Normal modes in structural biology

Charles H. Robert
CNRS Laboratoire de Biochimie Théorique
Institut de Biologie Physico Chimique
Paris

- What are normal modes?
- What are they good for?
- How do we obtain normal modes?
- How are they used in structural biology and drug design?

Why(or how)
II

Normal modes are used to describe the simplified dynamics of macromolecules

To proceed we must understand

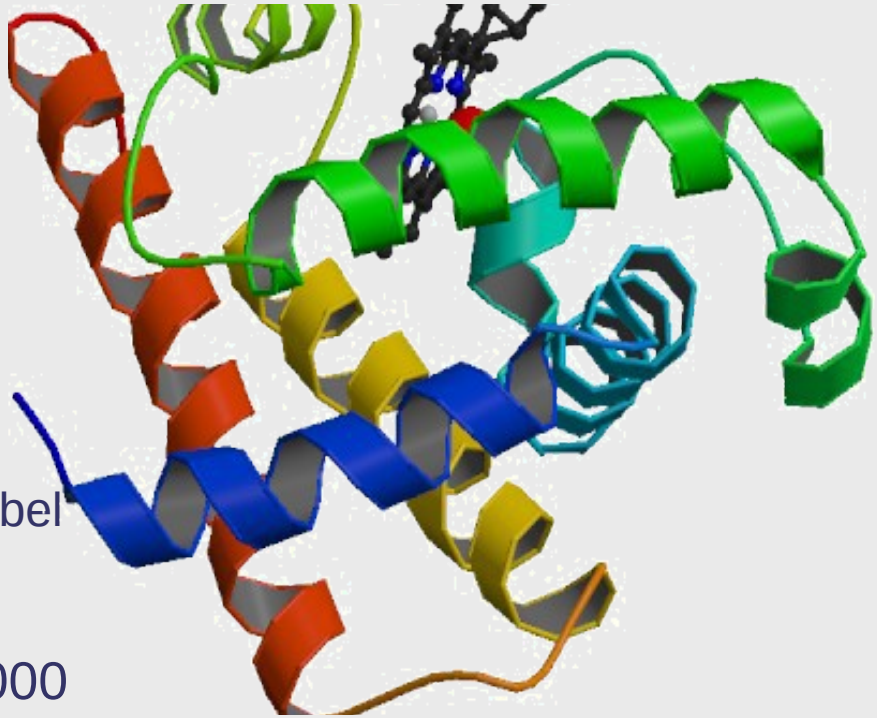
- Macromolecular structure
- Macromolecular dynamics and its role
- Standard approaches to understanding dynamics...
- … then we come back to simplified dynamics: normal modes

[plan]

- Macromolecular structure
- Macromolecular dynamics and its role
- Standard approaches to understanding dynamics...
- … simplif ed dynamics --> normal modes
- Use of normal modes in biology and pharmaceutical research

Macromolecular structure

- Organism cell macromolecules (proteins, DNA, RNA)
- Low to medium resolution (down to 3-4 A)
  - Electron Microscopy
- High resolution (1A):
  - X-ray crystallography (Chemistry Nobel to **Kendrew and Perutz 1962**)
  - Nuclear Magnetic Resonance (Chemistry Nobel to **Wütrich 2002**)
- Protein Data Bank (PDB) contains over 80 000 structures
  - PDB id = number + 3 letter-numbers

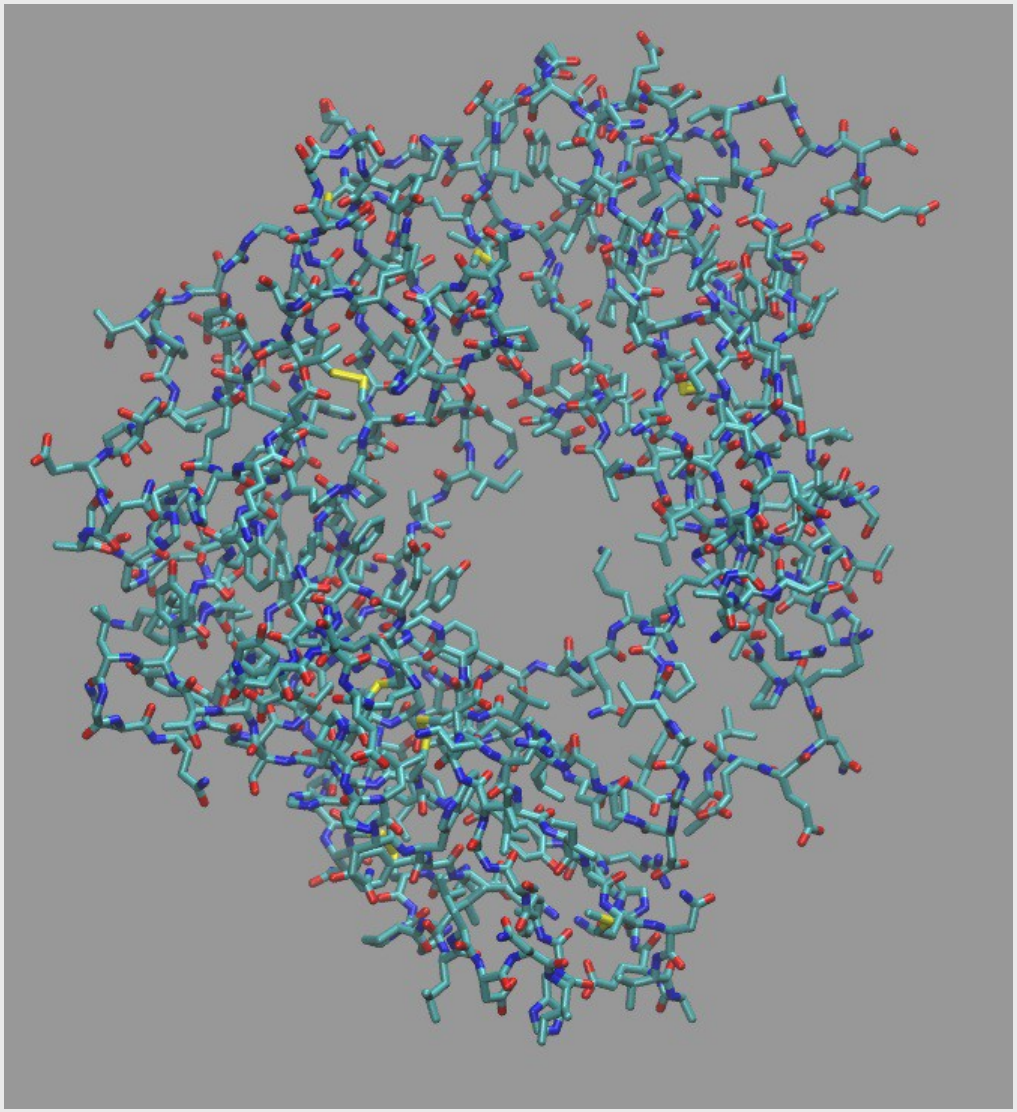


Myoglobin, PDB entry 1mbn
First protein crystal structure (Kendrew, 1958)

Example: interleukin receptor

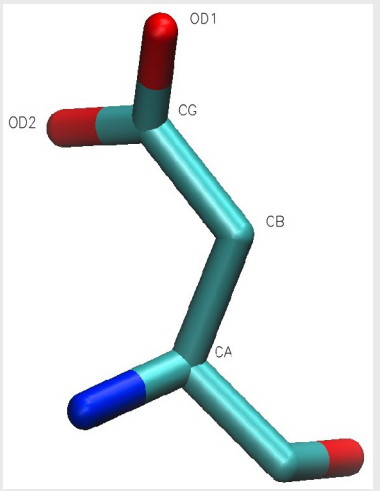
- Type 1 interleukin receptor (cytokine receptor)
  - inf ammation, medical stuff
- In the hypothalamus, IL-1 binding to the receptor increases fever
- Crystal structures of extracellular domain of receptor with interleukin, inhibitors, or agonist peptides

Focus on amino acids and chemical groups



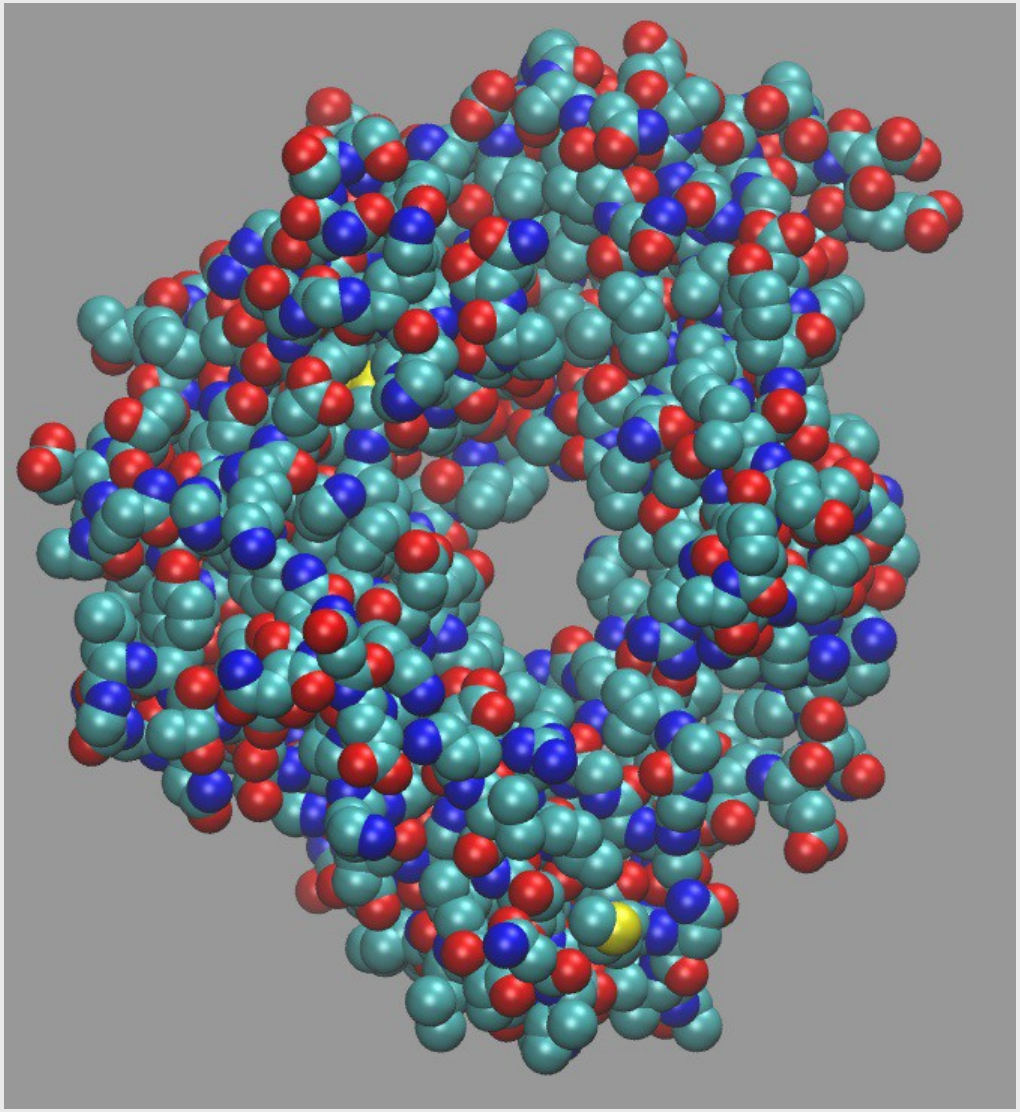
Interferon receptor (PDB id 1g0y)

- 3D coordinates x, y, z for each atom
- Distances are at the scale of Angstroms (1A = 0.1 nm)
- See the 20 amino acid types
- Hydrogen bonding, charge interactions



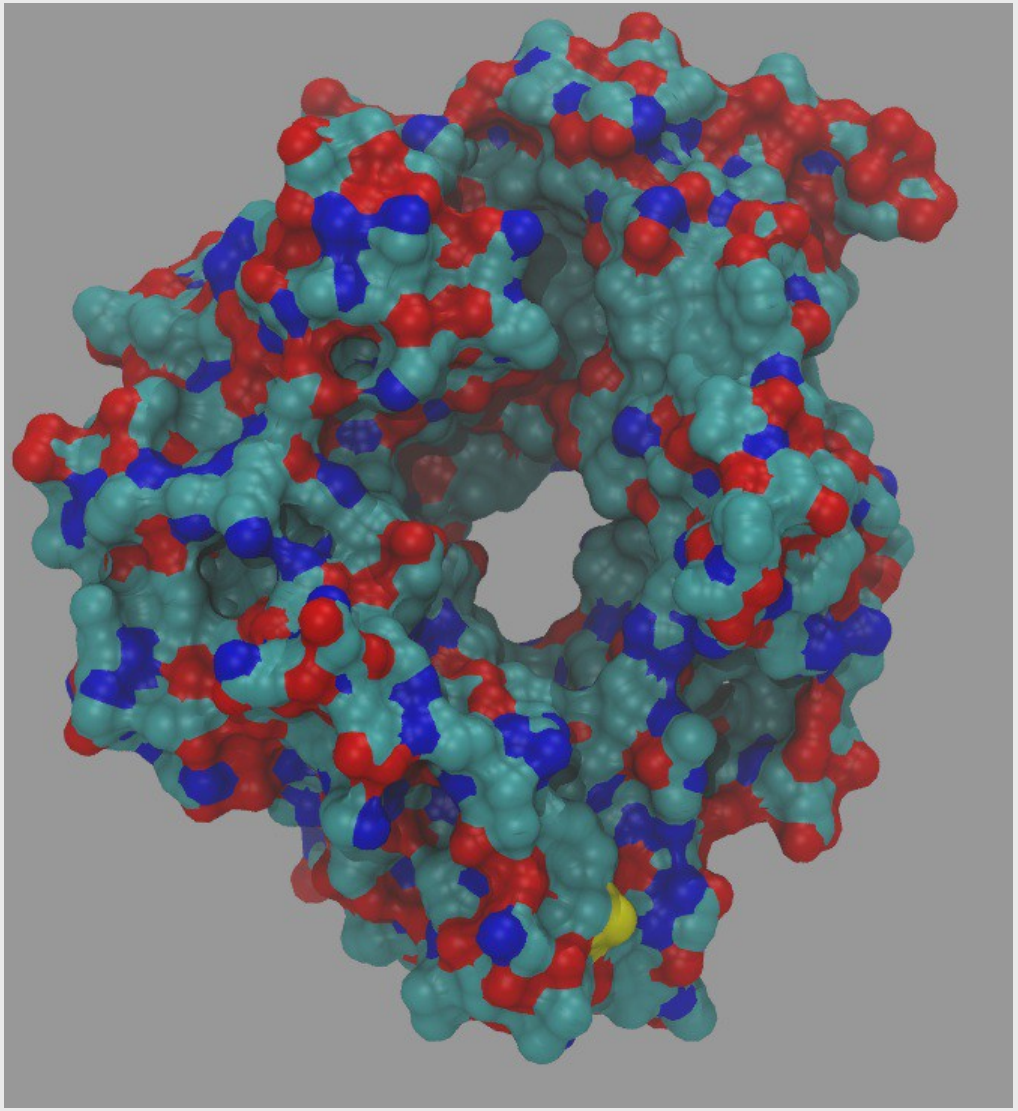
e.g., Aspartic acid

Atoms: type and positions in space



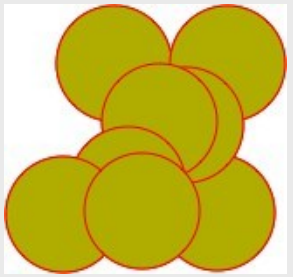
- Space filling representation
- Use van der Waals radii
  1-2 A
- Gives an idea of accessibility, excluded volume

Solvent accessible surface

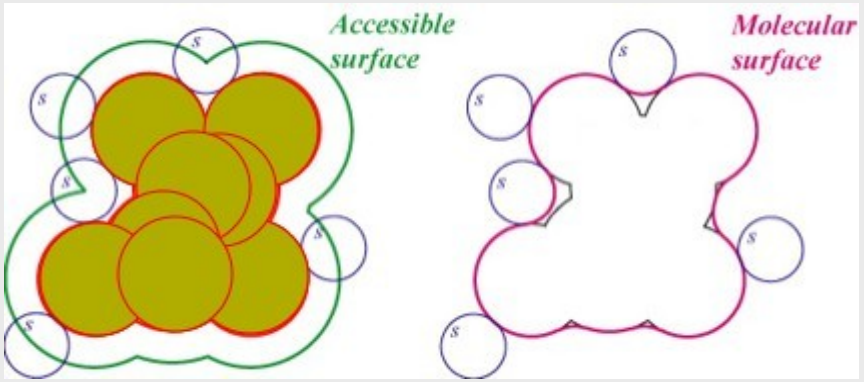


- Accessibility depends on specifying “to what”
- Accessible surface defined by contact with a spherical probe
- Typically, probe approximates a water molecule

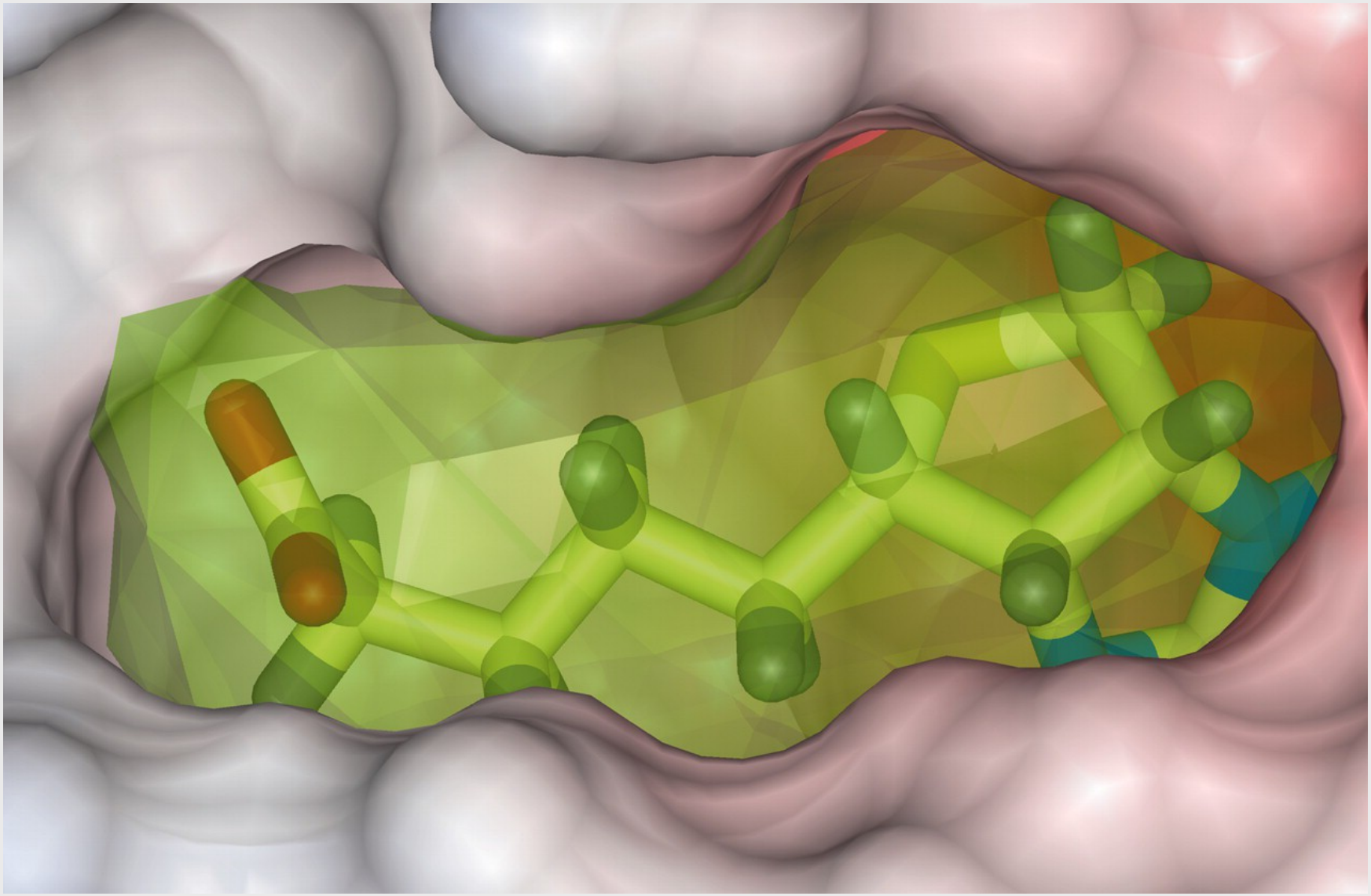
  sphere radius = 1.4 A



van der Waals surface of the union of balls representing the atoms

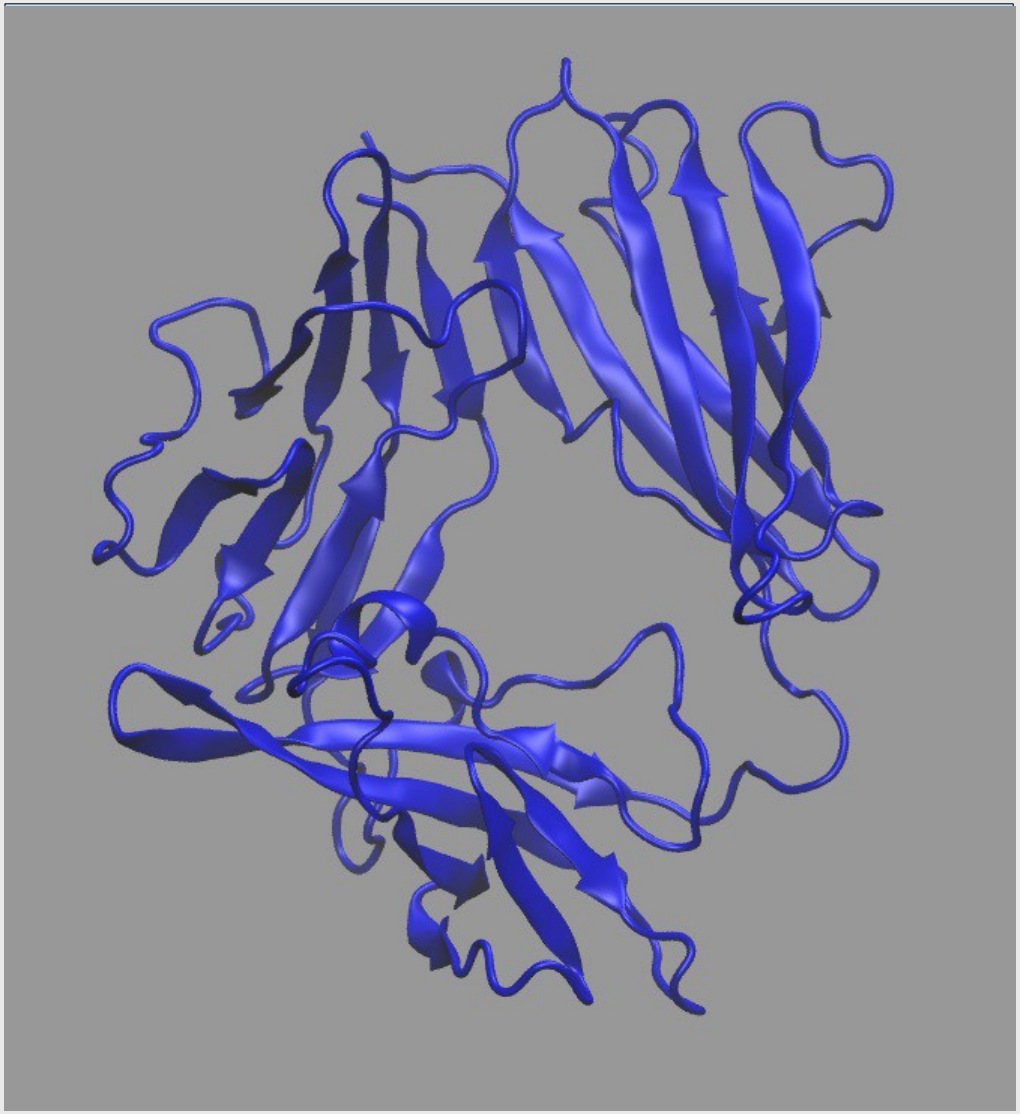


Facilitates identification of binding pockets



Biotin/avidin binding pocket, An et al., Mol Cell Proteomics (2005) 4:752-761

Focus on the backbone of the polypeptide chain



- Connectivity
- Conformation
- Domains
- Secondary structure
  - alpha helices
  - beta sheets

What can we do with a structure?

- Understand a biochemical reaction – catalysis
- See how proteins interact with other macromolecules, ligands, substrates
- Understand aff nity – e.g., burial of hydrophobic surface, H-bonds, ...
- Understand the effects of amino-acid point mutations
- Try to **block interactions** by tailoring a small molecule (drug) to bind instead

But a structure is not enough

- Proteins are not frozen in one form
- Atoms are in constant thermal movement (E = kT)
- The structure changes in time
- ...
- Structures (crystal, NMR, ...) are really *average* structures

[plan]

- Macromolecular structure
- Macromolecular dynamics and its role
- Standard approaches to understanding dynamics...
- … simplif ed dynamics --> normal modes
- Use of normal modes in biology and pharmaceutical research

Effects of atom movements

Protein folding to the native structure

Protein shape changes constantly

The shape of a binding site change can change constantly

Conformational change: there may be more than one stable structure!

