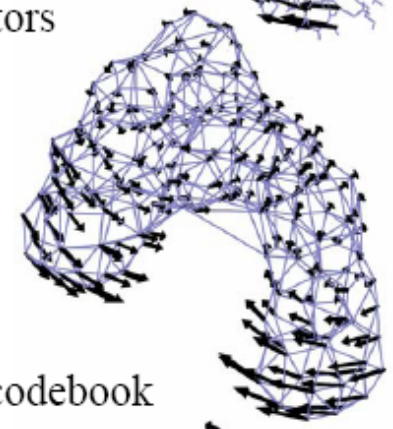
X-ray



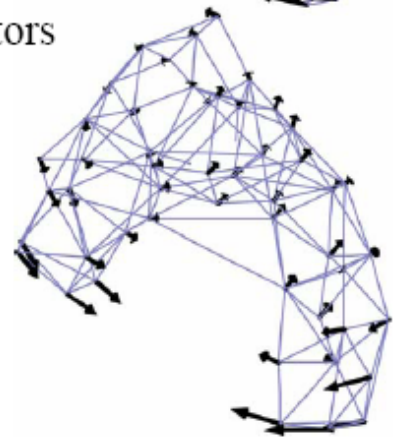
X-ray
structure



214 codebook
vectors

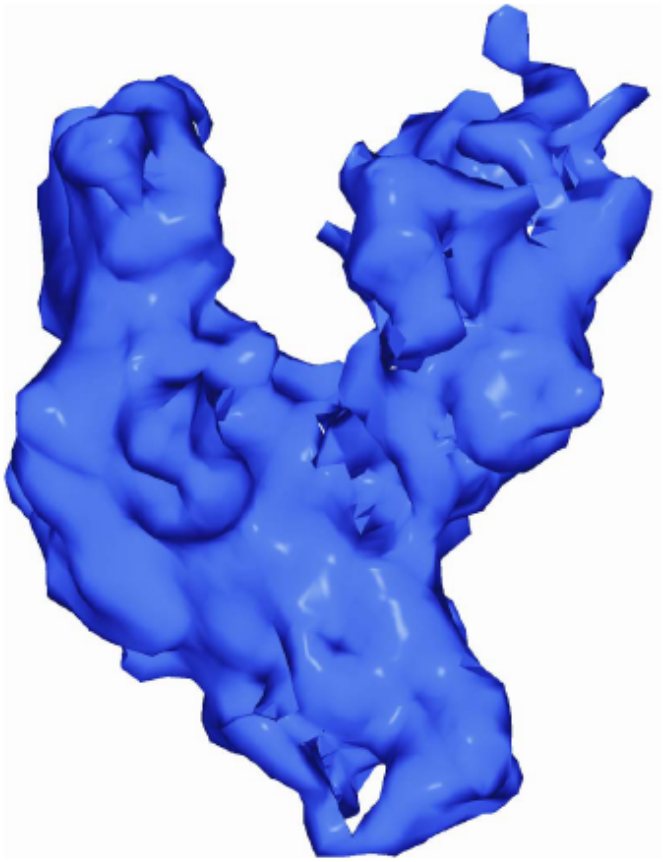


50 codebook
vectors



Application to EM Data

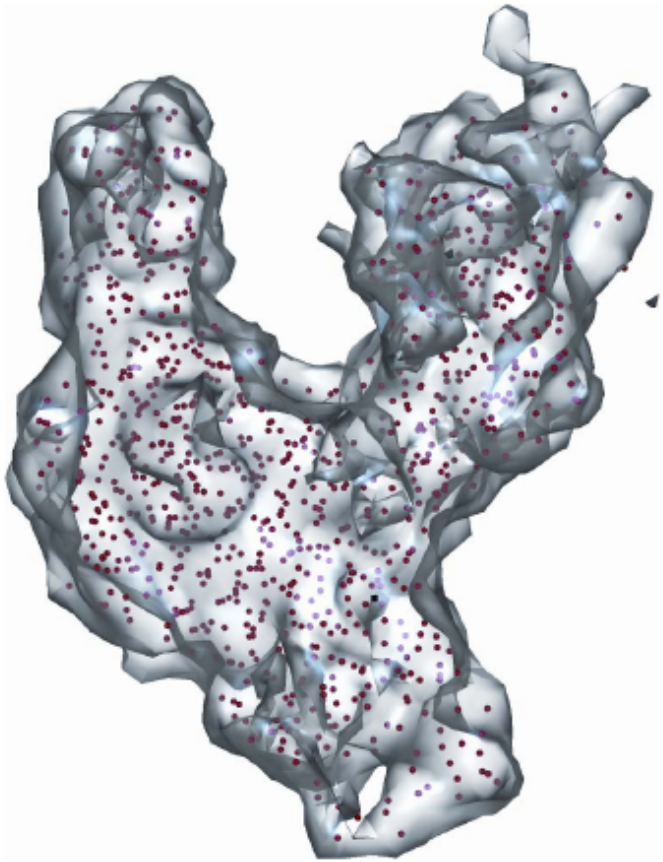
RNA Polymerase, S. Darst et al.



