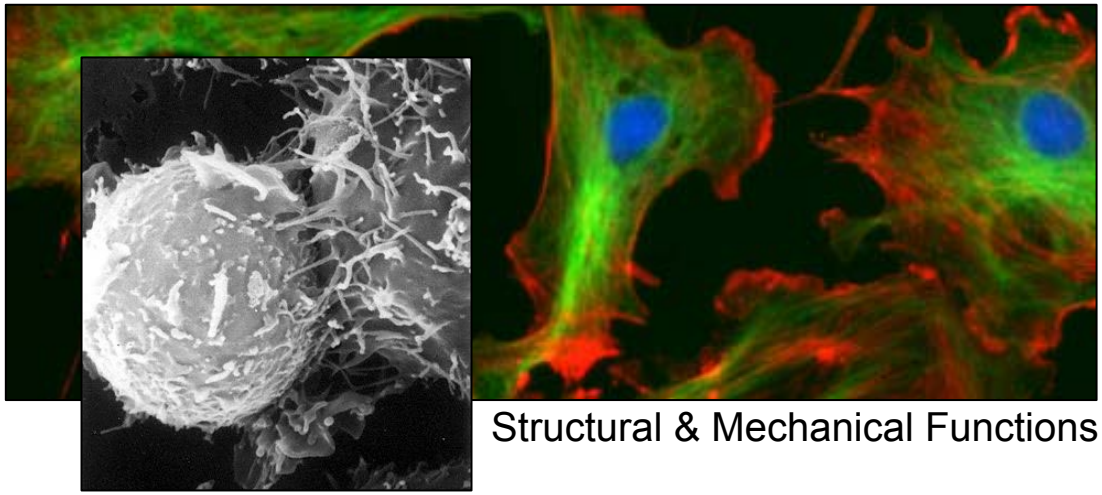
Modeling Biological Systems
Opportunities for Computer Scientists

Filip Jagodzinski
RBO Tutorial Series
25 June 2007

Protein: πρώτα, "prota“, of Primary Importance

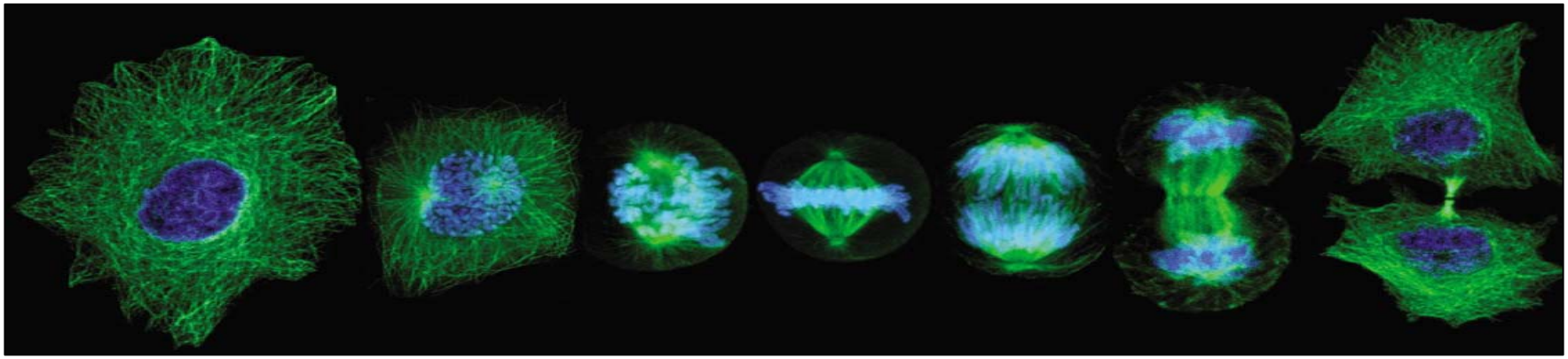
Mmmm … steak



Structural & Mechanical Functions

Immune Response

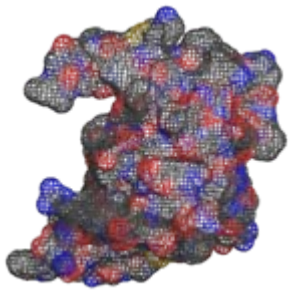
Cellular Division



Tutorial Overview

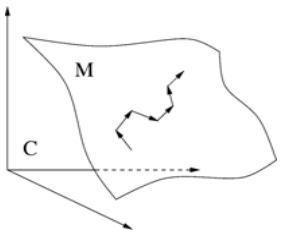
Proteins: Their structure, function, and dynamics

- Proteins as building blocks
- Protein dynamics
- Proteins and their functions
- The physics of protein motion



Modeling & analysis of proteins

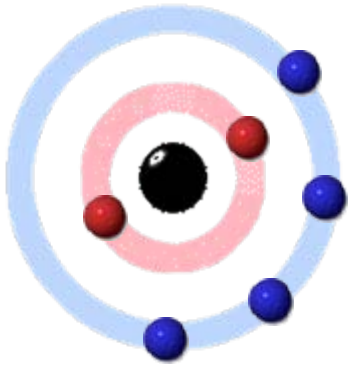
- Snapshots of a protein: Crystallography
- Numerical methods of protein motion: Molecular Dynamics
- Description of protein motion: Normal Mode Analysis and the Elastic Network Model



Opportunities for Improvement

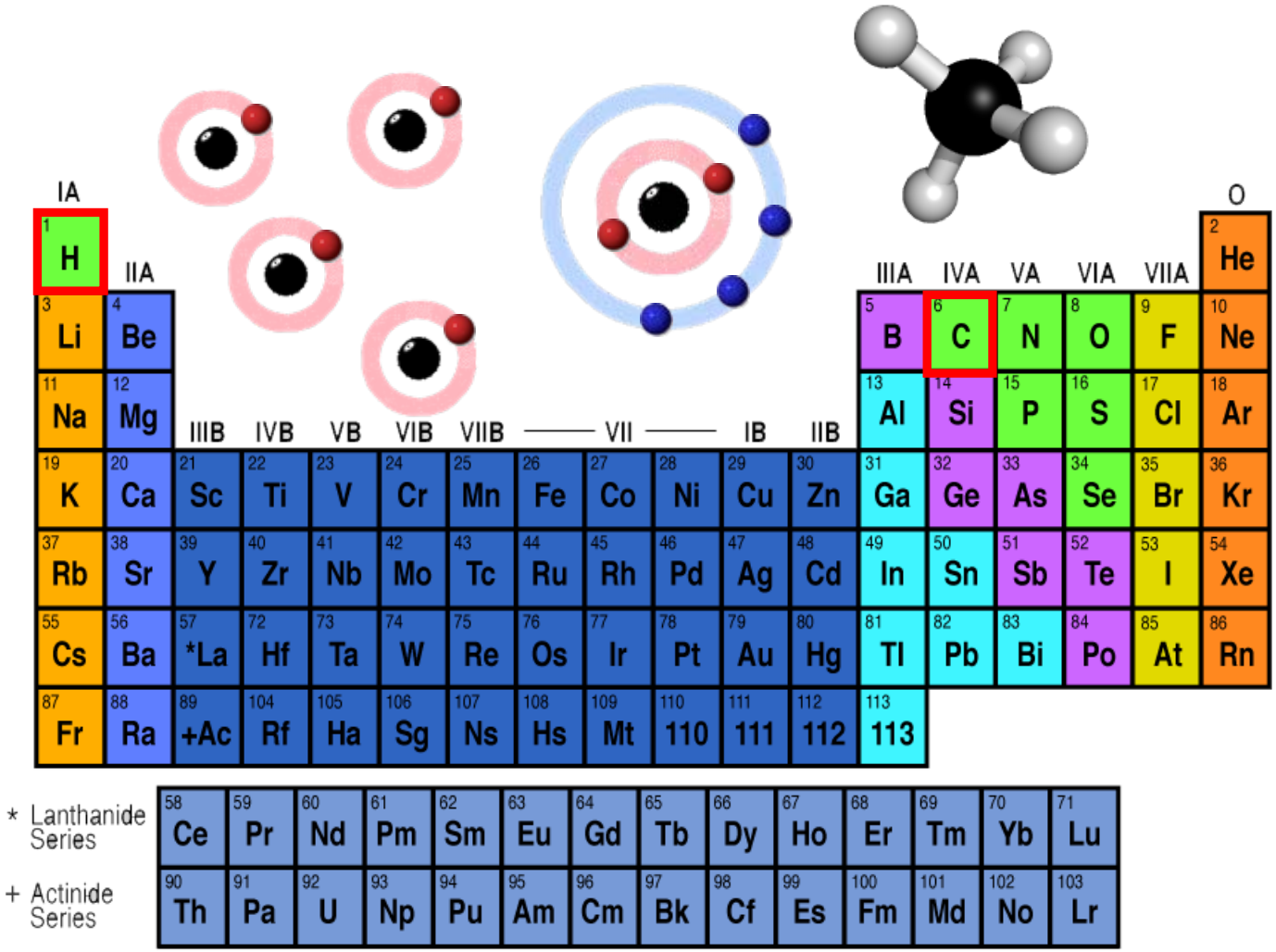
- How can we improve upon MD and NMA
- Are we asking the right questions?
- What features of protein motion should we be looking at?
- Utilize ideas from operational space control; kinematics from mechanics?