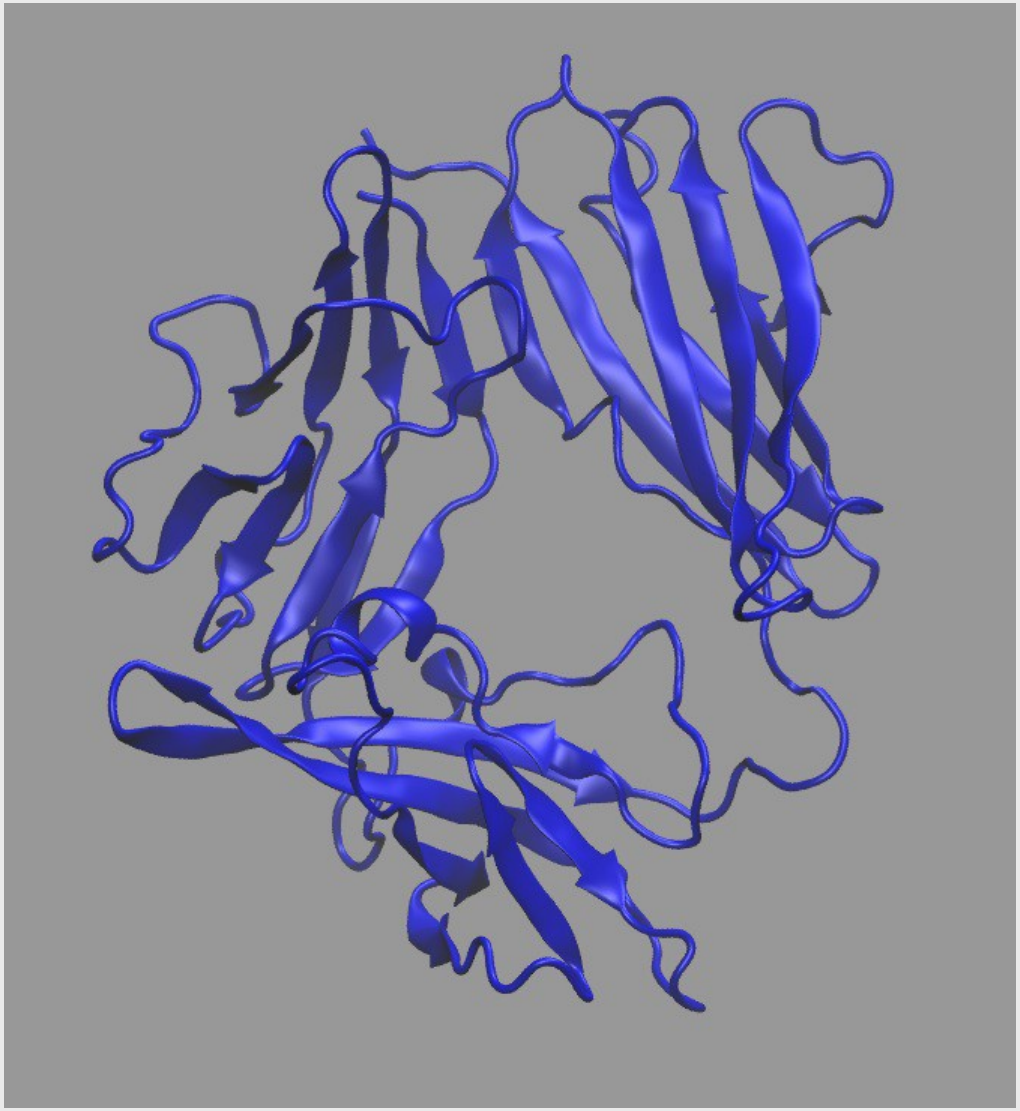
Basis of

- allostery
- cooperative binding
- signal transduction

...

... la biologie, quoi

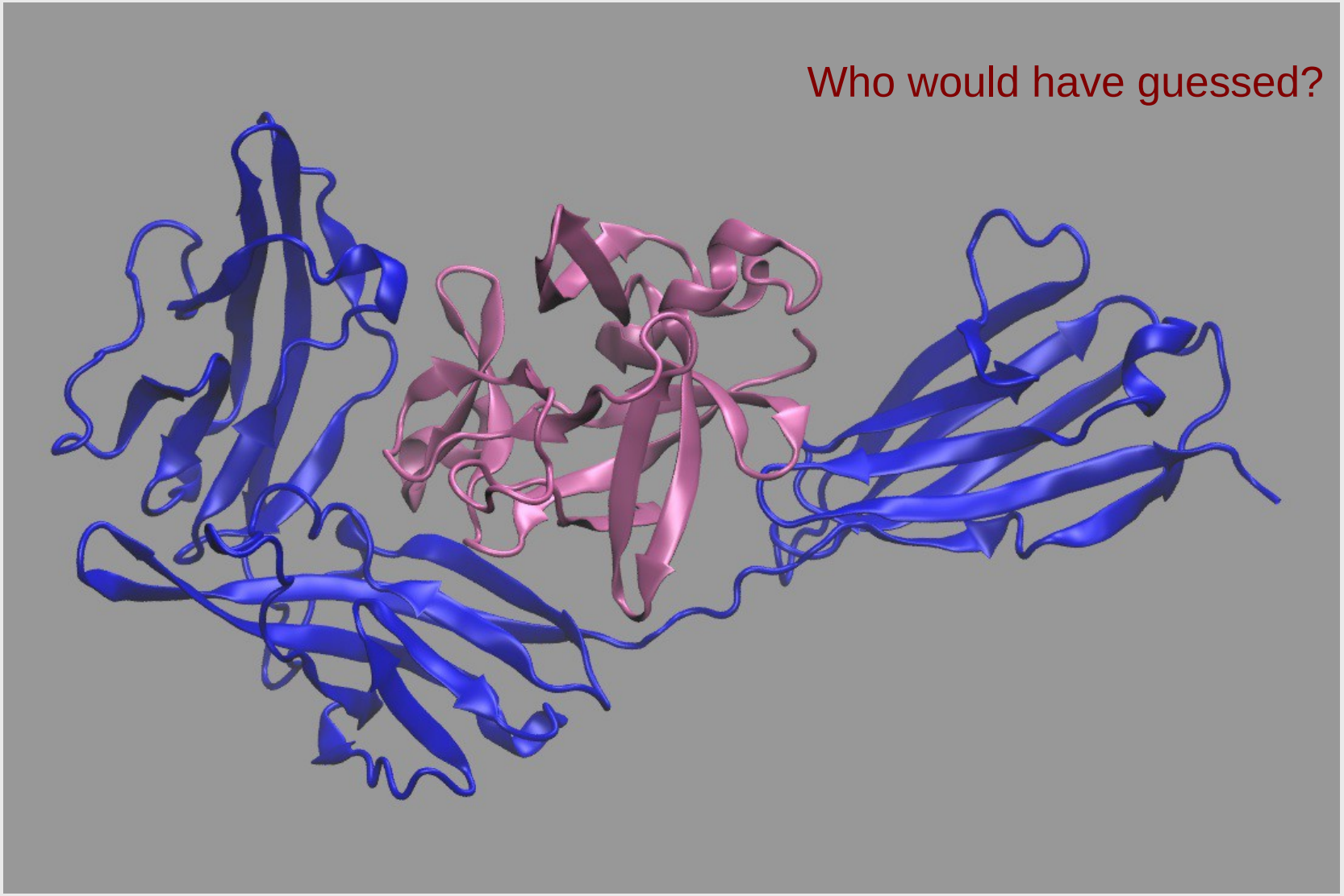
Interleukin receptor again



Interferon receptor (PDB id 1g0y)

Interleukin 1 (IL-1) binding to the receptor involves a substantial conformational change

Interleukin receptor conformational change



Interferon receptor with IL-1 antagonist protein (PDB id 1ira)

From structure to **dynamics**: how do proteins move?

- How “good” is the average structure?
- How does the structure change?
- Do changes occur on binding?

Energy

To understand dynamics we first need to know the energy

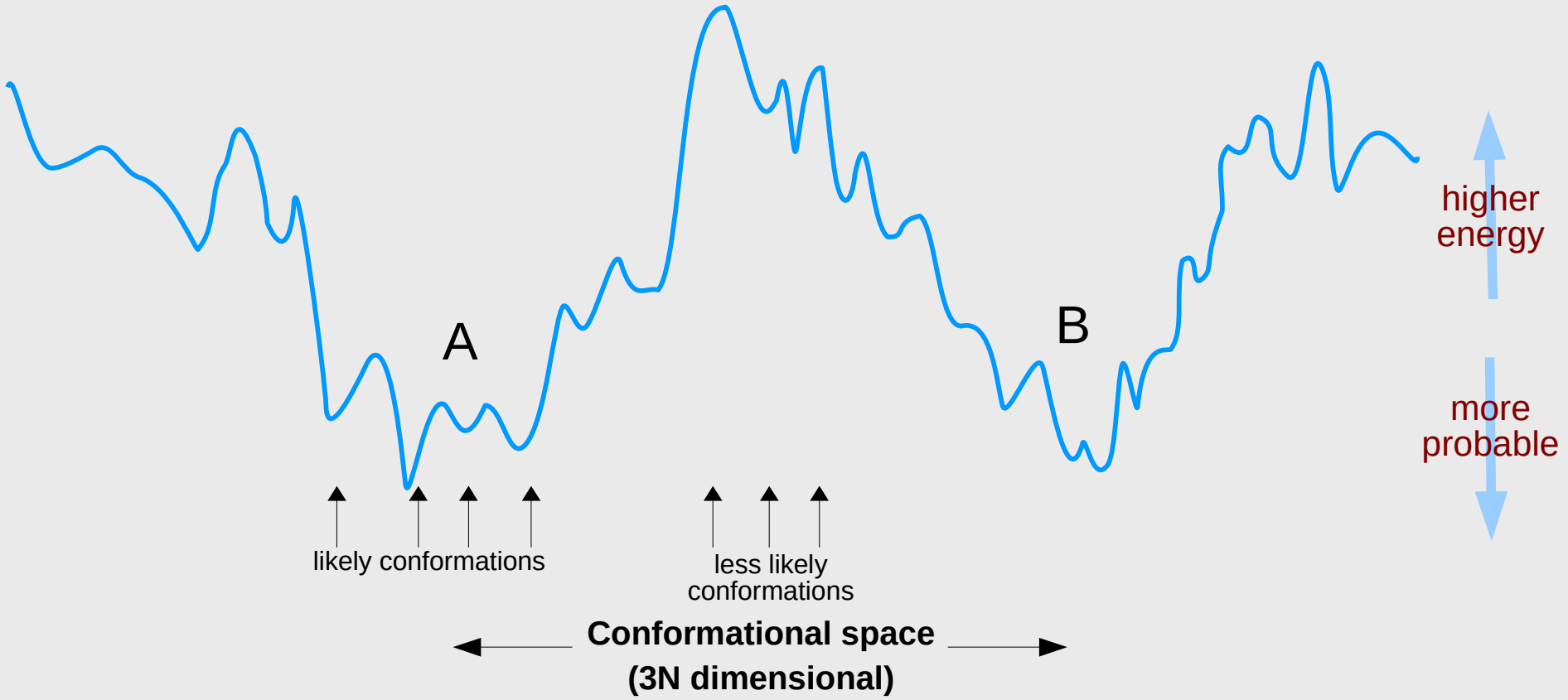
Why?

The energy of a particular configuration (conformation) X determines its probability (Boltzmann' law)

$$p(X) = \text{const} \cdot \exp(-E(X) / kT)$$

higher E(X) --> lower probability of X

Potential Energy surface (heuristic view)



- The molecule can explore the entire conformational space
  - Thermal energy allows it to cross energy barriers
- A broad, deep minimum (basin) indicates a stable structure
- Multiple conformations ref ect multiple basins A, B, ...

Importance of a “good” model of the energy

Energy model must include effects of all important forces at the atomic level

Example: ***protein folding***

Christian Anfinsen: Denature a protein, then renature it again: obtain the same native state

Dynamics of unfolded polypeptide chain direct its folding to the native (folded) state

- The native structure is at a **global energy minimum**

Levinthal's paradox (1960s)

Take an unfolded protein of 100 residues with 10 backbone states/residue

(e.g., phi, psi torsion angles in staggered positions)

Number of possible states 10^{100}

Try to find the native folded state?

If 1 ps/state, exhaustive search >> 10 x age of the universe

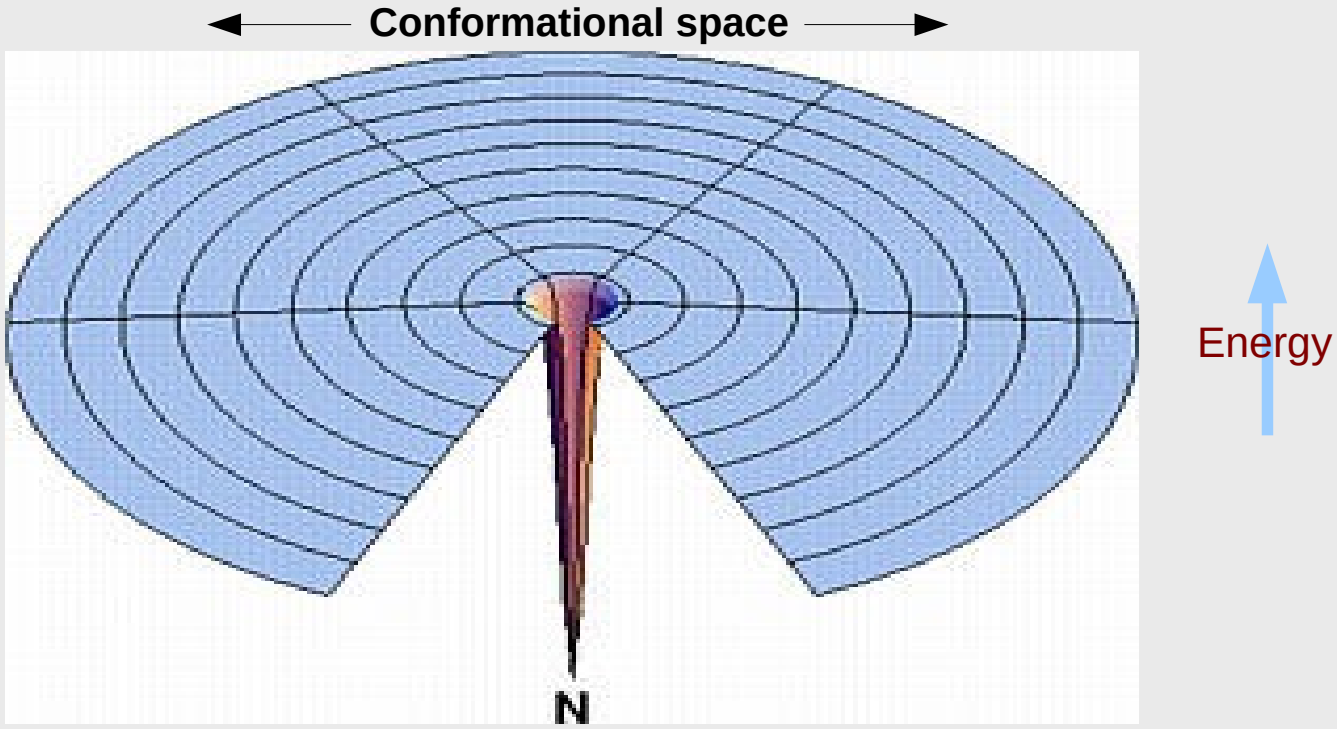
(4×10^{17} sec = 10^{11} years)

Yet proteins do find the native state, on μsec to sec timescale

Implicit model of the energy in early reasoning

- Exhaustive search assumes equal probabilities for non-native states
- This implies that the potential energy surface is f at

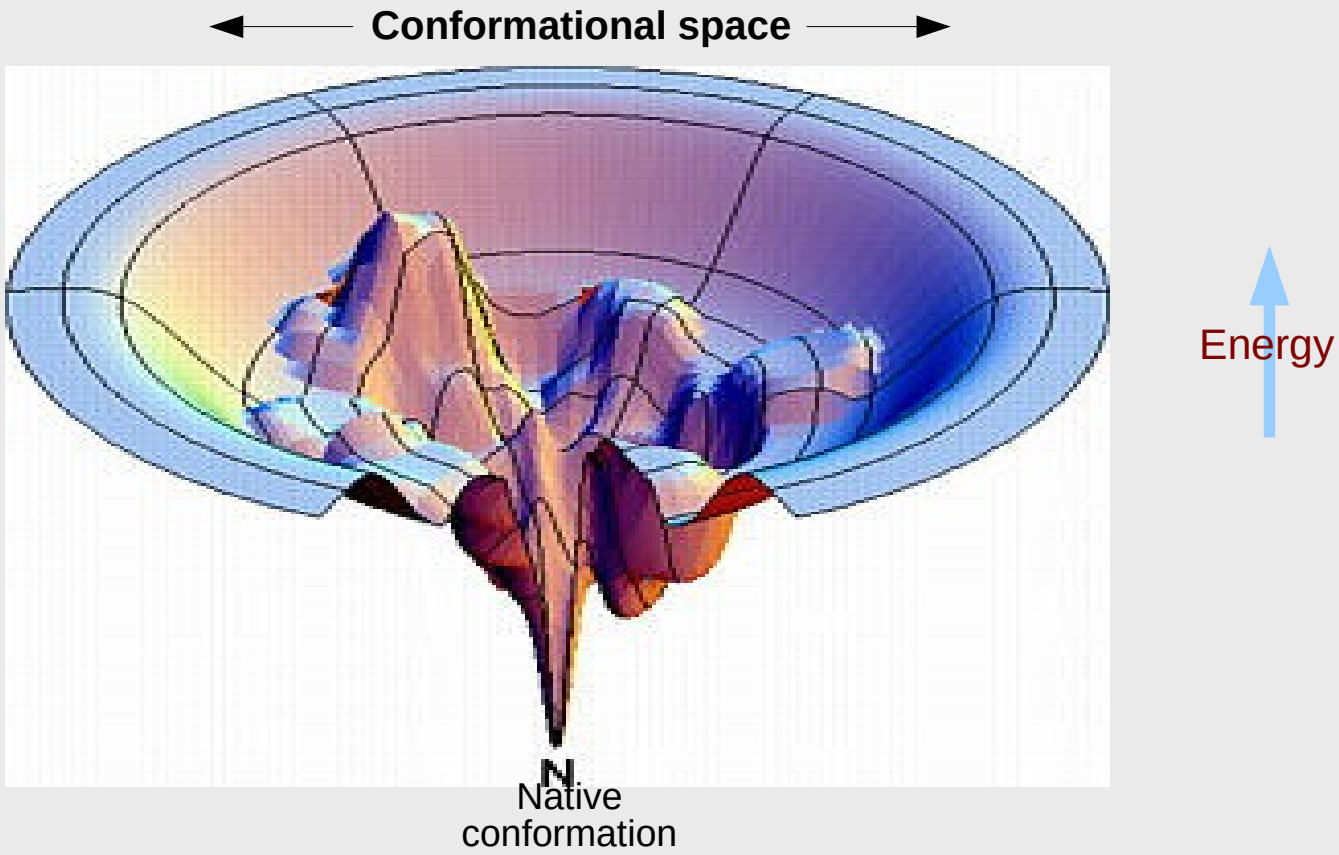
  AKA “the **golf-course model**”



Would be extremely unlikely to f nd the native conformation (like a “hole in one)

Resolution of paradox: **folding funnel**

- In a real protein, local interactions are quickly explored
- Native-like local interactions are **lower energy** than the alternatives
- Energy decreases as we approach the native state



Folding funnel or "New view" (for biochemists) -- Ken Dill, UCSF, early 1990s

A better golf course

Energy models- what do they include?

Energy function decisions

Need potential energy V as a function of atom coordinates

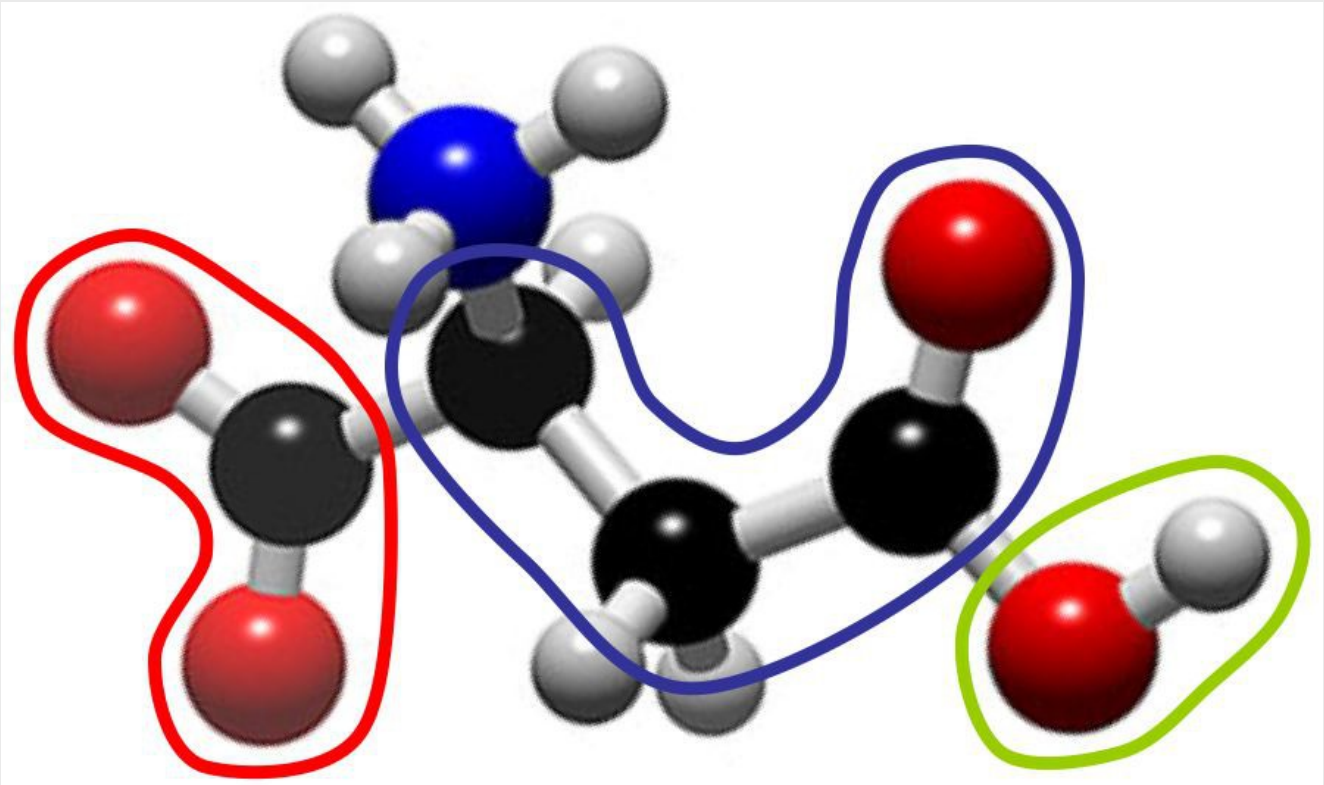
But what functions…

…and what atoms?

