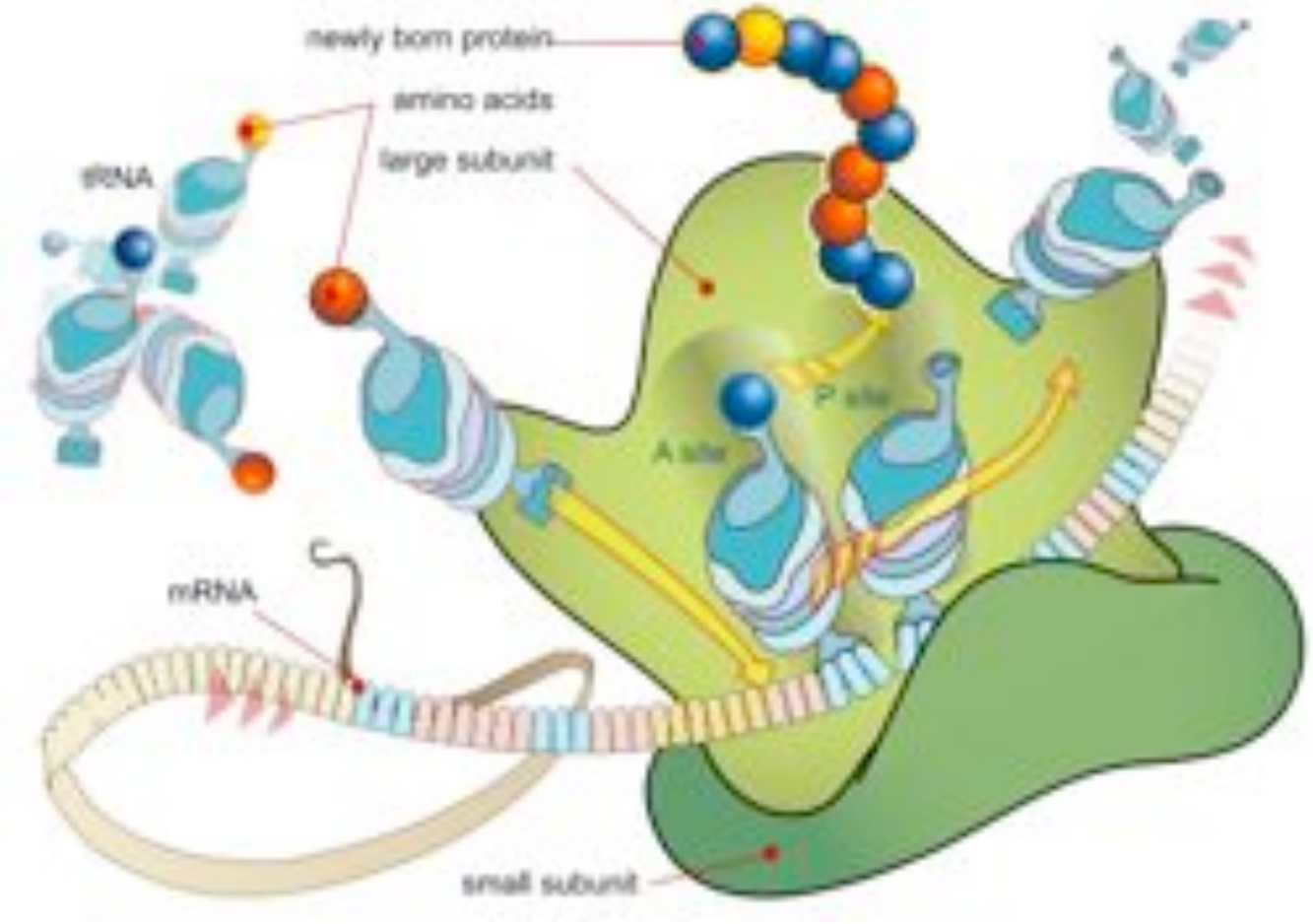
Dynamics of proteins in crystals or "Please hold still so we can take your picture!"

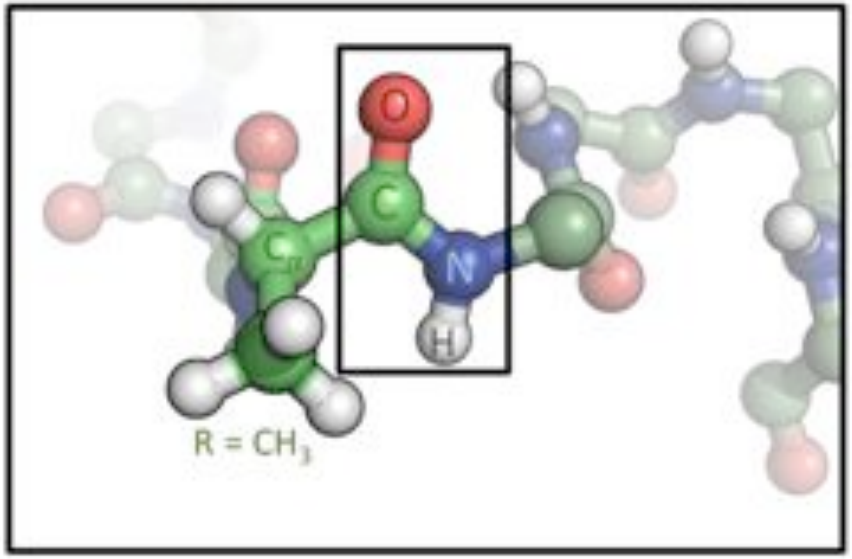
George N. Phillips Jr.

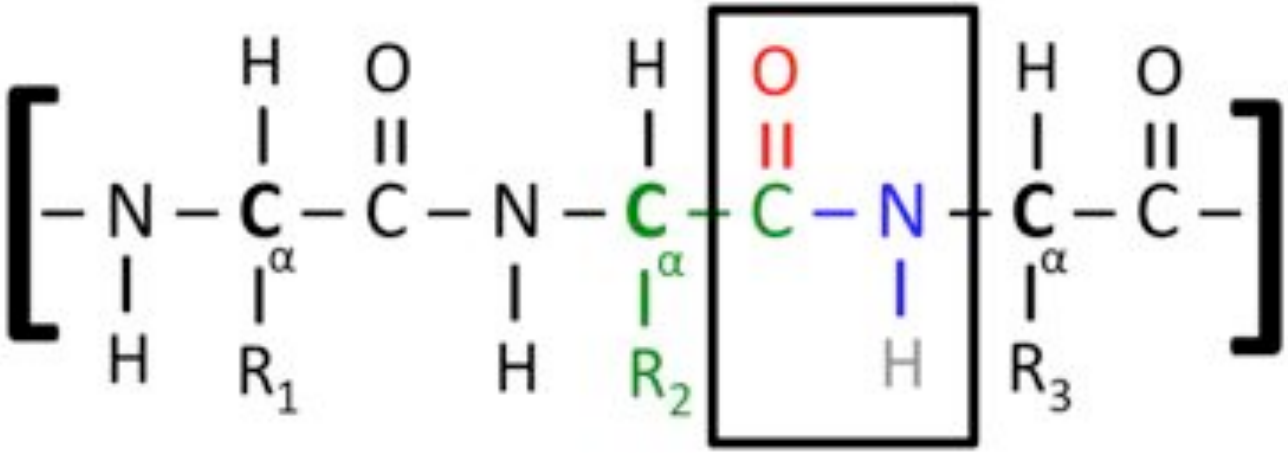
Department of Biochemistry and Cell Biology
Rice University