Part I: Protein Structure, function, and dynamics

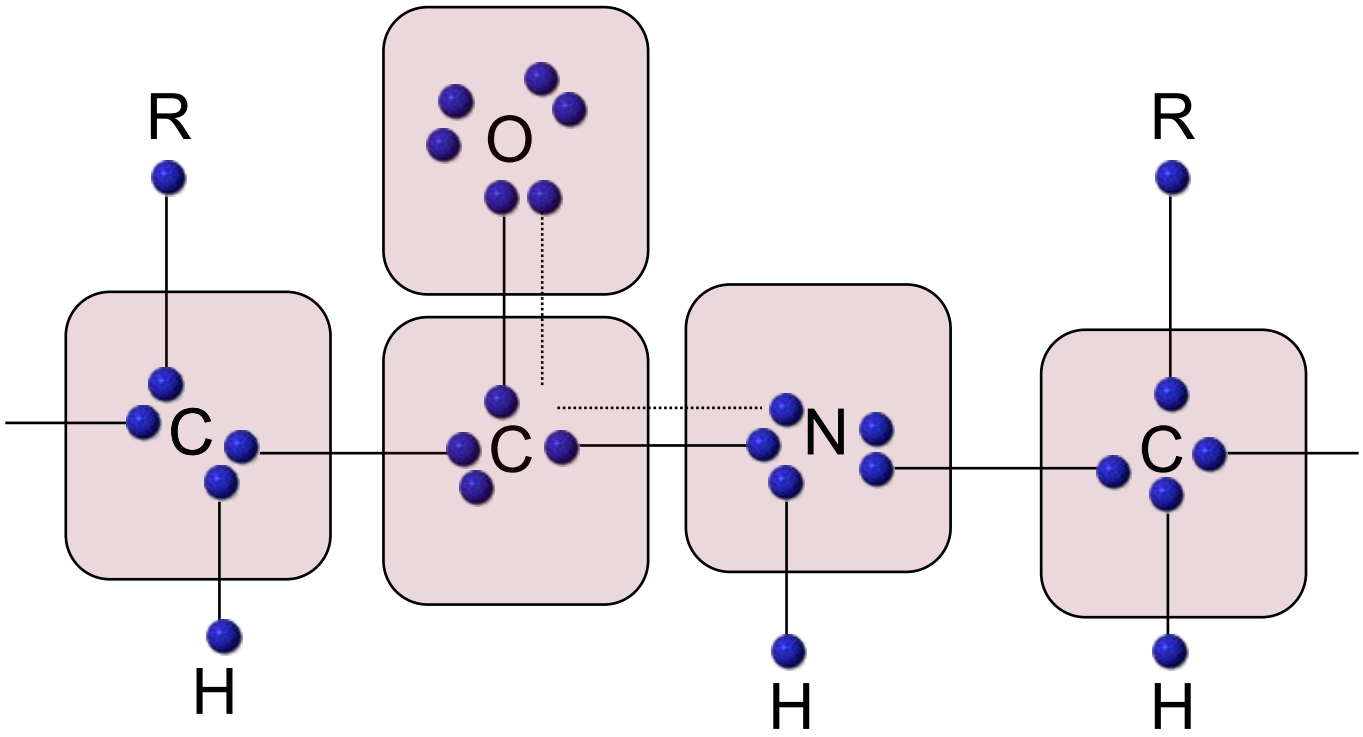


- Proteins as building blocks
- Protein dynamics
- Proteins and their functions
- The physics of protein motion

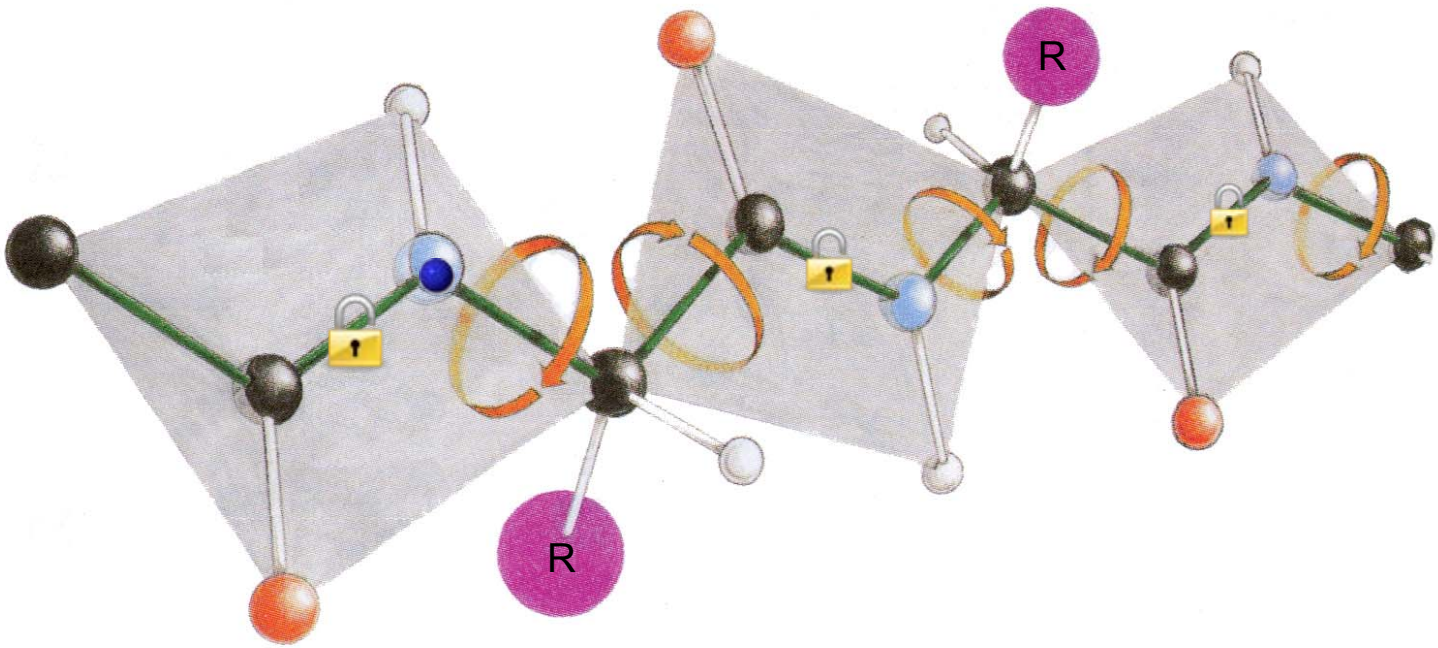
Proteins: Interaction on an Atomic Level…