Bonded interactions

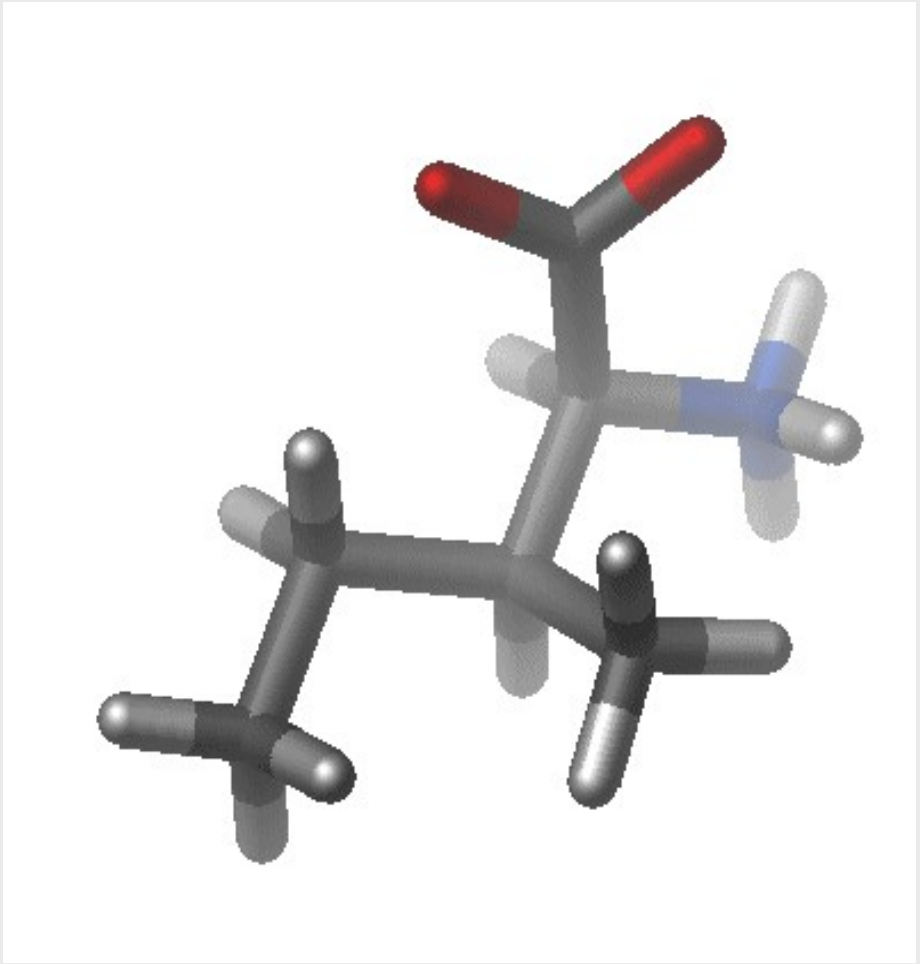
Energy of interaction for covalently bonded atoms

- Bonds (2 atoms)
- Valence angles (3 atoms)
- Torsion angles (4 atoms)
- Improper angle (4 atoms, planar groups)



After Patrice Koehl's course

Animated view of variations in bonded variables



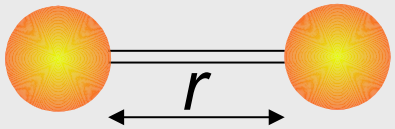
[from Stote and Dejaegere]

Describing the bonded energy

Specifying a conformation X specifies its conformational variables (distances, angles)

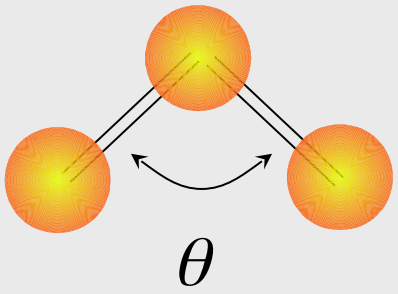
Conformational variables | **Energy terms**

- Bond:



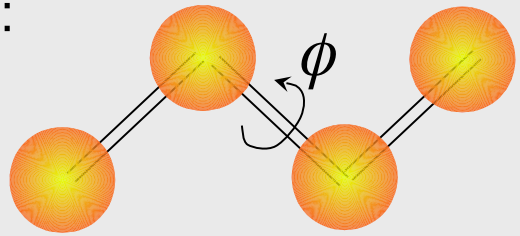
$$V_{bond} = \frac{1}{2} K_{bond} (r - r_o)^2$$

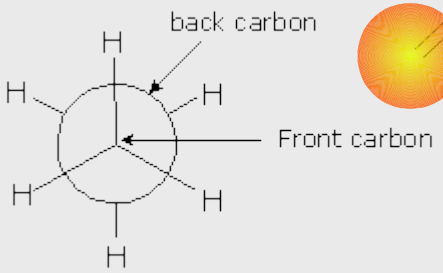
- Angle:



$$V_{angle} = \frac{1}{2} K_{angle} (\theta - \theta_o)^2$$

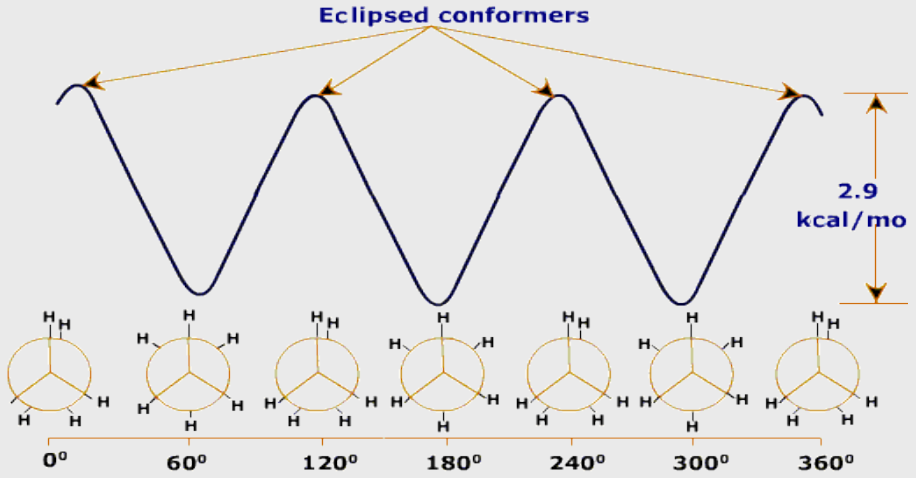
- Dihedral:





$$V_{dihedral} = K_{dihedral} (1 - \cos(n\varphi))$$

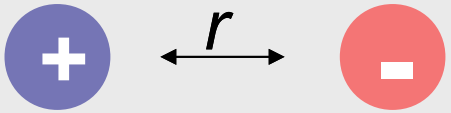
e.g., n = 3 for sigma overlap of sp3 orbitals

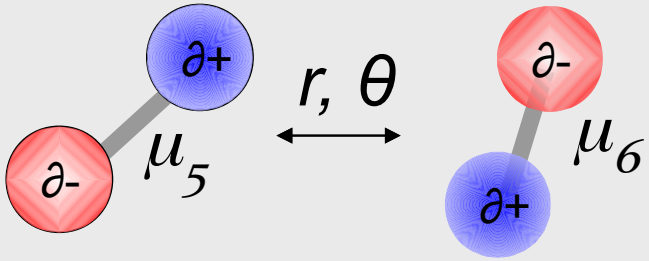


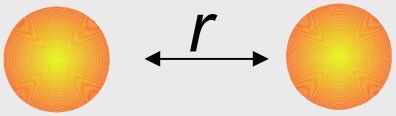
Describing the non-bonded energy

Energy of interaction for all other pairs of atoms

Conformational variables **Energy terms**

- charge-charge: 

$V_{Coulomb}$
- dipole: 

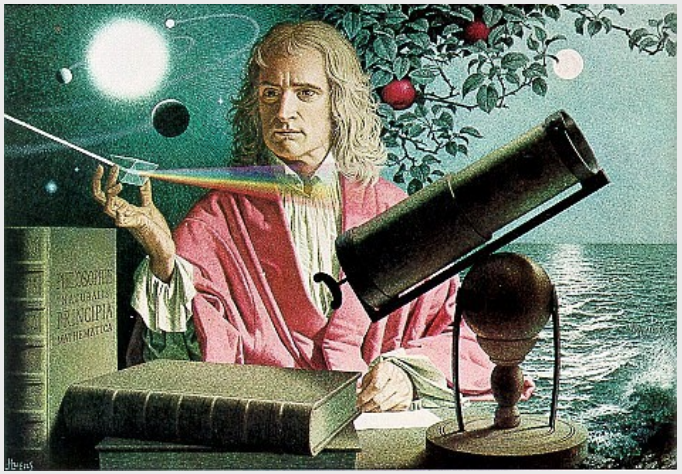
$V_{dipolar}$
- van der Waals: 

$V_{Lennard\text{-}Jones}$

[plan]

- Macromolecular structure
- Macromolecular dynamics and its role
- Standard approaches to understanding dynamics...
- … simplif ed dynamics --> normal modes
- Use of normal modes in biology and pharmaceutical research

Modelling the dynamics of a macromolecule: Newton



Jean-Leon Huens

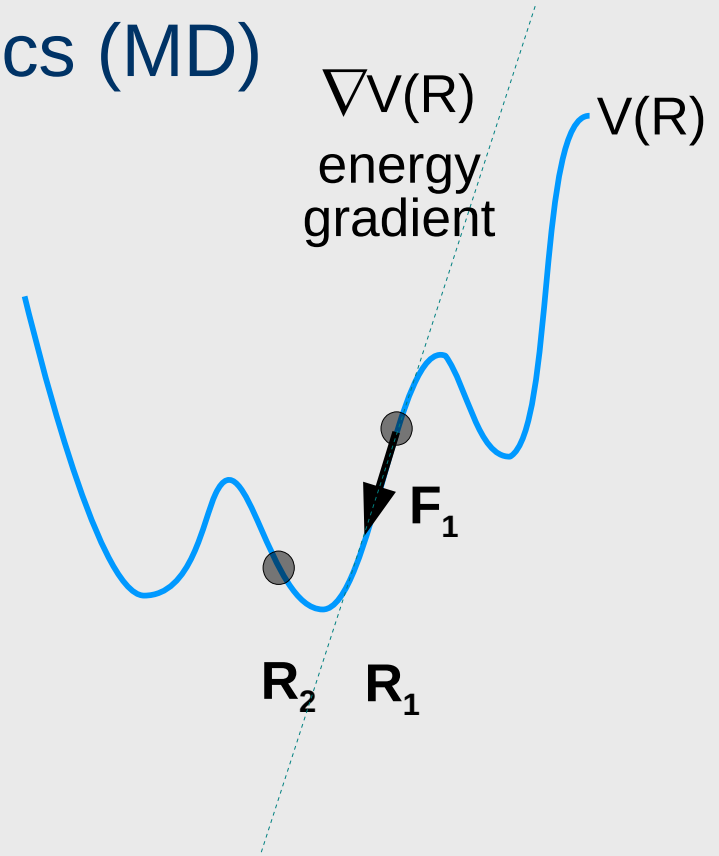
We have a structure, we have an energy model.

We can inject thermal energy, then find atom positions (structure) as a function of time *t*

→ Use classical mechanics for conformational dynamics

Use quantum mechanics for bond breaking and forming

Molecular Dynamics (MD)



Starting structure $\mathbf{R}_1$

- **R** is set of vectors, one $\mathbf{r}_i$ for each atom i
- **R** obtained from a xtal structure, model, …

(Locally minimize energy of the system)

- Like reducing temperature
- Find $\mathbf{R'}_1$ such that $\nabla V(\mathbf{R'}_1)$ 0
- Minimizes initial accelerations

Assign initial velocities

- inject thermal energy
- e.g., Maxwell-Boltzmann distribution

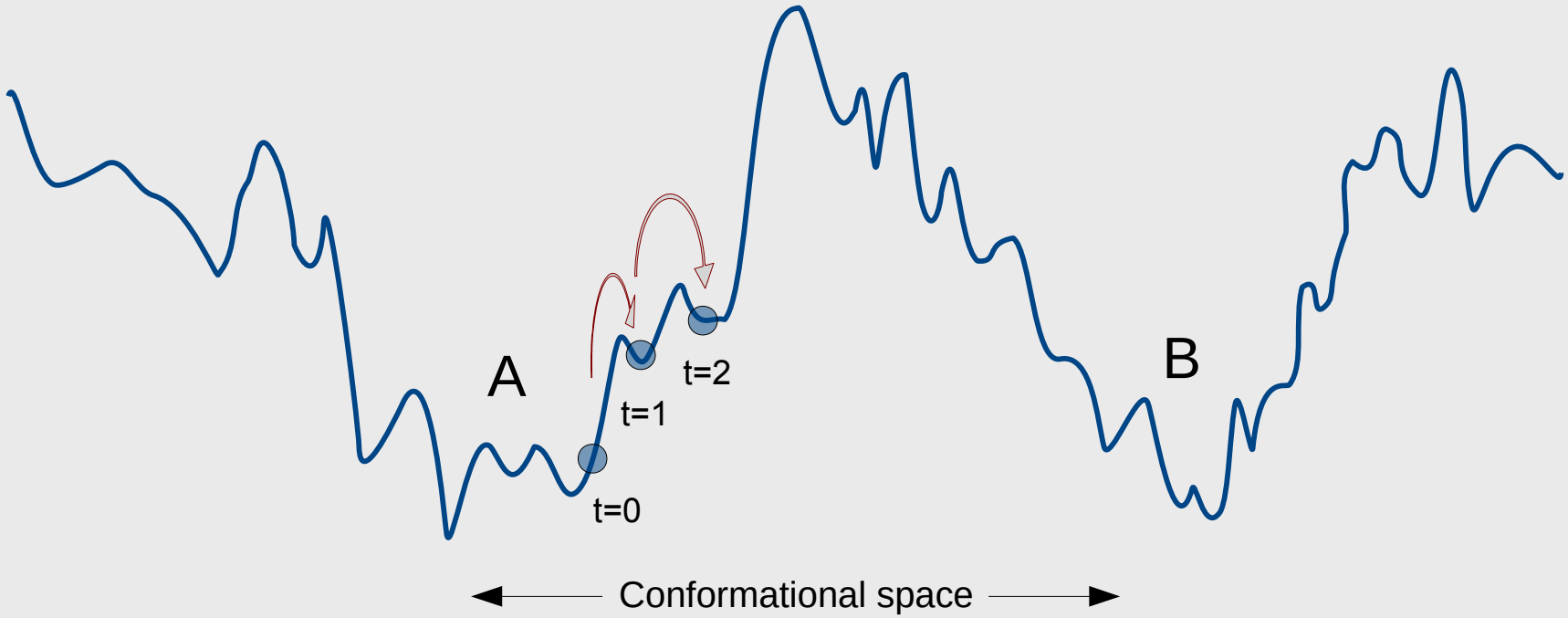
We have positions, velocities, forces (negative gradient of V) and a timestep ~ femtosecond = 10^{-15} s

$$\vec{F}_i = m_i \vec{a}_i$$

Solve Newton's equations of motion

- Use force to get new position $\mathbf{R}_2$ at time 2
- calculate new force, use new force to get new position $\mathbf{R}_3$ ... etc
- $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \ \ldots \ \mathbf{R}_{1,000,000}$ ***simulate the atom motions in time***

Potential Energy surface



- Thermal energy allows the molecule to change conformation by crossing energy barriers on the potential energy surface
- In principle, with long simulations the entire conformational space can be explored (ergodicity)

Importance of solvent

- Water screens electrostatic interactions because of its high dielectric constant *(bulk effect)*
- Local water interactions *(specif c water binding)* provide structural stabilization
- Finite-size effects (solvent exclusion) are important as well

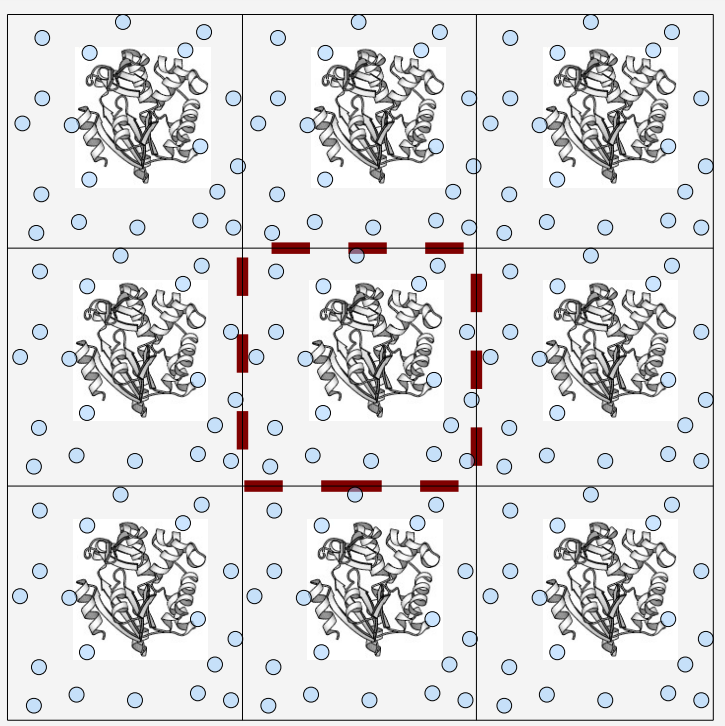
Explicit solvent with periodic boundary conditions: solvent model for typical MD simulations

Primary cell containing the molecule under study is repeated in 3D lattice

- Solid shape tiling 3D space
- Cube, rhombic dodecahedron, truncated octahedron
- Signif cant savings in solvent molecule number if quasi-spherical solid can be chosen

Suff cient solvent is necessary to extend beyond cutoff range (12 A)

Any atom leaving the primary cell is concomitantly replaced in its symmetry-related position



Advantage: Both bulk and specif c binding effects are taken into account

Disadvantage: Solvent atoms typically outnumber protein atoms (10:1)

Implicit solvent models - $W_{solvation}$

1. Hydrophobic effect
   - Solvation energy assumed proportional to exposed surface

2. Electrostatics
   - Water has a high polarizability (dielectric constant) – it “screens” charge-charge interactions
   - Can also calculate solvation energy using Poisson-Boltzmann equation (water+ions), but expensive-- used on individual structures or sets of structures

3. Heuristic model: dielectric “constant” assumed to vary with distance
   - e.g., $\varepsilon = d/1A$
   - For short distances (a few Angstroms): no bulk effect, dielectric constant small(on the order of 1)
   - At long distances: dielectric constant approaches bulk value (80), good screening

Advantage: Speed: only protein atoms are treated explicitly

Disadvantage: No specif c solvent binding effects

Explicit/implicit solvent approaches

Compare by looking at the partition function.

sum of probabilities over all conf gurations of the system, used to normalize the probability

Explicit solvent

All coordinates present -– protein and solvent N = **50 000** atoms

$$Z = \int \ldots \int exp^{-U(r_1,\ldots r_{Np},s_1,\ldots s_{N_s})/kT} dr_1 dr_2 \ldots dr_{Np} ds_1 \ldots ds_{N_s}$$

protein + solvent

Implicit solvent

Solvent coordinates integrated out, *only protein coordinates are left N* = **5 000** atoms

$$Z = \int \ldots \int exp^{-(U_p(r_1,\ldots r_{Np}) + W_{sol}(r_1,\ldots r_{Np}))/kT} dr_1 dr_2 \ldots dr_{N_p}$$

protein only

U is the potential energy (i.e., E)

W_{sol} is the solvation free energy as a function of the protein coordinates

W_{sol} is a free energy (a PMF) because solvent conf gurational sampling is included in its def nition

- But W_{sol} must be derived or def ned by a model

Example:

Molecular dynamics simulation of a small G protein

Studying protein dynamics using MD

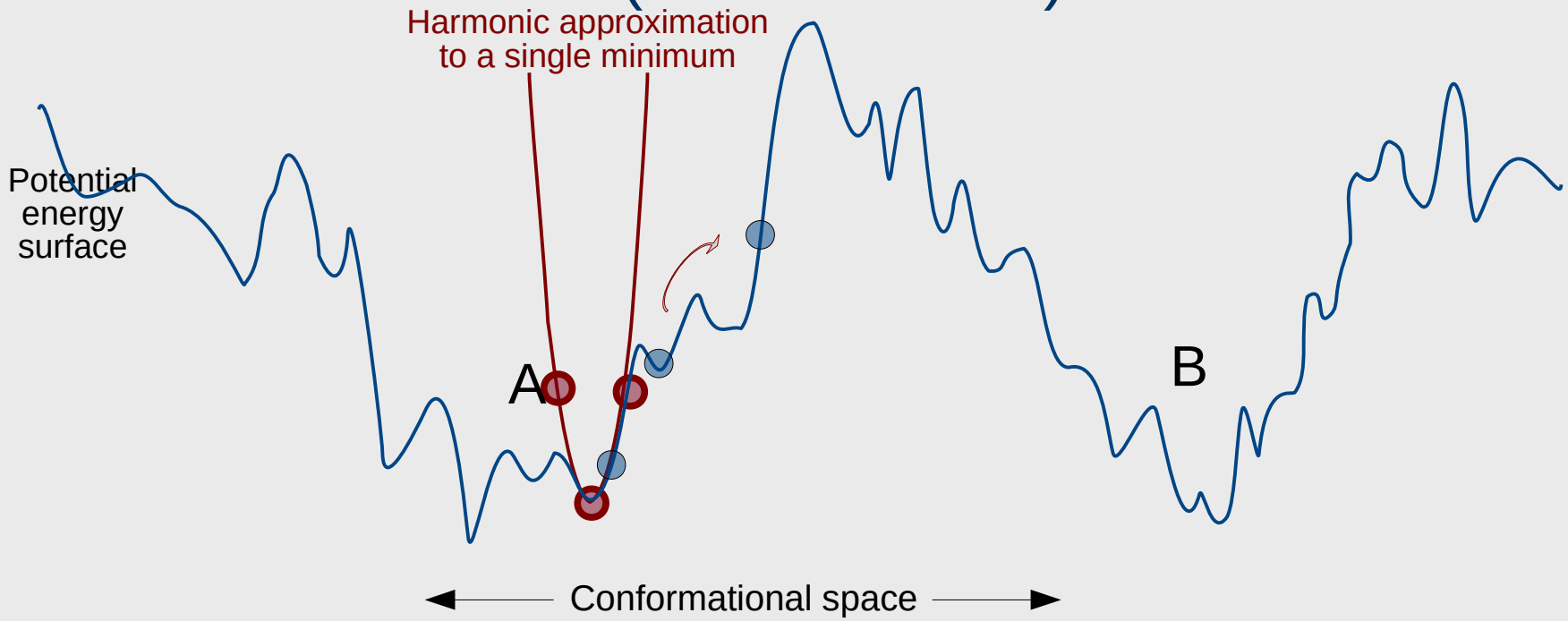
- Pros:
  - Lots of detail
  - Realistic simulation of atom movements
  - Movements may suggest mecanistic models
- Cons:
  - Lots of detail
  - Signif cant computational effort
  - Results tend to be anecdotal – signif cant analysis effort required to ascertain large-scale principles

[plan]

- Macromolecular structure
- Macromolecular dynamics and its role
- Standard approaches to understanding dynamics...
- **… simplif ed dynamics --> normal modes**
- Use of normal modes in biology and pharmaceutical research

What are normal modes?

Normal Mode versus Molecular Dynamics (heuristic view)



In **MD** the molecule can explore all possible structures (in principle)

- Thermal energy allows it to cross energy barriers

In N**ormal Modes** motion is restricted to a harmonic approximation of a single minimum

- Thermal energy produces vibrational deformations about a stable structure

In the low-temperature limit, NM is equivalent to MD

- Thermal movement becomes harmonic as cooled structure is trapped in an energy minimum

Harmonic approximation

Represent *N* atoms each with coordinates *(x,y,z)* by a single vector of *3N* coordinates

$$\mathbf{R} = [(x_1, y_1, z_1), (x_2, y_2, z_2), ..., (x_N, y_N, z_N)]$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & ... & x_{3N} \end{bmatrix}$$

Expand potential energy V about a point x_o

$$V(\mathbf{x}) = V(\mathbf{x_o}) + \sum_{i}^{3N} \left(\frac{\partial V}{\partial x_i} \right)_{\mathbf{x_o}} (x_i - x_{io}) + \frac{1}{2!} \sum_{i,j}^{3N} \left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_{\mathbf{x_o}} (x_i - x_{io})(x_j - x_{jo}) + ...$$

- We specify our starting conformation x_o to be a minimum of V: first derivatives are zero
- Harmonic approximation: keep 2nd order terms only

Normal Mode Dynamics

Analytical solution to the equations of motion for harmonic potential

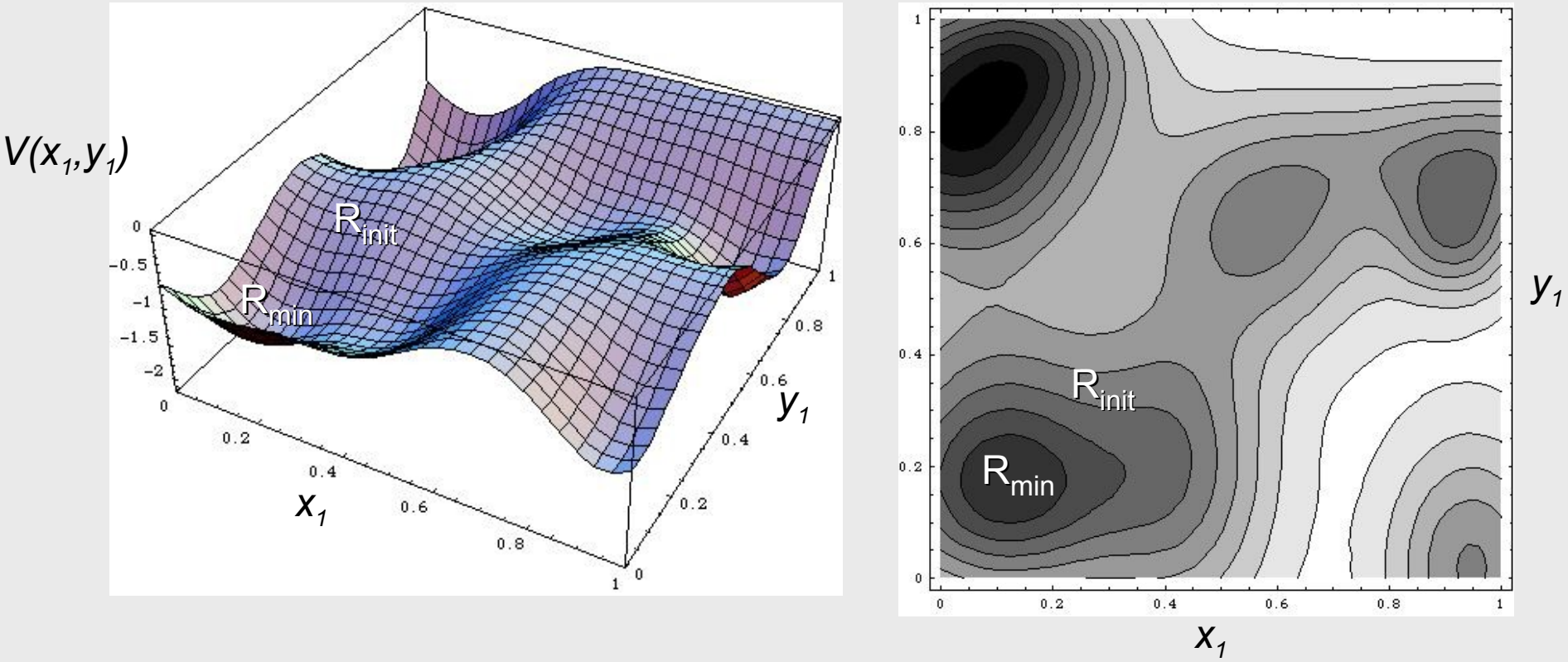
- Eigensystem analysis
- Periodic solutions = vibrations

How do we calculate normal modes?