Deposition of Density Map →

Application to EM Data

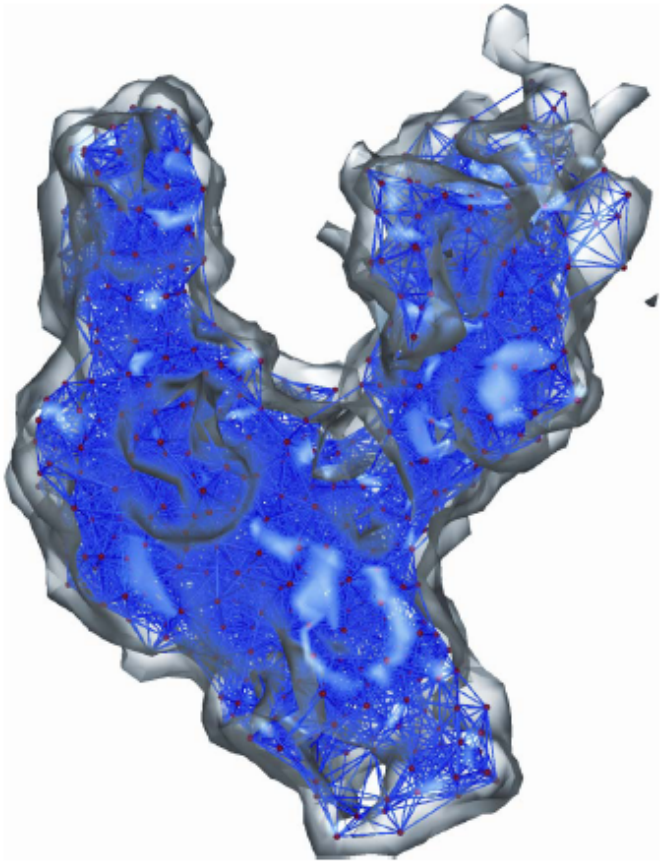
RNA Polymerase, S. Darst et al.

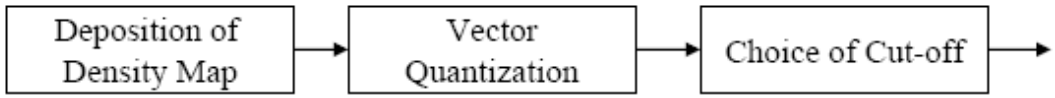


Deposition of Density Map → Vector Quantization

Application to EM Data

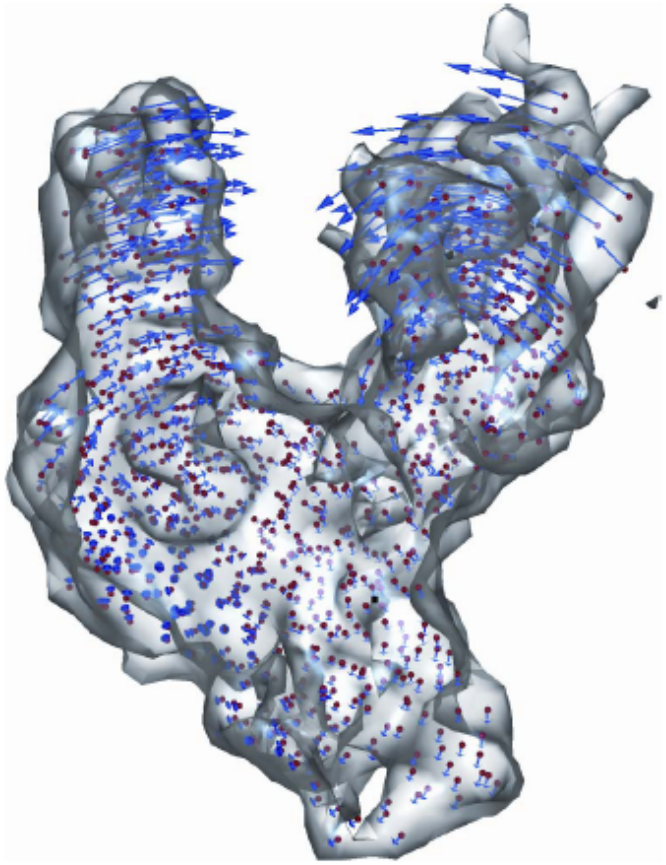
RNA Polymerase, S. Darst et al.

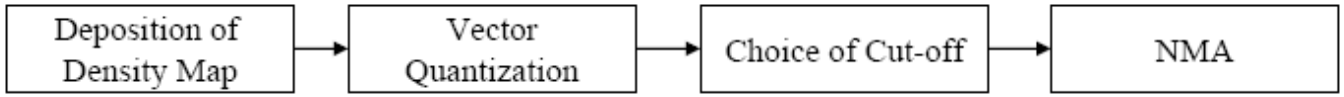




Application to EM Data

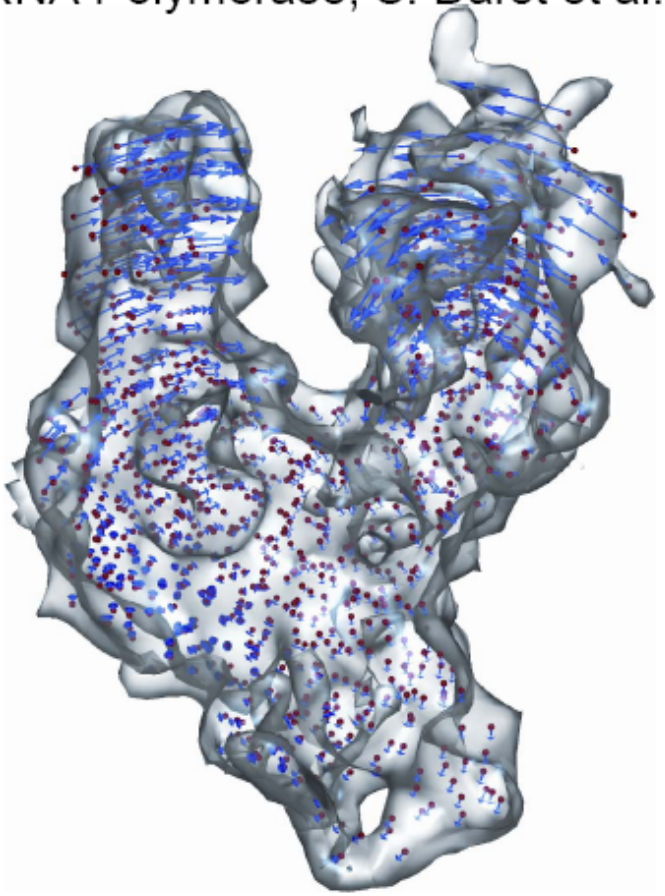
RNA Polymerase, S. Darst et al.