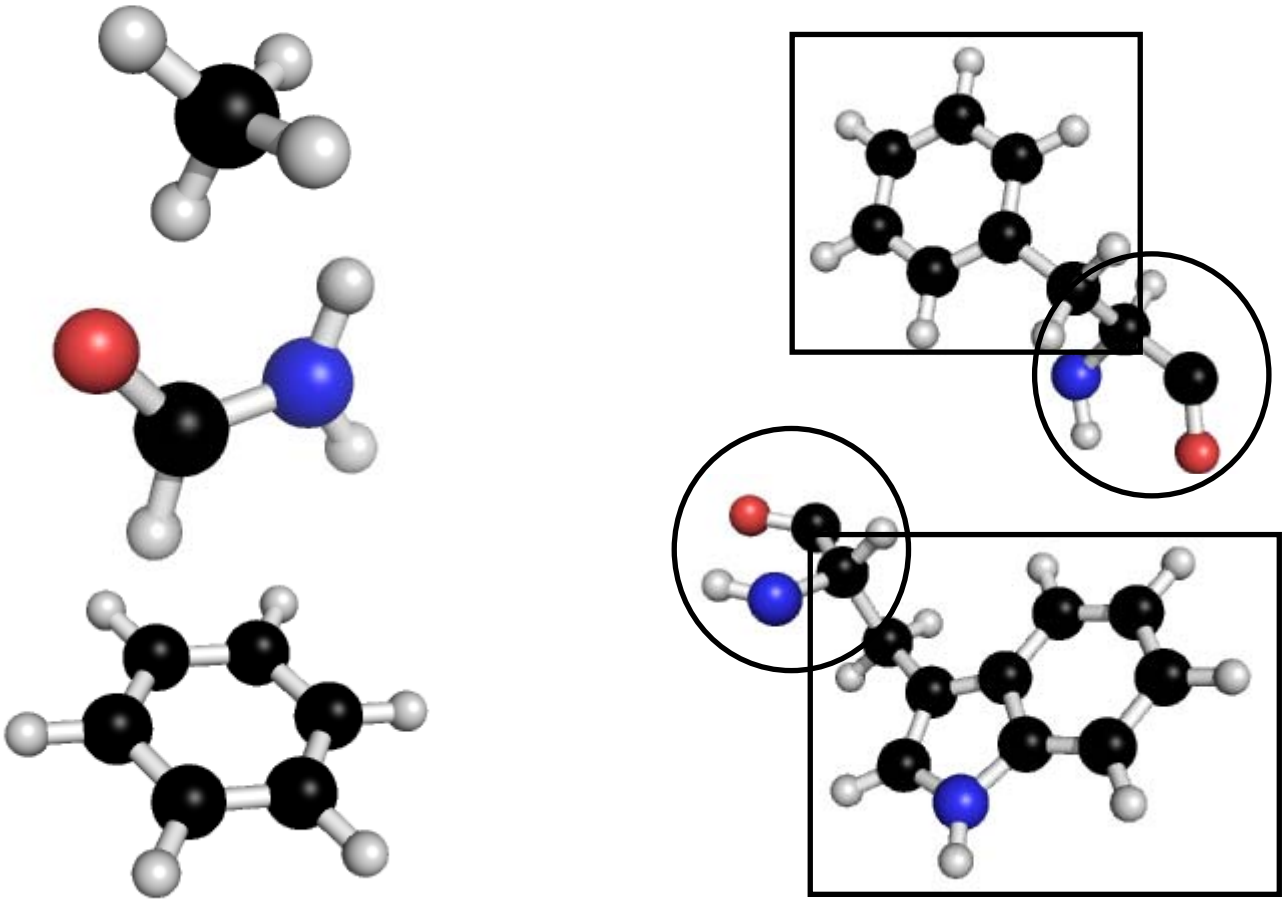
Proteins: Elements Combine



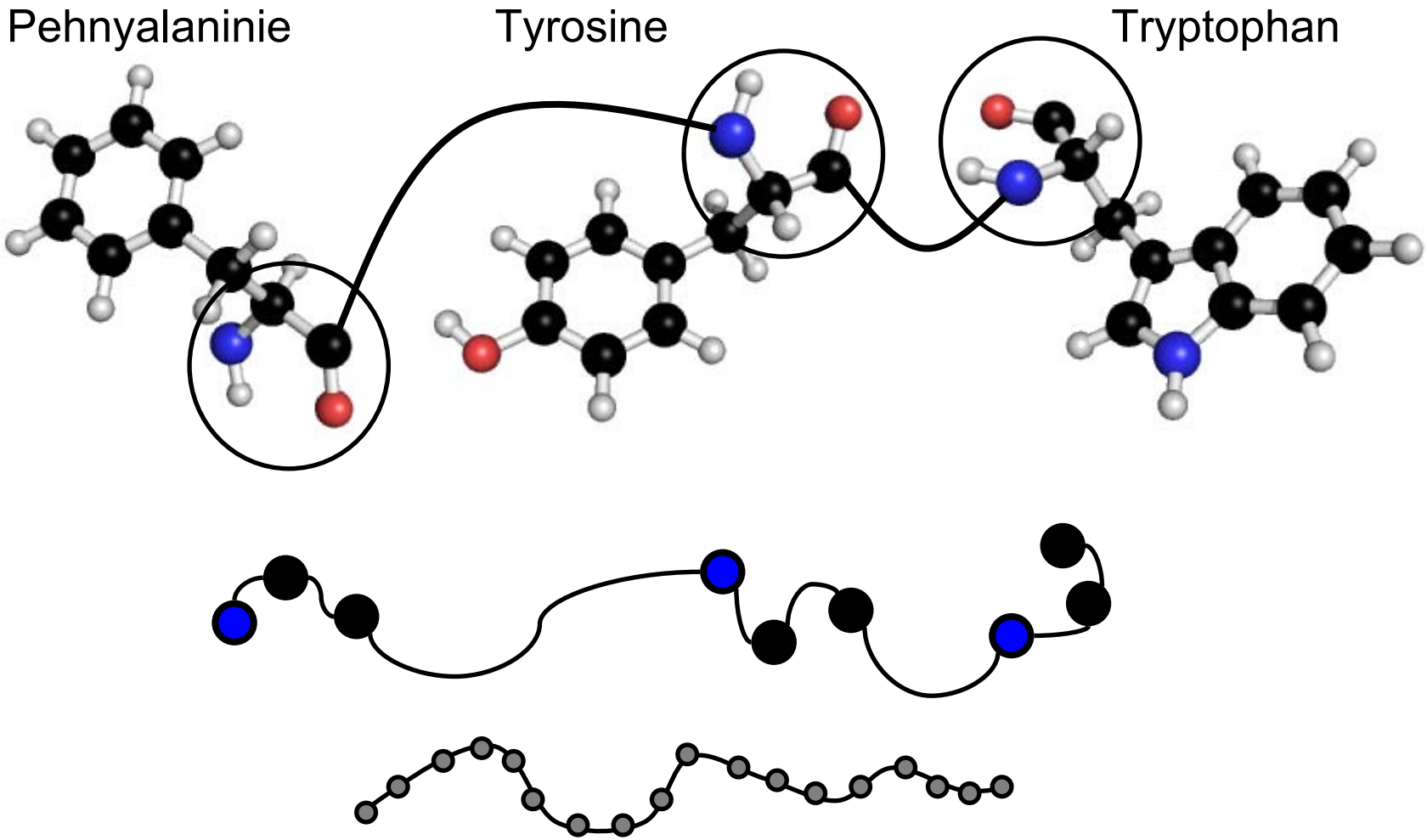
Proteins: Backbone Constraints



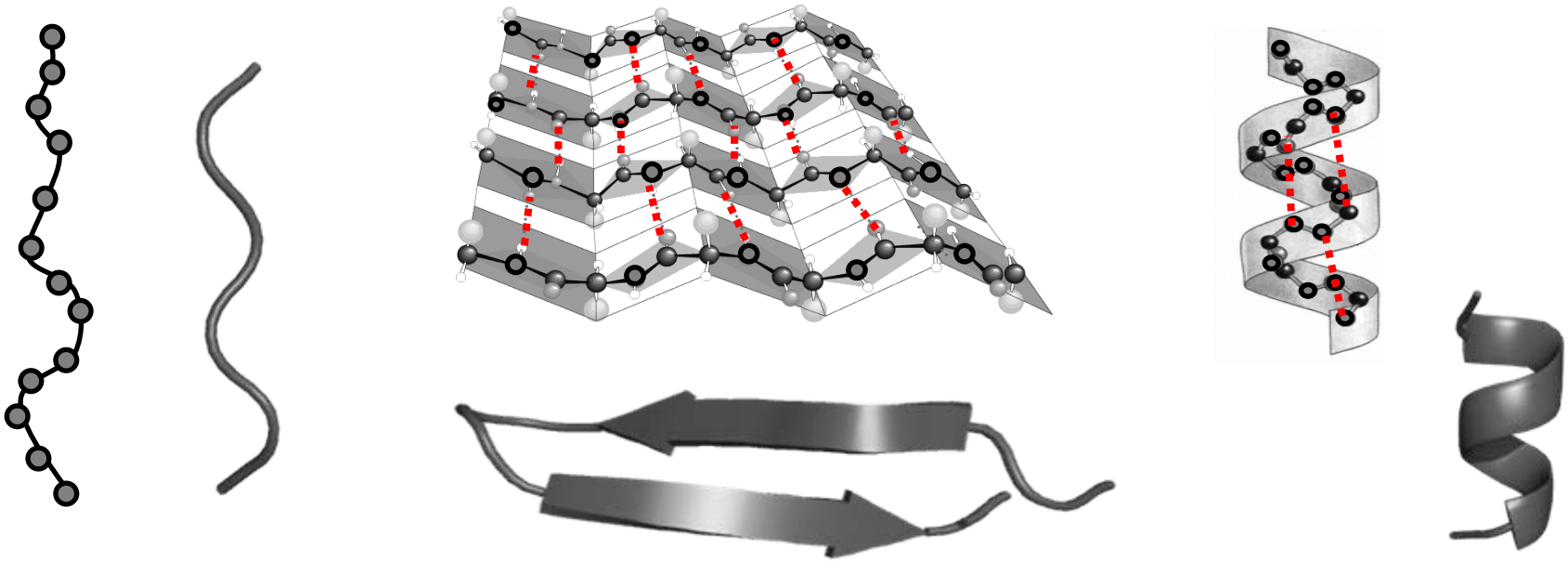
Proteins: Elements Combine



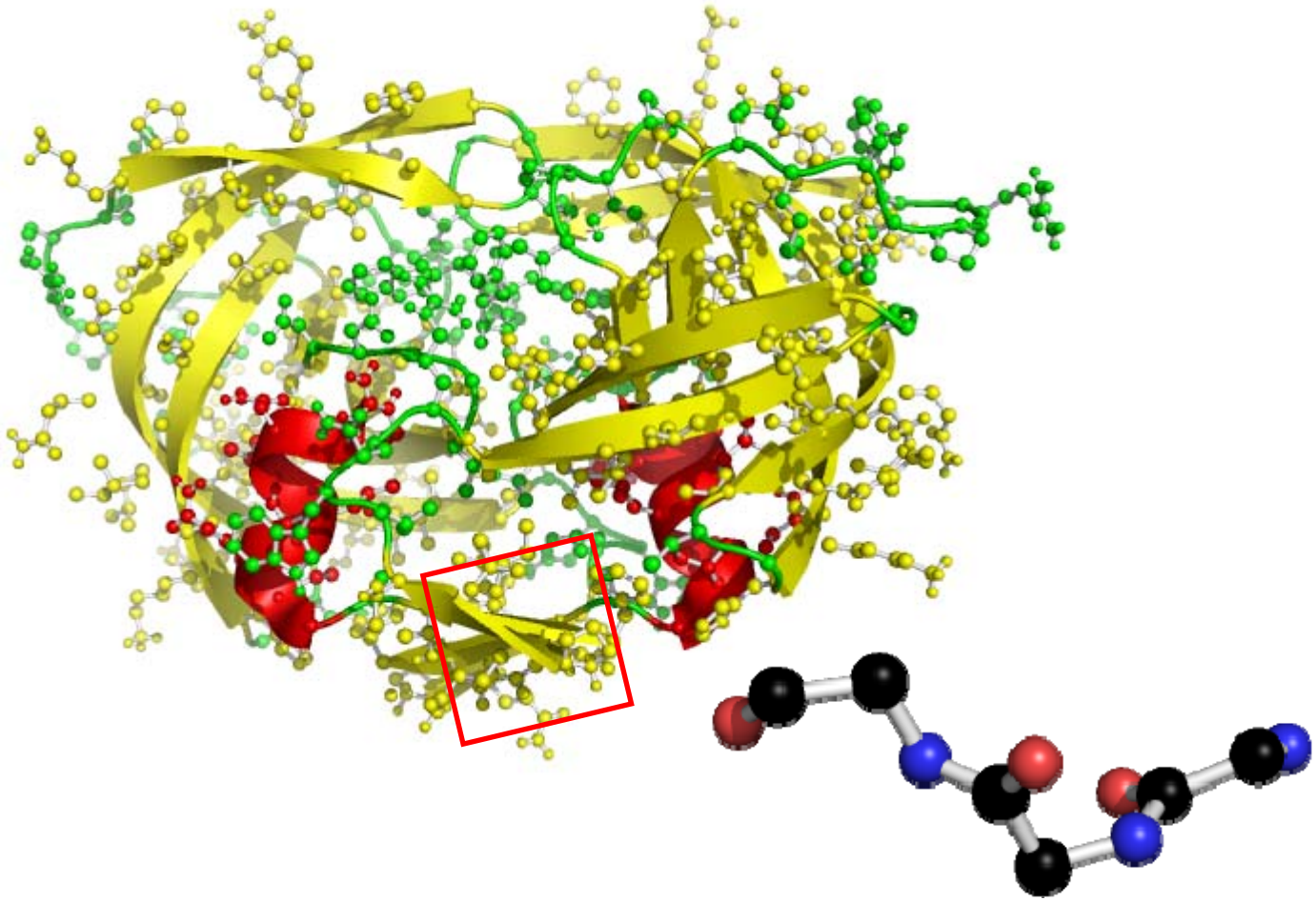
Proteins: From Amino Acids ...



Proteins: to Secondary Structure

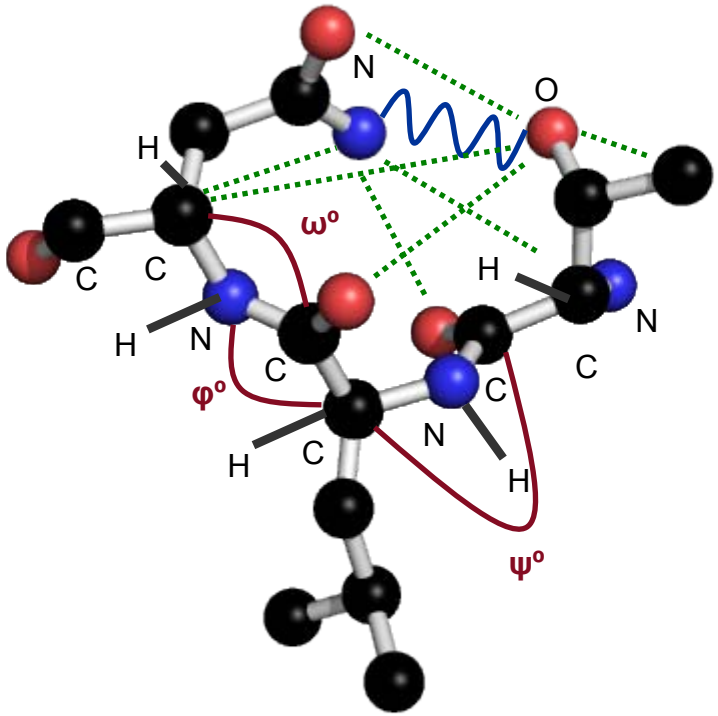


Proteins: MANY! Atom-Atom Interactions