Energy minimization

- Typically start with a crystal structure of a macromolecule
- Adjust the conformation to reduce the energy
- Removes steric clashes, optimizes bond lengths, ...
  Crystal structure is ignorant of our energy model!
- Potential energy surface V has multiple minima – we will only look for the nearest local minimum

Extremely simple example : two conformational variables

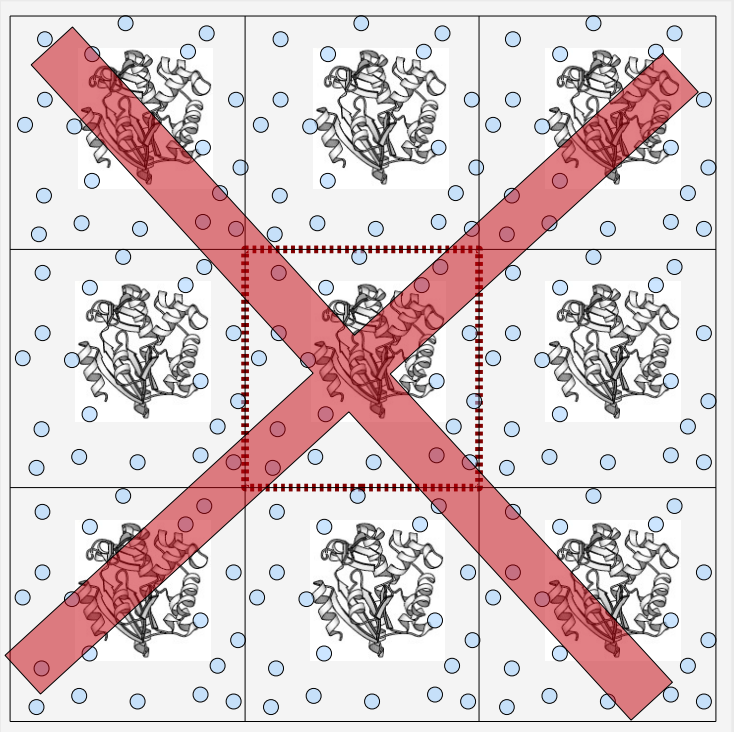


Explicit solvent with periodic boundary conditions is incompatible with normal modes calculations

Lots of solvent used for MD

Extensive energy minimization of system necessary for NM calculation

- Bulk water freezes!
- Protein movements become highly restricted, high frequency, unrealistic

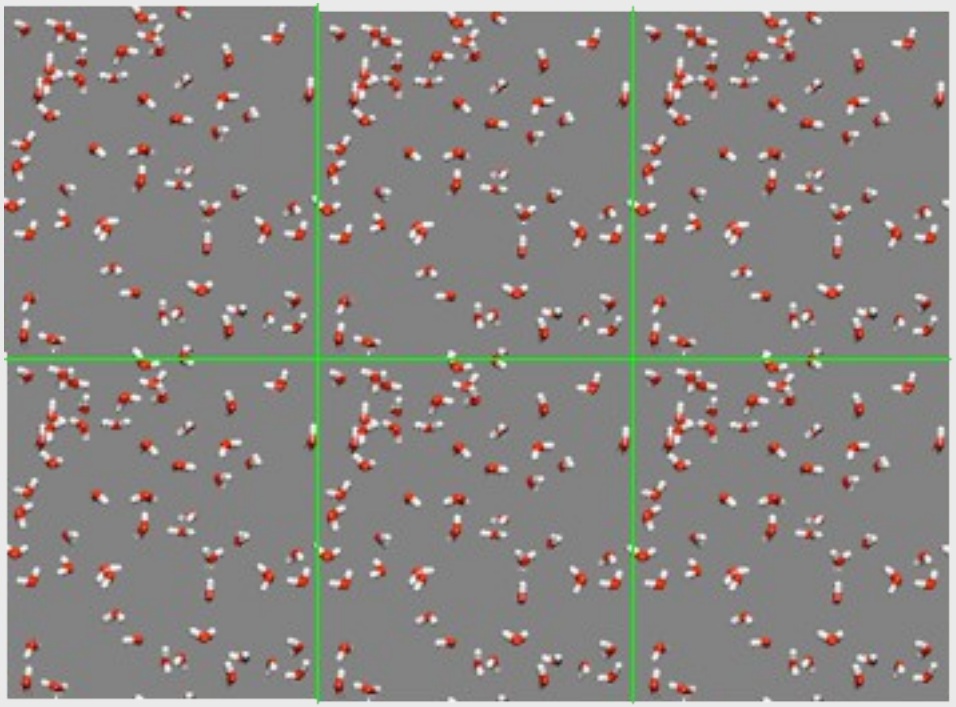


Use of a hydration shell

Take starting structure from MD simulation

Cut away water farther than a given cutoff (e.g., H-bonding distance) from protein

Remaining water layer is included with the protein during energy minimization and normal modes calculation



But bulk water is important too...

Implicit solvent models – W_{solv} again

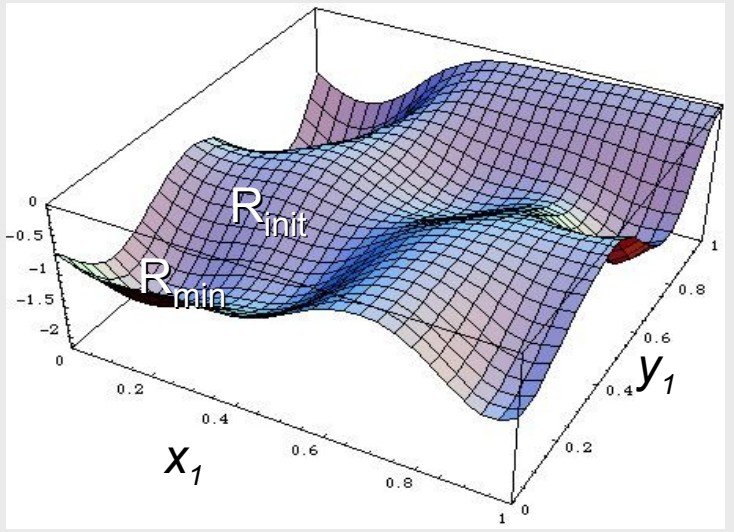
1. Solvation energy proportional to exposed surface area
   - Very good for hydrophobic effect
   - Diff cult to integrate with normal modes (need 2nd derivatives)

2. Continuum electrostatics
   - Water has a high dielectric constant – it “screens” charge-charge interactions
   - Calculate solvation energy using Poisson-Boltzmann equation (water+ions)
   - Expensive, used on individual structures or sets of structures
   - Diff cult to integrate with normal modes

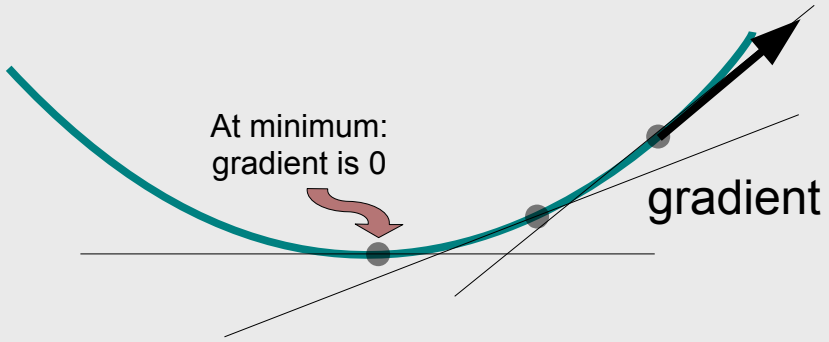
3. Heuristic model: dielectric “constant” assumed to increase with distance
   - e.g., $\varepsilon = d/1A$
   - For short distances (a few Angstroms): no bulk effect, dielectric constant small(on the order of 1)
   - For long distances: dielectric constant approaches bulk value (80), good screening
   - Well adapted to use with normal modes

Techniques for energy minimization

Extensive minimization is required for calculation of Normal Modes



- 1st derivative (energy gradient) approaches
  - steepest descent (gradient of V)
  - conjugate gradient (list of productive directions)
- 2nd derivative (curvature of energy surface) approaches
  - use curvature matrix (Hessian)
  - *Would* find minimum in one step if surface were quadratic
  - For a real surface, very useful once we are near the minimum
  -



Quality of minimization judged by magnitude of residual forces

force proportional to gradient: should be **zero** at minimum

Normal modes calculation

Vibrational energy

Total vibrational energy is conserved

$$E = T + V$$

Sum of kinetic energy (T) and potential energy (V) of the macromolecule

Kinetic energy is a function of the **velocities** (time derivative of the positions)

$$T(\ \ )$$

The Potential energy is a function of the **positions**

$$V(\boldsymbol{R})$$

Vibrational energy

Represent *N* atoms each with coordinates *(x,y,z)* by a single vector of *3N* coordinates

atom 1 atom 2 ... atom N

$$\mathbf{R} = [(x_1, y_1, z_1), (x_2, y_2, z_2), ..., (x_N, y_N, z_N)]$$

$$\boldsymbol{x} = [x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6 \quad ... \quad x_{3N}]$$

Vibrational energy is

Harmonic approximation
x_o is $\mathbf{R}_{min}$ -- a minimum of V

$$\frac{1}{2}\sum_{atoms,a}^{N} m_a(v_x^2 + v_y^2 + v_z^2) \; + \; \frac{1}{2}\sum_{i,j}^{3N}\left(\frac{\partial^2 V}{\partial x_i x_j}\right)_{x_o}(x_i - x_{oi})(x_j - x_{oj})$$

Kinetic energy
sum over atoms

Potential energy
sum over Cartesian coordinates

But the equation has an inconvenient form...

Root-mass weighting

Change to root-mass-weighted coordinates

$$\xi_i = \sqrt{m_i}(x_i - x_{io})$$

Vibrational energy can now be written

$$E_{\mathrm{vib}} = \frac{1}{2}\sum_{i}^{3N} \dot{\xi}_i^2 + \frac{1}{2}\sum_{i,j}^{3N} \left(\frac{\partial^2 V}{\partial \xi_i \partial \xi_j}\right)_{\mathbf{0}} \xi_i \xi_j$$

kinetic energy + potential energy

.... or even more compactly as a matrix equation (H is the mass-weighted Hessian or force constant matrix)

$$E(\boldsymbol{\xi}) = \frac{1}{2}\dot{\boldsymbol{\xi}}^t \dot{\boldsymbol{\xi}} + \frac{1}{2}\boldsymbol{\xi}^t \mathcal{H} \boldsymbol{\xi}$$

Rewrite in diagonal form to obtain vibrations

$$E(\boldsymbol{q}) = \frac{1}{2}\dot{\boldsymbol{q}}^t\dot{\boldsymbol{q}} + \frac{1}{2}\boldsymbol{q}^t\boldsymbol{L}\boldsymbol{q}$$

$\boldsymbol{L}$ is the result of finding matrix **A** that diagonalizes the mass-weighted Hessian *H*

$$\boldsymbol{L} = \boldsymbol{A}^t\mathcal{H}\boldsymbol{A}$$

The equation has **periodic** time-dependent solutions for each degree of freedom j with amplitude $\boldsymbol{b_j}$ depending on T

$$q_j(t) = b_j \cos(\ _j t + \ \ _j)$$

$\boldsymbol{L}$ contains the squared **vibrational frequencies** (eigenvalues)

$$\boldsymbol{L} = \begin{bmatrix} \omega_1^2 & 0 & 0 & \dots & 0 \\ 0 & \omega_2^2 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & \dots & \omega_{3N}^2 \end{bmatrix}$$

6 zero eigenvalues, **number of NM vibrations is *3N - 6***

translations of the CM of the protein in x, y, z are not periodic (3 dof)

rigid rotations about x, y, z axes are not periodic (3 dof)

Note: angular frequencies (radians/sec) are typically converted to cm^{-1} using $\bar{\nu} = \omega/2\pi c$

Diagonalization provides the normal mode coordinates

$$\boldsymbol{L} = \boldsymbol{A}^t \mathcal{H} \boldsymbol{A}$$

The columns of $\boldsymbol{A}^t$ are the eigenvectors -- ***aka*** the normal mode vectors.

Each normal mode is a linear combination of the root-mass weighted Cartesian coordinates. The jth normal mode coordinate is defined from the jth eigenvector:

$$q_j = \sum_{i=1}^{3N} A_{ij} \xi_i$$

Normal mode vector can be used to describe the vibrational movement of each atom

$$\delta x_{ij} = \frac{A_{ij}}{\sqrt{m_i}} b_j \cos(\omega_j t + \phi_j)$$

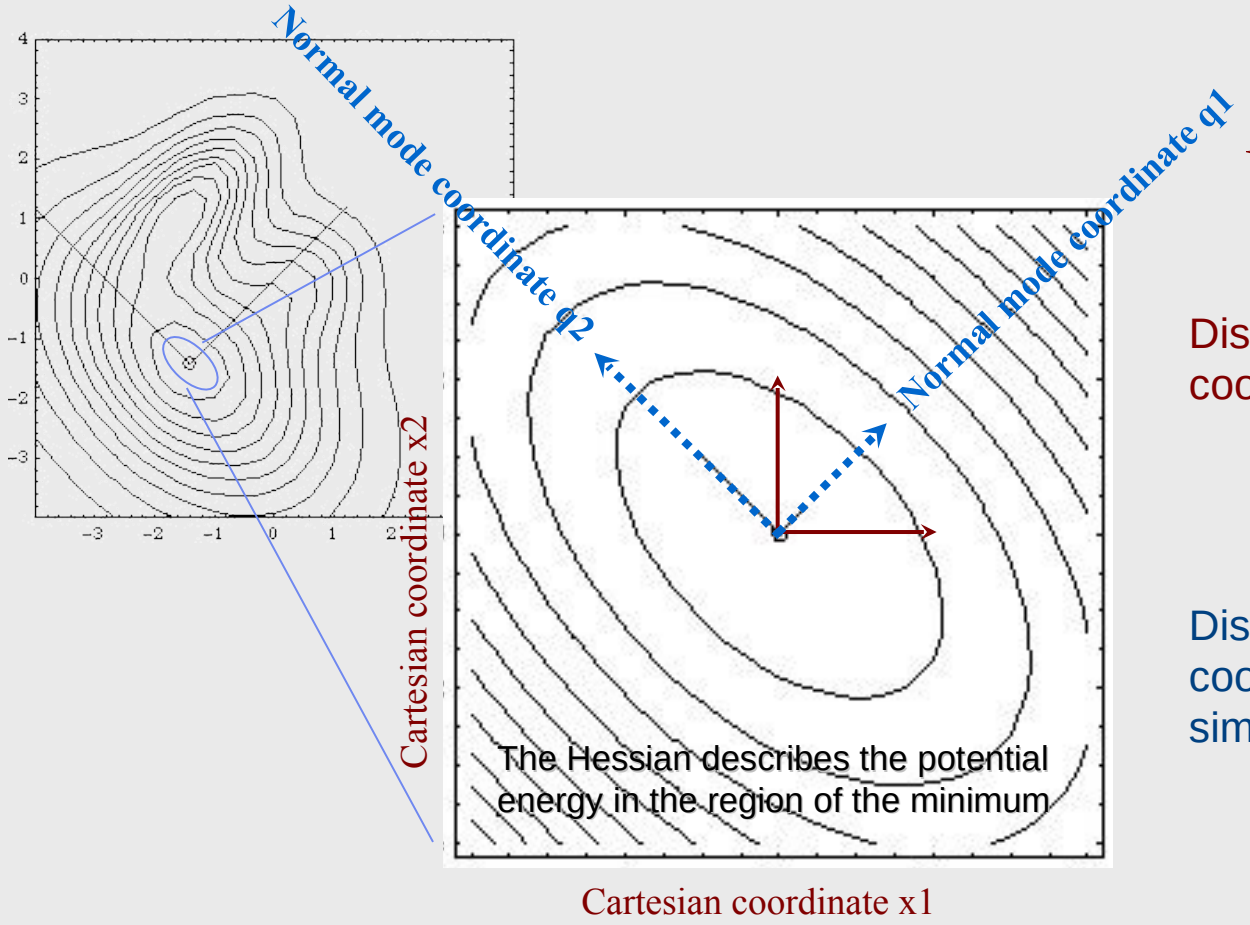
Amplitude of the motion

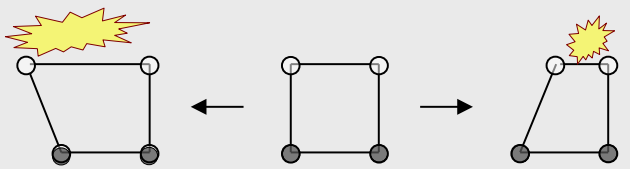
Normal modes directions allow for “intelligent” deformation of a structure

Changing the shape of a structure (macromolecule) involves changing the atom coordinates

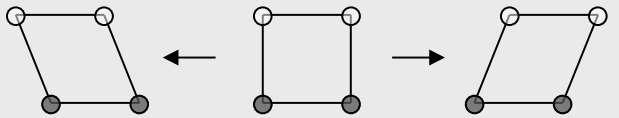
In general, the Cartesian coordinate axes are not aligned with the principal axes of the hyper-parabola described by the mass-weighted Hessian

Normal coordinates are “natural” coordinate axes





Displacement along a single Cartesian coordinate



Displacement along a single normal mode coordinate moves all coupled atoms simultaneously

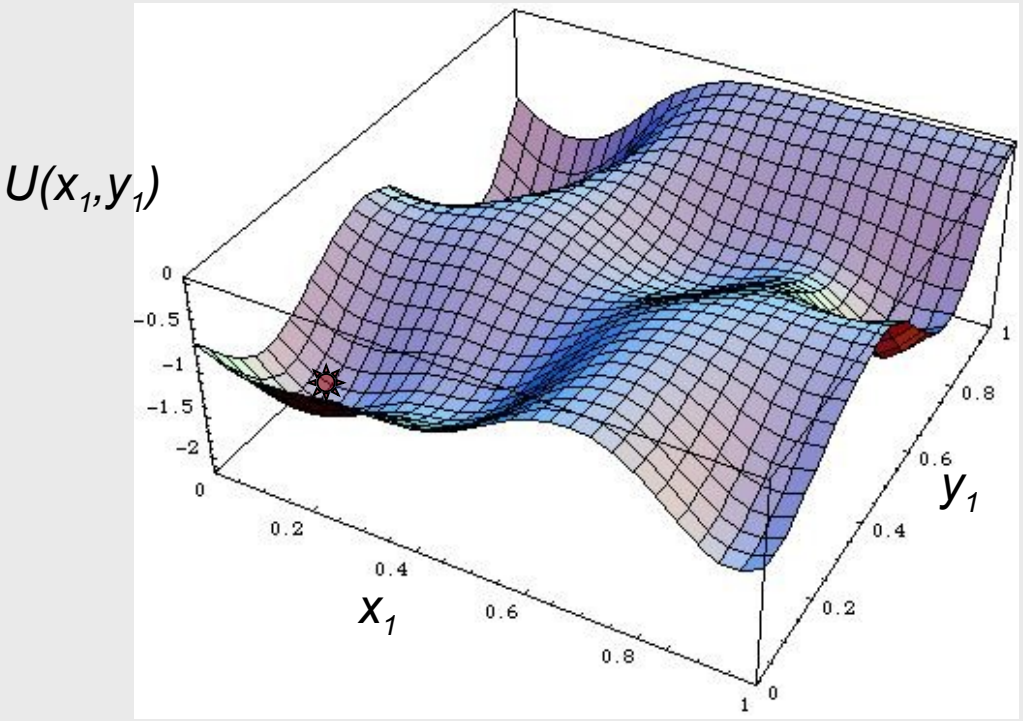
Normal Mode calculation, summary

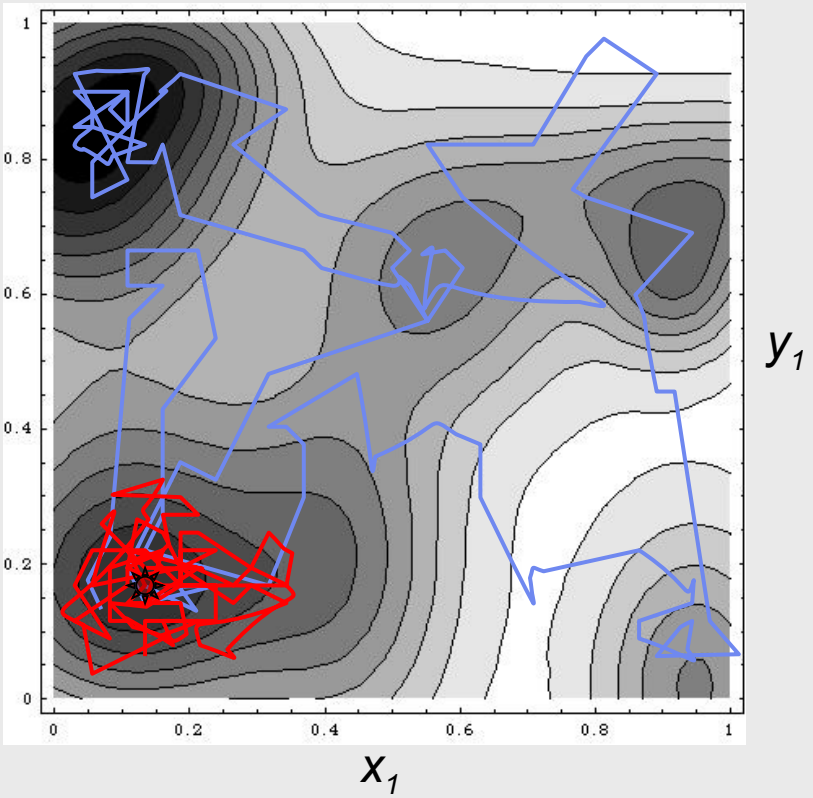
- Minimize the potential energy V
- Change to mass-weighted coordinates, expand V to quadratic terms (def nes the Hessian matrix)
- Diagonalizing the mass-weighted Hessian matrix gives vibrational solutions
  - Eigenvalues: (squared) frequencies of vibration
  - Eigenvectors: coordinates q = directions of vibration, implicating all atoms.
  - Each eigenvector is a linear combination of Cartesian atom coordinates
- Normal coordinates correspond to the directions of natural vibrational movement of the structure near the minimum

Normal modes provide dynamic information without MD

With MD, in principal one can explore entire conformational space (3N dof)

- In practice one is often confined to starting region
- Harmonic approximation is not as artificial as it might appear!