Application to EM Data

RNA Polymerase, S. Darst et al.



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

What are the Limitations of NMA (I)?

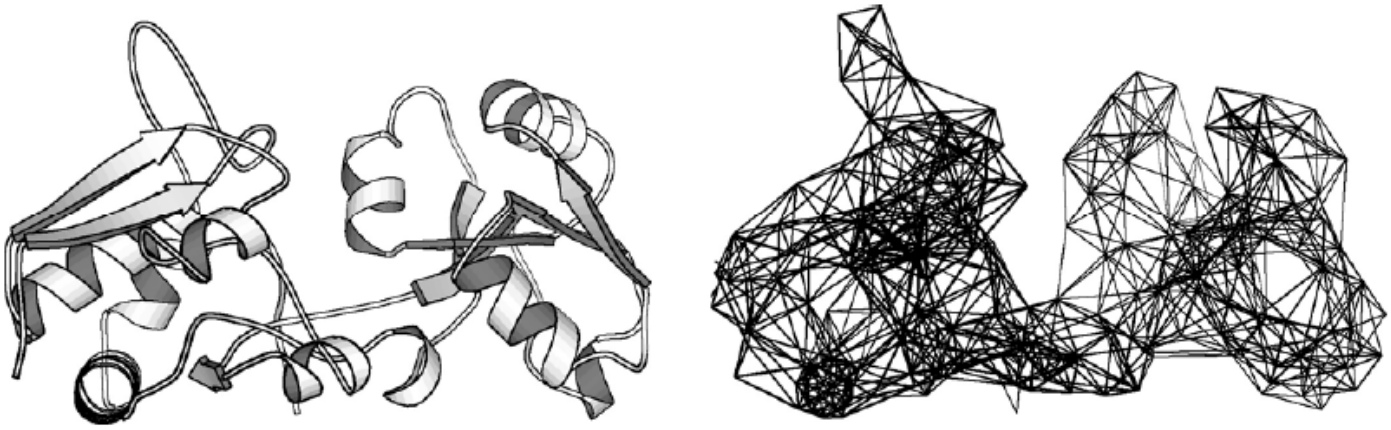
- We do not know *a priori* which is the relevant mode, but the first 12 low-frequency modes are probable candidates.
- The amplitude of the motion is unknown.
- NMA requires additional standards for parameterization, i.e. a screening against complementary experimental data to select the relevant modes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

Elastic Network Model

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

Simplified force-field: no MM, already minimized

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



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

Principle Components Analysis (PCA)

- Similar to NMA, but the Hessian matrix is replaced by a covariance matrix.
- Covariances are calculated from superimposed coordinates from a MD simulation.
- PCA compresses and summarizes MD.

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.

20.3 Enzyme catalysis and conformational changes

A story of how conformational changes in an enzyme contribute to enzyme catalysis.

E=dihydrofolate reductase=DHFR

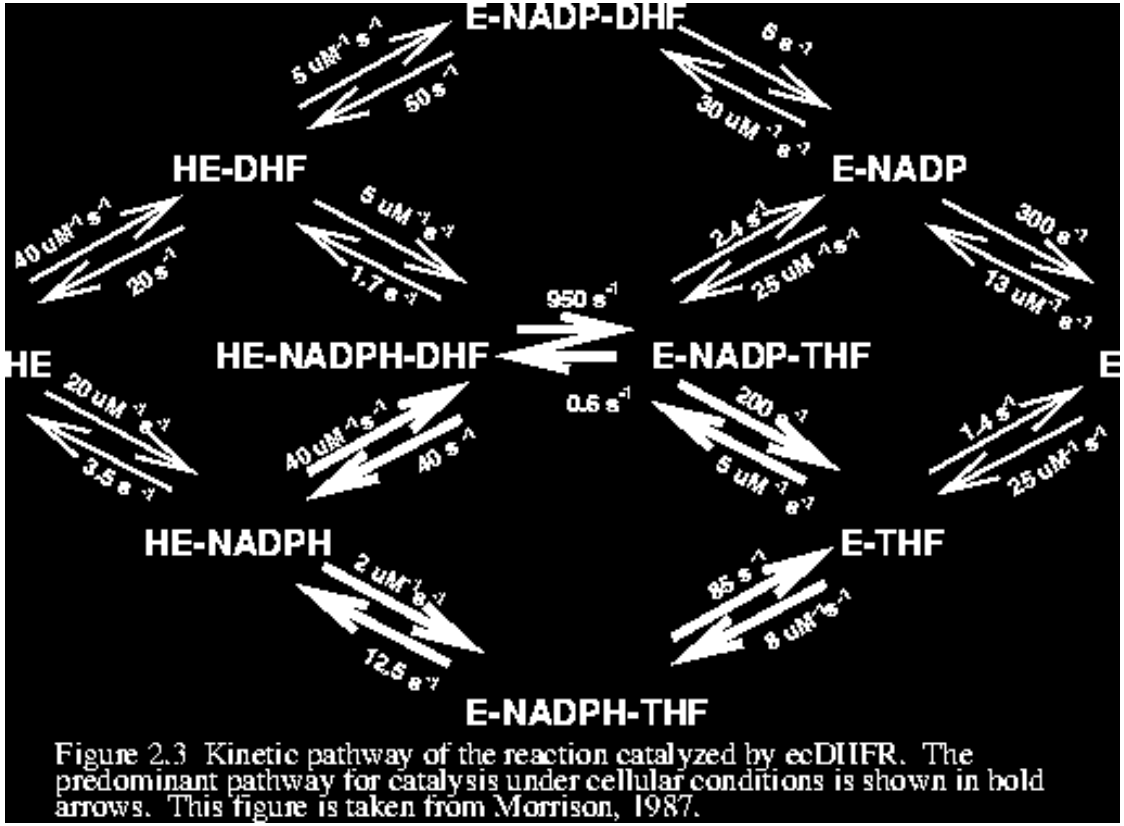


Figure 2.3 Kinetic pathway of the reaction catalyzed by ecDHFR. The predominant pathway for catalysis under cellular conditions is shown in bold arrows. This figure is taken from Morrison, 1987.