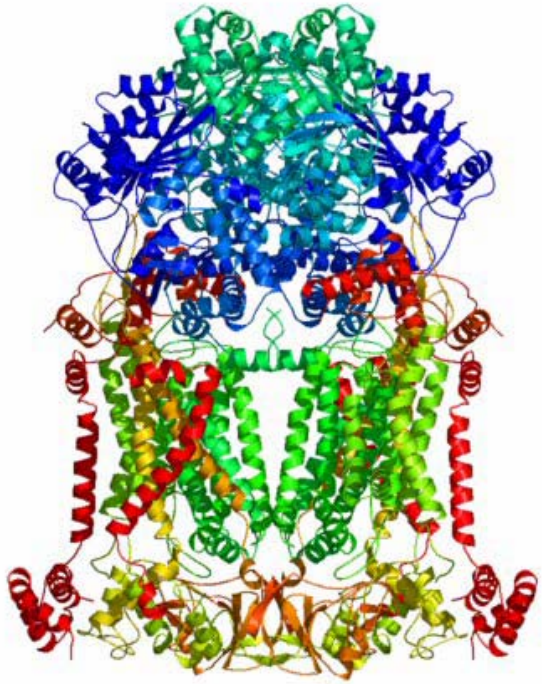
Proteins: Physical Constraints

- Backbone angles
- Sidechain attraction
- Sidechain repulsion
- Bending
- Twisting
- Rigidity
- Non-Twisting

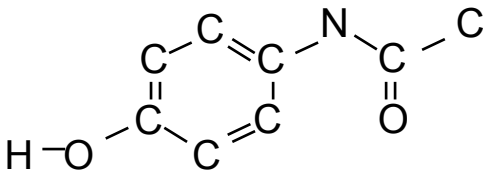


Protein motion

- Is there movement?
- Where does movement occur?
- What is the rate of change?
- Are there rigid regions?
- Can we make atom substitutions?