Molecular Biology 101



Covalent structure of Proteins





Our Nobel Prize-Winning Founders

1915 WH Bragg and WL Bragg

Use of X-rays to determine crystal structure

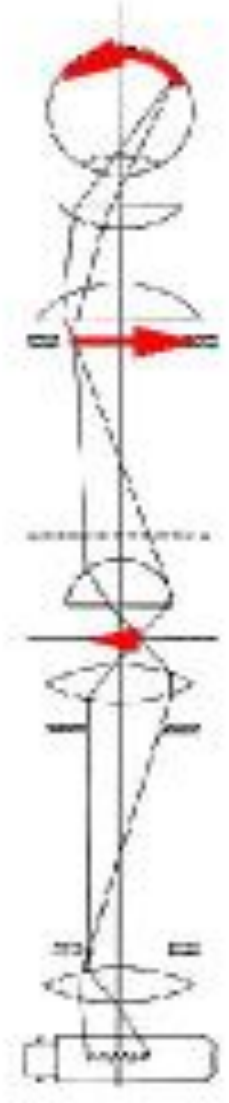
1914 M von Laue

Diffraction of X-rays by crystals

1901 WC Röntgen

Discovery of X-rays

Light microscope

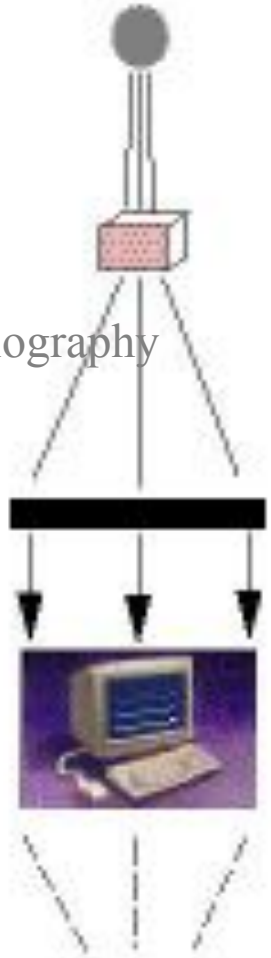


2,000 X

Electron microscope

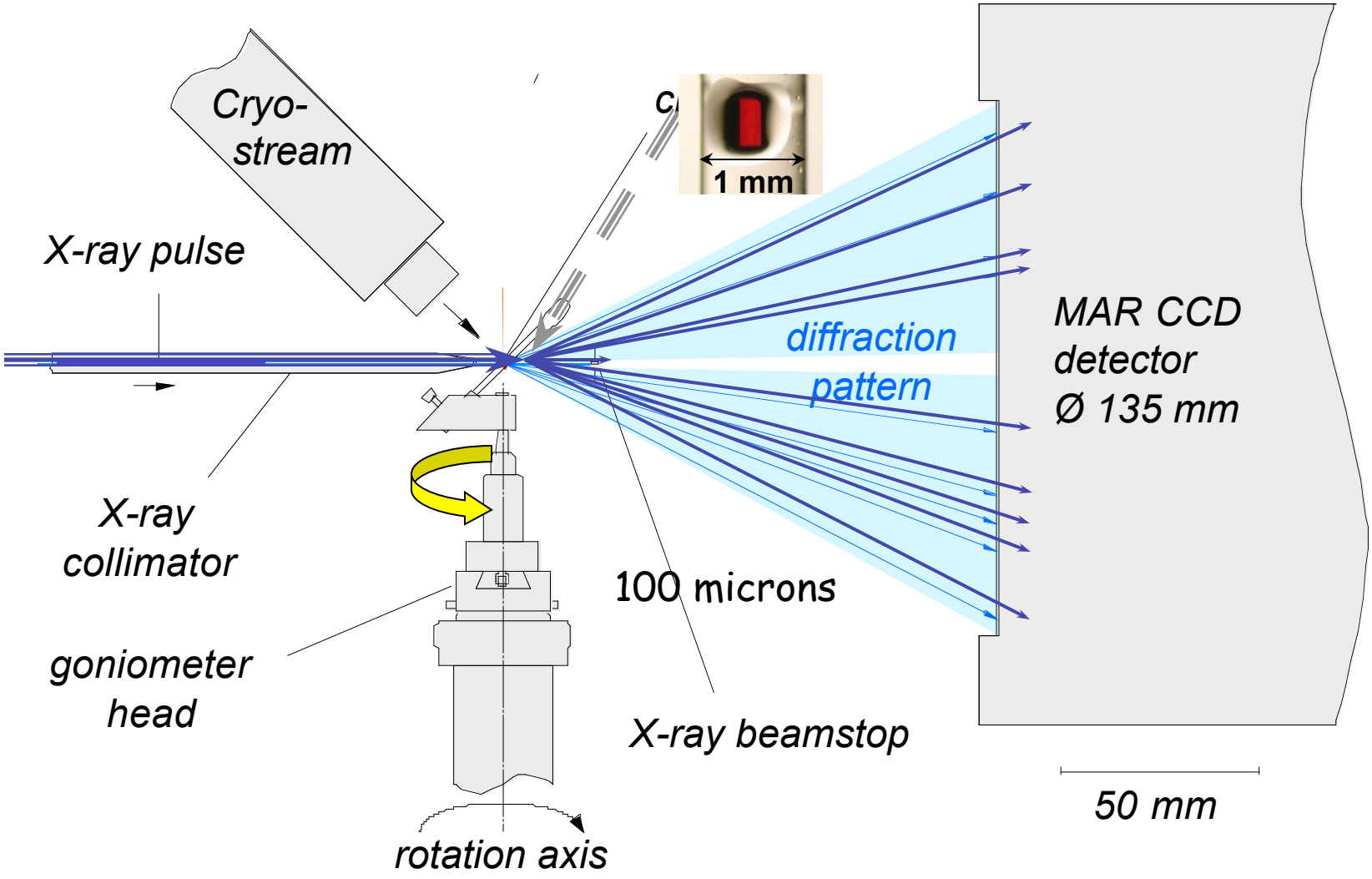
1,000,000 x

X-ray
Crystallography

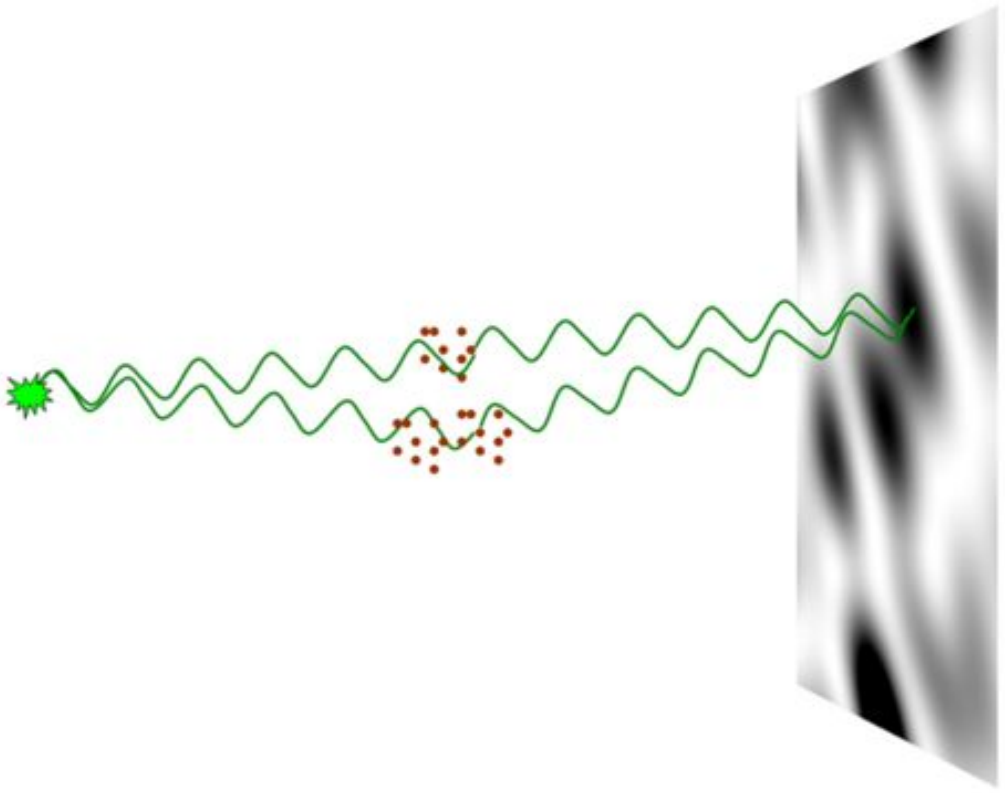


100,000,000 X

Intro to Crystallography

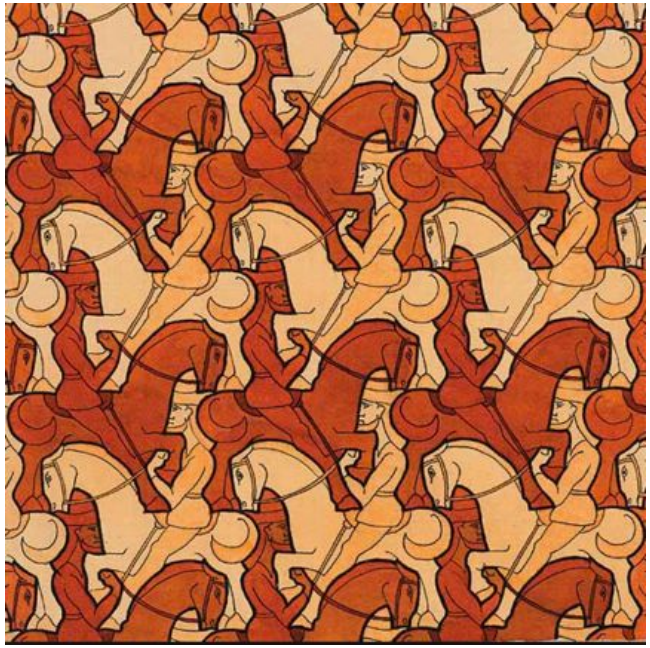


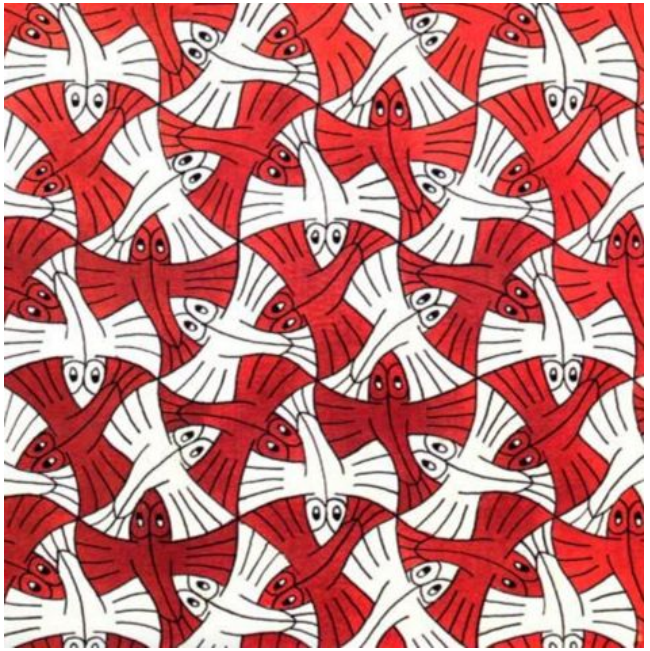
Scattering by several electrons



James Holton

Periodicity and Symmetry

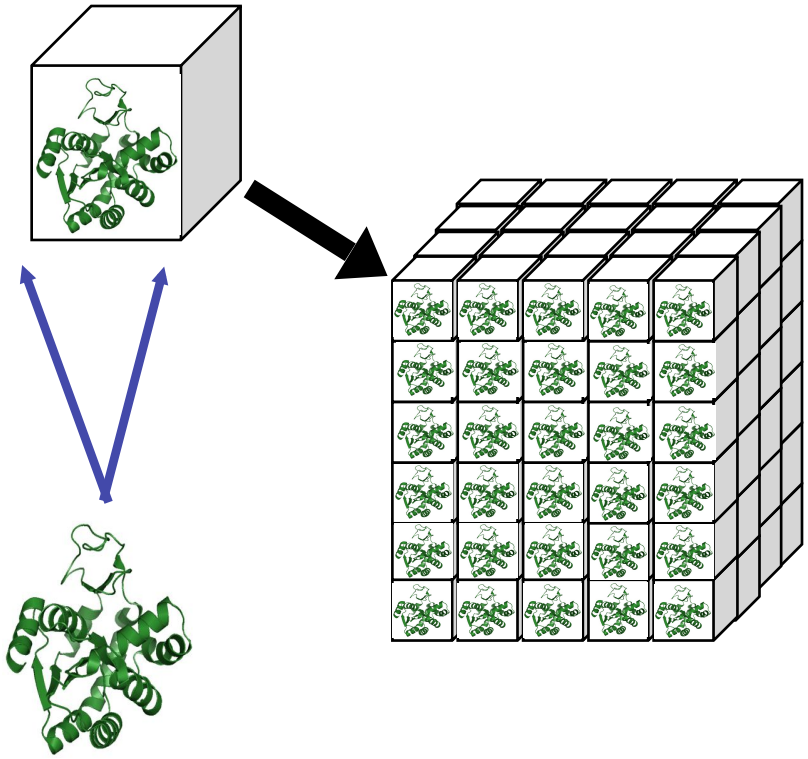


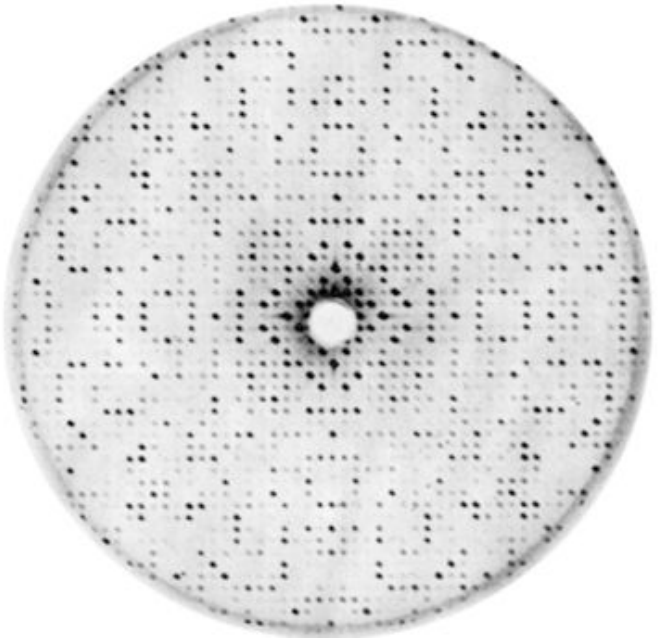


M.C. Escher

Convolution Theorum

$$FT(\rho_{molecule} \otimes L_{\inf}) = FT(\rho_{molecule}) \times FT(L_{\inf})$$





Kinematic Level Theory

General diffraction expression

$$I(\mathrm{Q}) = \sum_{lk} \sum_{l'k'} f_{kQ} e^{-(\mathbf{Q}^T \langle \mathbf{u}_{lk} \mathbf{u}_{lk}^T \rangle \mathbf{Q})/2} f^*_{kQ} e^{-(\mathbf{Q}^T \langle \mathbf{u}_{l'k'} \mathbf{u}_{l'k'}^T \rangle \mathbf{Q})/2} e^{i\mathbf{Q}\cdot(\mathbf{r}_{lk} - \mathbf{r}_{l'k'})} e^{\mathbf{Q}^T \langle \mathbf{u}_{lk} \mathbf{u}_{l'k'}^T \rangle \mathbf{Q}}$$

By application of periodicity and with isotropic displacements of the atoms

$$I(\mathrm{H}) = \sum_{k} \sum_{k'} f_{kH} e^{-(2\pi H)^2 \langle u_k^2 \rangle /2} f^*_{k'H} e^{-(2\pi H)^2 \langle u_{k'}^2 \rangle /2} e^{i2\pi \mathbf{H}\cdot(\mathbf{r}_k - \mathbf{r}_{k'})} e^{(2\pi H)^2 \langle u_k u_{k'} \rangle}$$

A nasty inverse problem

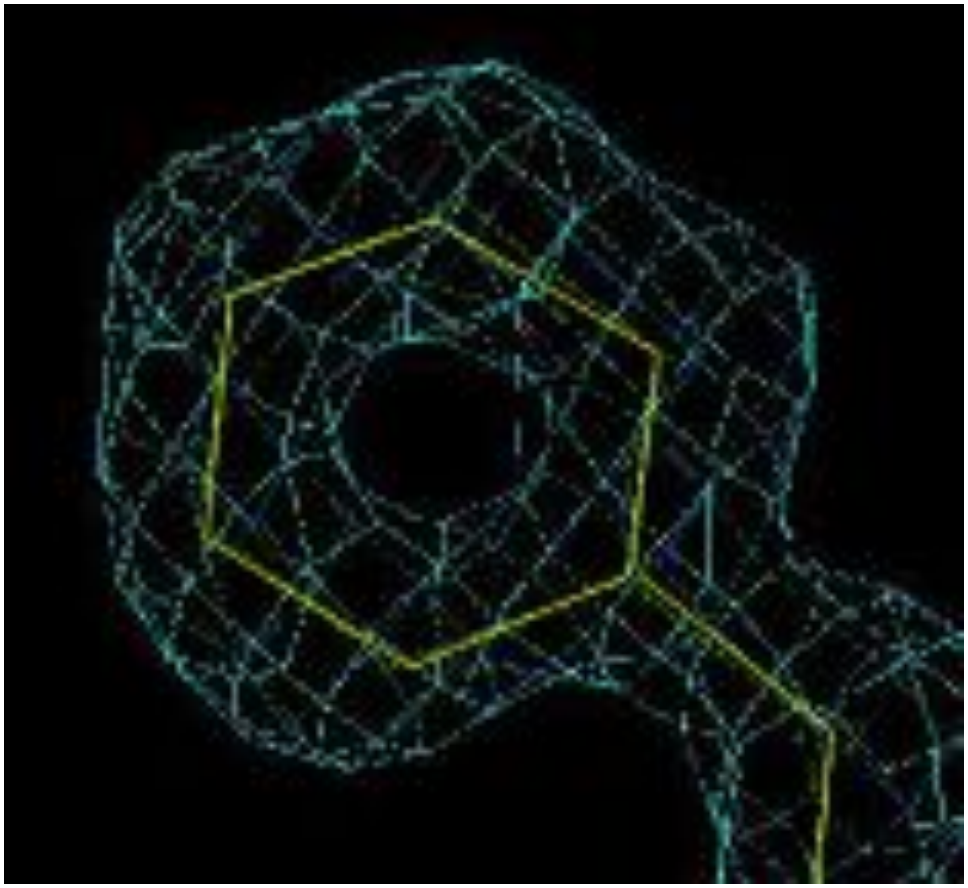
Requires experimental or other estimation of the real versus complex parts of thousands of measured structure factor amplitudes.