Watch the DHFR movie http://chem-faculty.ucsd.edu/kraut/dhfr.html

Bystroff C & Kraut J. (1991). Crystal structure of unliganded Escherichia coli dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding. *Biochemistry* 30, 2227-39.

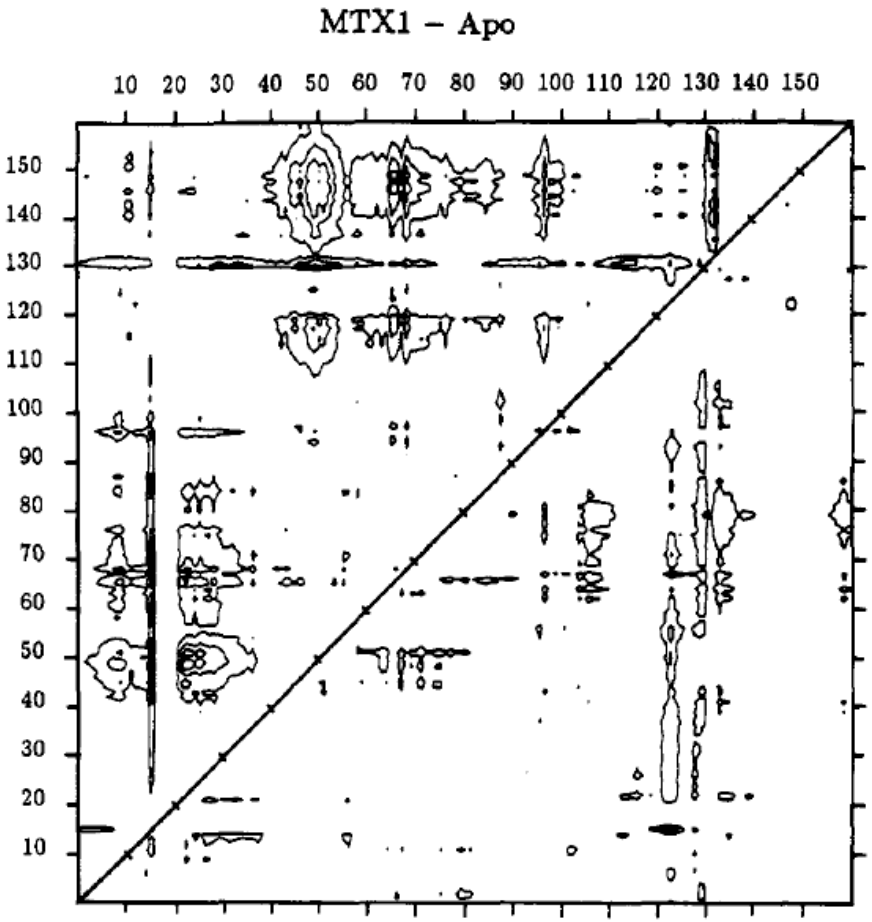


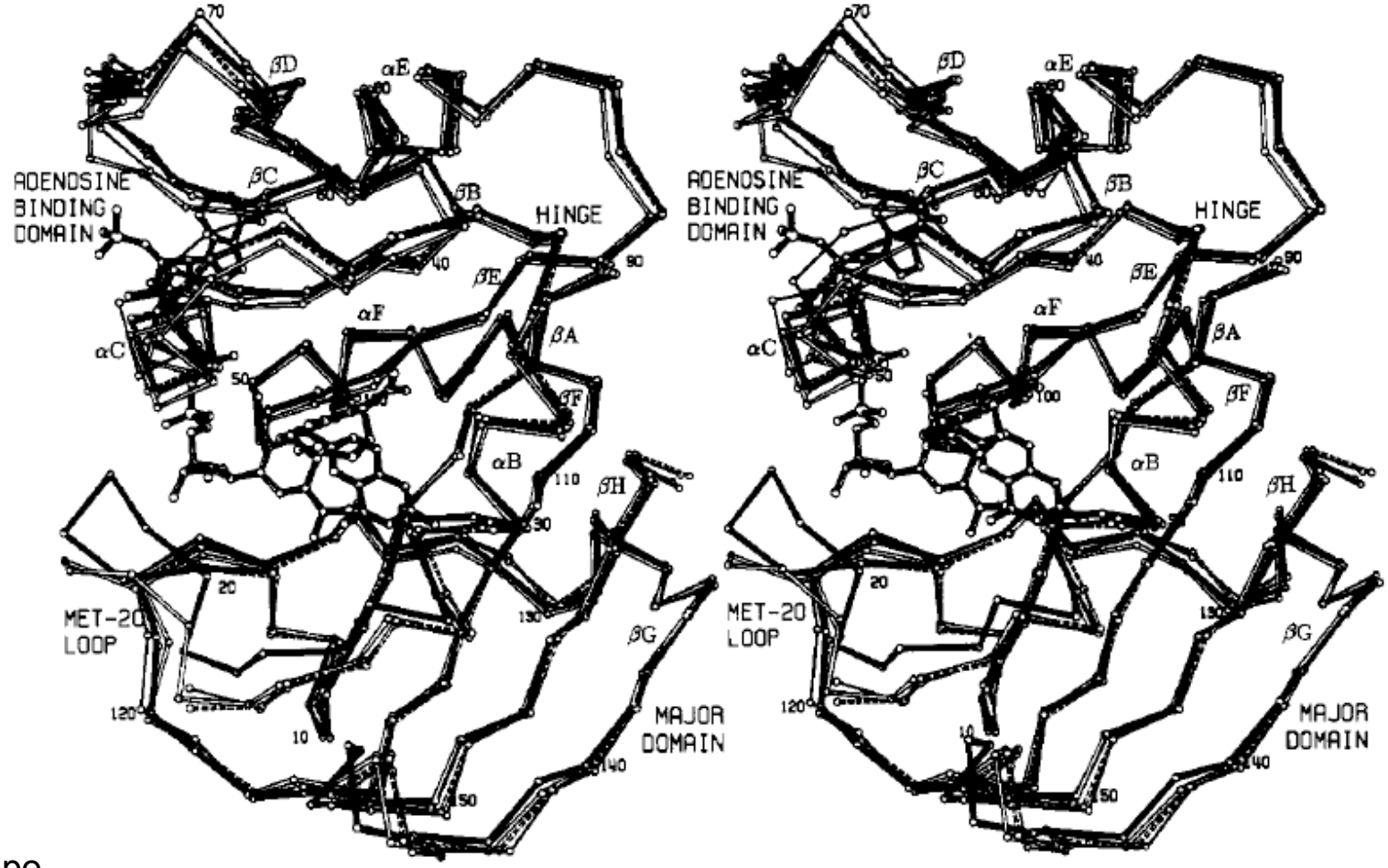
FIGURE 1: Example of distance difference (ΔD) plot for comparison between α-carbon positions of two *E. coli* DHFR crystal structures, in this case the MTX binary complex minus apoenzyme. Positive values are contoured at 0.5, 1.5, and 2.25 Å in the lower right triangle. Negative values are contoured at 0.5, 1,5, and 2.25 Å in the upper left triangle. Typical features of these plots are described in the text.