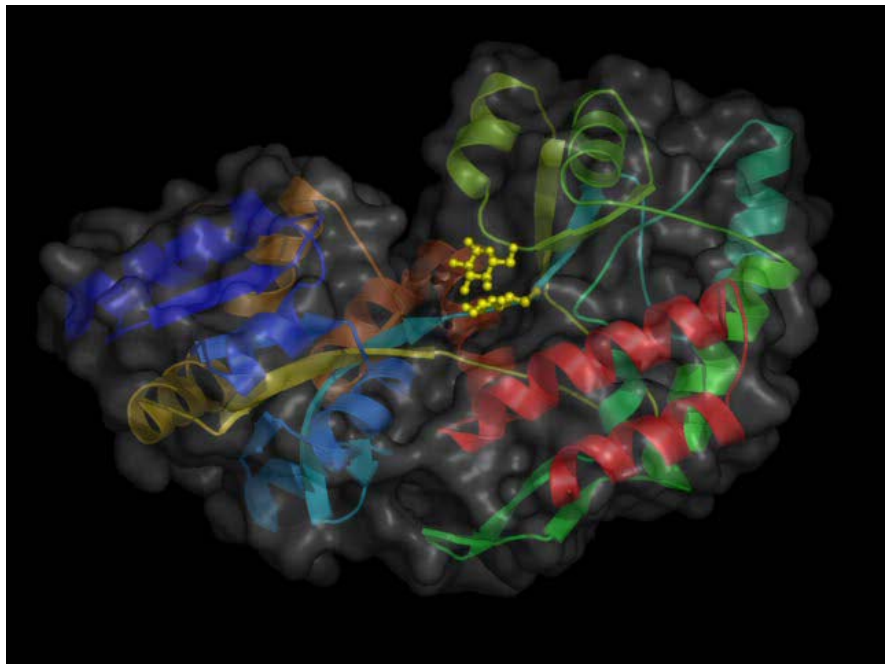


Protein Ligand Compatibility



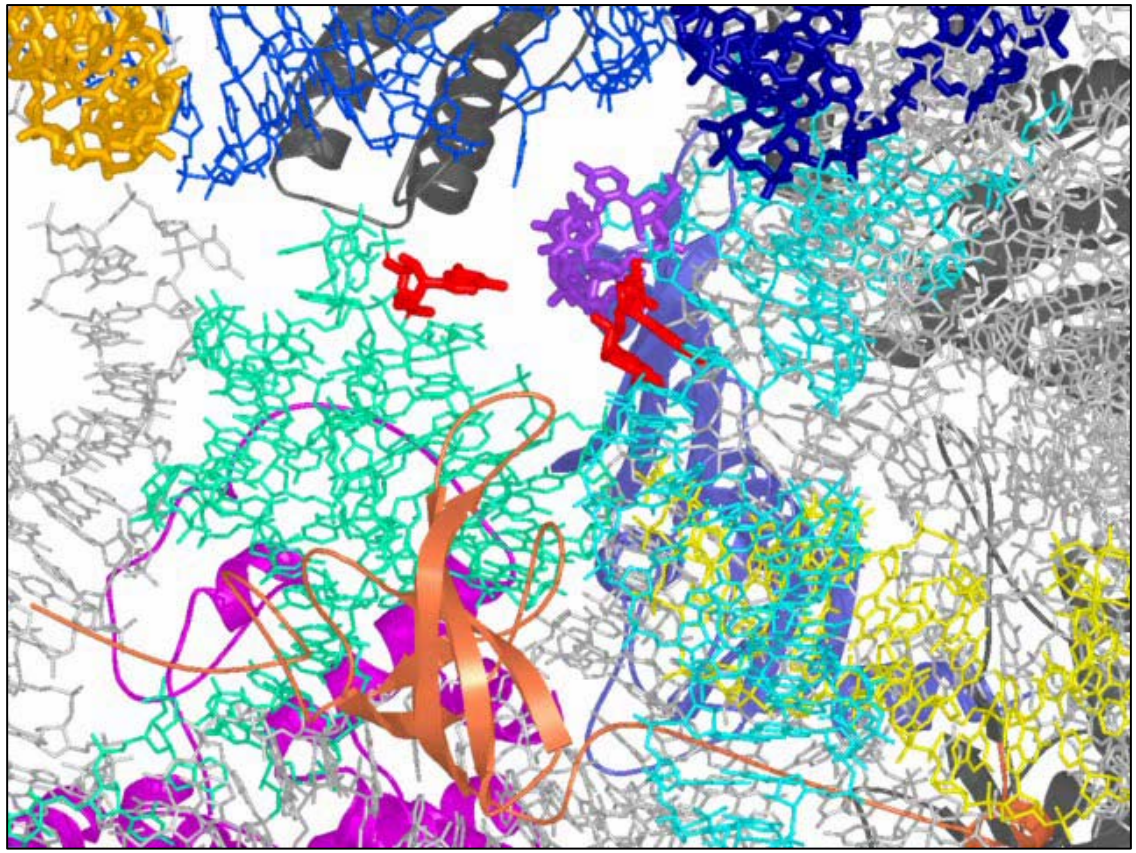
Tylenol

- Which ligands are effective?
- Does a ligand fit into a cleft?
- Is a ligand compatible?

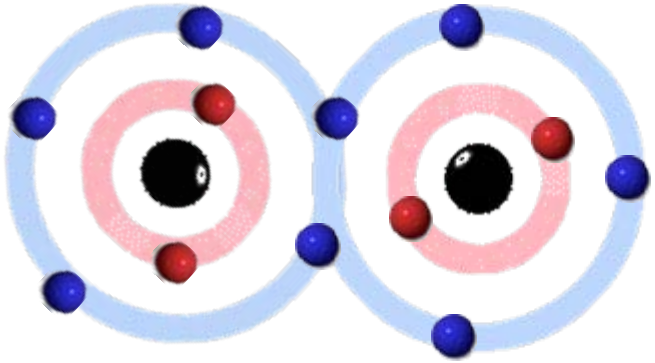


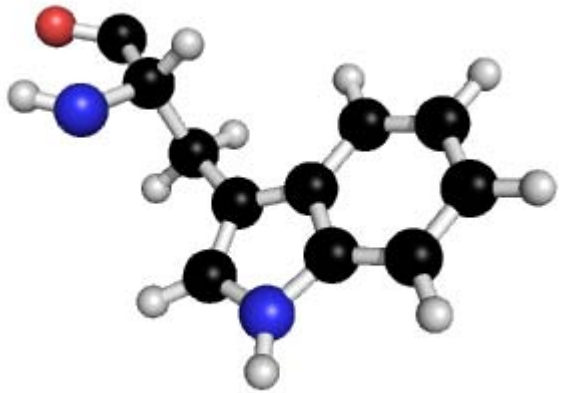
Protein Ligand Interaction

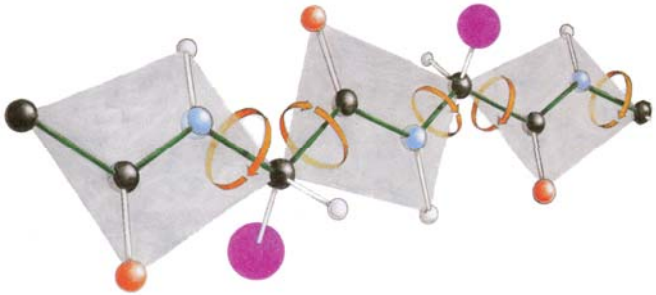
- Is there a feasible approach path?
- Can we modify a known effective ligand and will that new ligand still be effective?

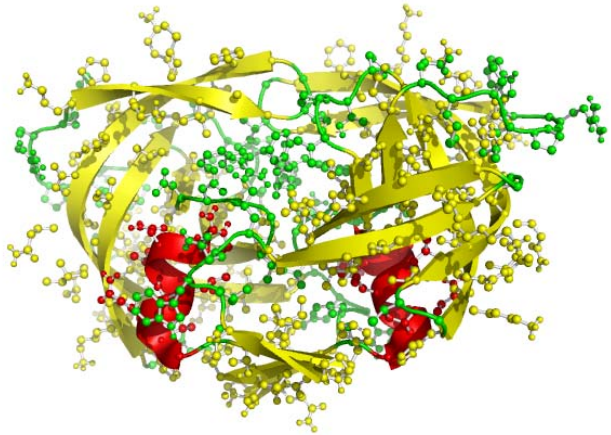


Protein Structure, function and dynamics

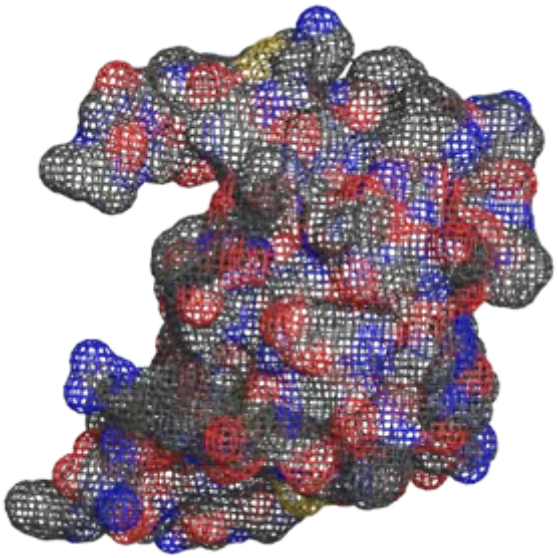






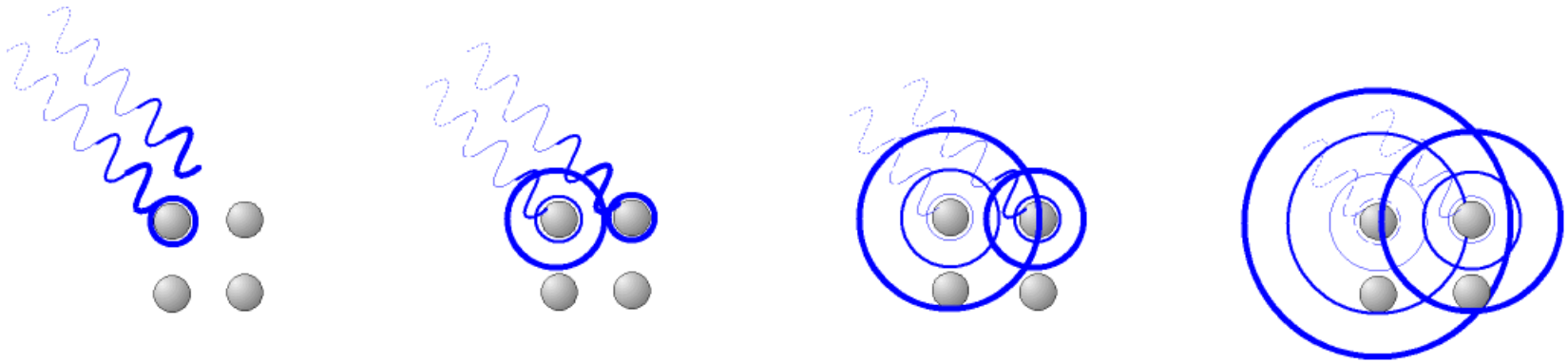


Part II: Modeling and Analysis of Proteins



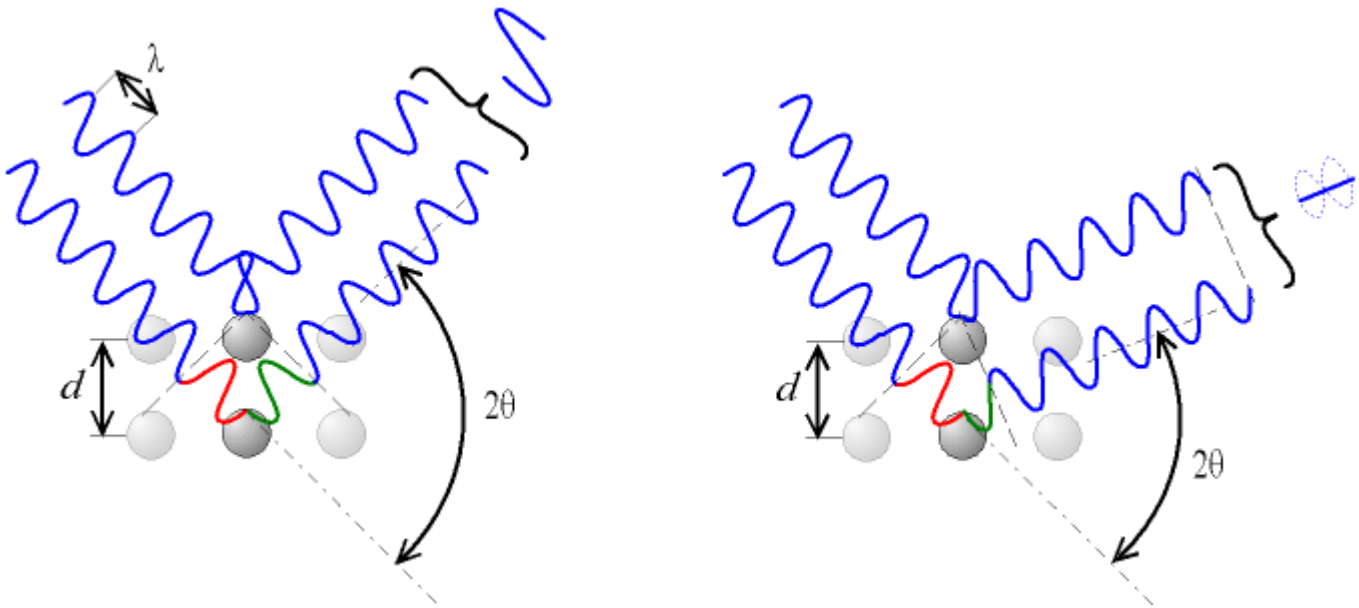
- Snapshots of a protein: Crystallography
- Numerical methods of protein motion: Molecular Dynamics
- Description of protein motion: Normal Mode Analysis and the Elastic Network Model