Use of NM for proteins

1977

- BPTI Molecular Dynamics [McCammon, Gelin, Karplus (1977) Nature]

1982

- BPTI Normal Modes [Noguti and Go (1982) Nature; Levitt et al. (1985) JMB]

1990's

- Simplified normal modes (ENM, GNM)
- NM projections on conformational differences
- Simplified normal modes (ENM, GNM)
- Mode exploration

2000's

- Crystal structure refinement
- Flexible docking
- Reaction path estimation

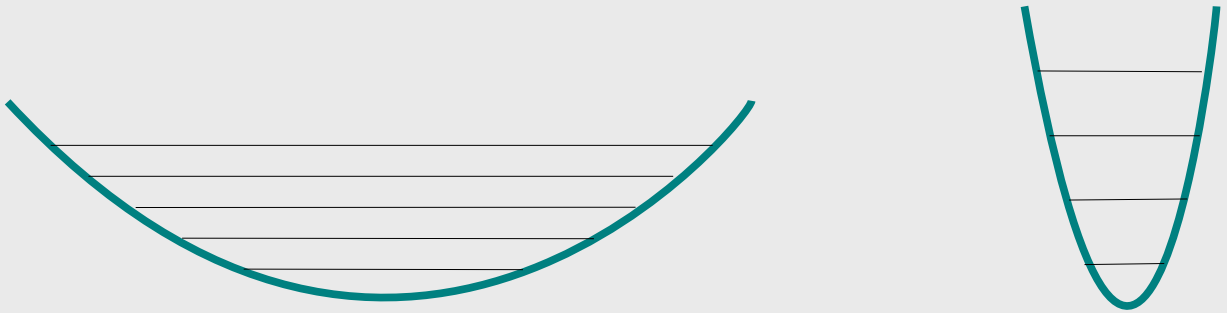
Entropy estimation

Important for rationalizing or predicting free-energy changes

- Binding affinity calculations require entropy of bound and free states

In the harmonic approximation (normal modes), the entropy is calculated from the volume of the potential energy well

- Larger entropy for larger-amplitude (lower frequency) vibration (larger box)
- Analytical calculation



Energy levels are spaced closer together in the shallow potential -- larger number of occupied states -- higher entropy

Consequences of normal coordinate description

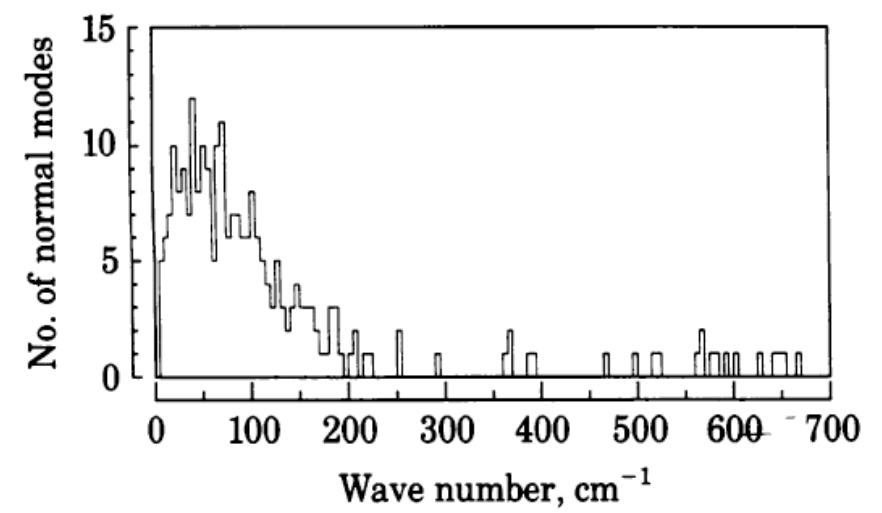
For a given vibrational mode, all atoms move in a particular direction, and at the same frequency

Independence of vibrational modes:

- Exciting one normal mode doesn't excite another
- Can speak of the vibrational energy of a mode, associated with a vibrational frequency (cm-1)
- Higher frequency = more localized motion
- Lower frequency = more “collective” motion -- modes often called “collective movements”

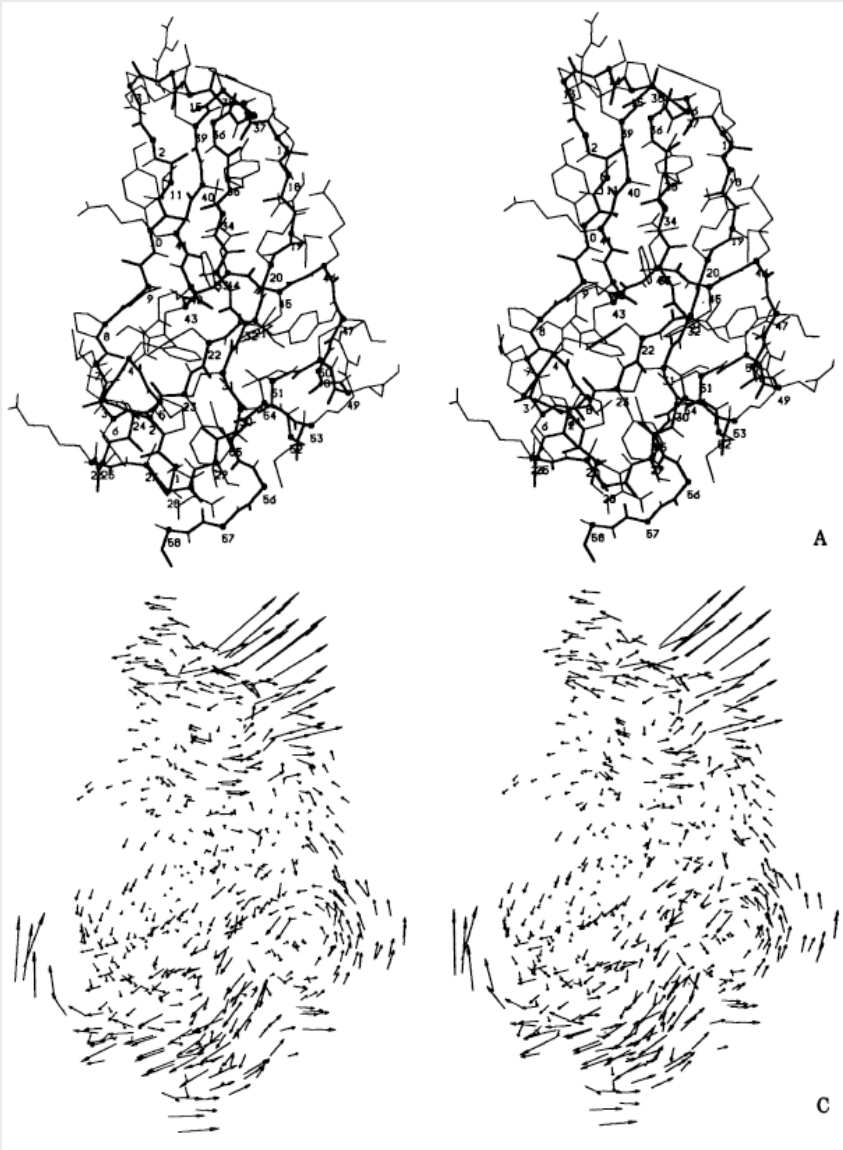
Vibrations in actual solvent are overdamped (friction)

- Absolute frequencies (eigenvalues) used mainly for ordering
- Directions of vibration (eigenvectors) are more physically meaningful



Frequencies of vibration for BPTI

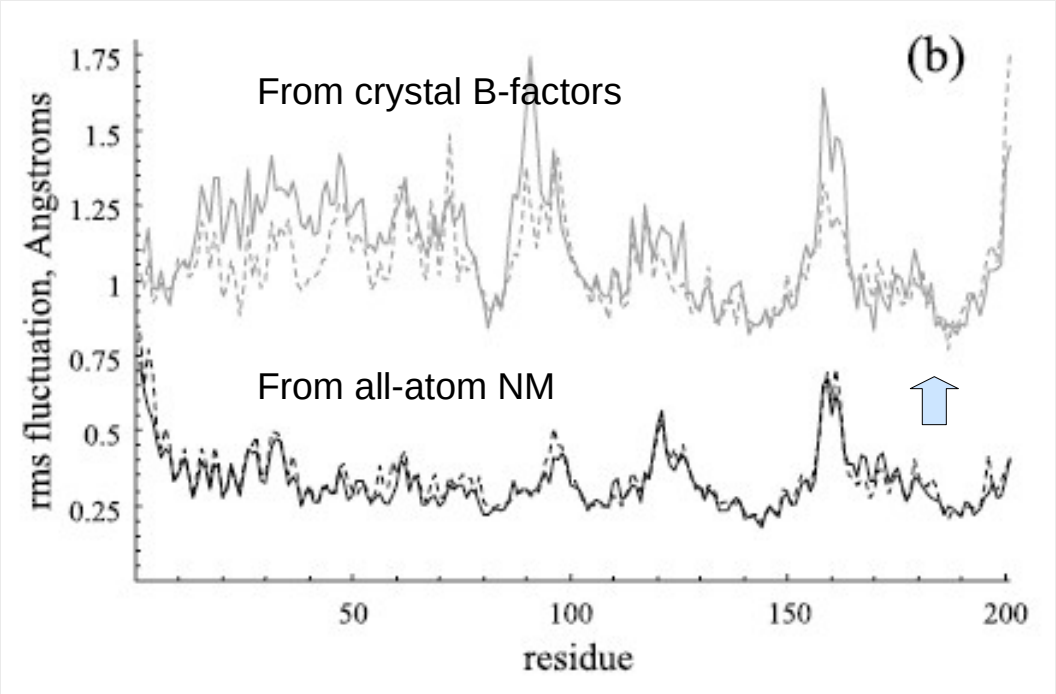
Go et al (1982) PNAS 80, 3696



Atom displacements in 1st low-frequency mode

Compare NM dynamics to experiment – xtallo

B factor → $B_i = \frac{8\pi^2}{3}\langle \delta_i^2 \rangle$ ← RMS fluctuation



Fluctuations for a nucleotide exchange factor [Robert et al. (2004) JMB]

Crystal structures provide isotropic temperature factors (B factors) describing atom dynamics in the crystal

- Like a standard deviation (écart type) around the average atom position
- Temperature (B-) factor is the surface of sphere containing a given probability of finding the atom center

We can calculate the same quantity using vibrational contributions from all NM's

(for fluctuation of a single atom i=j)

$$\langle \delta_i \delta_j \rangle = 2kT \sum_{k=1}^{3N-6} \frac{q_{ik} q_{jk}}{\omega_k^2}$$

Sum over modes

Typically compare alpha carbon (CA) fluctuations

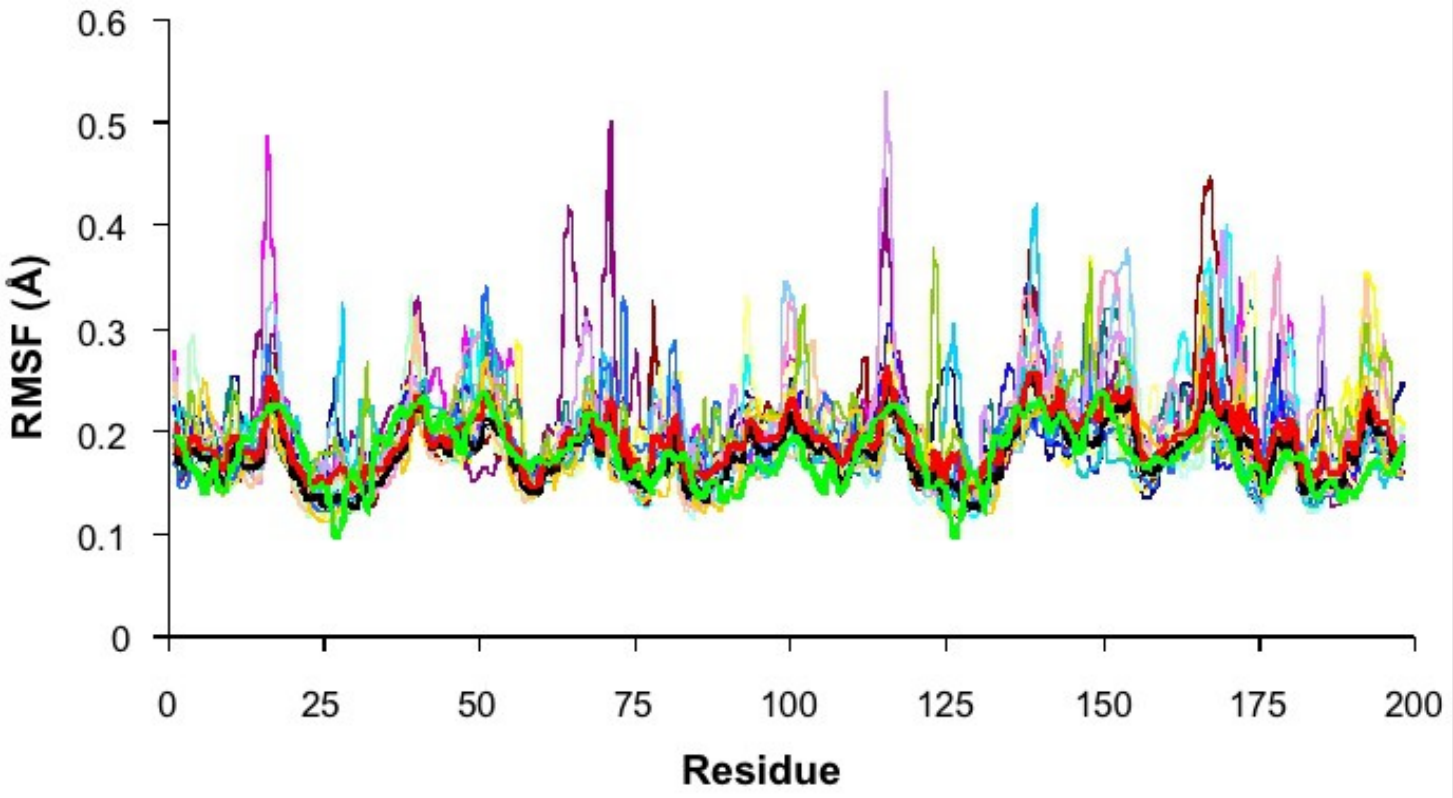
Crystal B-factors are larger because they include other factors

- Crystal disorder
- Model imprecision

Compare NM for different (but related) minima

Compare NM determinations for 20 different structures (snapshots) obtained from MD

Compare to fluctuations derived from crystallographic B-factors (in red) and MD simulation (in black)



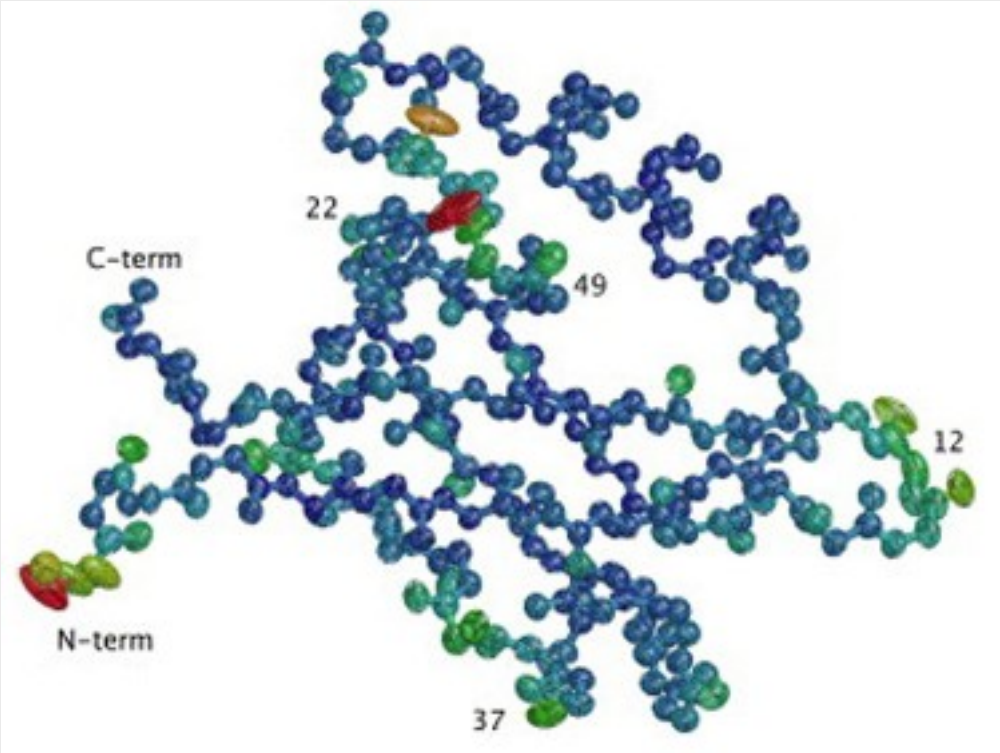
Variability among NM results depending on structure chosen to analyze!

PR Batista, CH Robert, M Ben Hamida-Rebaï, JD Maréchal, and D Perahia (2010) Phys Chem Chem Phys

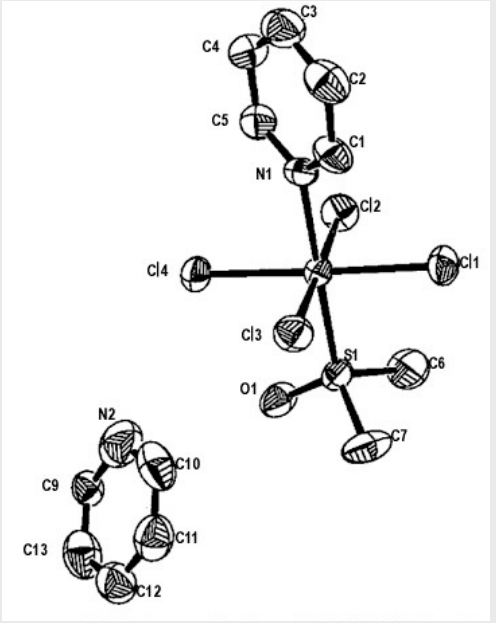
Compare to experiment – xtallo (II)

Excellent quality X-ray crystal structures provide ***anisotropic*** temperature factors

Compare movements from NM to principal axes of the probability ellipsoids for 83 high-resolution protein crystal structures

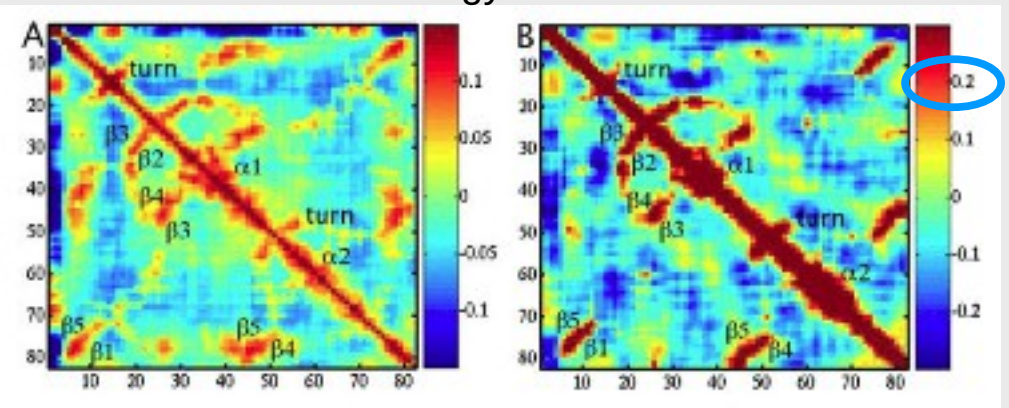


90% probability ellipsoids for CA atoms from the syntenin PDZ2 domain (resolution 0.73 A) colored by degree of anisotropy



Example 50% probability ellipsoids from high-resolution small-molecule crystal structure [DMSO, de Paula et al (2000)]

Atom movement correlation – two different energy models



ENM Charmm Block-normal modes

Kondrashov et al. (2007) Structure 15, 169.

Pertinence of normal modes for understanding conformational changes

- Superpose structures of a protein in two different conformations

  A = closed, B = open

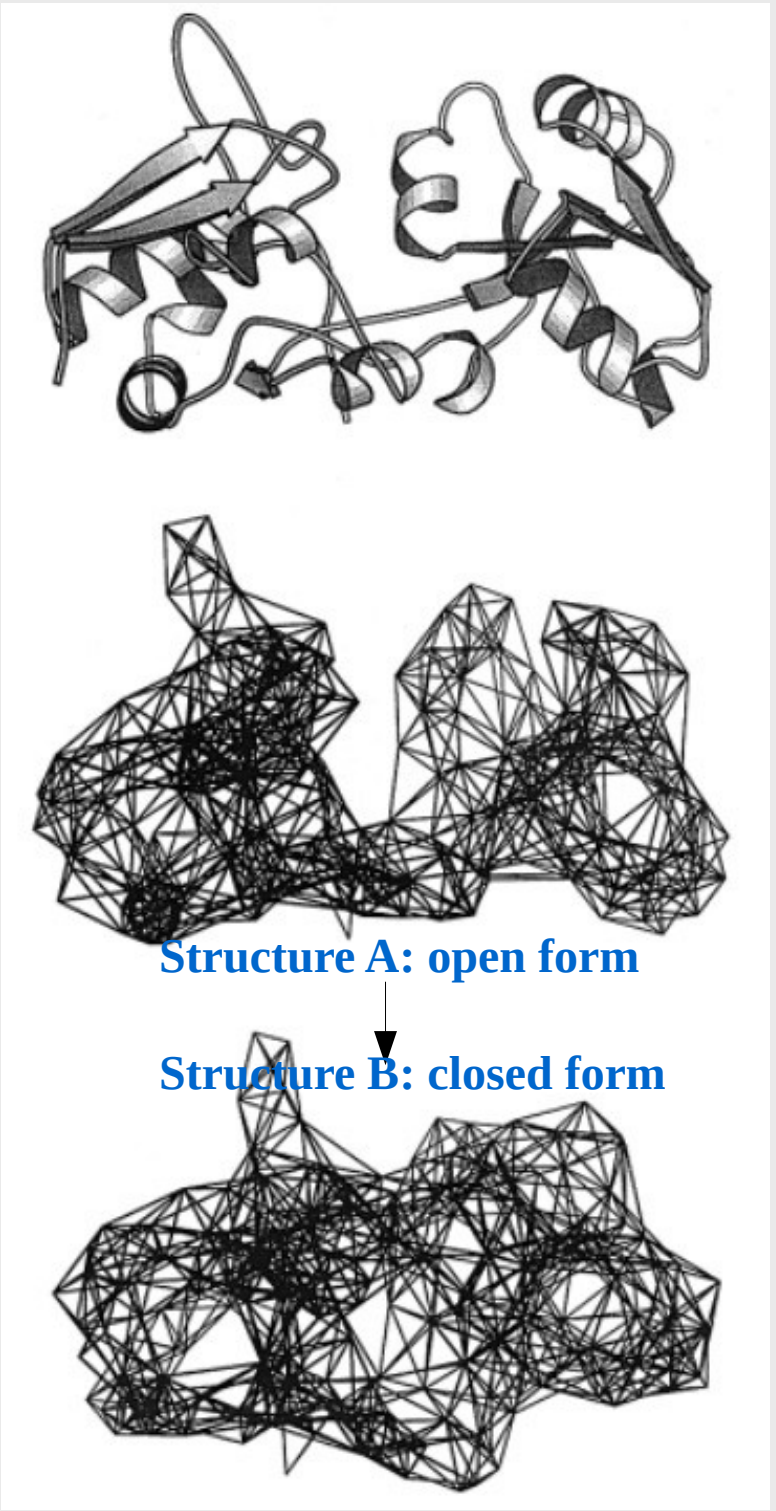
- Calculate the conformational difference vector

$$\Delta \boldsymbol{x} = \boldsymbol{x}_{\text{closed}} - \boldsymbol{x}_{\text{open}}$$

- Mode involvement (or overlap) of a mode can be def ned as the ***projection*** of the mode vector onto the conformational difference vector

$$\text{projection}_i = \frac{\boldsymbol{v}_i \cdot \Delta \boldsymbol{x}}{\|\boldsymbol{v}_i\| \|\Delta \boldsymbol{x}\|}$$

Here $\boldsymbol{v}_i = \boldsymbol{M}^{-1/2} \boldsymbol{q}_i$, but can use v_i or q_i if CA-only



Tama and Sanejouand (2001) Prot Eng 14, 1

Normal-modes-related approaches

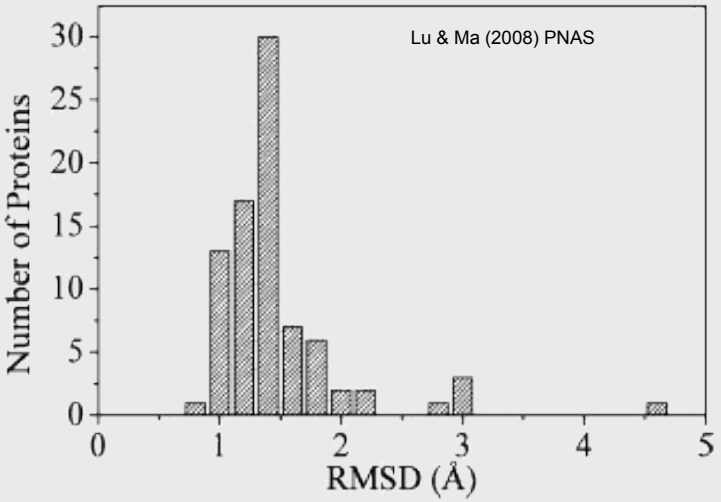
Normal Modes – Pros and Cons

Advantages

- Less computationally demanding than MD
- Identify correlated motions
- Analytical

Disadvantages

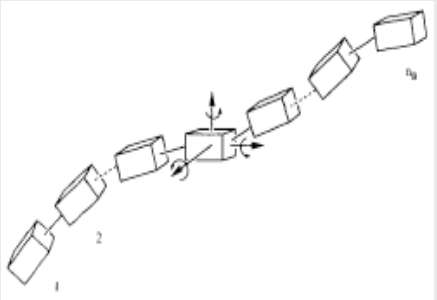
- Extensive minimization needed
  - Can be costly
  - Structural deviation
- Diagonalize large matrices (3Nx3N)
  - Memory/time
- Dependence on initial structure
- Solvent effects poorly incorporated
  - Single solvent conf guration if any
  - Heuristic distance-dependent dielectric
- Strict application only to small displacements about minimum
  - Large-amplitude movements are nonsensical
  - Disadvantage for conformational searching

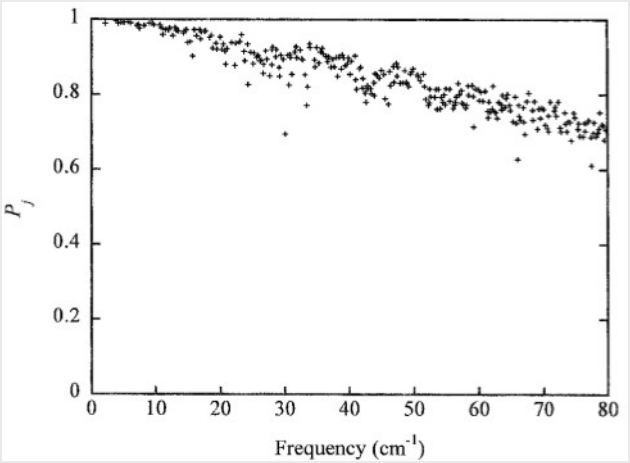


Structural deviation from crystal structure upon energy minimization of 83 proteins

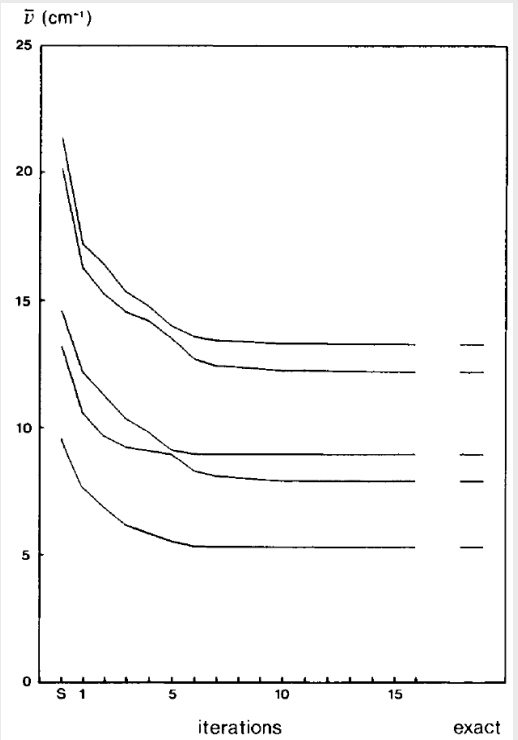
Rotation-translation blocks (RTB)

- All-atom NM calculations are memory-intensive for large systems
- Partition the polypeptide chain into blocks (e.g., a single amino acid)
- Combine rotations and translations of blocks to calculate approximate low-frequency modes (matrix dimension = 6 x number of blocks)
- **Allows treating assemblages of arbitrary size**
- Improved method (Block Normal Modes) includes relaxations within the blocks [Li & Cui (2002) Biophys J 83, 2457]





Quality of RTB mode vectors versus exact



Frequencies converge to exact values

Durand&Sanejound (1994) Biopolymers 34, 759; Tama et al (2000) Proteins 41, 1.