Electron density equation

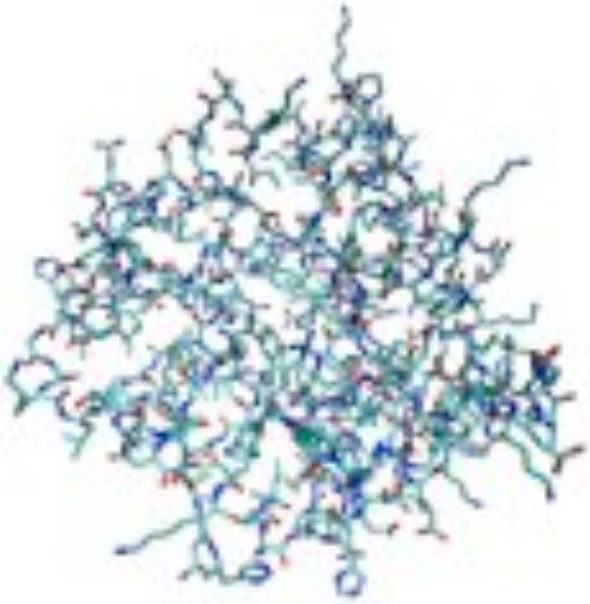
$$\rho(\mathbf{x}) = \frac{1}{V}\sum_{\mathbf{h}} \mathbf{F}(\mathbf{h}) \exp(-2\pi i \mathbf{h}\cdot\mathbf{x})$$

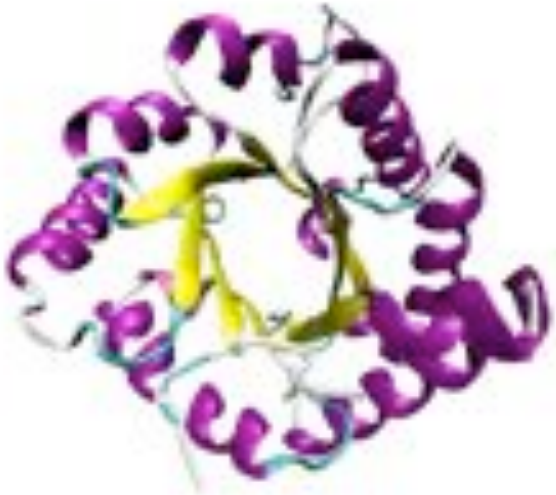
x is a vector with x,y,z fractional components in real space
h is a vector with h,k,l components in reciprocal space
F(h) is the complex structure factor
V is the unit cell volume

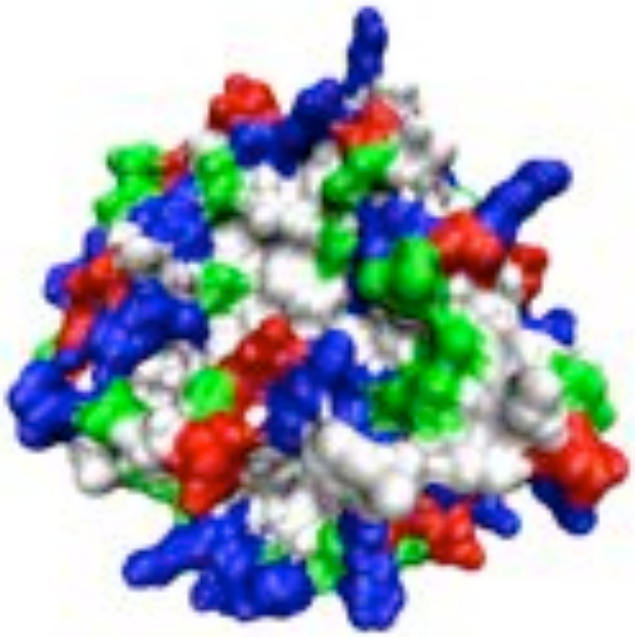
Electron density map



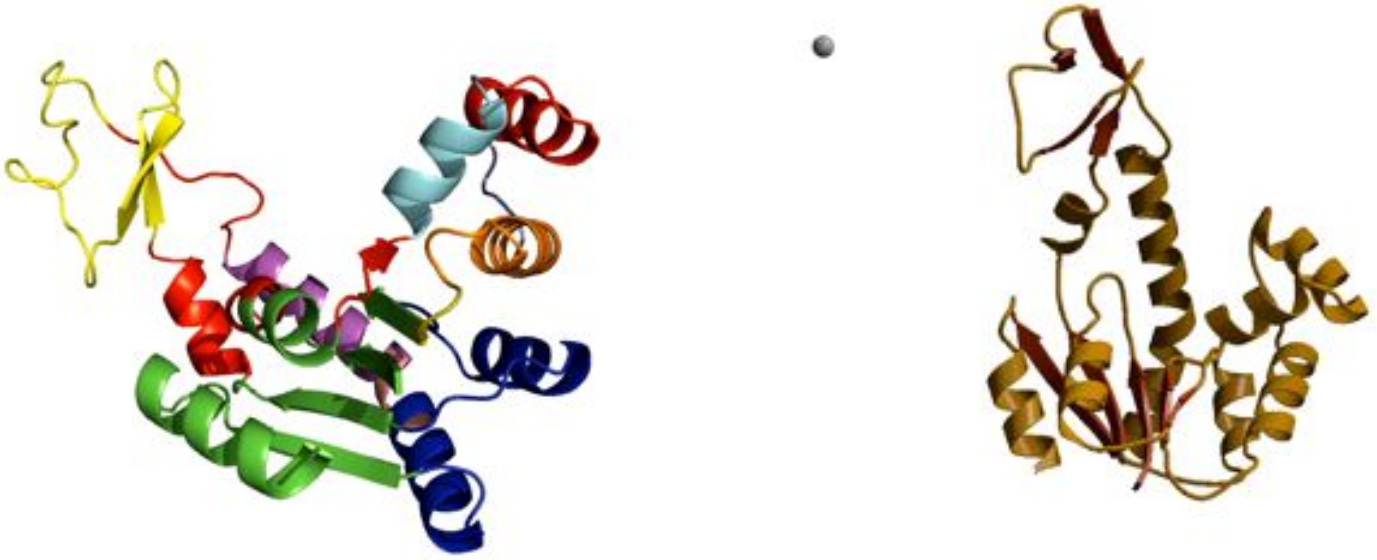
Representations of protein molecules





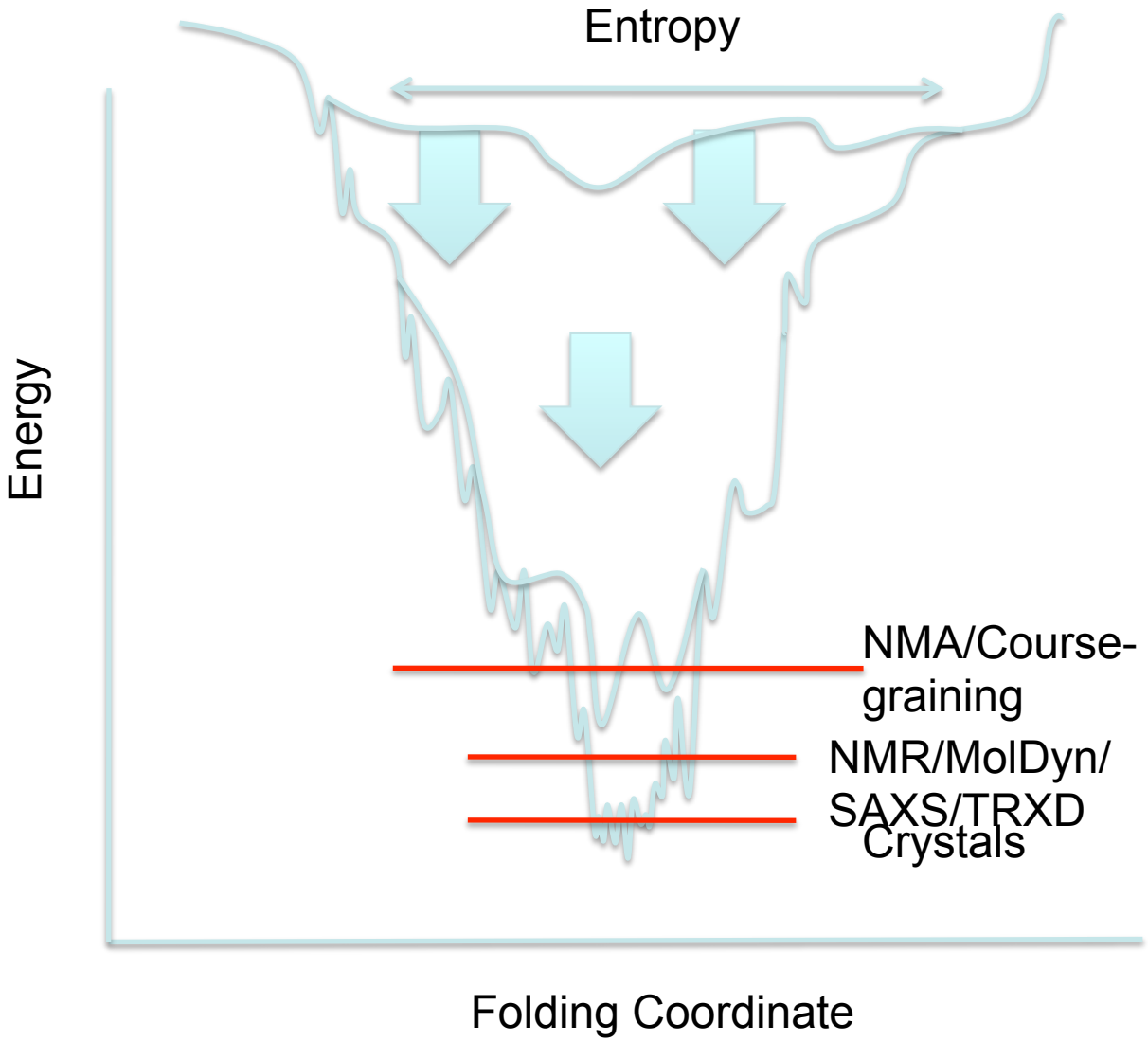


Adenylate kinase motions



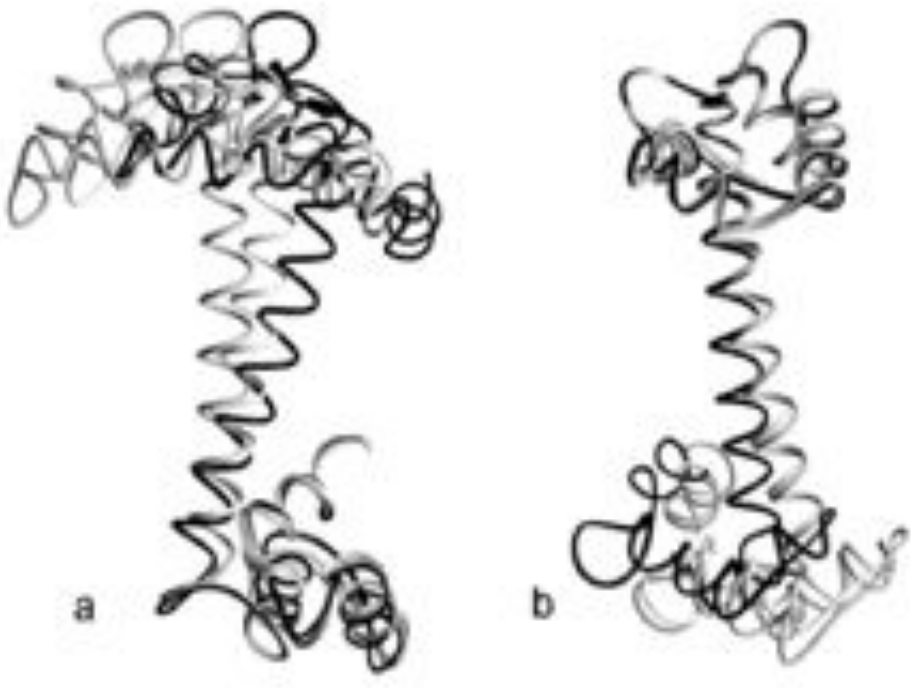
Schulz et al. and Berry and Phillips *Proteins* 1998

Ensembles at Multiple Levels



Crystal's effect on Structure?

Troponin C

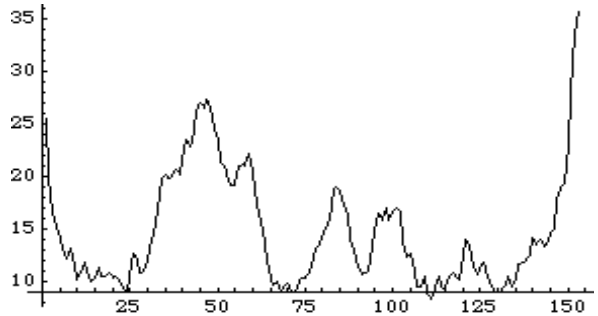


Soman, Tao, Phillips *Proteins* 1999

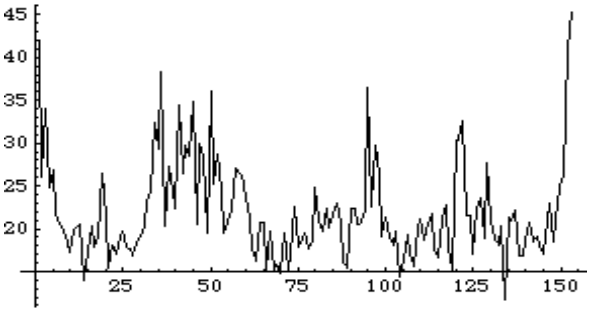
The protein is variable in structure

- Crystallography (usually) confuses the space and time averages.
- Dynamic behavior remains--There IS temperature dependence, both kT-ish and landscapes more shallow
- The crystal lattice constrains the ‘dynamics’ to varying degrees

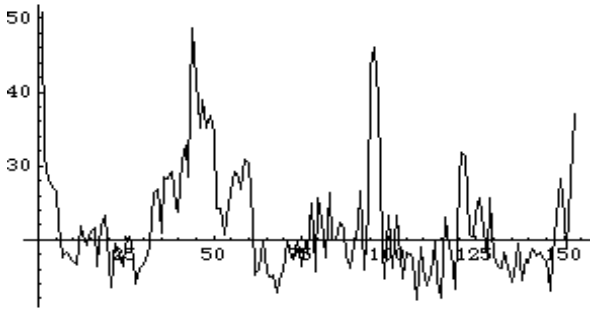
Experimental B-factors of myoglobin in five crystal forms