Crystal structures were solved for DHFR in 3 liganded states, and unliganded (apo).

The difference distance matrix shows two domains. The N and C terminal segments of the chain move together relative to the middle domain (the domain that binds the adenosine moiety of NADPH).

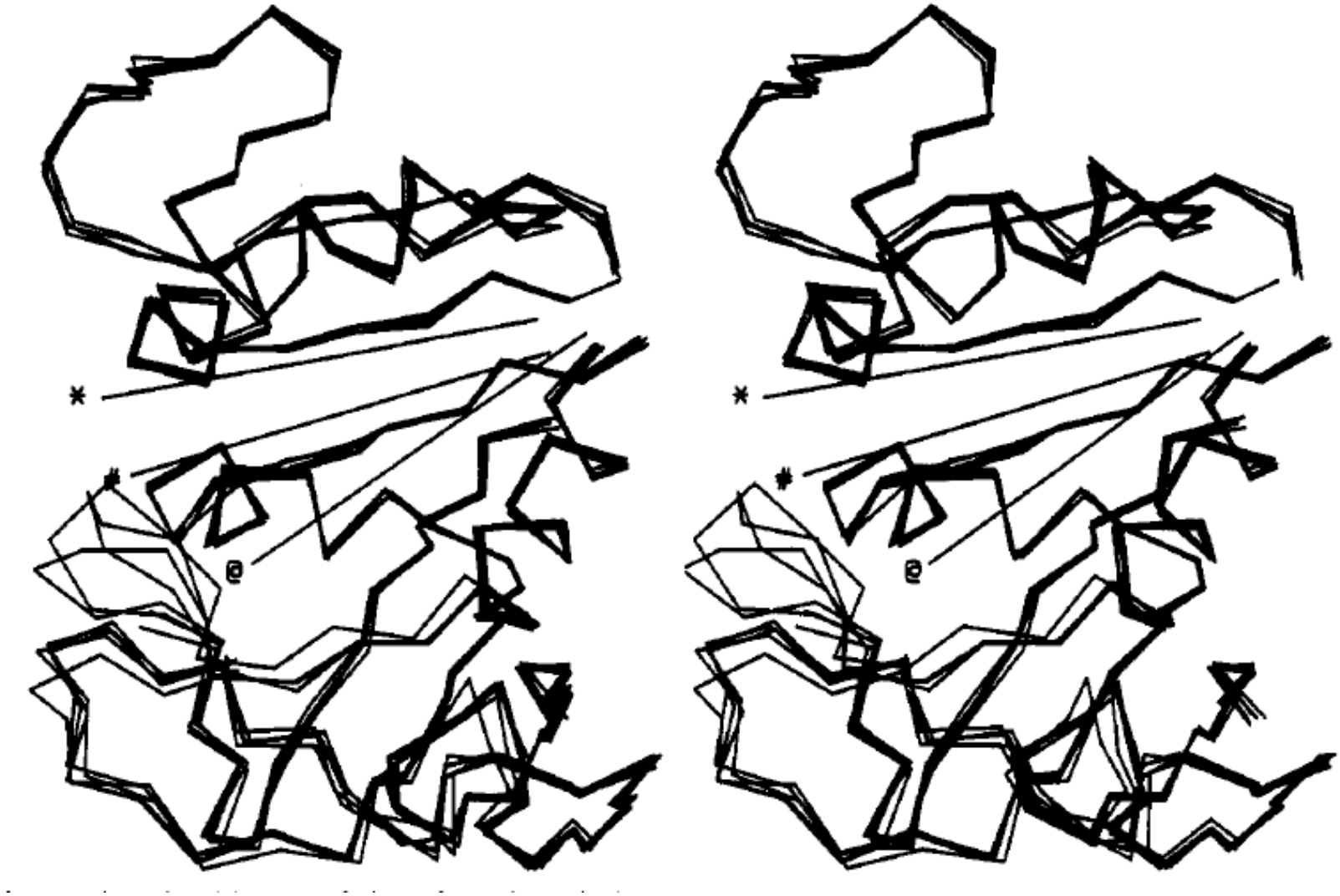
The shift is caused by MTX binding.

Superposed backbone traces for five bound states of DHFR, showing domain rotation.