Elastic Network Model (ENM)

Normal modes using a simplif ed potential based on analysis by Tirion (1996) Phys Rev Lett 77, 1905.

Join atom pairs by Hookean springs

Use distance cutoff to limit pairs to a reasonable number

$$V_{\mathrm{ENM}} = \sum_{i,j} C_{ij}(d_{ij} - d_{ij}^{o})^2 \quad \forall \; d_{ij} < d_{\mathrm{cutoff}}$$

Reference distances are original distances in (crystal) structure

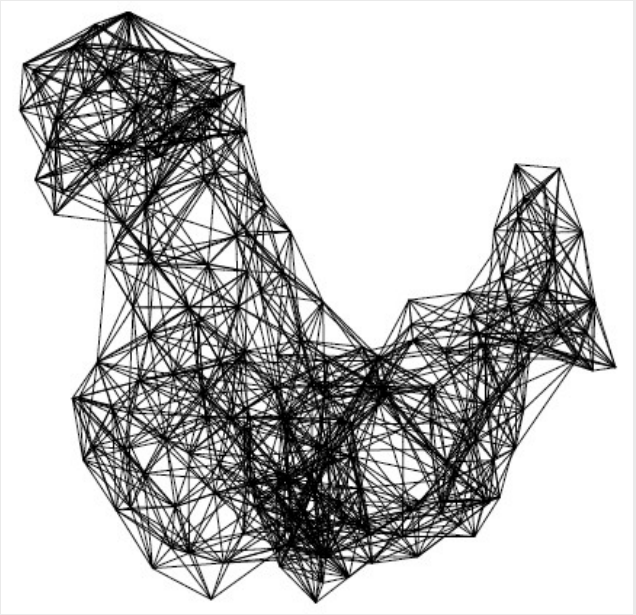
- **No energy minimization necessary!**
- **No structural deviation!**

Typical conventions

- All force constants (Cij) are set equal
- Use residues (e.g., CA atoms) instead of all-atom representation
- Cutoff at 8-13 A

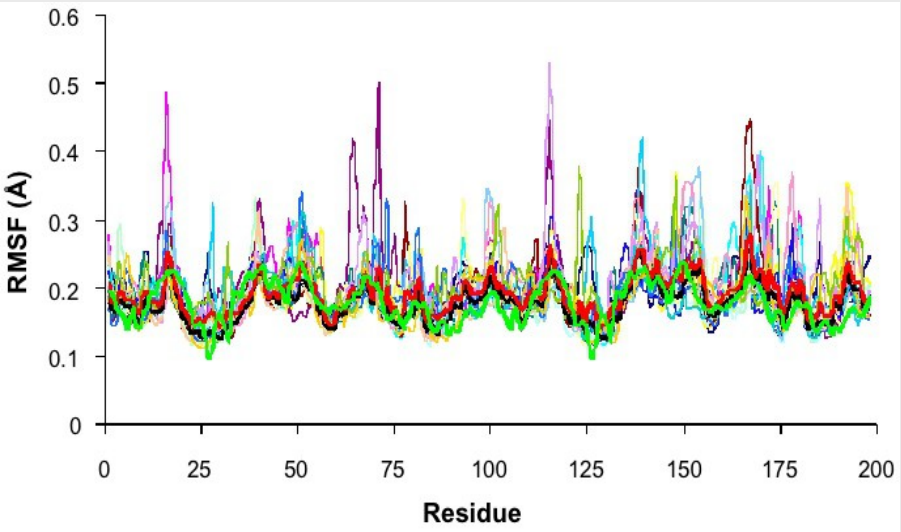
Can use inverse weighting instead of cutoff

[Yang et al (2009) PNAS 106, 12347]



Adenylate kinase showing pairwise distances with d_{cutoff} = 10 A [Sanejouand (2006)]

Consensus Normal Modes



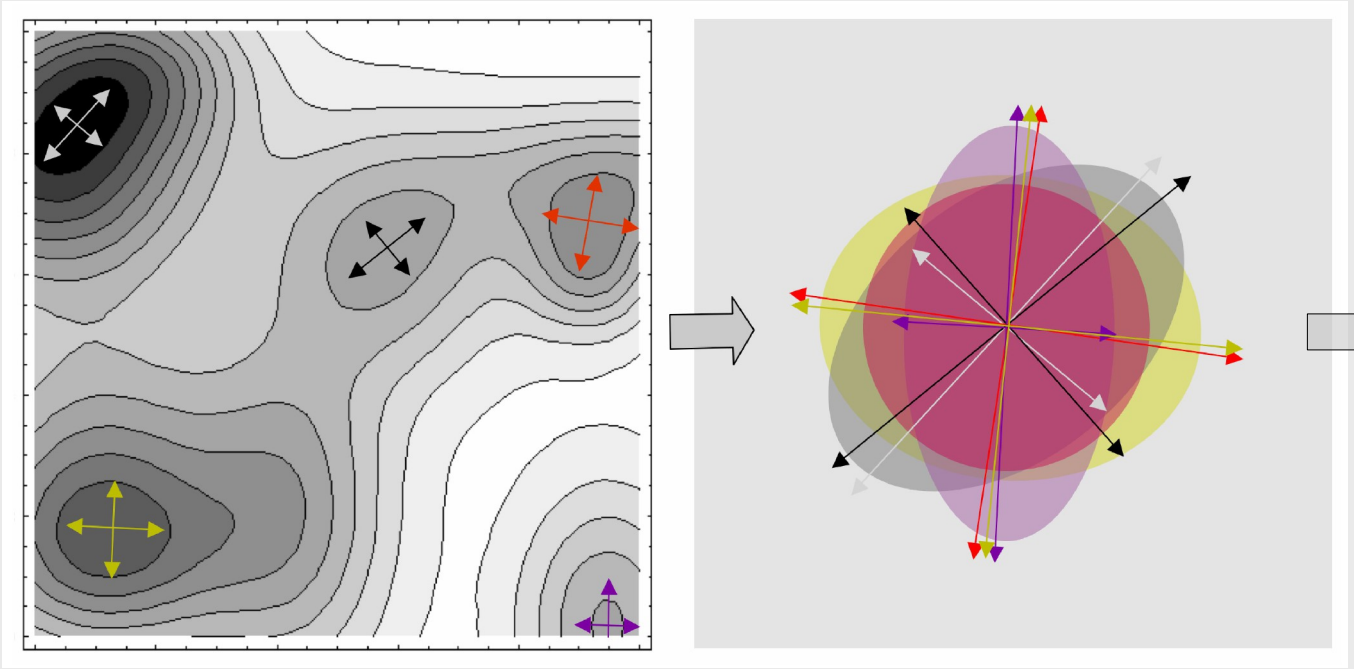
Different local minima have different properties

Combine information for several different minima on the energy surface via the **covariance matrix** (inverse of the Hessian matrix of 2nd derivatives of the energy surface)

Average the covariance matrix over these minima

Calculate normal modes using the averaged covariance matrix

Lessens bias from any single structure determination -- more robust



PR Batista, CH Robert, M Ben Hamida-Rebaï, JD Maréchal, and D Perahia (2010) Phys Chem Chem Phys

Principal Component Analysis (PCA)

Also known as factor analysis, data mining, ...

Requires collections of related structures

- Multiple crystal structures of same protein (e.g., HIV protease)
- Structure determination by NMR provides a family of structural solutions
- MD sampling

Calculate the covariance matrix directly from the sampled structures

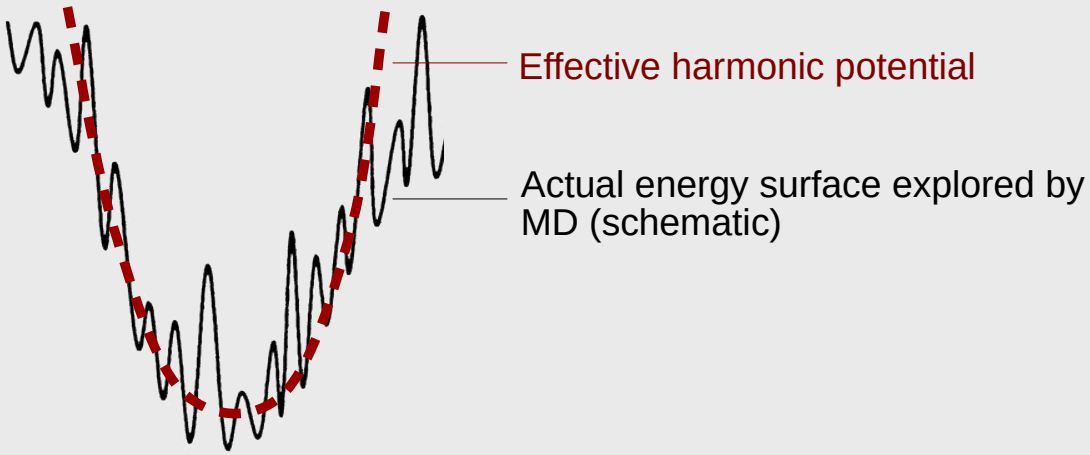
- Superimpose the structures
- Diagonalize the covariance matrix
- Components (eigenvectors) are displacement vectors (like normal modes)

Quasiharmonic modes

- Like PCA, but structures are extracted at regular intervals from MD simulation and mass weighted as in normal modes

- Expect MD sampling of a single basin (single average structure) to follow some normal distribution

normal = quadratic deviations about the mean



- Diagonalize the covariance matrix (inverse of an effective Hessian) to find effective modes

[plan]

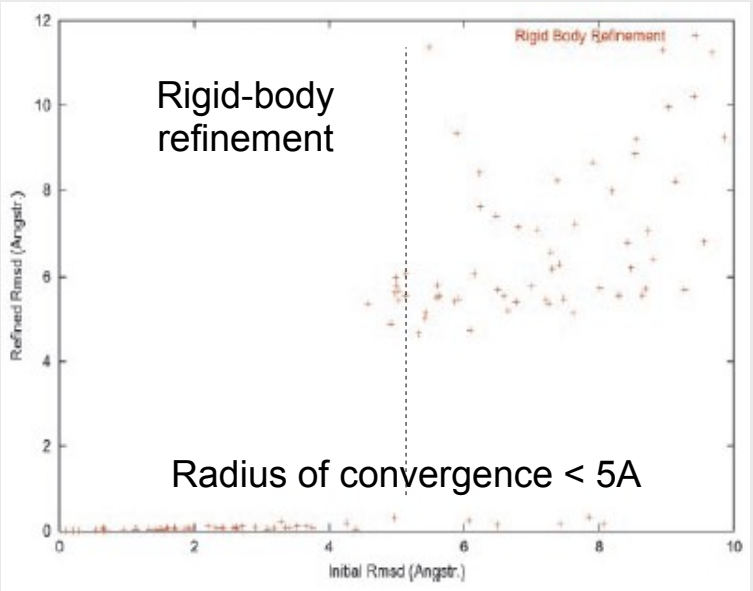
- Macromolecular structure
- Macromolecular dynamics and its role
- Standard approaches to understanding dynamics...
- … simplif ed dynamics --> normal modes
- Use of normal modes in biology and pharmaceutical research

Normal modes in structural biology and drug research

Refining crystal structure data

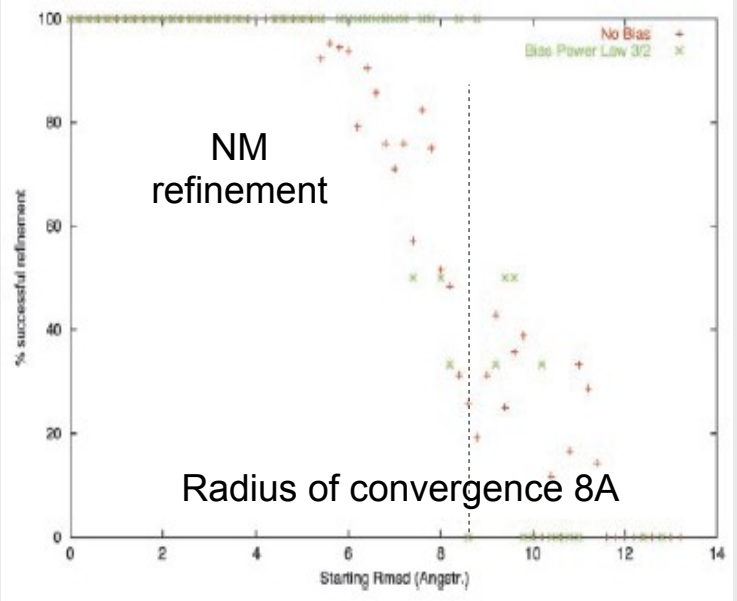
Traditional molecular replacement refinement of crystal structures explores a limited space of rigid-body displacements of a known structure

- Calculate model structure factor, compare to experimental data
- Rotate/translate
- Recalculate structure factor
- ...wash, rinse, repeat...
- --> Minimize R-factor (difference between calculated and experimental structure factors)



Extend this space to include NM displacements of the starting model

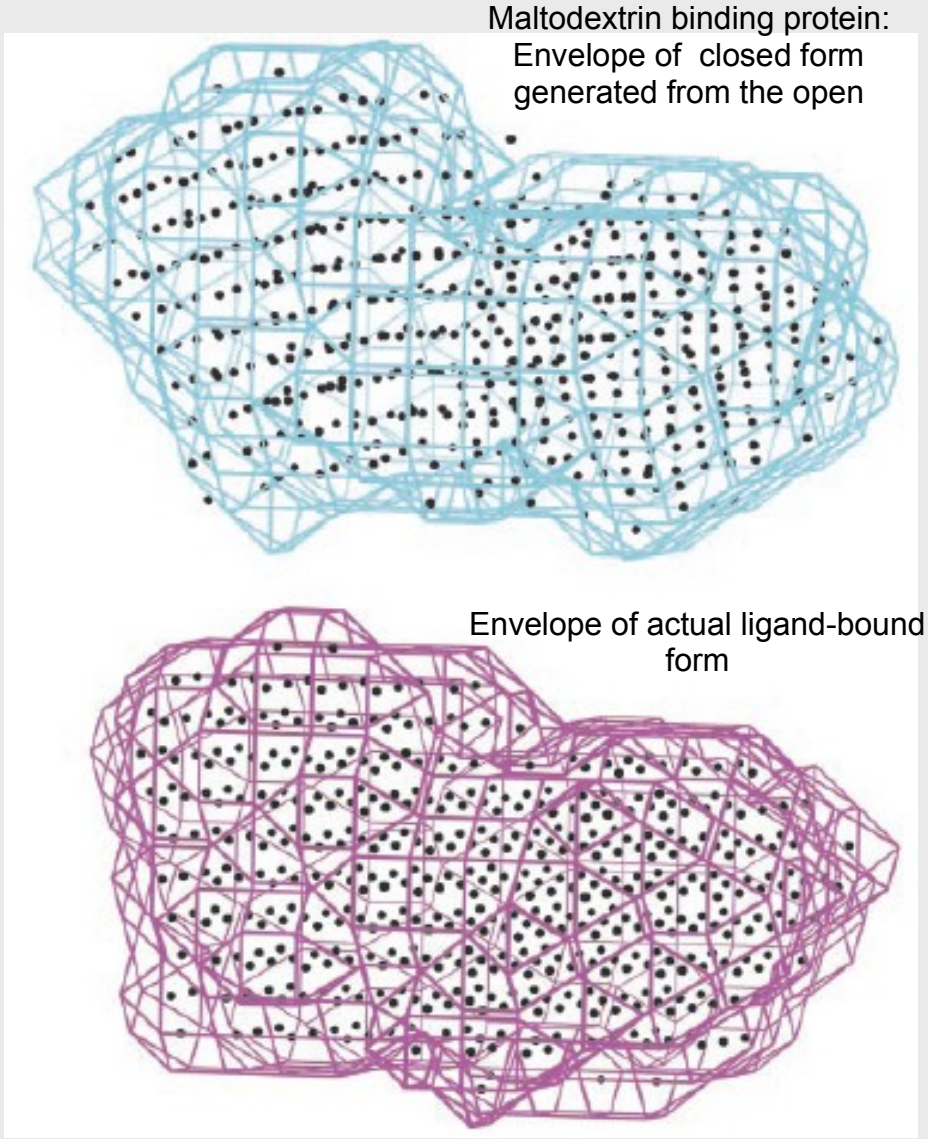
- Pointless to use all-atom model (it's not refined!) Use CA-only (ENM approach)
- Limit search to the first 5-20 normal modes + rotation/translations
- Model "relaxes" to fit data



Use of normal modes increases radius of convergence for successful refinement

Delarue & Dumas (2004) PNAS 101, 6957

Interpreting low-resolution structural data



Electron Microscopy (EM)

- Obtain an “envelope” at 5-10 A resolution
- Often one has high-resolution data for related conditions (e.g. xtal structure for an apo form = no ligand)

Apply an even coarser grained model than ENM !

- Calculate NM for the grid points near CA atoms inside the molecular envelope
- Refine structure factor for this envelope to match data
- Reconstruct atomic model
- rmsd drops from 3.8 to 1.8 A in the example shown

Related applications with Small Angle X-ray scattering (SAXS)

Identifying correlated motions

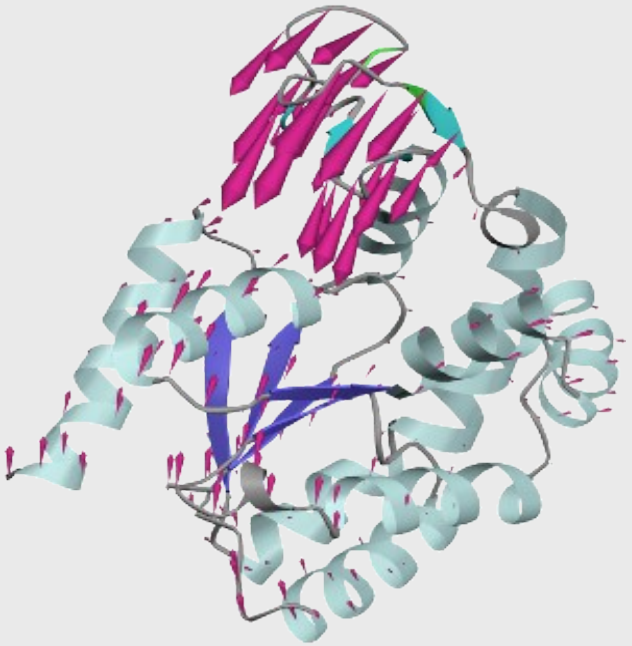
Atoms are not independent – for one atom to move, others must get out of the way

- Atoms are bonded
- Dense medium

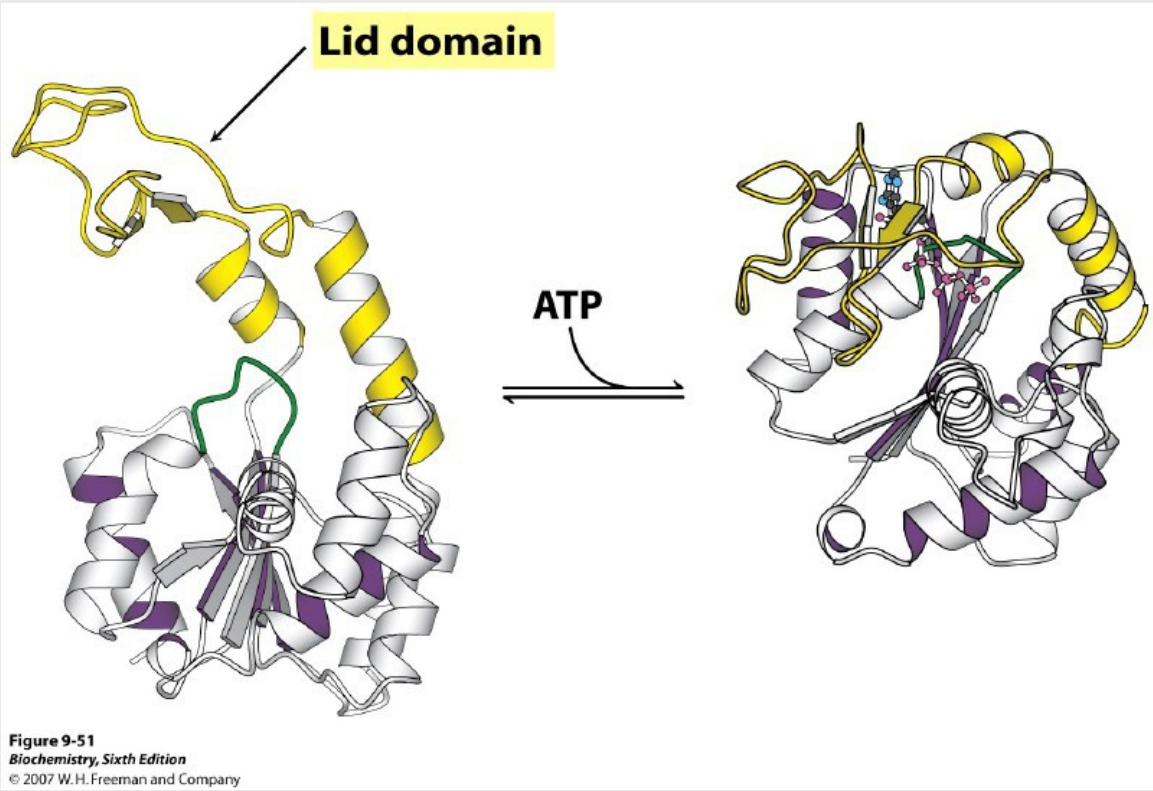
Can identify correlated atom motions

- Sidechain rotamer changes
- Domain movements

Important for interpreting mechanisms



Adenylate kinase. Low-frequency NM movement (Chapman, *Structure* 2007)

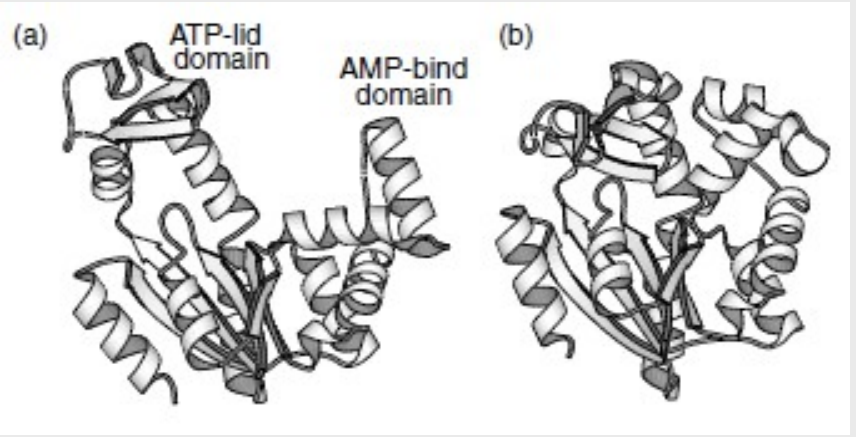


Adenylate kinase. Crystal structures of open and closed forms.