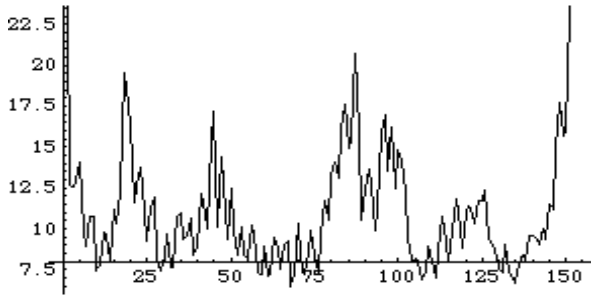
P6 AmSulfate pH 9



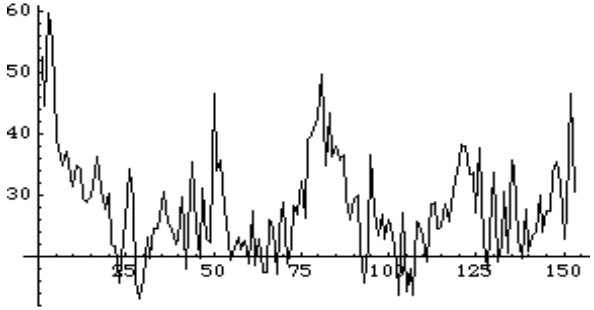
$P2_12_12_1$ - PEG
Imd pH 7



$P2_12_12_1$
2.5 M AS pH 8



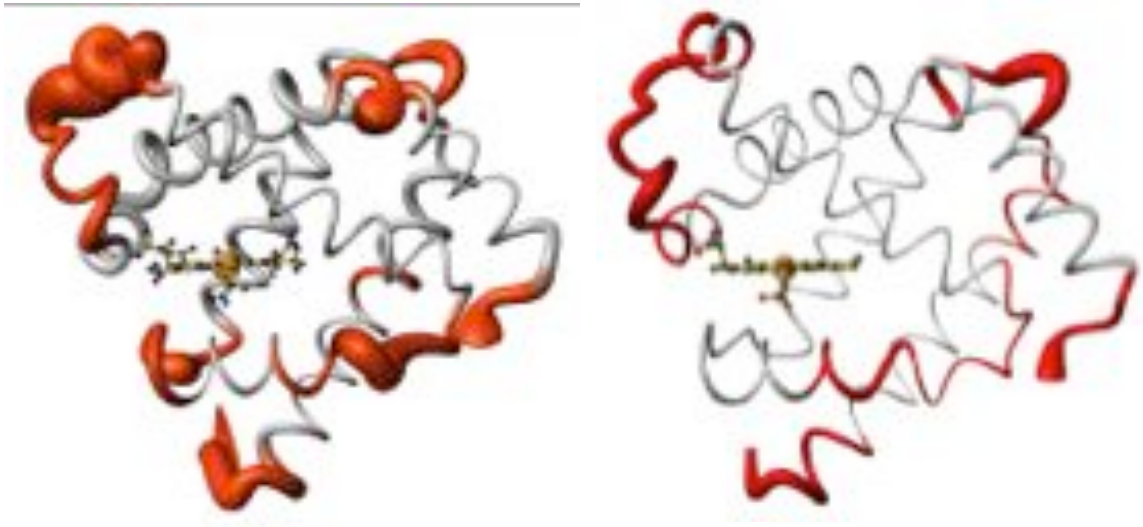
$P2_1$ AmSulphate pH 7



$P6_122$ citrate

Phillips Biophys J. 1990
Kondrashov, Zhang, Aranda, Stec, Phillips *Proteins* 2007

NMR and Crystallography: comparison of backbone dynamics

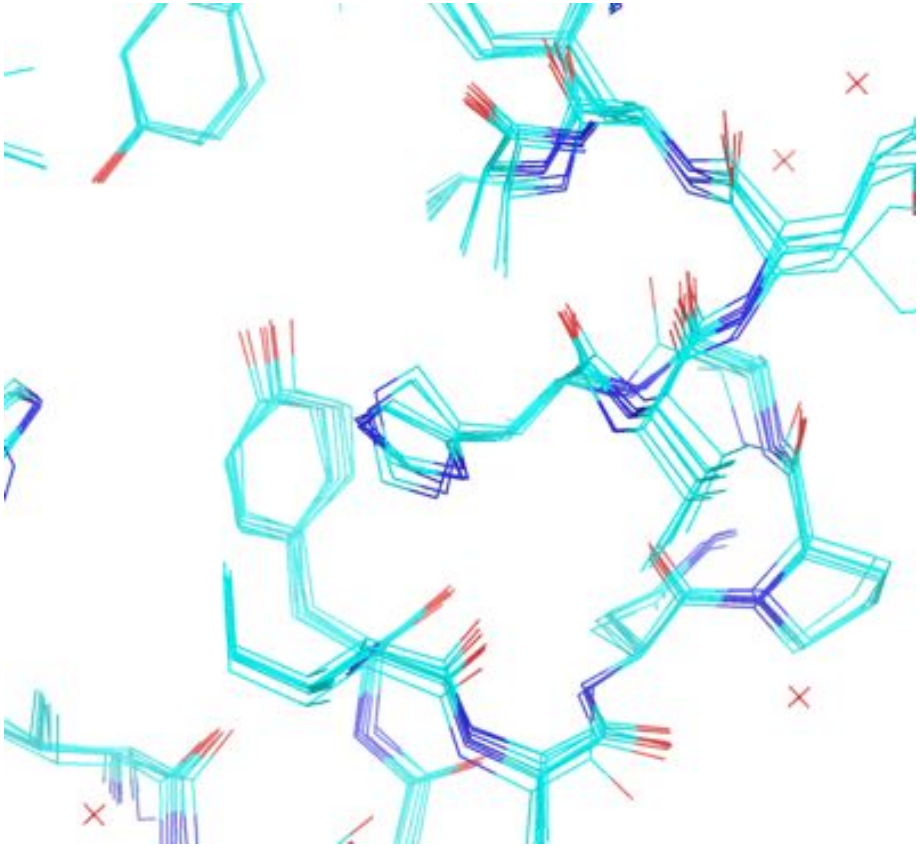


Main chain variations from NMR ensemble and various crystal forms of myoglobin.

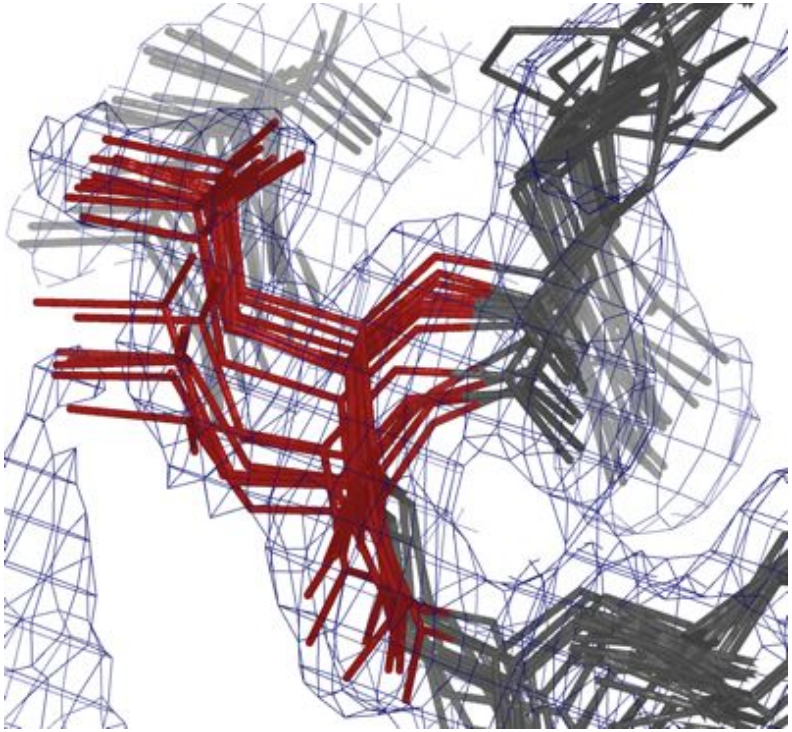
Kondrashov, Zhang, Aranda, Stec, and Phillips *Proteins* 2008

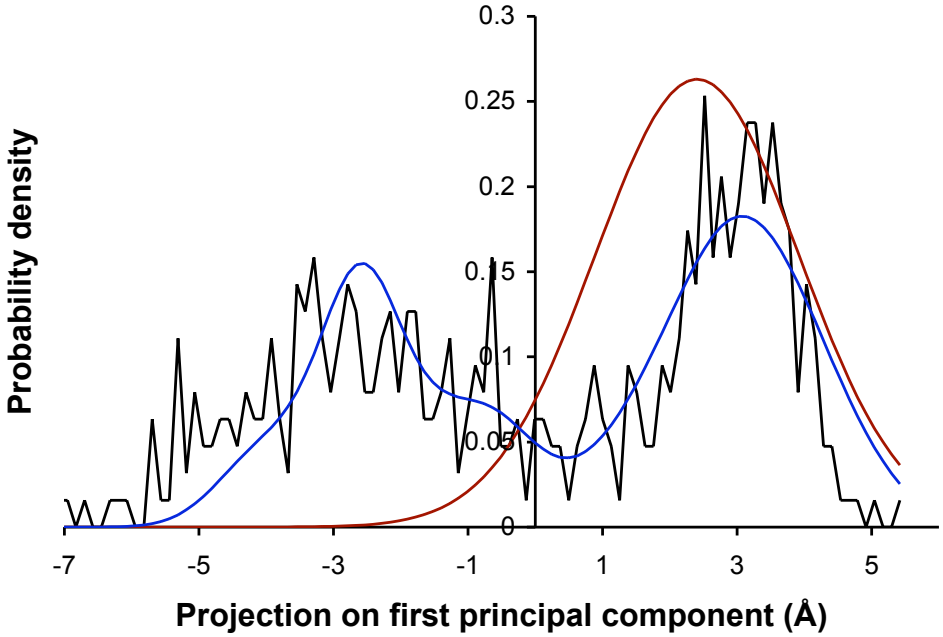
Ensemble Refinement

- Refine several copies of the entire protein simultaneously.
- Each copy has a fractional occupancy and does not interact with the other copies.