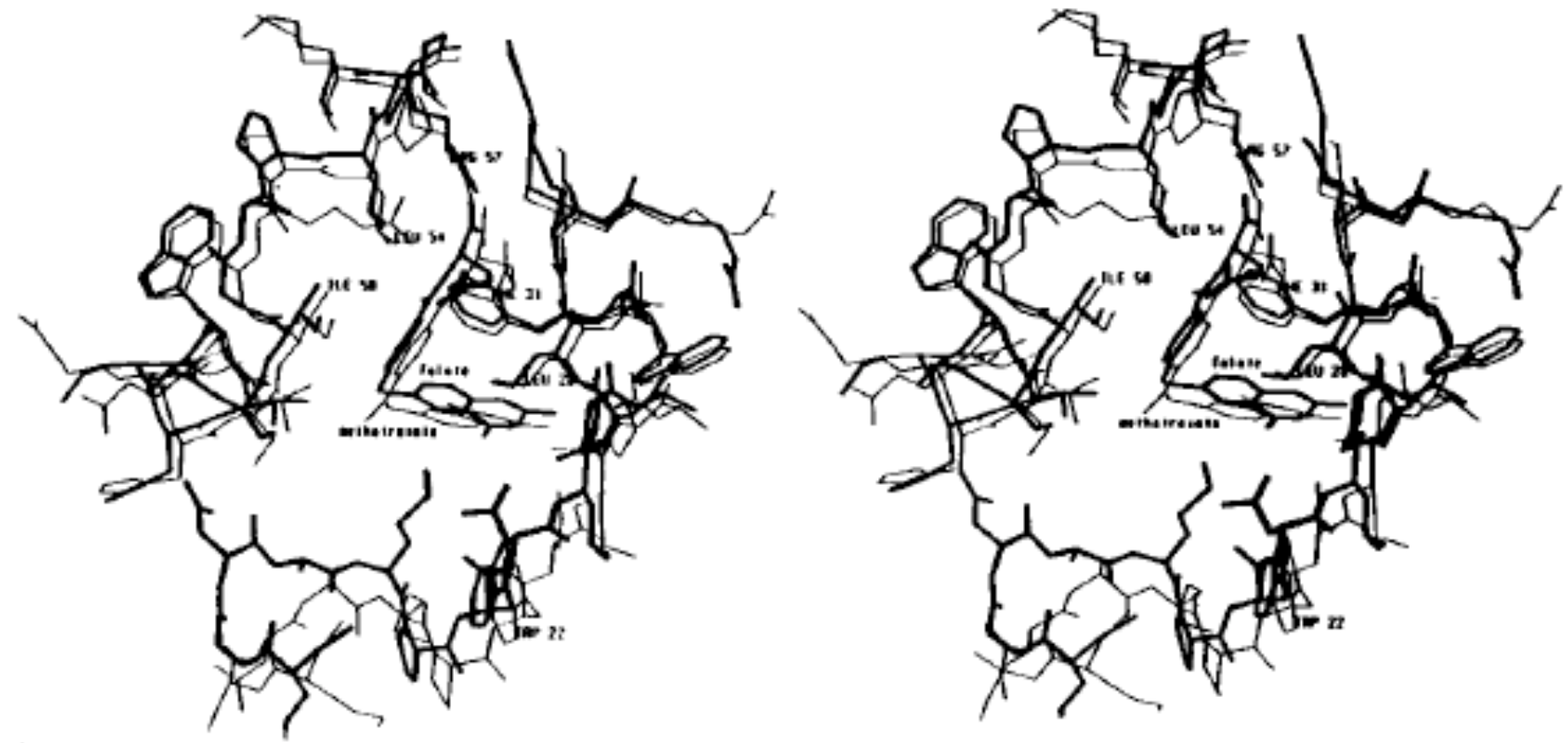


1. Apo,
2. Holo (NADP bound),
3. Ternary (folate and NADP bound),
4.-5. methotraxate (tight-binding inhibitor, 2 structures)