Observing the behavior of atoms…

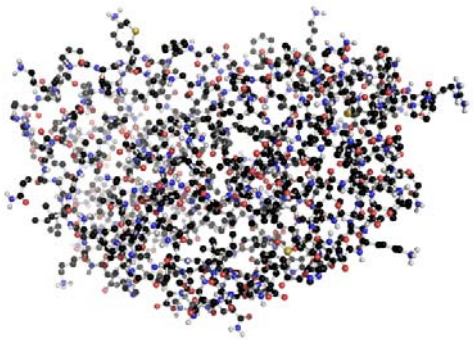


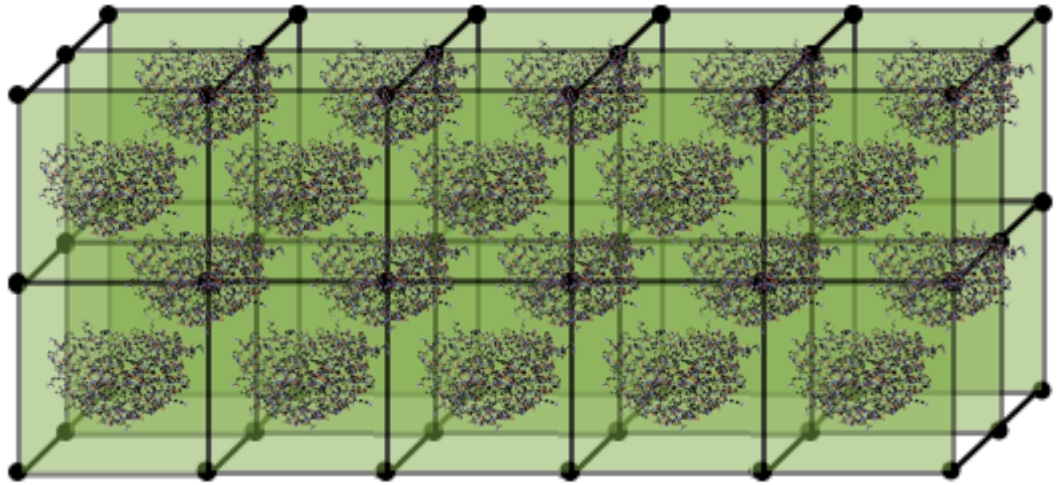
Bragg Scattering

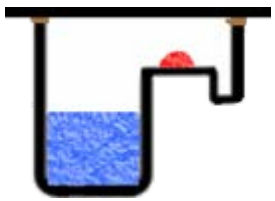
$$2d \sin \theta = n\lambda$$

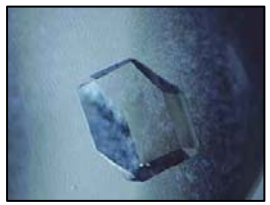


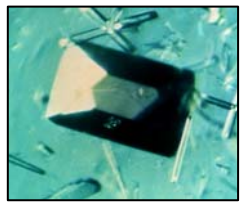
Crystallography – making the crystal

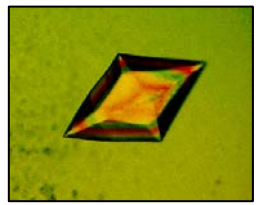


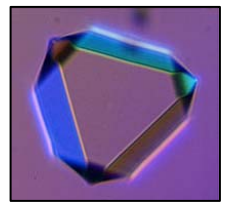


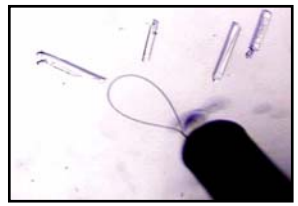




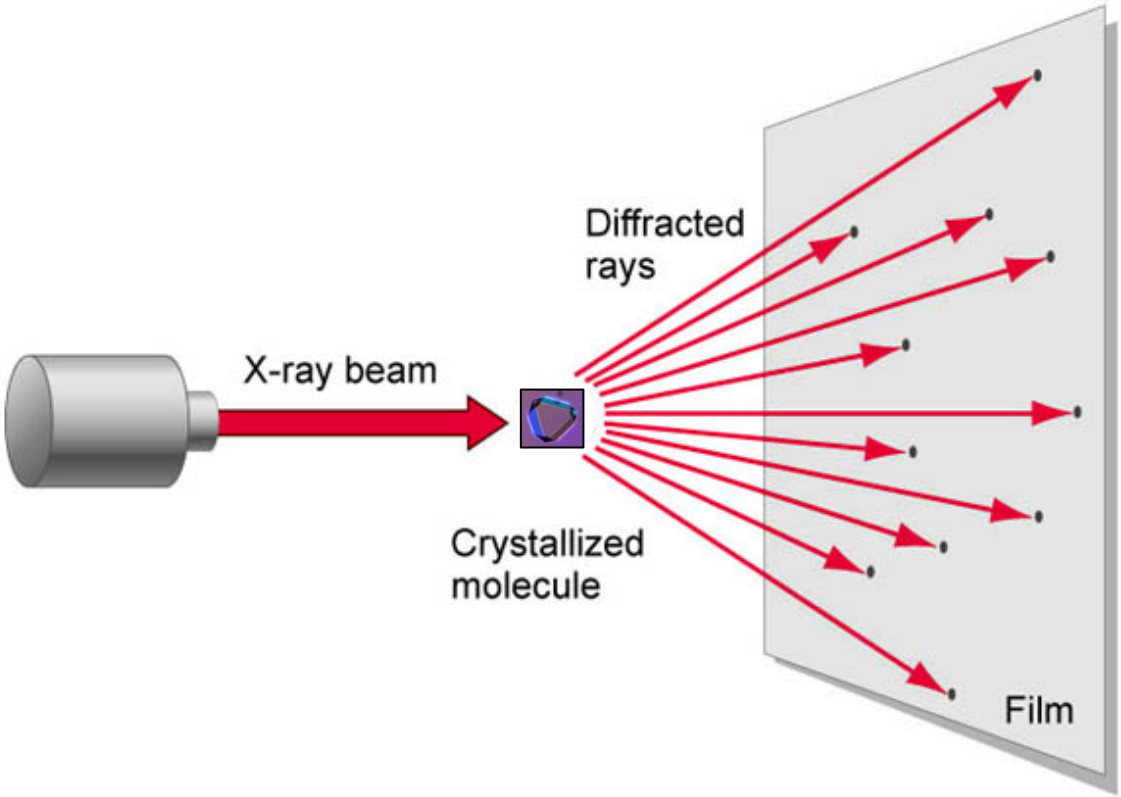




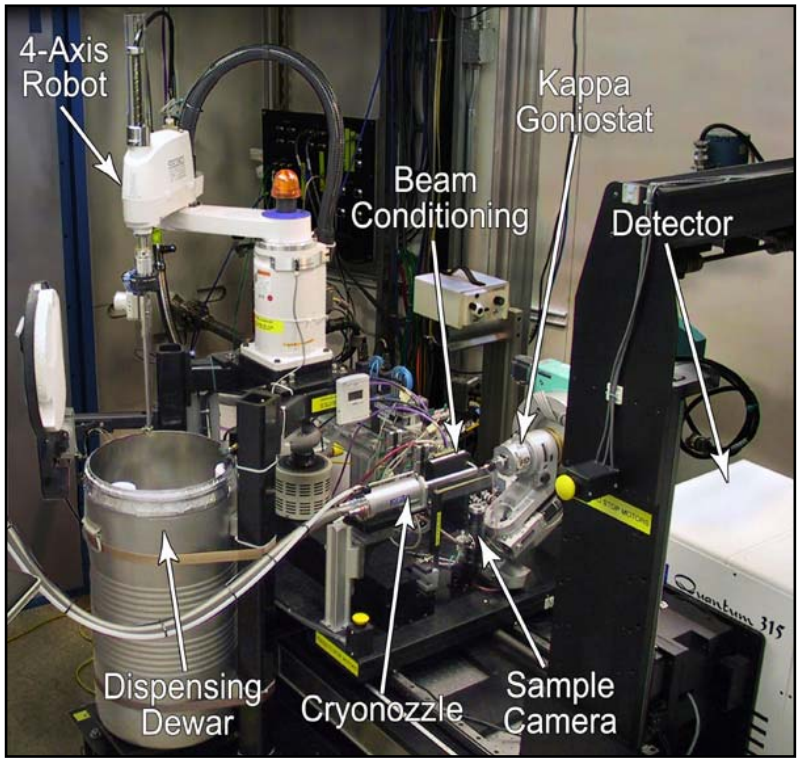


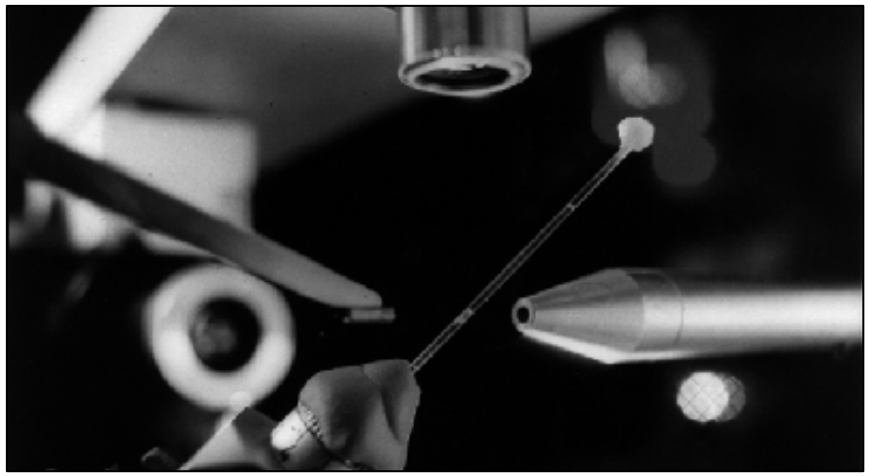


Crystallography – Experimental Setup