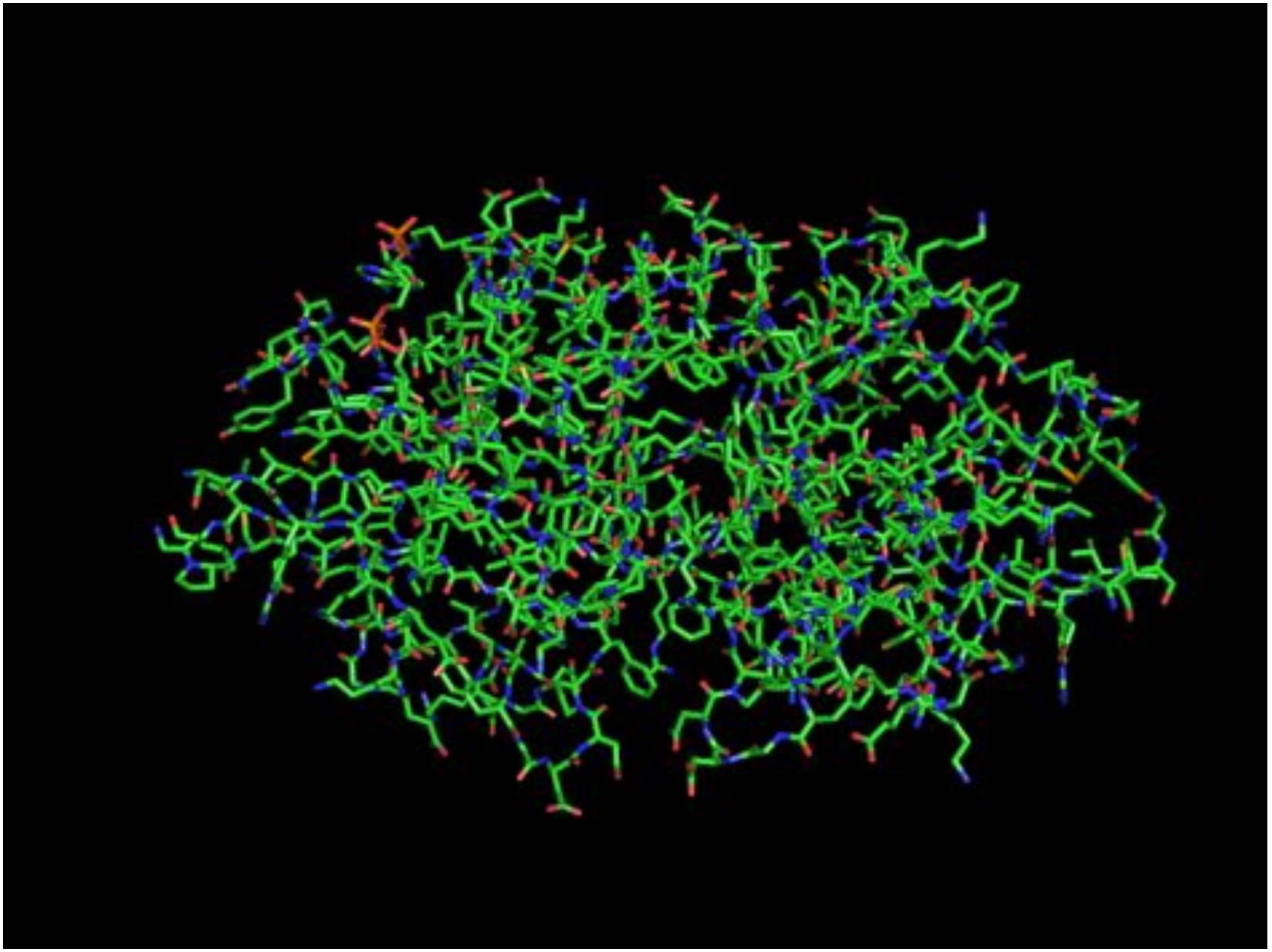


Levin, Kondrashov, Wesenberg, Phillips, Structure, 2007

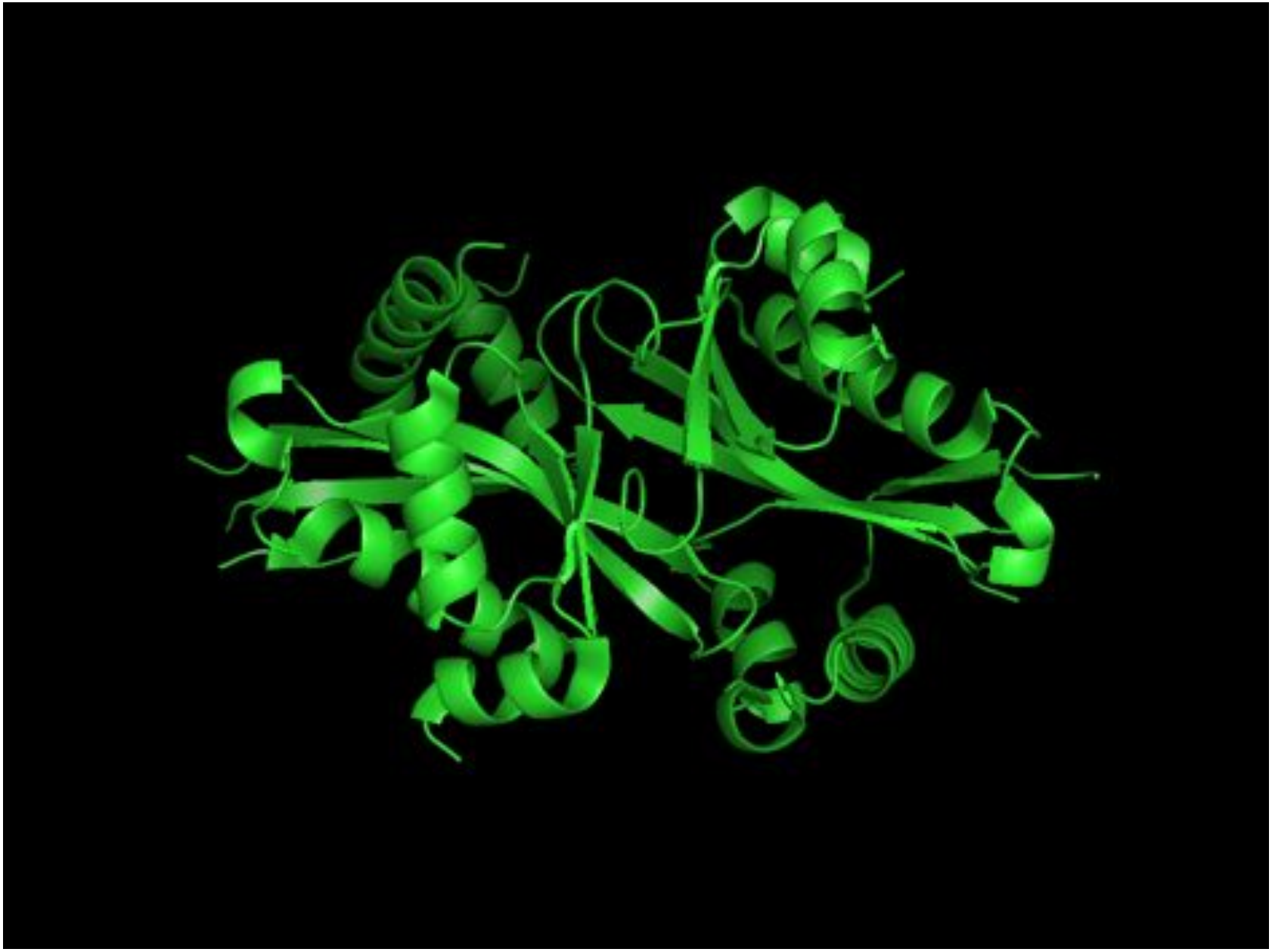



Probability density
0.3
0.25
0.2
0.15
0.1
0.05
0
-7
-5
-3
-1
1
3
5
Projection on first principal component (Å)