Examples of mode involvement

Domain closing on substrate binding

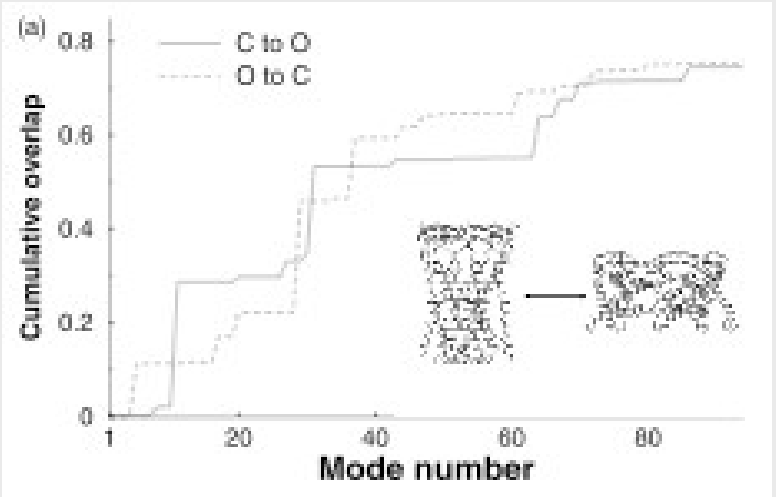
- Adenylate kinase binds ATP+AMP
- Project of each mode onto conformational difference vector
- P_{max} = 0.53 (full all-atom potential)
- P_{max} = 0.62 (simplified ENM)

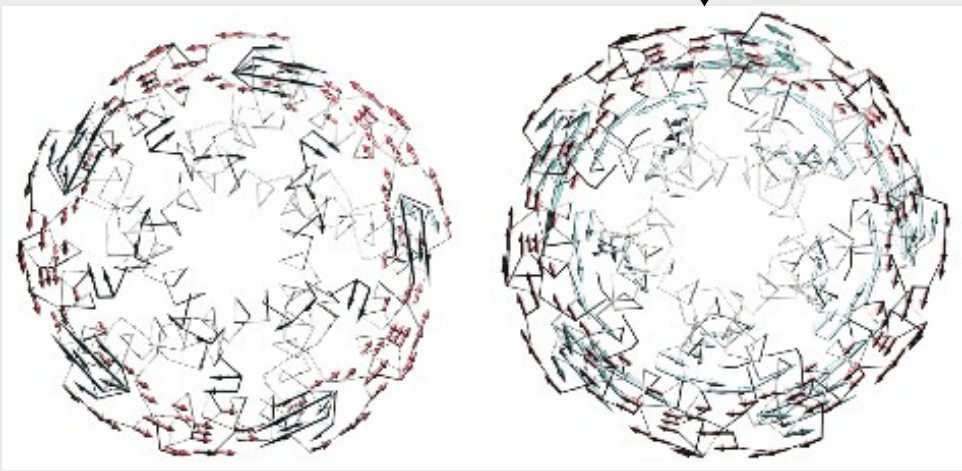


Tama & Sanejouand (2001) Prot Eng 14, 1

Mechano-sensitive channel (MscL) opening

- Closed structure known
- Homology-modelled open structure
- P_{max} = 0.25 (ENM)





Valadie et al. (2003) JMB 332, 657

Examples of mode involvement

Small G-protein activation (Ras, Rho, Arf ...)

- e.g., Arf1-GDP --> Arf1-GTP
- Catalyst is an exchange factor (GEF)
- Crystal structures

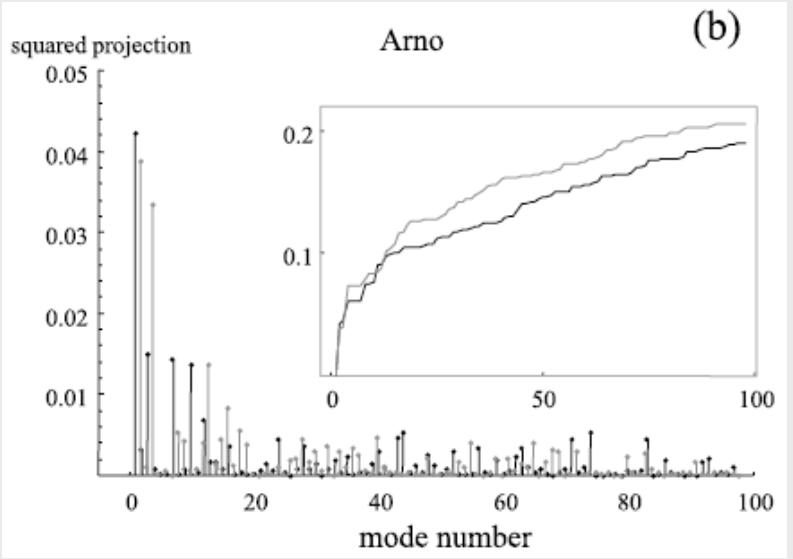
  Four structures of two free GEFs

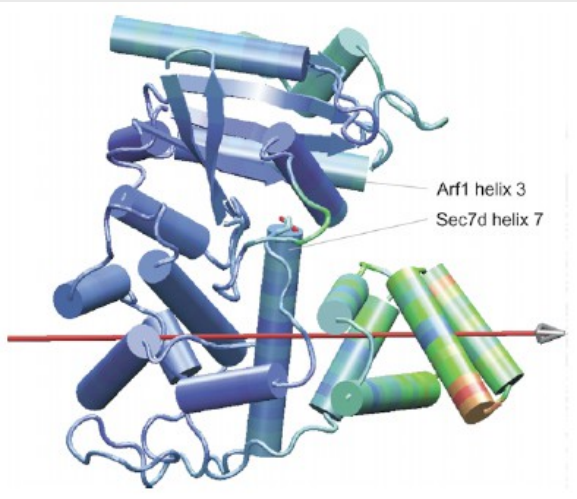
  Nucleotide-free Arf1-GEF complex

Mode movements in free GEFs

- GEF hydrophobic groove closes on extracting “switch 1” region of G-protein
- open --> closed movement in GEF
- Low-frequency modes give large projections on this conformational-difference vector

Low frequency twisting modes in Arf1-GEF complex may help expel GDP





Robert et al. (2004) JMB

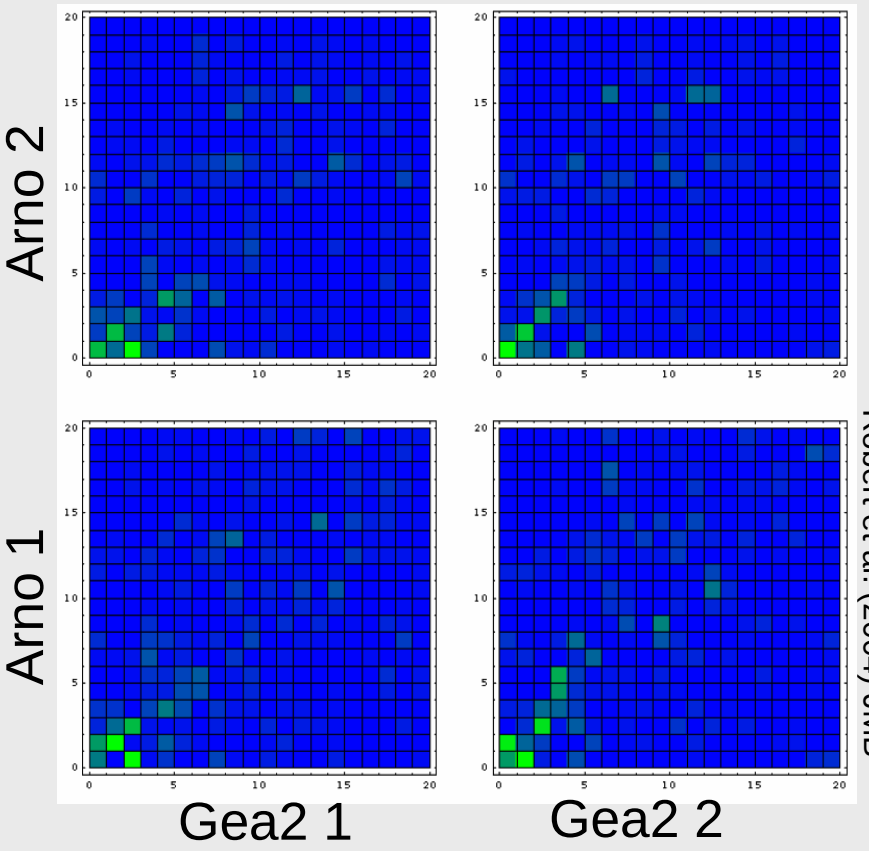
Modes for homologous structures

Do different but homologous proteins give similar movements?

- Pairwise projections of mode vectors for two different structures of two homologous GEFs
  - Arno (human)
  - Gea2 (yeast)
  - 37% sequence identity
- Low-frequency modes have fairly high similarity

Justif cation for ENM approaches

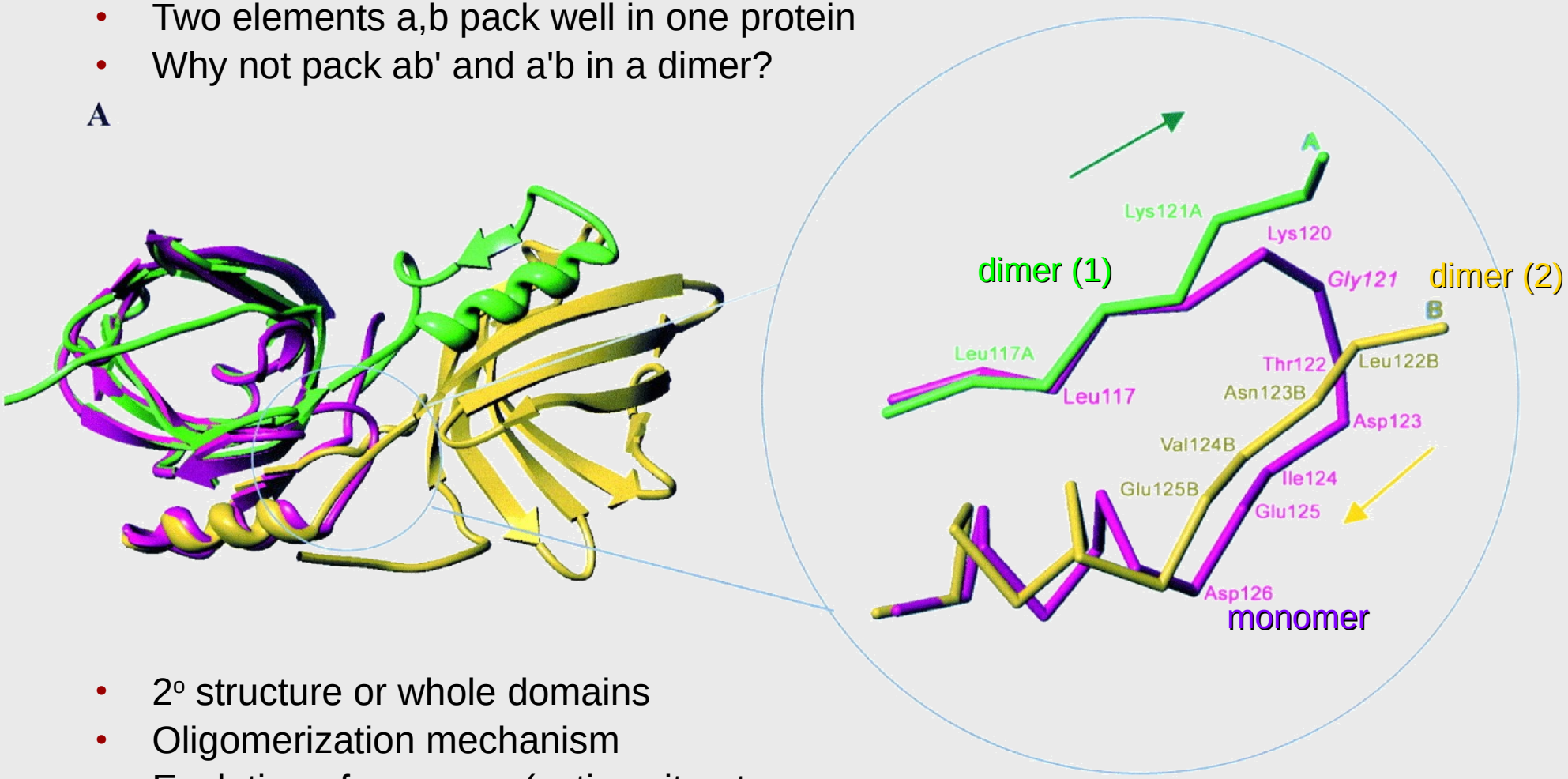
- structural topology is one of the most important factors -- ignores amino-acid sequence



Flexibility in protein protein interactions

Example: Domain swapping

- Two elements a,b pack well in one protein
- Why not pack ab' and a'b in a dimer?



- 2^o structure or whole domains
- Oligomerization mechanism
- Evolution of enzymes (active site at interface)

Herenga&Taylor, Curr Op Struc Biol (1997)

Olfactory Binding Protein (Tegoni et al., 2002)

Flexible protein-protein docking

Prediction of the structure of protein-protein complexes

- **Protein interactions are essential in biology**
- But free structures outnumber complexes more than 20:1
- Similar problem for multidomain proteins

Predicting is fairly easy for rigid-body complexes

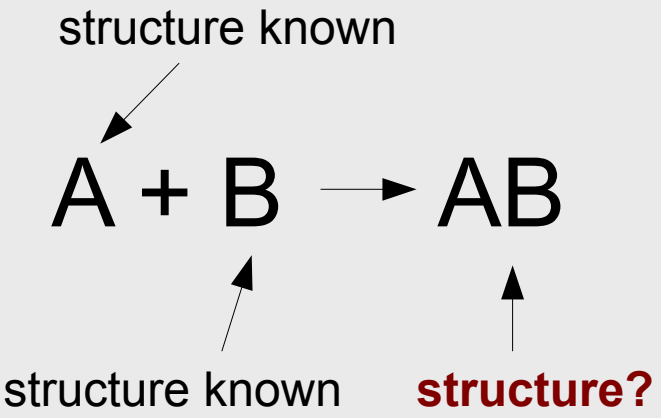
in which structures deviate by < about 1 Angstrom

6 degrees of freedom

= 3 rotations, 3 translations

Association often involves conformational changes

3 (N_1 + N_2) degrees of freedom



Example of flexible docking: FlexDock

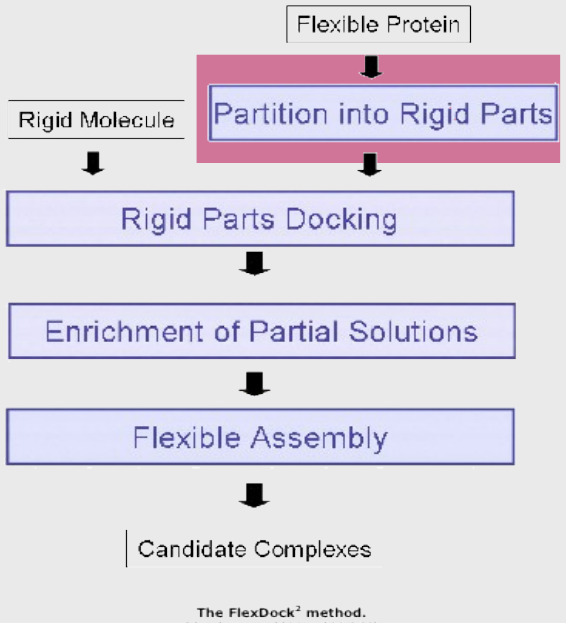
Vibrational analysis (Gaussian Network, related to ENM)

Identify hinges between rigid parts of a flexible protein

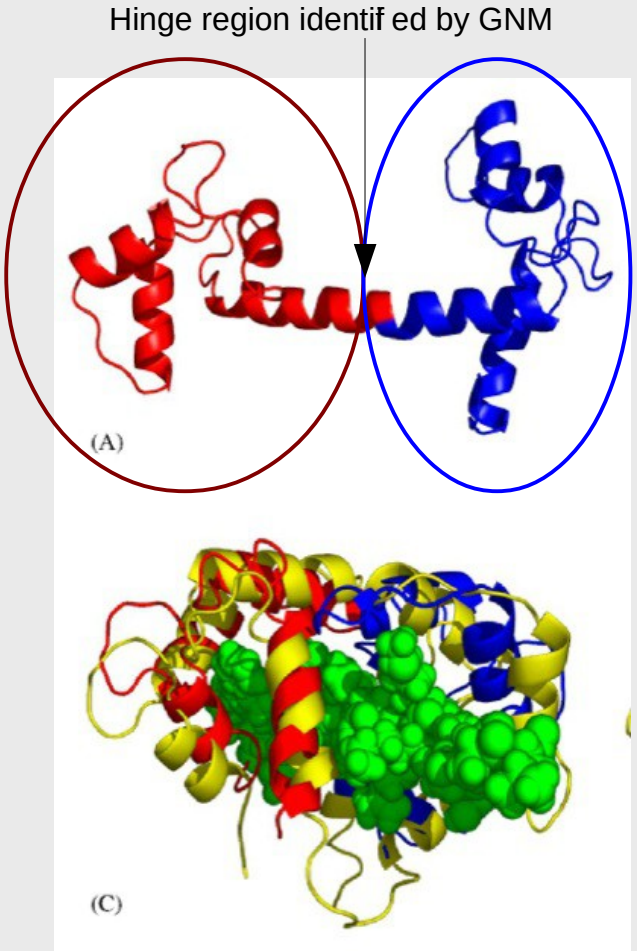
Dock rigid parts separately

Use geometric hashing to filter the resulting docked complexes

Successful prediction of calmodulin docking to nitric oxide synthase peptide



The FlexDock² method.



Wolfson (2005, 2007) Proteins

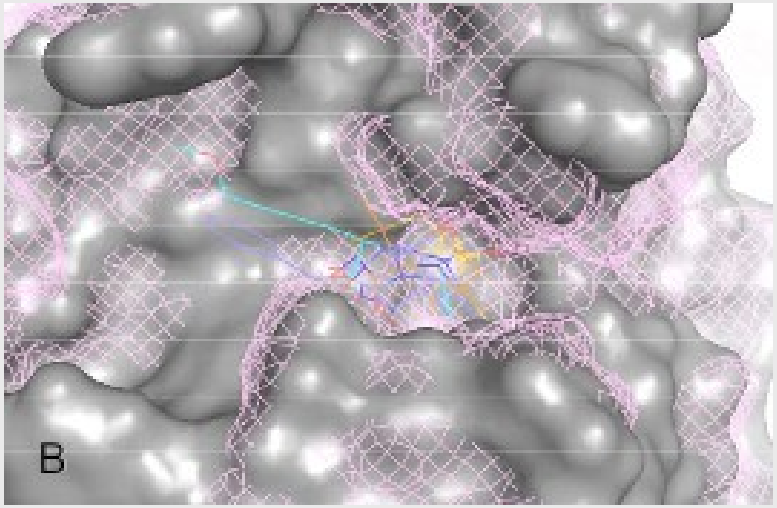
Flexible docking of small molecules to proteins

Example: Inhibitor docking to a matrix metallo-proteinase (MM3)

Only have structure of the protein receptor bound to a small-molecule ligand
Use for docking a novel inhibitor

Remove bias from known ligand by energy minimizing the apo receptor structure...
... but the binding pocket closes

Low-frequency NM allows opening of the binding pocket and successful docking

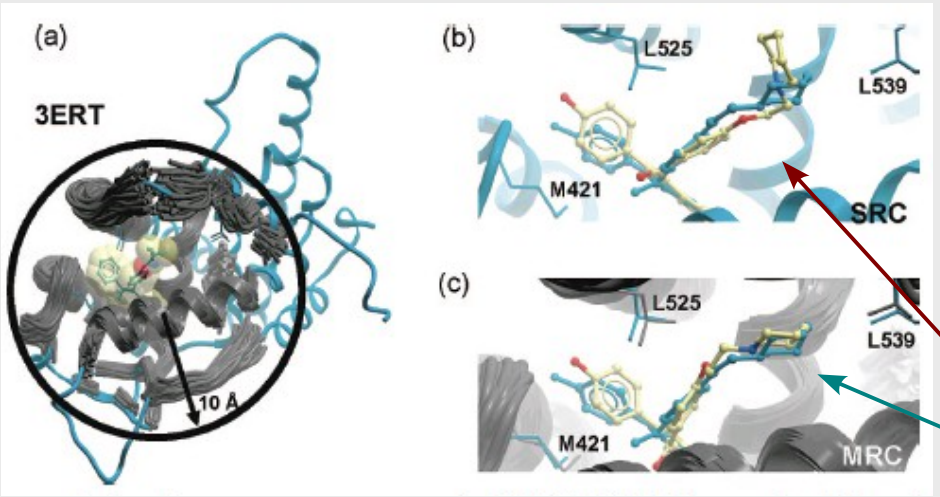


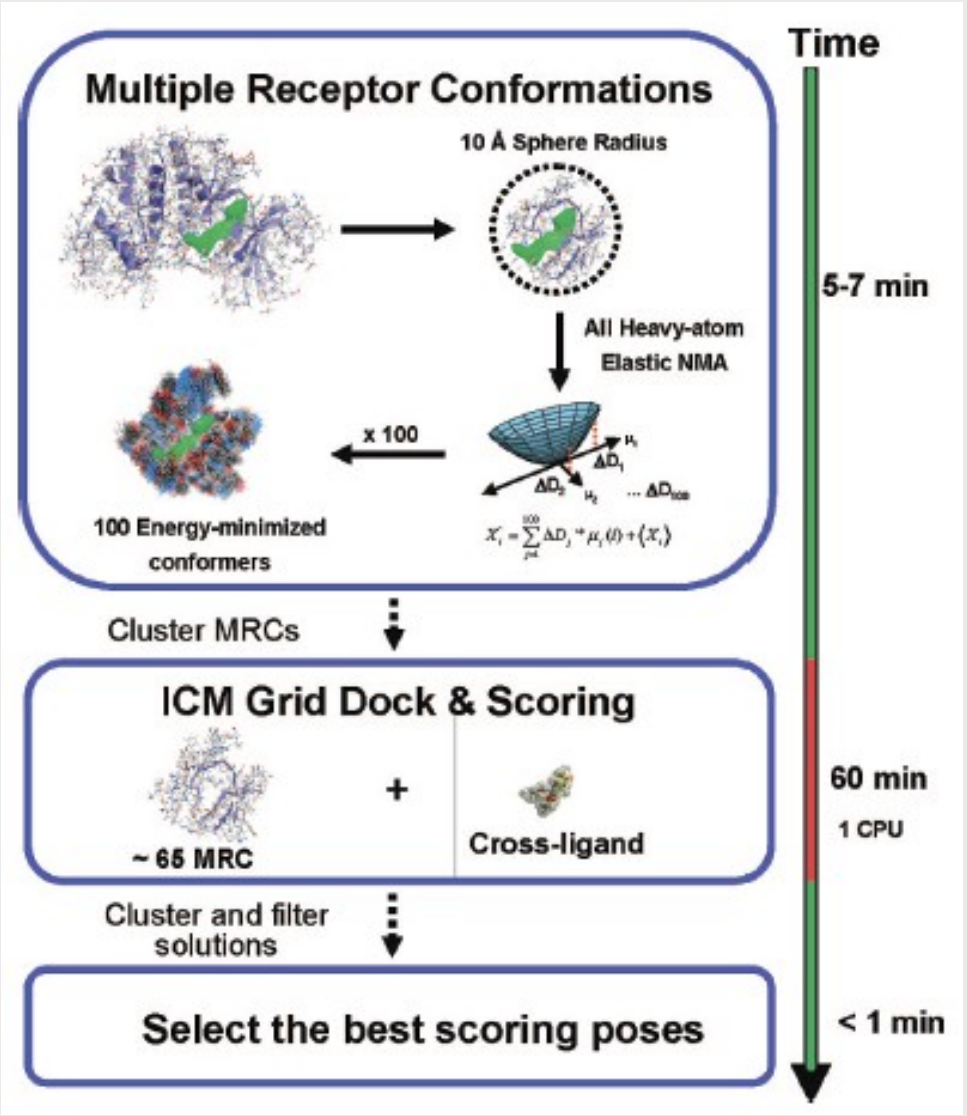
Binding site in the apo structure opened by following a low-frequency normal mode. Closed pocket indicated in pink wire mesh.