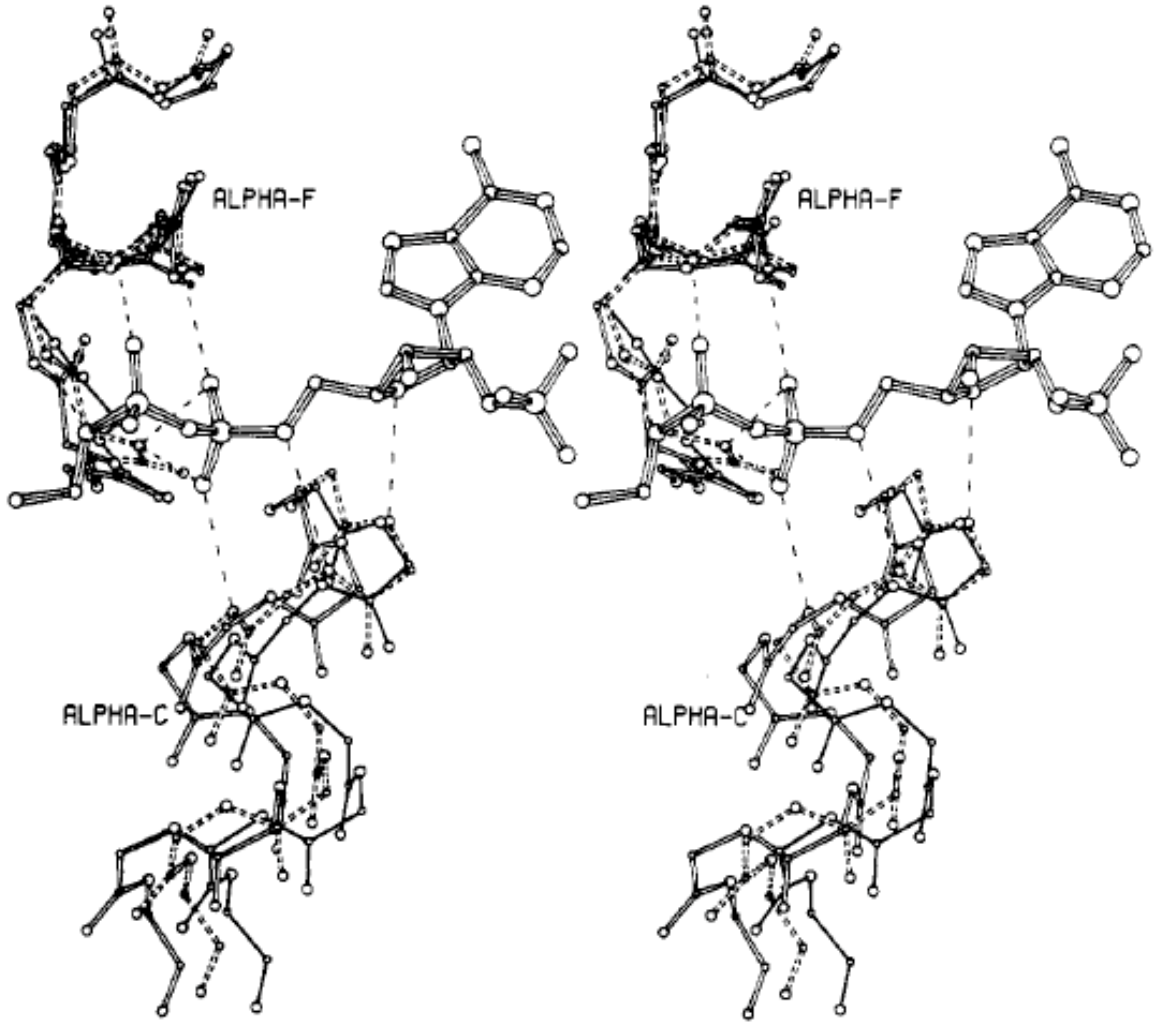
Approximate axes of the domain rotations



(*) binding of NADP to apoenzyme, (#) binding of methotrexate to apoenzyme (molecule **2)** (@) binding of methotrexate to apoenzyme (molecule 1)

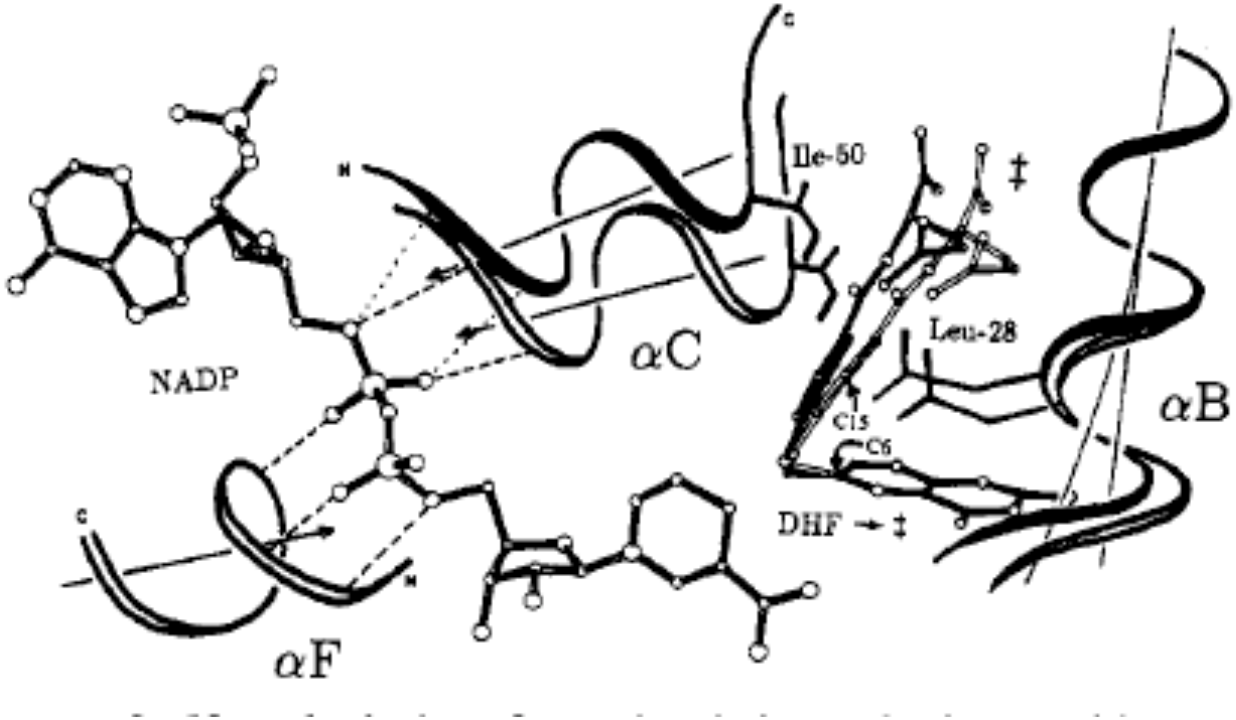


Methotrexate (MTX) binds to DHFR with the pterin group flipped upside down, causing the tail group to shift relative to folate. MTX binds much more tightly than folate (or DHF).