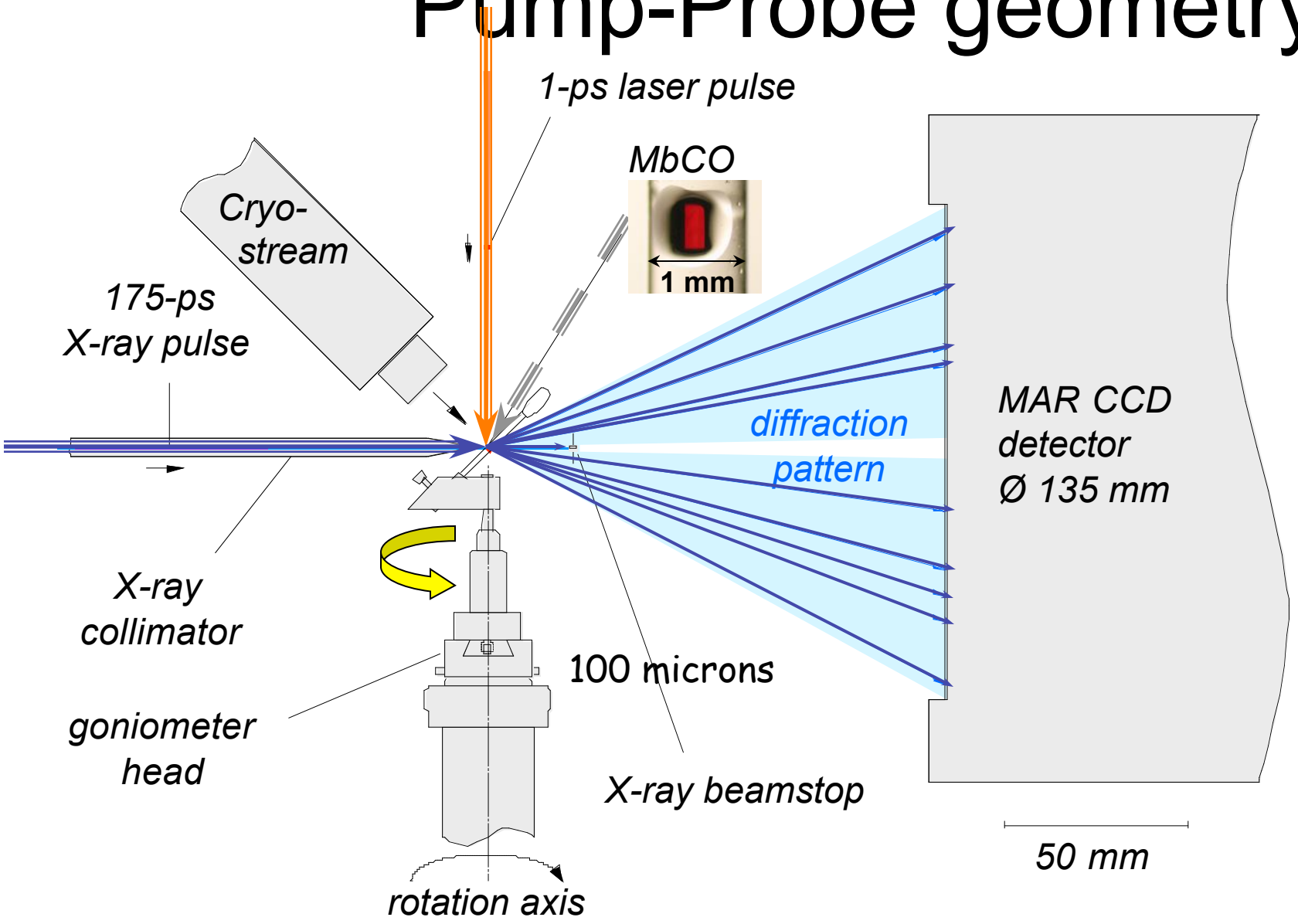
Entire Dimeric Protein



Protein Cartoon with Larger Scale Variations

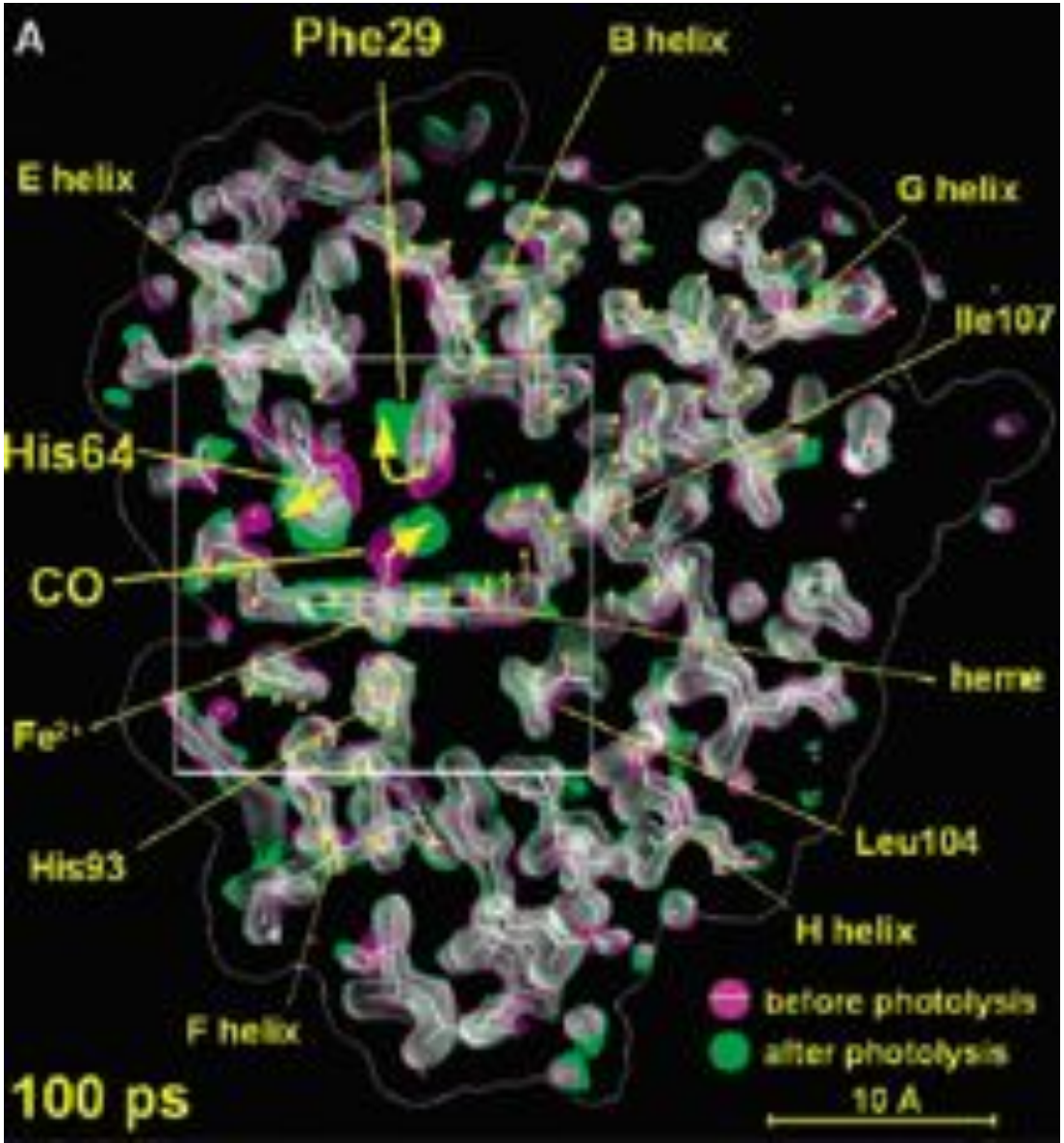


Pump-Probe geometry

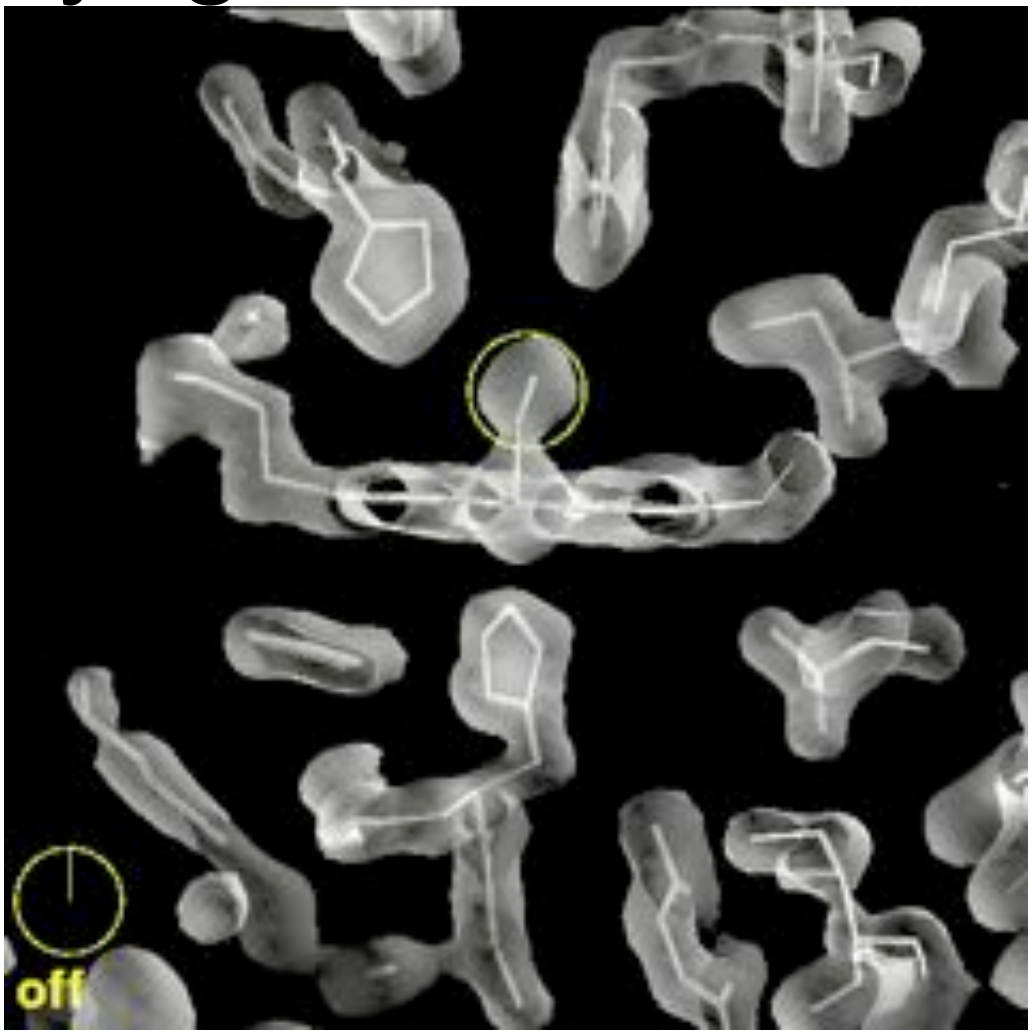


Schotte, Lim, Jackson, Smirnov, Soman, Olson, Phillips, Wulff, Anfinrud, *Science* 2003

Guide to the “actors”

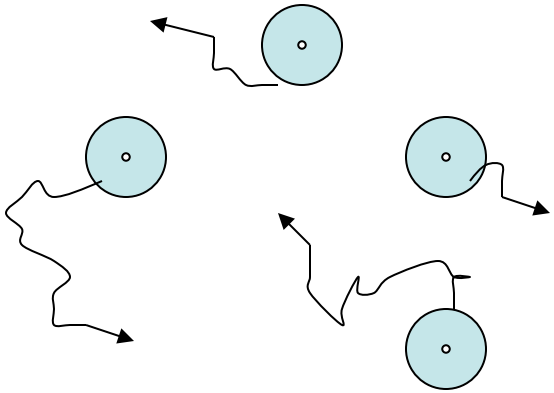


Myoglobin: The movie



Molecular Dynamics Simulations

F = m a = - grad V, where V is the potential



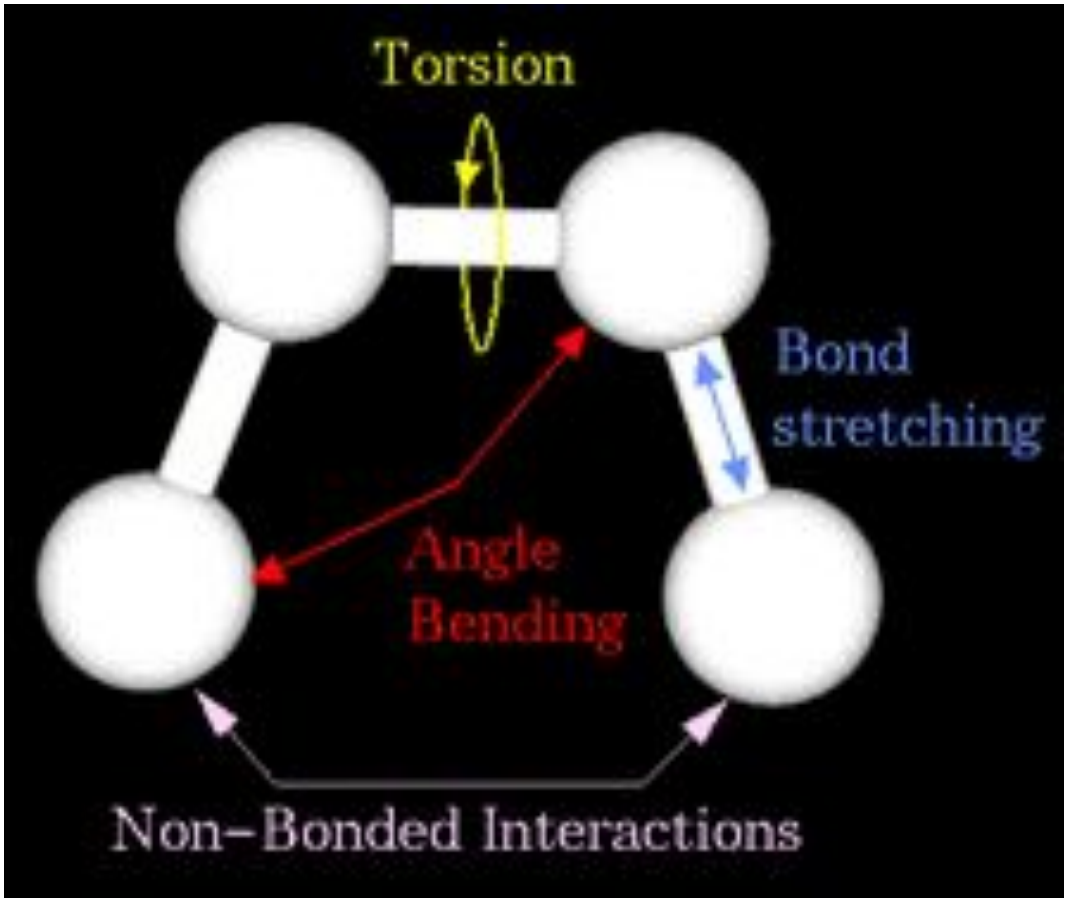
All atoms are moving

Forces between atoms are complicated functions of time

ANALYTICAL solution of x(t) and v(t) is impossible!
This is an N-body problem.

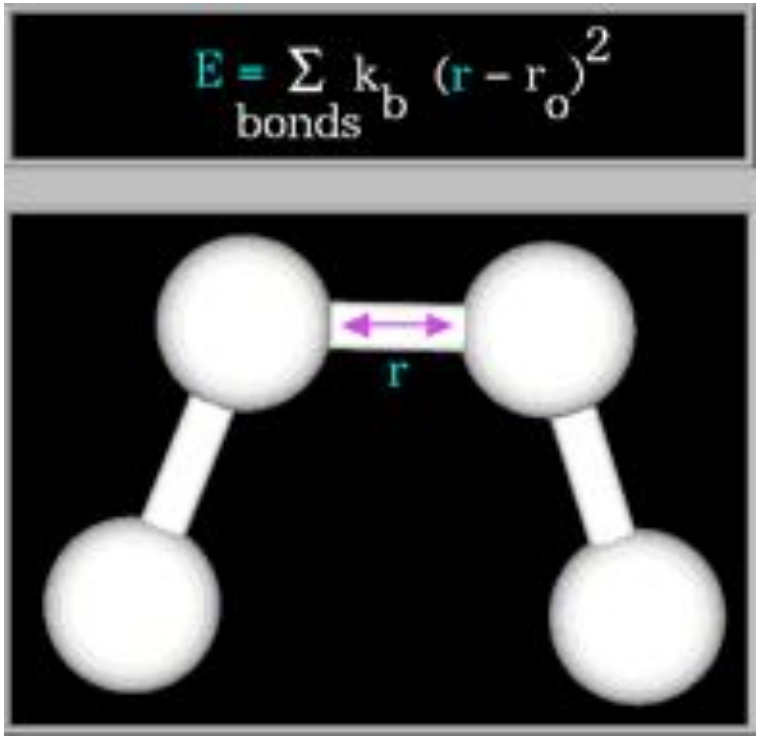
NUMERICAL solution is possible but expensive.
(use short time steps and assume independence)

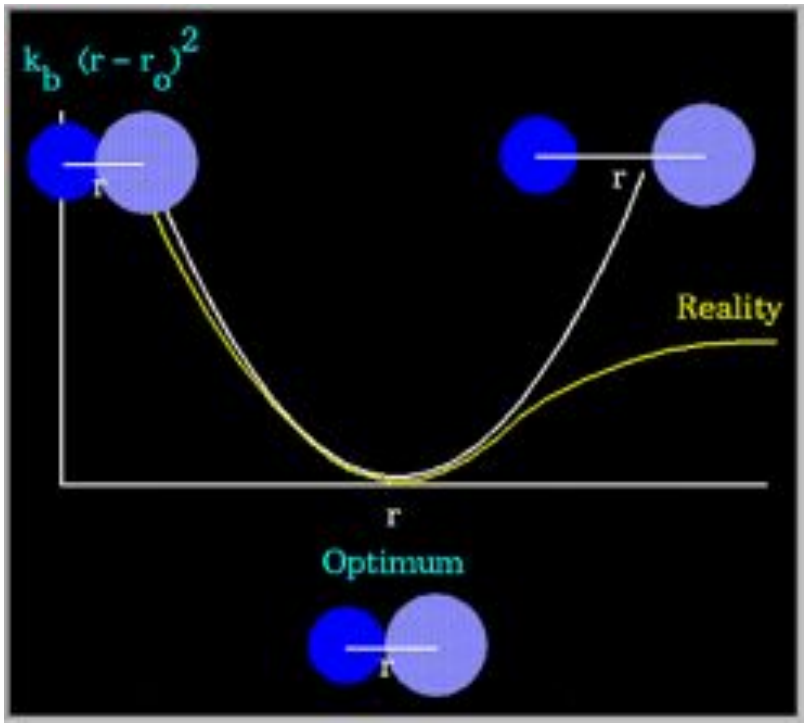
Force field



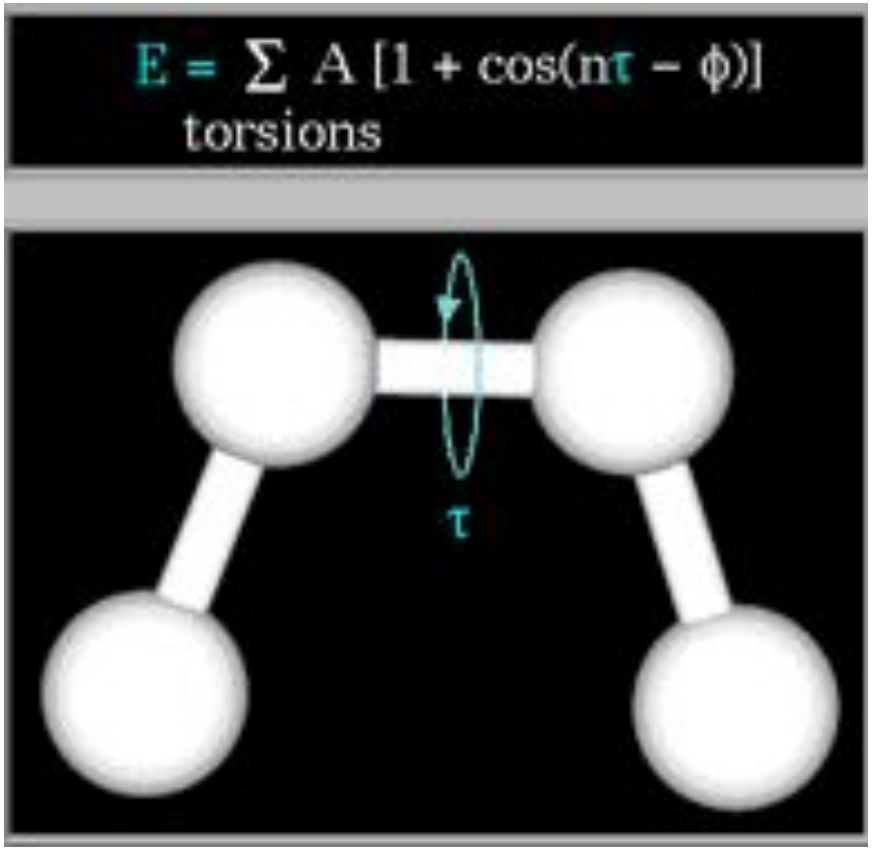
http://cmm.info.nih.gov/modeling/guide_documents/molecular_mechanics_document.html