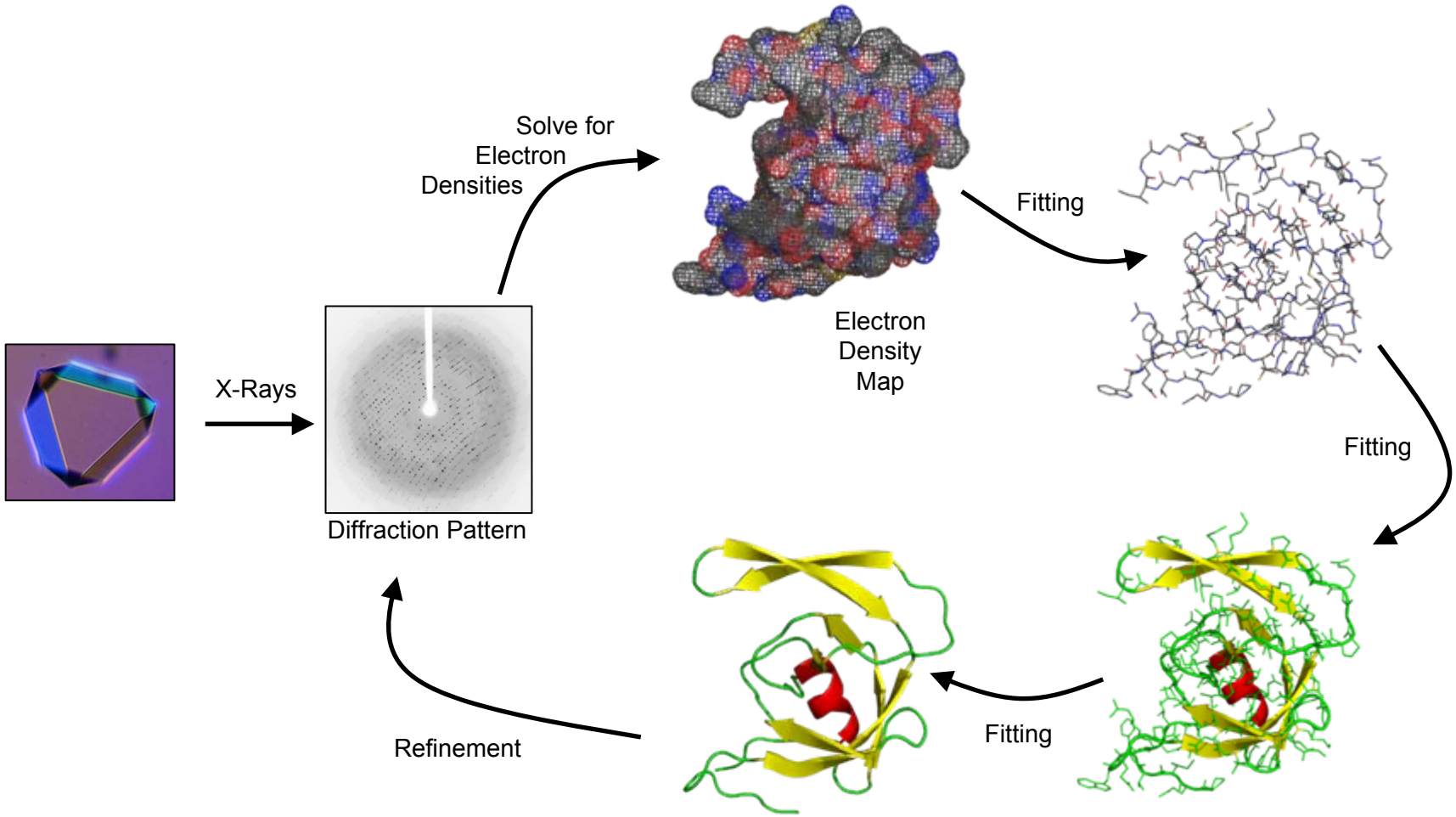


Crystallography – Experimental Setup

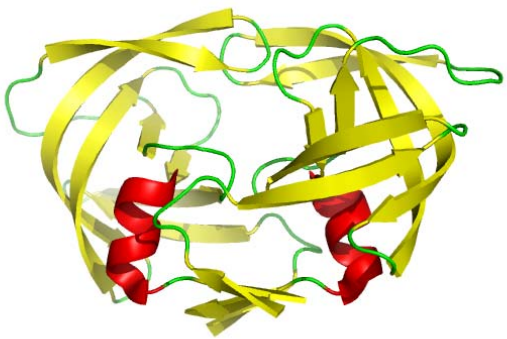




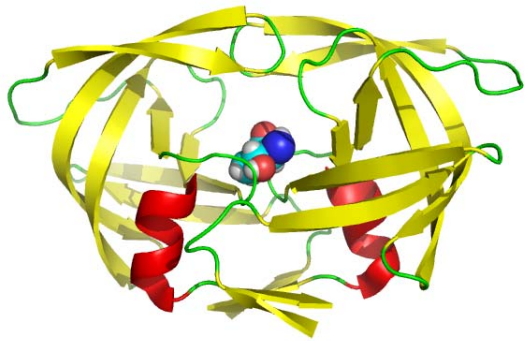
Crystallography



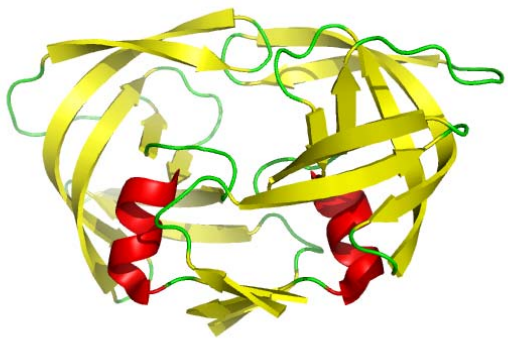
Molecular Dynamics



Protein Rigidity,
Protein Flexibility

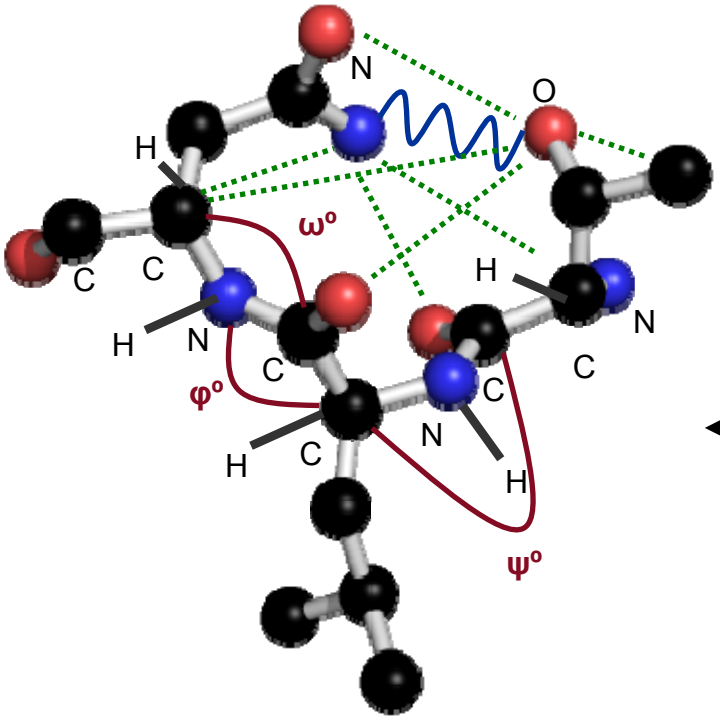


Binding Site /
Ligand Compatibility



Binding Site /
Ligand Interaction

Molecular Dynamics



Atomic Positions & Velocities

Statistical

Mechanics

Pressure
Energy
Temperature
Heat capacities

Molecular Dynamics provides the means to solve the equations of particle motion and evaluate these mathematical formulas.