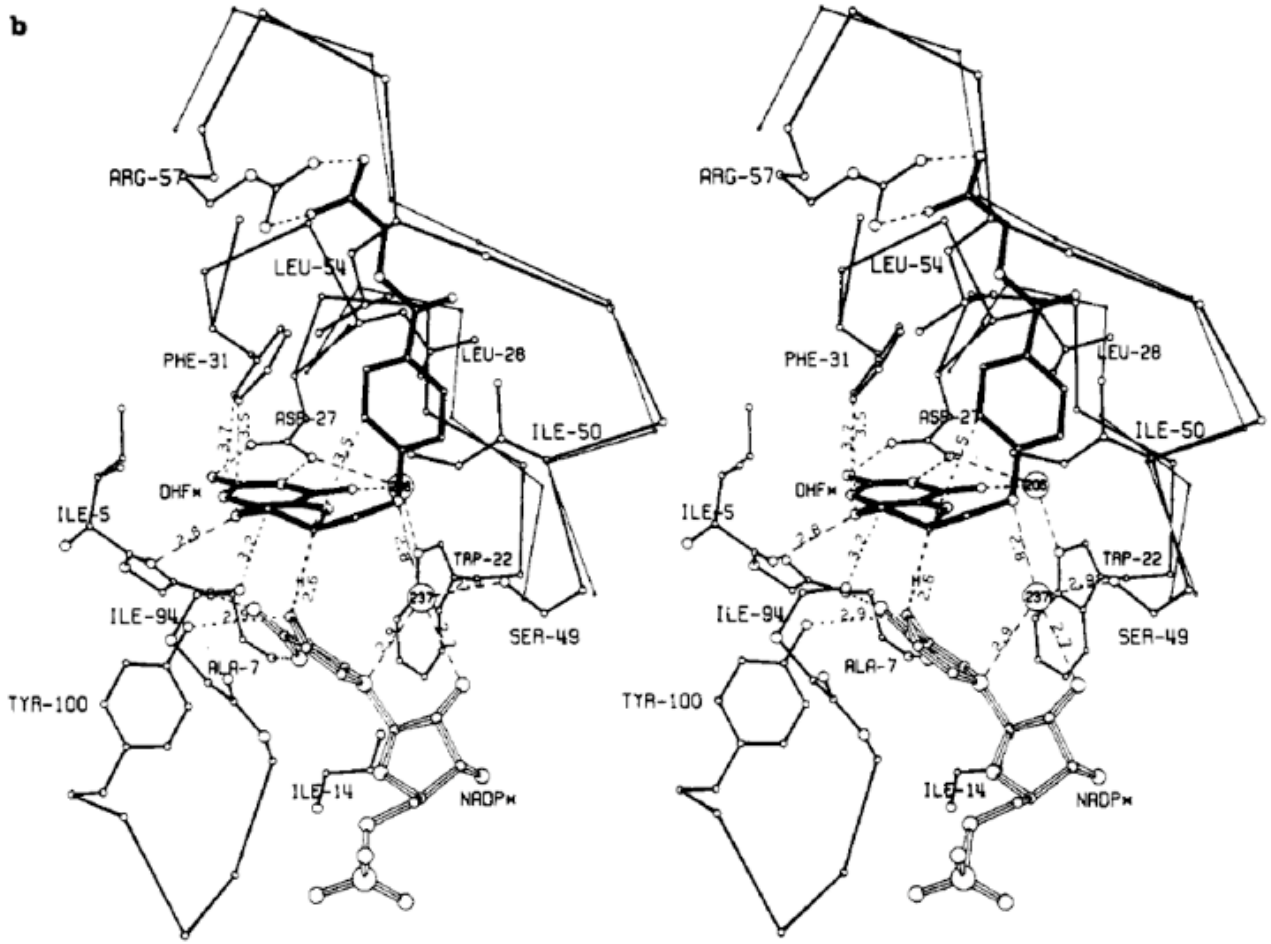


The +-end of two alpha helices make H-bonds to the pyophosphate group of NADPH. Lining up the H-bonds is only possible if the C-helix shifts relative to the F-helix. Making H-bonds “pays for” the conformational change.

Functional significance of conformational change

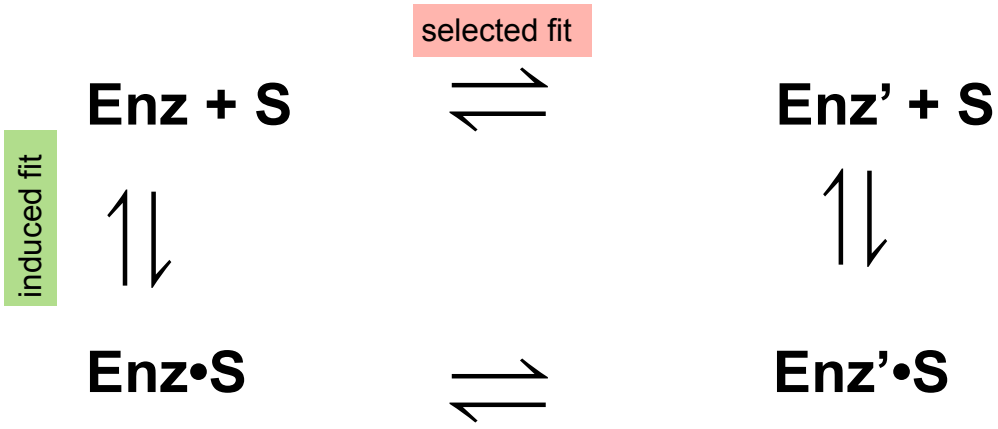


DHFR adds a hydrogen to C6 making it a sp3 carbon and changing the angle of the “tail” para-aminobenzoate group, with the pterin. Binding a tail with a sharper angle favors the product (THF) over the substrate (DHF).


b
ARG-57
LEU-54
PHE-31
LEU-28
ASP-27
ILE-50
DHFM
ILE-5
TRP-22
ILE-94
SER-49
ALA-7
TYR-100
ILE-14
NADPM

Induced fit

- Thermodynamically, a conformational change following binding and a conformational change preceding binding are the same, energetically. But one route may be faster than the other.
- The route starting with binding, followed by conformational change, may be called "**induced fit**".
- The route starting with conformational change, followed by binding, may be called "**selected fit**".



Do exercise 20

- Normal mode analysis of DHFR
- 4dfr, 5dfr, 6dfr or 7dfr