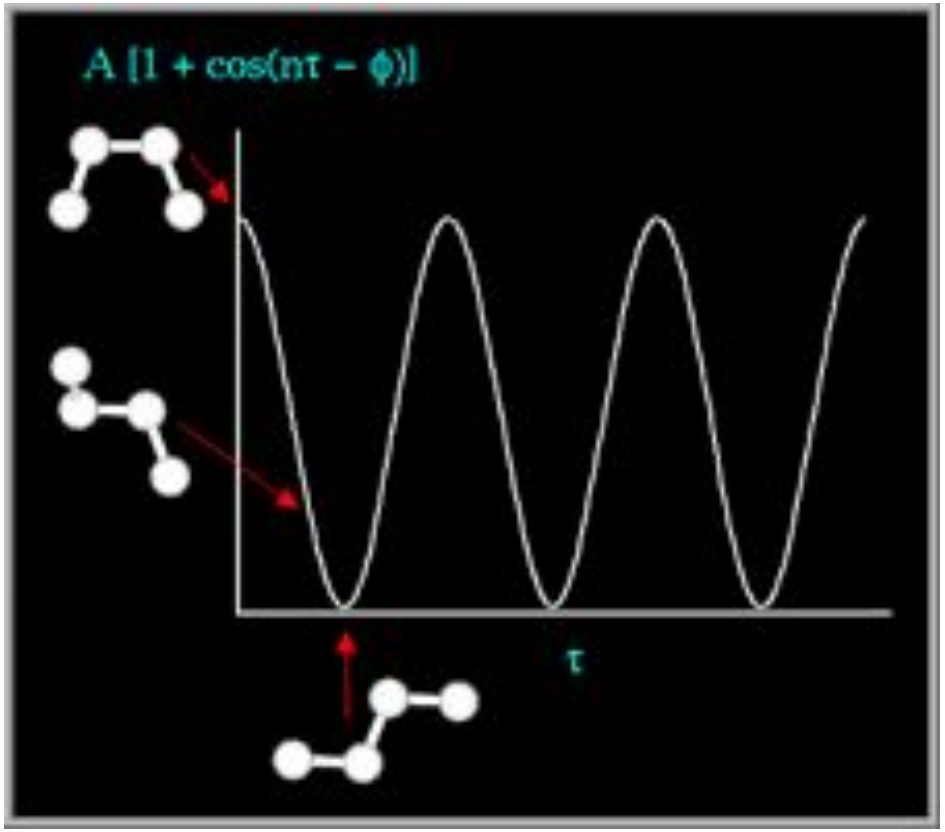
Bonds





Dihedrals

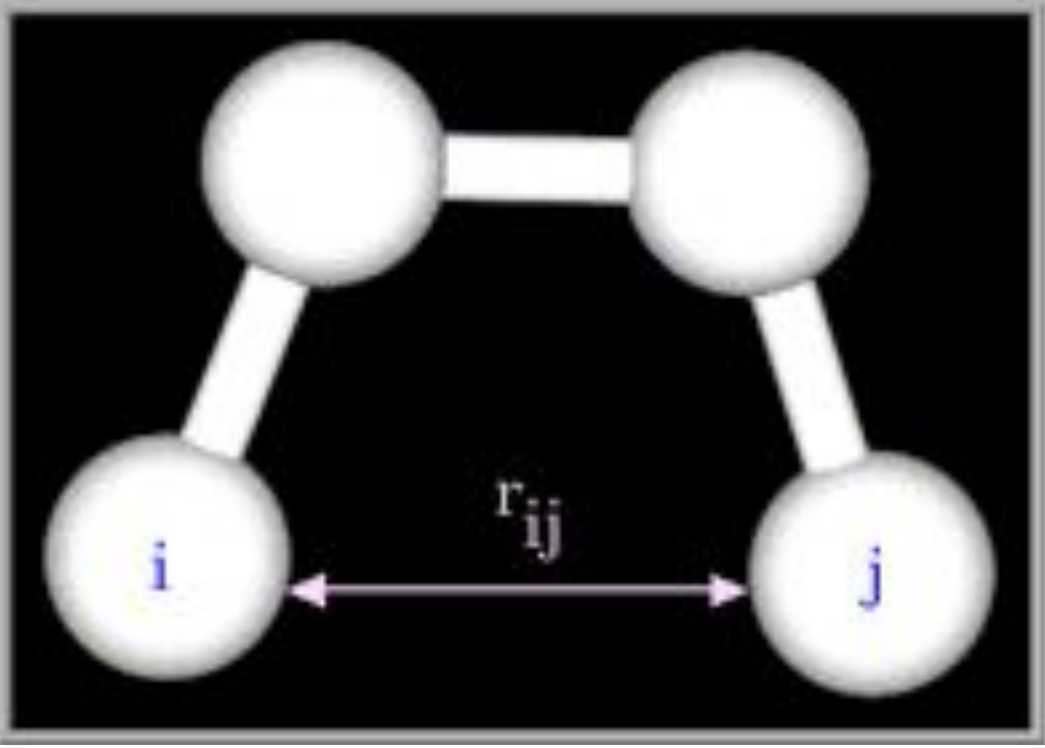




Non-bonded interactions

$$E = \sum_i \sum_j \frac{-A_{ij}}{r_{ij}^6} + \frac{B_{ij}}{r_{ij}^{12}} + \sum_i \sum_j \frac{q_i q_j}{r_{ij}}$$

van der Waals term Electrostatic term



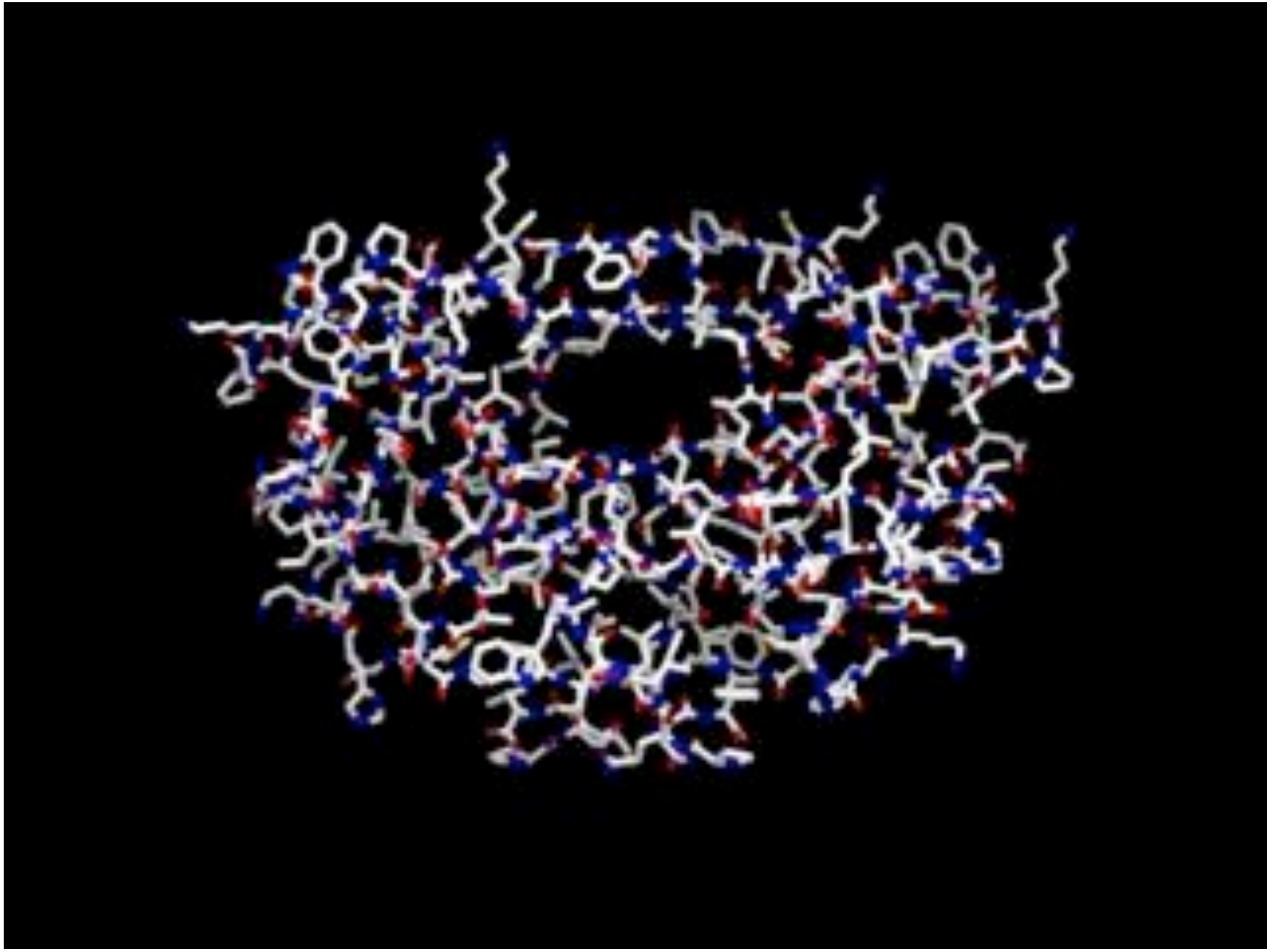
Time component

1 solve for a_i at t using: $-\frac{dE}{dr_i} = F_i = m_i\, a_i(t)$

2 update v_i at $t + \Delta t/2$ using: $v_i(t + \Delta t/2) = v_i(t - \Delta t/2) + a_i(t)\, \Delta t$

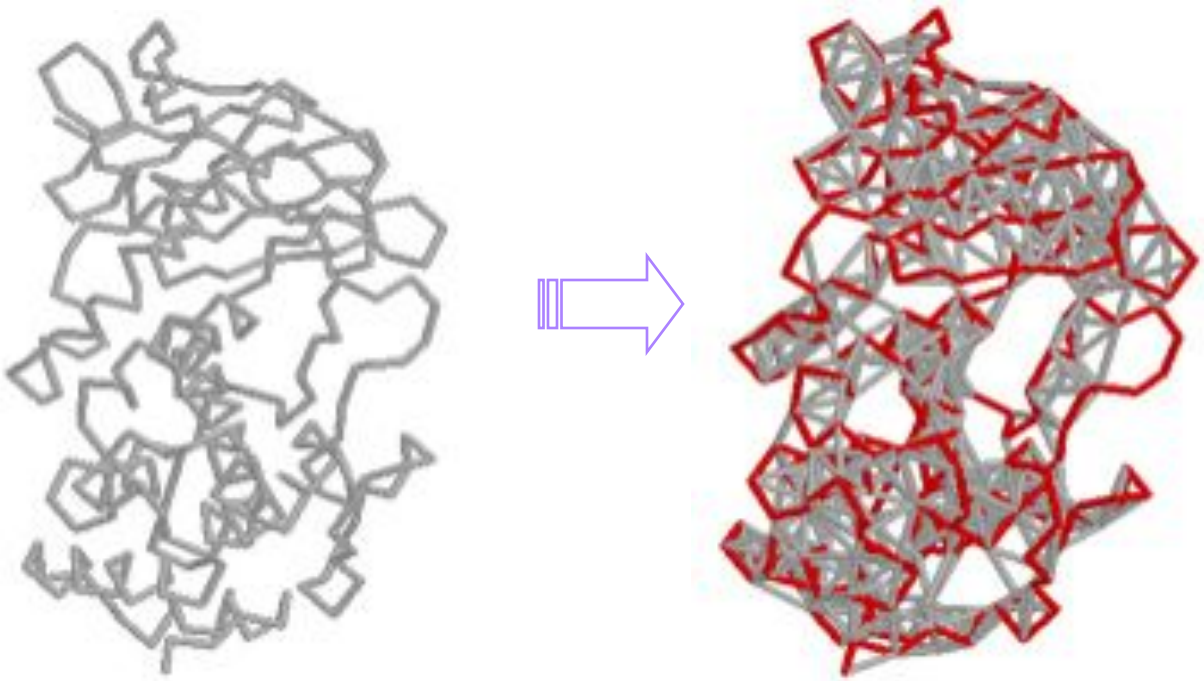
3 update r_i at $t + \Delta t$ using: $r_i(t + \Delta t) = r_i(t) + v_i(t + \Delta t/2)\, \Delta t$

Leap frog algorithm



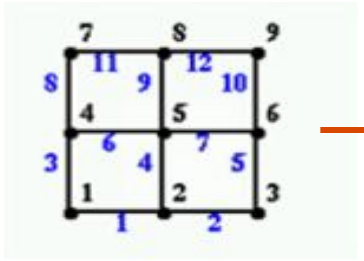
Gaussian Network Model

- Model assumes harmonic “springs” between segments (represented by Cα locations) within a certain cutoff distance (~7 Å), forming an elastic network



- Each Cα atom forms a node in the network and represent a single residue. Edges correspond to the springs.
- (After M.M. Tirion and I. Bahar et al, who popularized the method)

Formulation of GNM



$$\Gamma = \begin{cases} -1 & \text{if } i \neq j \text{ and } R_{ij} \leq r_c \\ 0 & \text{if } i \neq j \text{ and } R_{ij} > r_c \\ -\sum_{i,i\neq j} \Gamma_{ij} & \text{if } i = j \end{cases}$$

- Build a matrix (Kirchhoff, from graph theory, or Laplacian matrix)
- Mobility of Cα atom depends on the inverse of the matrix, which is related to the number of neighboring Cα atoms i.e, their connectivity and contact map
- Being an “elastic network” of springs, the model provides dynamic information from static crystal structures

Relating GNM to atomic displacements

- Eigen analysis or SVD to get psuedo-inverse

$$\Gamma^{-1} = \sum_{k=1}^{n-1} \lambda^{-1} q_k q_k^T \qquad \Gamma^{-1} = V^T M_D^{-1} S$$