Floquet et al. (2006) FEBS Lett

Flexible docking of small molecules to proteins

- ENM calculated in limited region around binding site for speed
- Blind search of NM directions in multiple conformation docking trials
- Improves cross-docking results (test different ligands for same protein)
- Improved docking poses compared to single-receptor trials





High energy, 1.7 A rmsd. (single receptor)

Low energy, 1.2 A rmsd. (NM search)

Rueda et al. (2009) J Chem Inf Mod 49, 716

NM in binding aff nity calculations

Predicting the free energy for a binding process

A+B AB

Exact methods for calculating G:

- long MD simulations in explicit solvent (Free energy perturbation, PMF, ...)
- Expensive in terms of computation time

Simpler approach:

- MD of complex alone, explicit or implicit solvent, to sample both free and bound degrees of freedom
- Estimate solvation free-energy contribution (e.g., MM-PBSA)

How to calculate remaining entropy change (unbound A and B)?

- Normal modes -- vibrational entropy:

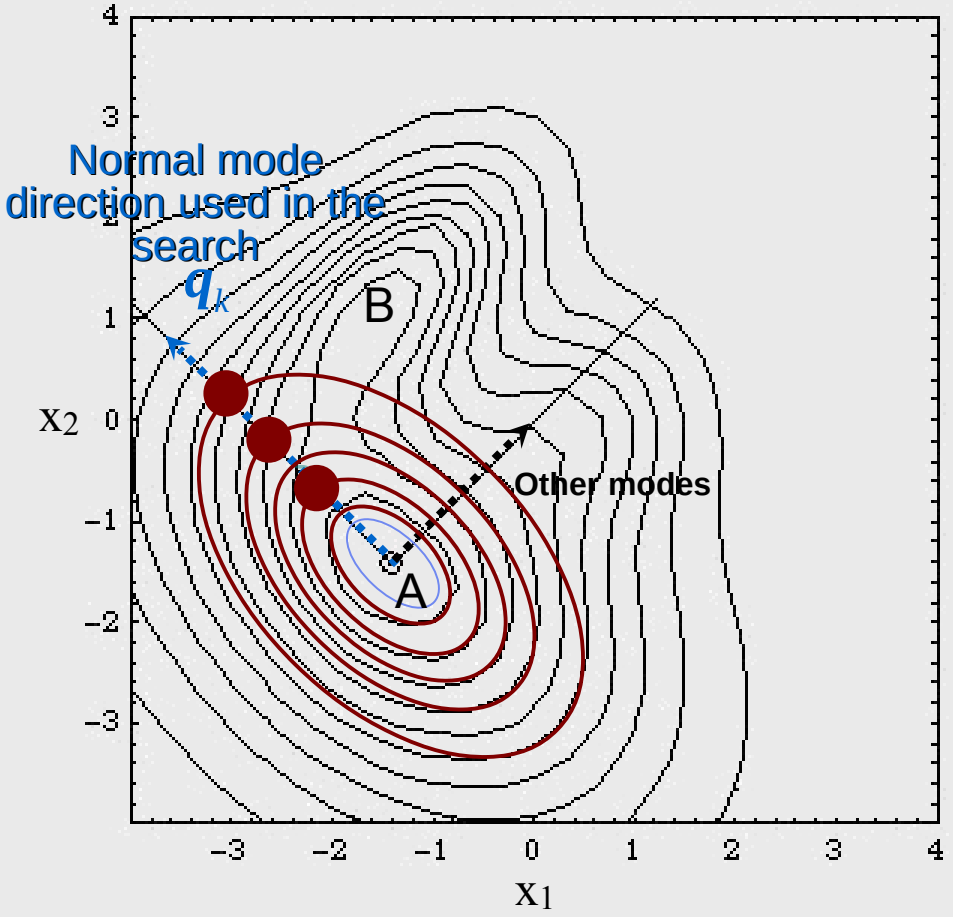
$$S_{vib} = \sum_{j=1}^{3N} \left(\frac{h\nu_j}{kT}\right) \frac{1}{exp^{h\nu_j/kT} - 1} - \ln(1 - exp^{-h\nu_j/kT})$$

- Larger vibrational amplitude (lower frequency)--> larger entropy contribution
- Energy $h\nu = kT$ for $\nu = 417$ cm^{-1}

Enhanced conformational sampling using normal modes

Goal: Change the conformation of our protein by displacing the starting conformation along a chosen normal mode $\boldsymbol{q}_k$

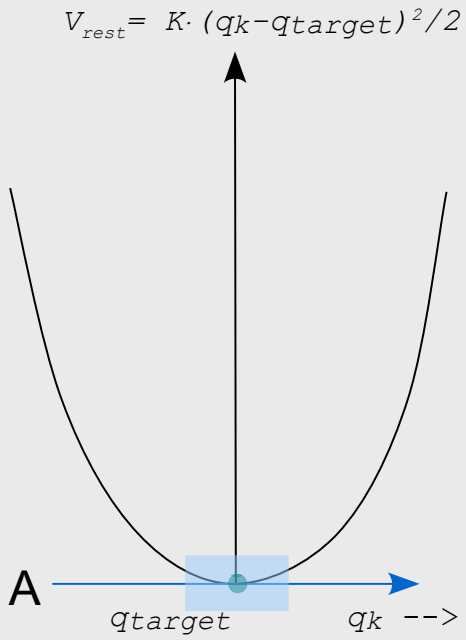
- Displace the structure (●) by moving all the atoms along $\boldsymbol{q}_k$
- Displacements along other modes $\boldsymbol{q}_{m\ k}$ are set to zero: becomes a line search
- E(q_k) increases rapidly as we leave the harmonic (minimum) region of A
- To avoid this, the search should occur in space of all coordinates – this is rarely done

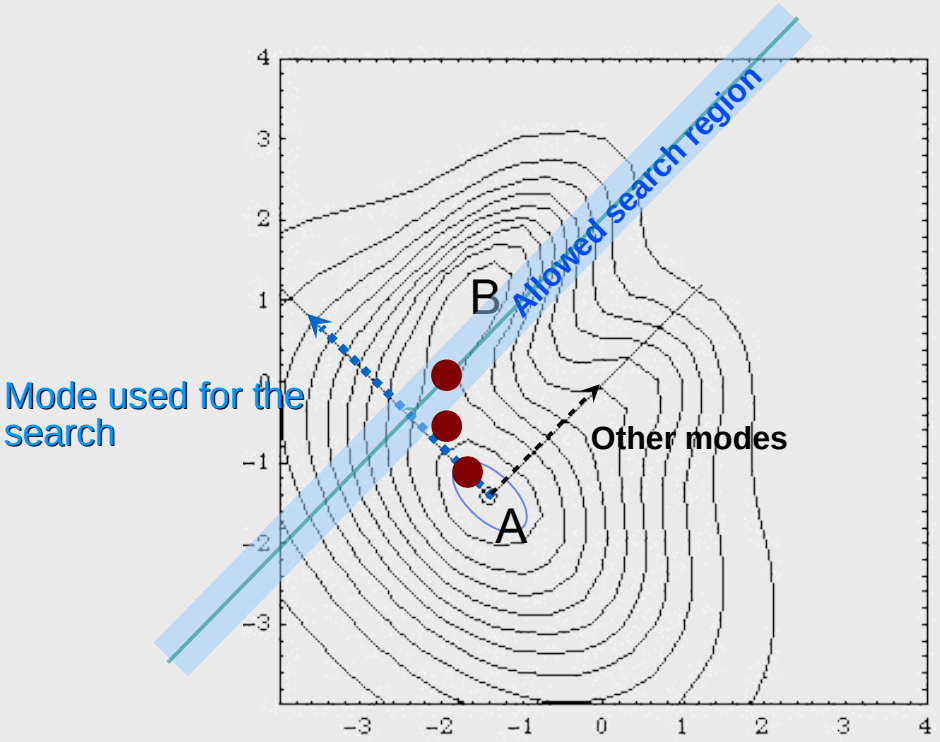


Enhanced conformational sampling

Another approach: restraining the mode coordinate

- Can calculate the coordinate q_k along the vector $\boldsymbol{q}_k$ by projecting the position vector onto the mode vector
- Add restraint term to potential energy in terms of this projection
- Only one degree of freedom is restrained (instead of $3N\text{-}6\text{-}1$ in previous case)
- PMFs can be calculated along mode directions
- Now included in Charmm (VMOD) [Perahia&Robert (2003,2006)]

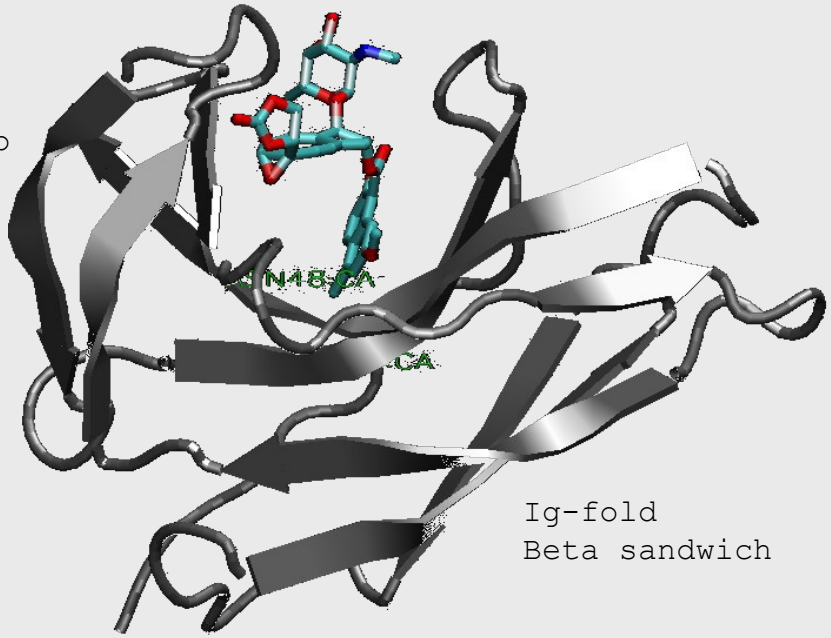


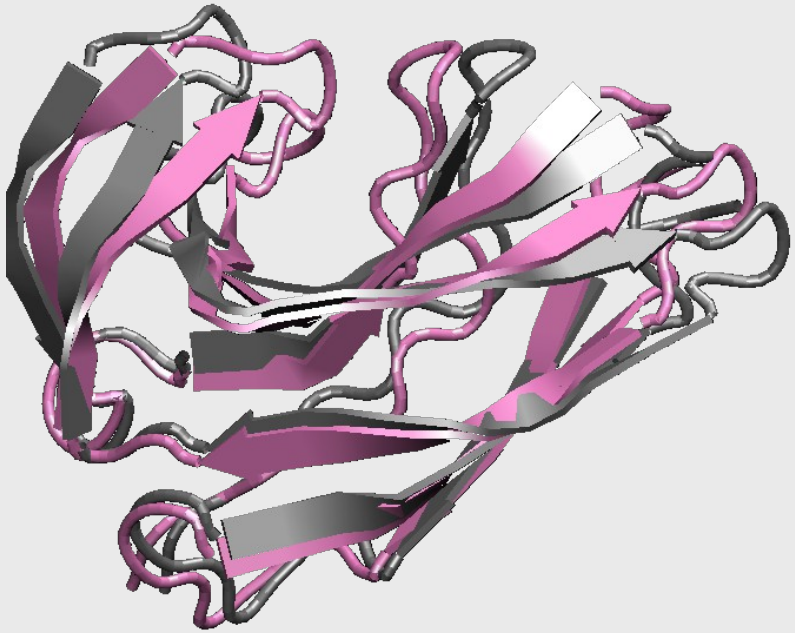


Example: opening/closure of ligand binding pocket

- Neocarzinostatin (NCS): protein + enediyne ligand
  - Ig-fold beta sandwich + double beta ribbon « thumb »
  - Antitumoral activity (digestive system cancers - Japan)
  - Ligand is synthesis target with NCS apoprotein as transporter

- Structure: Beta-sandwich + thumb = binding crevice
  - Directed mutagenesis to bind new ligands
  - Computational protein design to find mutants to optimize binding of a given ligand?

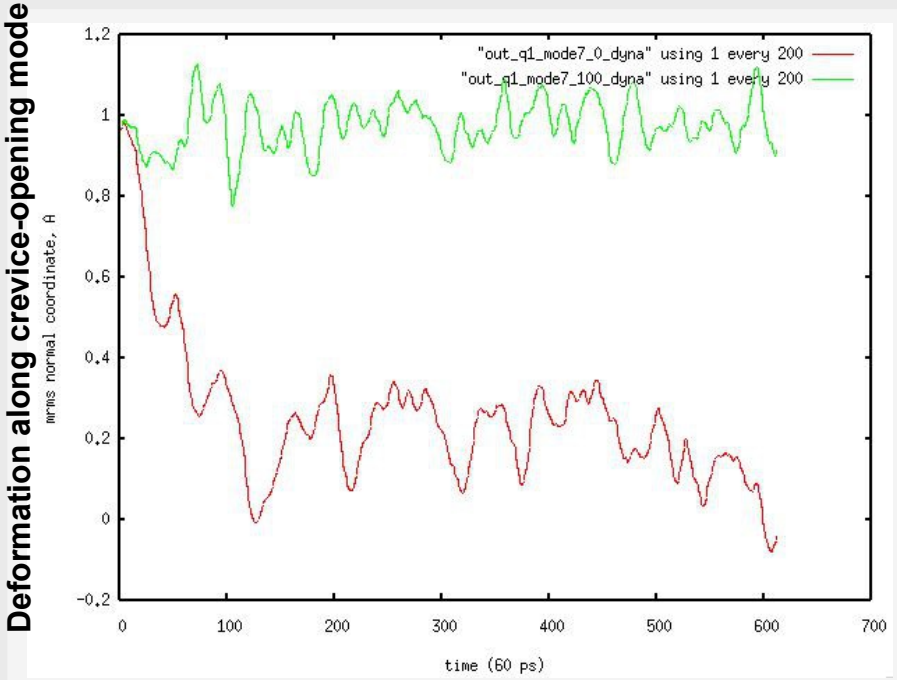


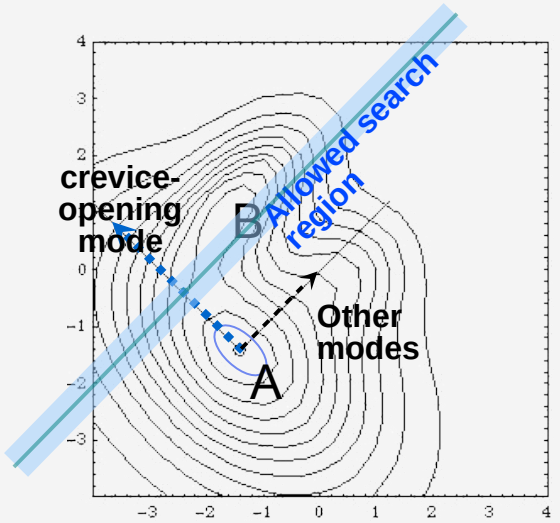


Low frequency normal mode describes opening/closing of binding pocket

Restrained search along a normal mode direction: crevice opening

Restraint is necessary — displacement along mode direction followed by MD returns it to the original structure if mode restraint not present!



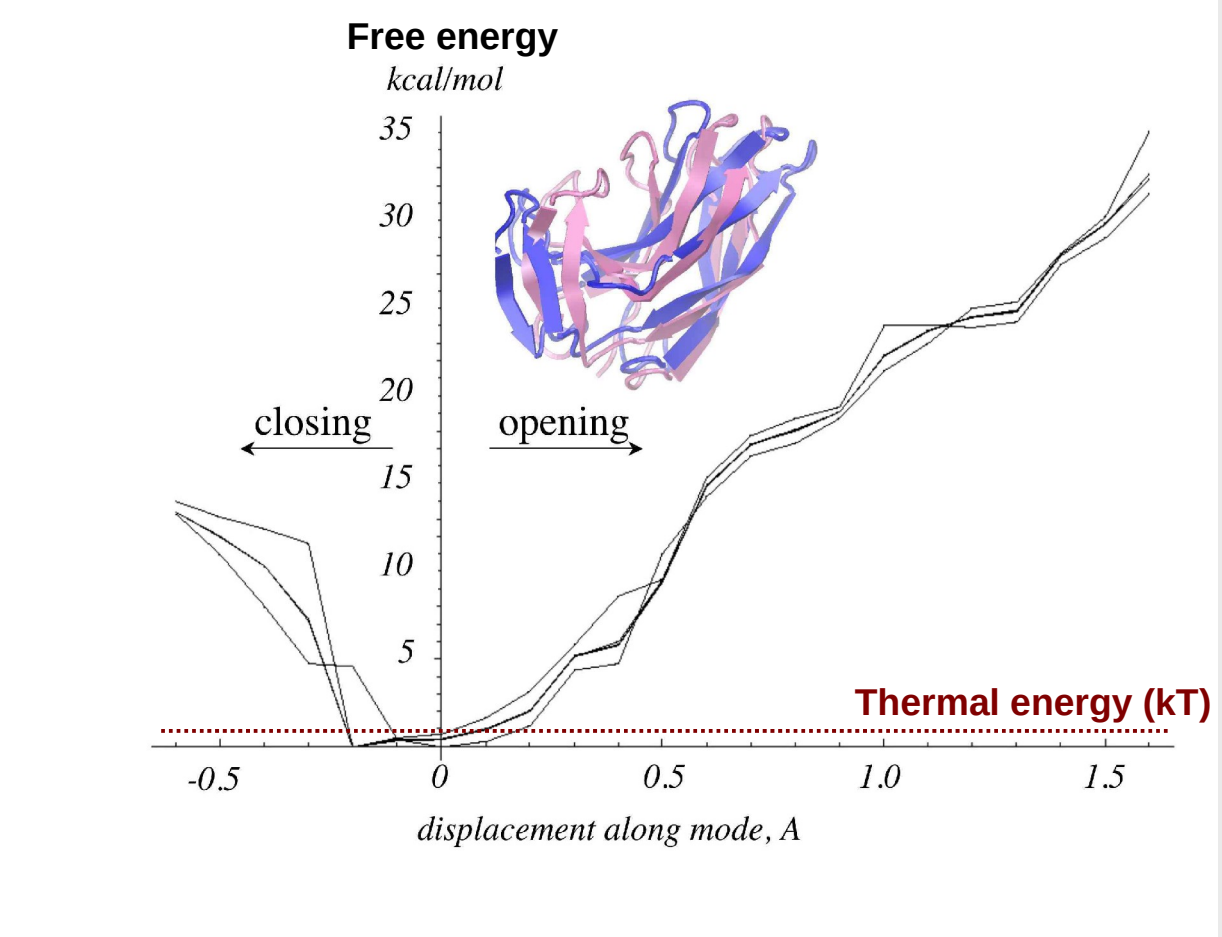


Umbrella sampling along the mode coordinate allows calculation of Potential of Mean Force (PMF)

- estimation of free energy profile along opening/closing coordinate

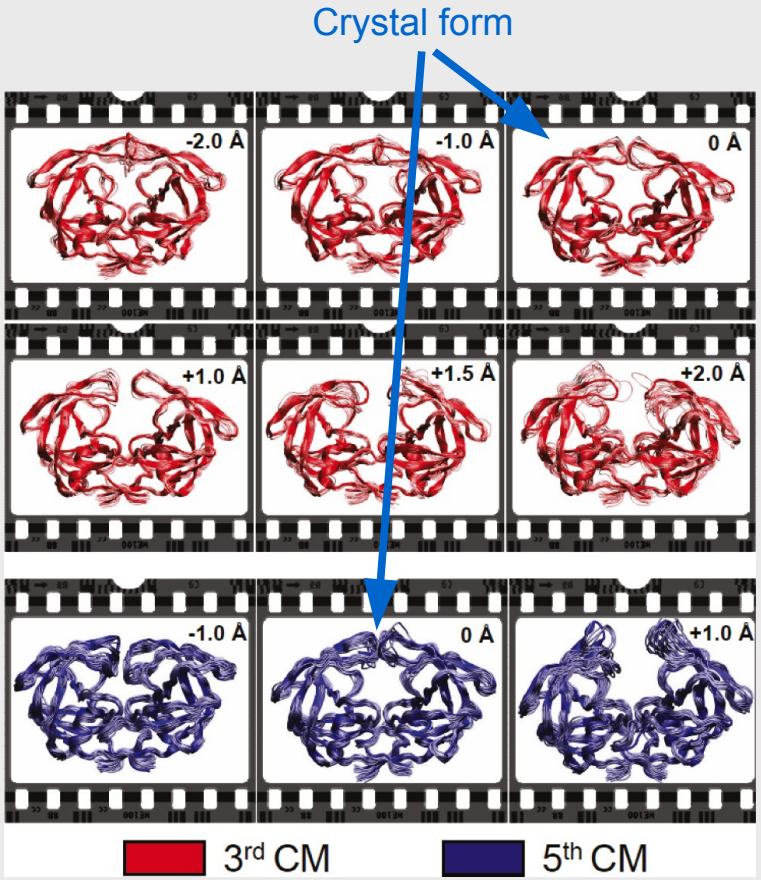
Constrained search along a normal mode direction: crevice opening

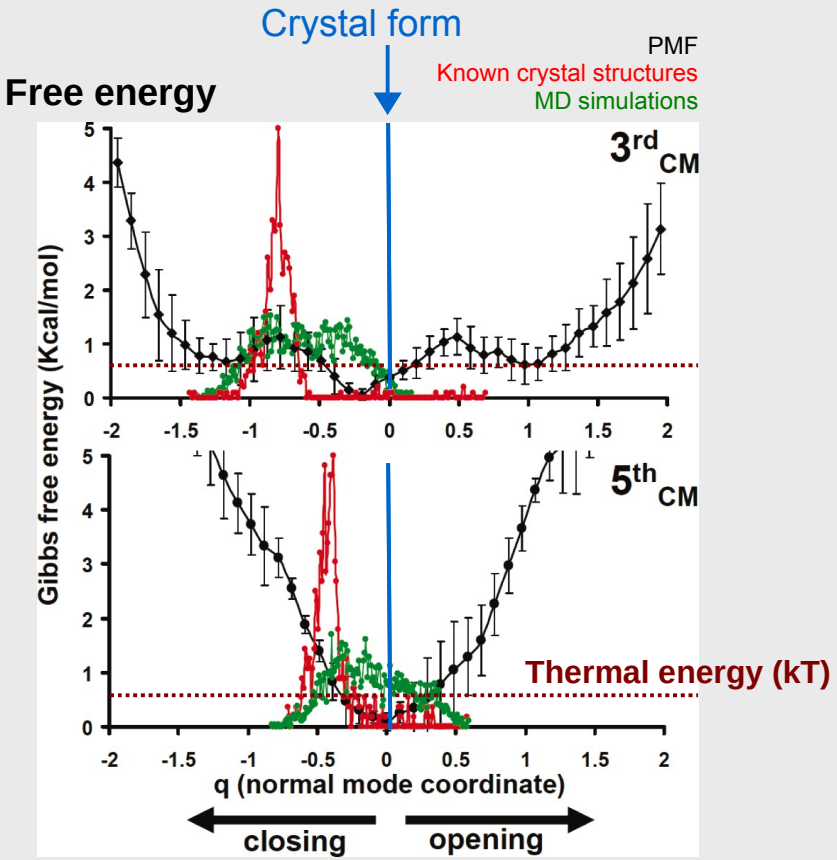
Calculation of Potential of Mean Force (PMF) along the opening coordinate suggests intrinsic flexibility of the binding site



Constrained search along a normal mode direction allows opening the binding site of HIV-1 protease

PMF calculated along two different **consensus normal mode** (CM) coordinates





- HIV-1 protease can open suff ciently to bind substrate with relatively small overall energy change
- Opening has been indicated by NMR but not yet observed in long MD simulations

Batista, Pandey, Pascutti, Bisch, Perahia & Robert (2011) J Chem Theory Comput 7, 2348-52.

Summary

Rationalizing, **predicting**, and **modifying** biological macromolecular function requires more than structure – it requires understanding dynamics

Normal modes analysis

- provides an faster alternative to standard MD approaches
- directly accounts for the highly correlated nature of atomic movements
- facilitates structural data refinement
- facilitates functional model building
- improves predictions of protein-ligand and protein-protein interactions (therapeutic applications...)

Overall Summary

WHAT	Biochemistry/Biophysics
WHY	Simple Models Modelling detailed dynamics
Next:	Tools for intuition: VMD, PyMol
	Biopython Biskit, MDanalysis ...Flexbase

FIN

Acknowledgements

D. Perahia (ENS Cachan)
M. Guharoy, A. Pal (IBPC, Paris)
M. Ben Hamida-Rebaï
F. Cazals (Here!)
P. Chakrabarti, S. Dey (Bose Institute, Calcutta)
H.Van Tilbeurgh lab (FAAM, IBBMC, Orsay)
A. Bonvin, P. Kastritis (Utrecht University)
I. Moal, P. Bates (Cancer Research UK, London)
H. Hwang, Z. Weng (University of Massachusetts, Worcester)