- Mean square fluctuation (variance and co-variance)

$$< u_i u_j >= (3k_B T / \gamma)[\Gamma^{-1}]_{ij}$$

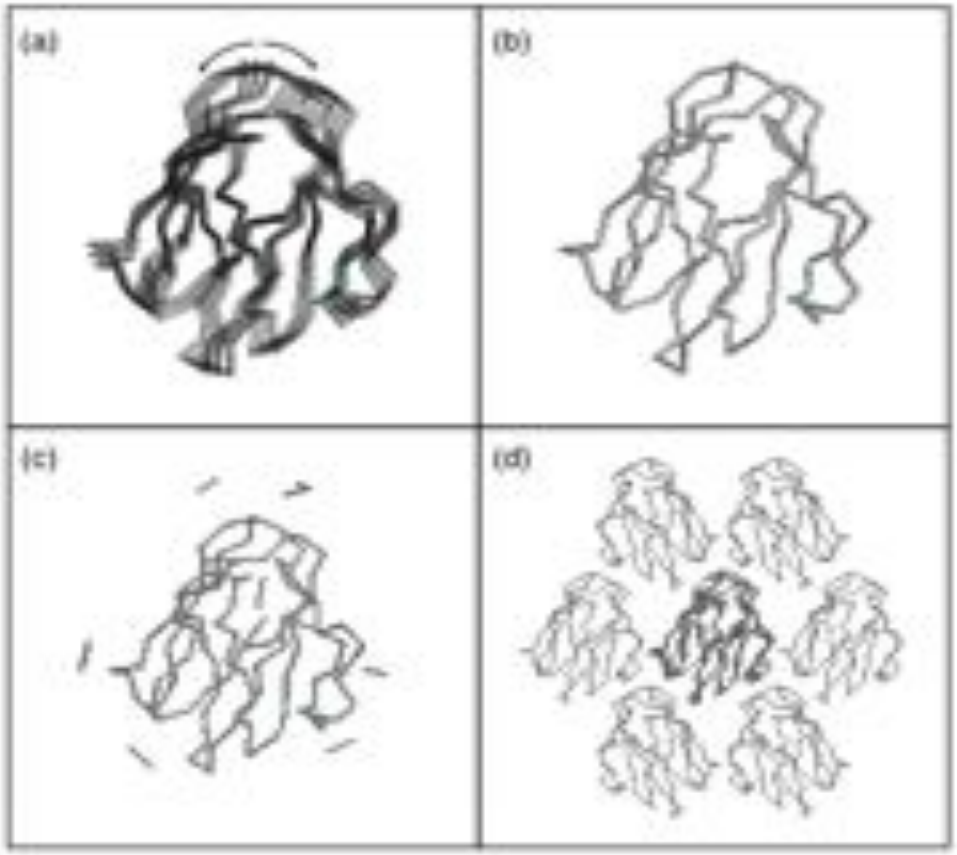
- Calculation of crystallographers' B-factors

$$B_i = 8\pi^2 \langle u_i^2 \rangle$$

Visual description of different model systems

Libration

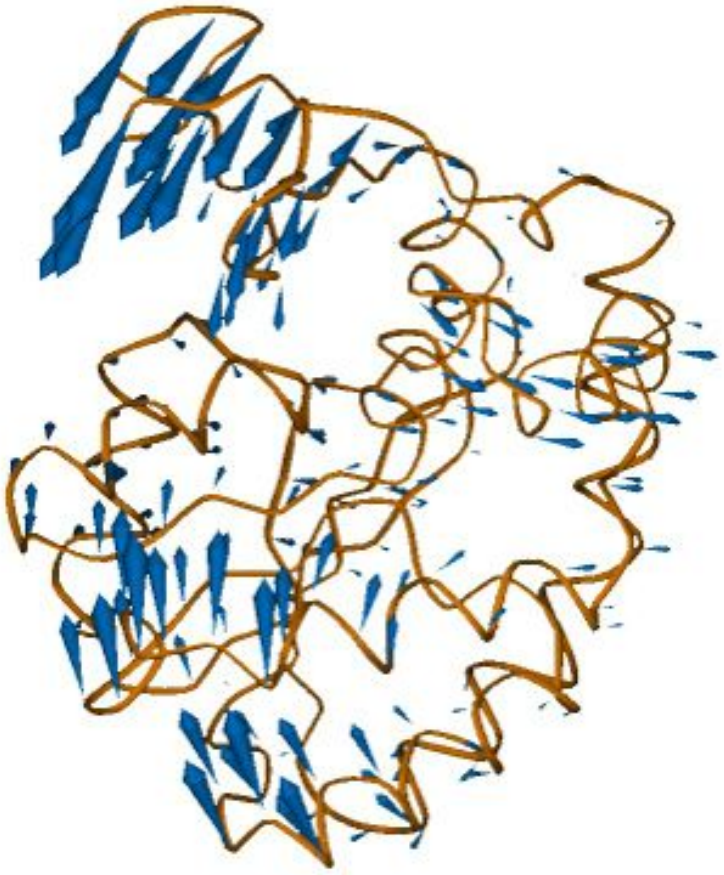
Contact atoms



Isolated molecule

Neighbor molecules

Normal mode analysis with elastic network models



Adenylate Kinase

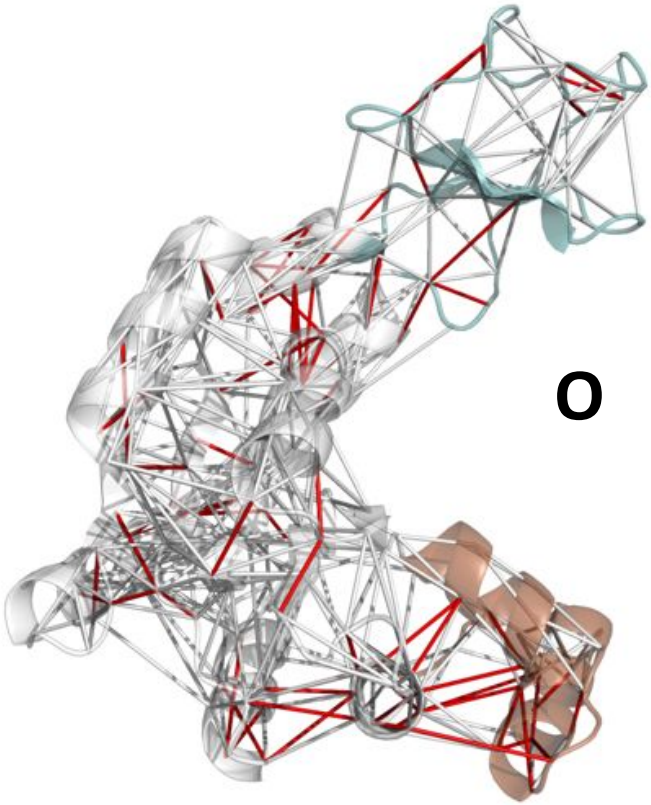
- One of adenylate kinase's major motions can be seen in its lowest mode
  - Orange = α-carbon backbone
  - Blue = Movement vector

Other Coarse-grained Gō-like models

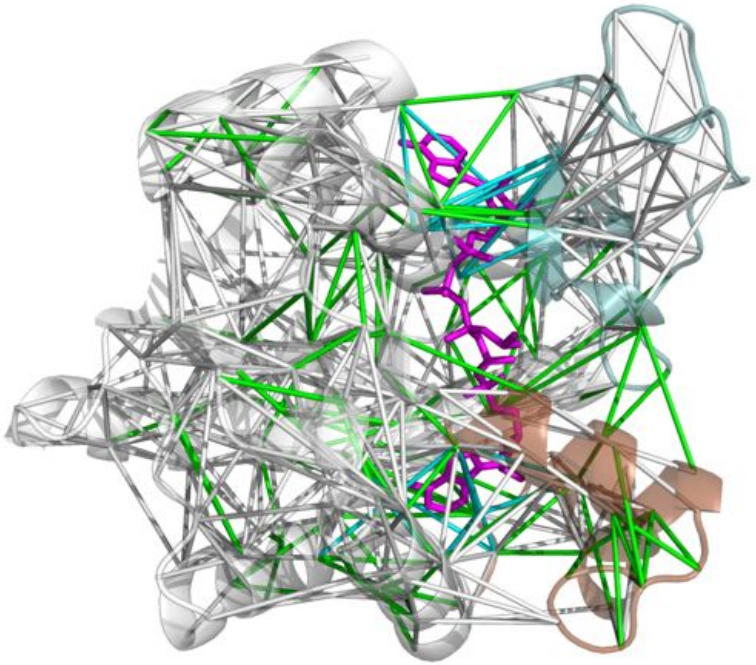
- Can simulate large-scale structural transitions without constraints
- One bead (C_α) per residue
- Harmonic bond potential
- Dihedrals
  - statistical based on sequence of residues *i*-1,*i* , no structural info
- Bond angles (some implicit φ,ψ)
  - generic: allow both α-helix and β-sheet
- Contacts
  - native: Lennard-Jones 12-10 potential (increase curvature)
  - non-native: LJ repulsion only

refs: Karanicolas & Brooks (2002), Best et al. (2005)
Daily, Phillips, Cui, J. Mol. Biol. (2010)

AKmeso O and C native contacts



O



C

LID

NMP

Common contacts

Unique to O

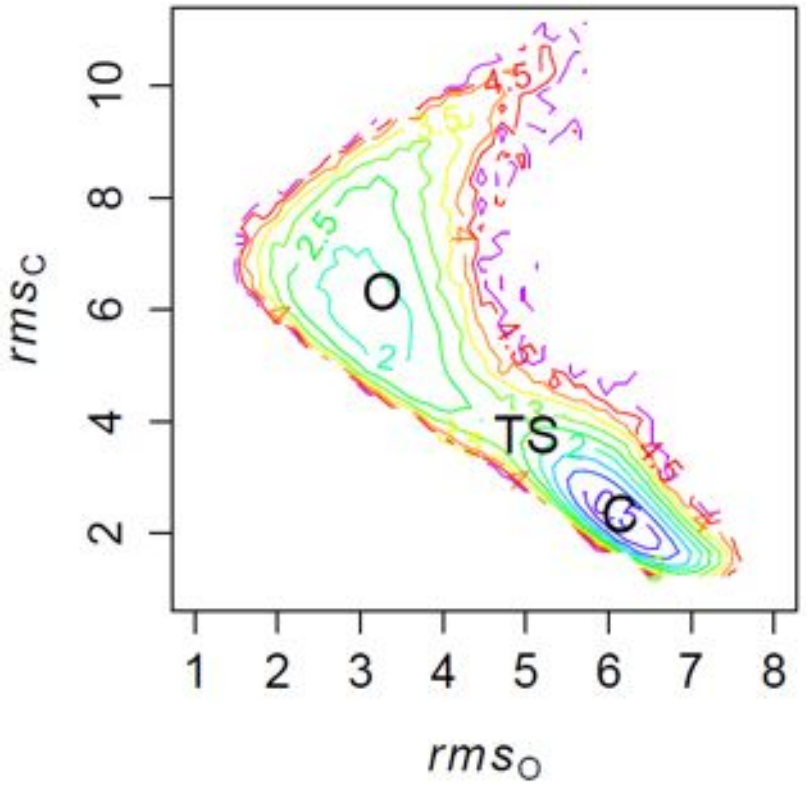
Unique to C

Substrate

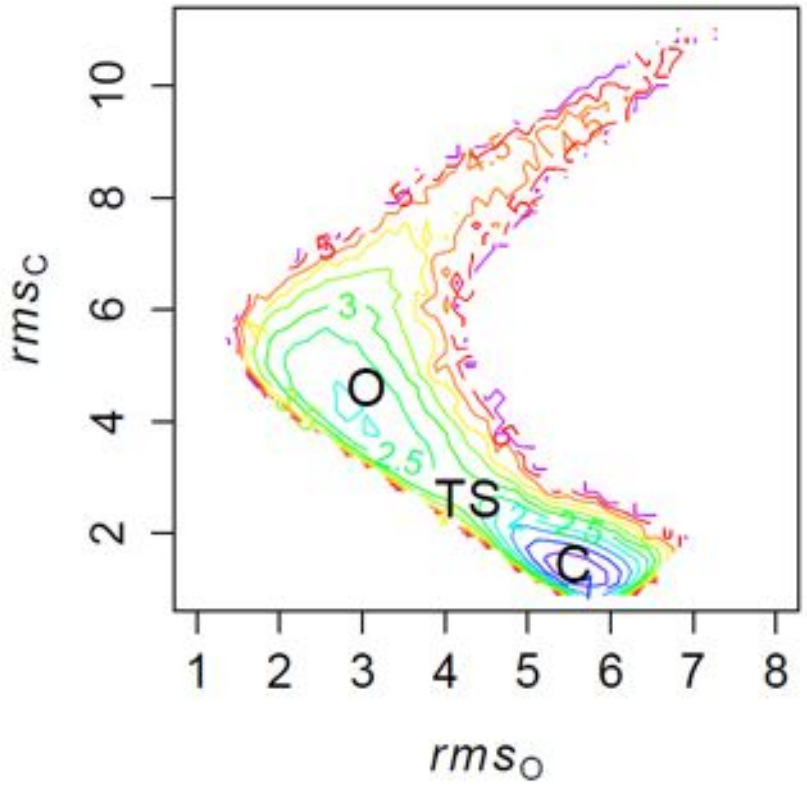
Ligand-mediated contacts

AKmeso and AKthermo simulations in rmsd space

AKmeso



AKthermo



Very similar PMFs, thermo slightly more stable in rms_C

Summary

- Crystals allow average structures of large molecules to be determined
- The crystal symmetry is only an approximation, however
- Motions of proteins are critical parts of their fitness for their functions
- While we can start to make ‘movies’ of proteins, to understand the motions, they are primitive

Acknowledgements

Contributors

- Elena Levin
- Dmitri Kondrashov
- Jason McCoy
- Ryan Bannen
- Roman Aranda
- Andre Francis
- Friedrich Schotte
- Philip Anfinrud
- Anand Kolatkar
- Mitch Miller
- Gary Wesenberg
- Craig Bingman
- Ed Bitto
- Michael Wall
- Wei Zhang
- Bog Stec
- Sibsankar Kundu
- Euiyoung Bae
- Demian Riccardi
- Ragothaman Yennamalli
- All other members

PSI Protein Structure Initiative

NSF