Potential Energy; Classical Mechanics

$$V = \langle V \rangle = \frac{1}{M} \sum_{i=1}^{M} V_i$$

M is the number of configurations in the molecular dynamics trajectory and V_i is the potential energy of each configuration.

Classical Mechanics

$$F_i = m_i a_i$$

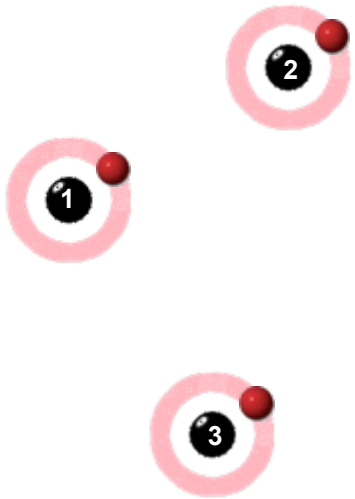
$$F_i = -\nabla_i V$$

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

Relate the potential energy of a system to the changes in position as a function of time.

Why using energy function? This makes sure that we maintain the conservation of energy.

Molecular Dynamics



$$\vec{F}_1(X) = m_1\vec{a}(t)_1 \qquad \boldsymbol{F}(X) = \boldsymbol{M}\boldsymbol{a}(t)$$

$$\vec{F}_2(X) = m_2\vec{a}(t)_2 \qquad \boldsymbol{F}(X) = \boldsymbol{M}\dot{\boldsymbol{V}}(t)$$

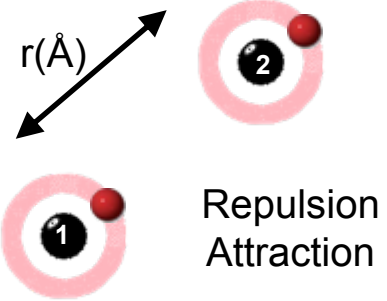
$$\vec{F}_3(X) = m_3\vec{a}(t)_3 \qquad \boldsymbol{F}(X) = -\nabla V(X)$$

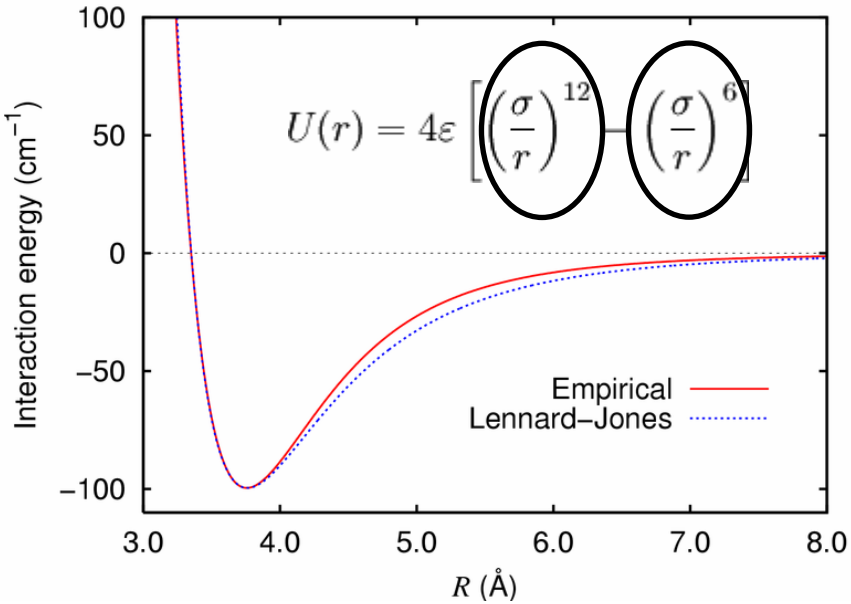
$V(X)$ = of the system is a function of the particle coordinates X
= "potential" in Physics
= "force field" in Chemistry

Given initial positions (X-ray Crystallography) and velocities (randomized Gaussian) of the atoms, we can calculate all future (or past) positions and velocities for all atoms.

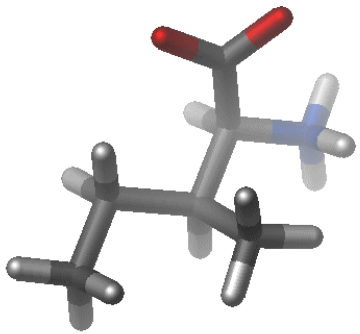
Molecular Dynamics – Force Fields

V(*X*) = description of the terms by which the particles in the system interact





Molecular Dynamics - Motion



$$V(R) = E_{bonded} + E_{non-bonded}$$

$$E_{bonded} = E_{bond-stretch} + E_{angle-bend} + E_{rotate-along-bond}$$

$$E_{non-bonded} = E_{van-der-Waals} + E_{electrostatic}$$

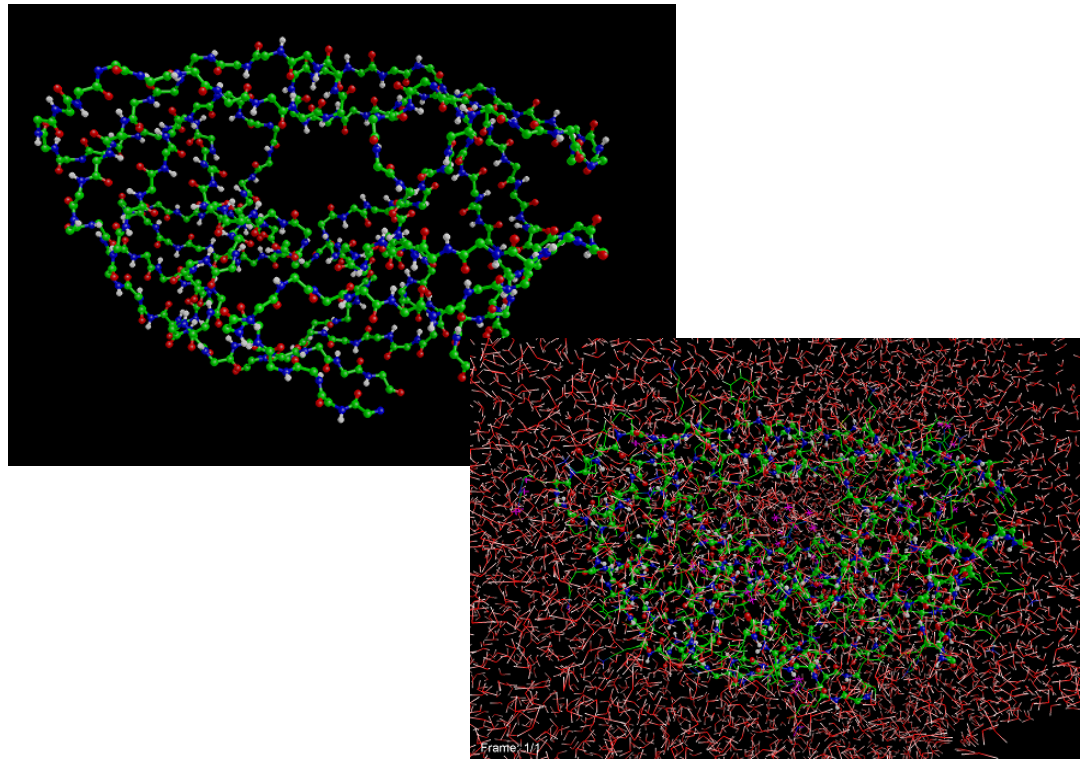
$$E_{van-der-Waals} = \sum_{\substack{nonbonded \\ pairs}} \left(\frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^{6}} \right)$$

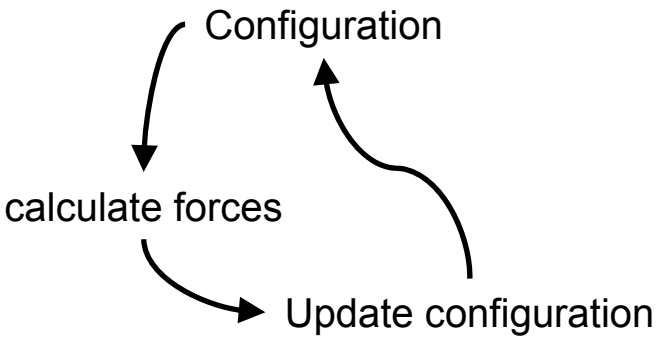
$$E_{electrostatic} = \sum_{\substack{nonbonded \\ pairs}} \frac{q_i q_k}{D r_{ik}}$$

Molecular Dynamics – Sample Run

1. Atom Coordinates
2. Add hydrogen bonds
3. Solvate Protein
4. Energy Minimization
5. Dynamics of water
6. Dynamics of protein

```
ATOM      1  N   PRO A   1      -12.735  38.918  31.287  1.00 39.83
ATOM      2  CA  PRO A   1      -12.709  39.097  29.830  1.00 39.29
ATOM      3  C   PRO A   1      -13.575  38.051  29.162  1.00 39.78
ATOM      4  O   PRO A   1      -14.097  37.126  29.753  1.00 38.67
ATOM      5  CB  PRO A   1      -11.243  39.010  29.398  1.00 37.79
ATOM      6  CG  PRO A   1      -10.636  38.128  30.469  1.00 38.69
```





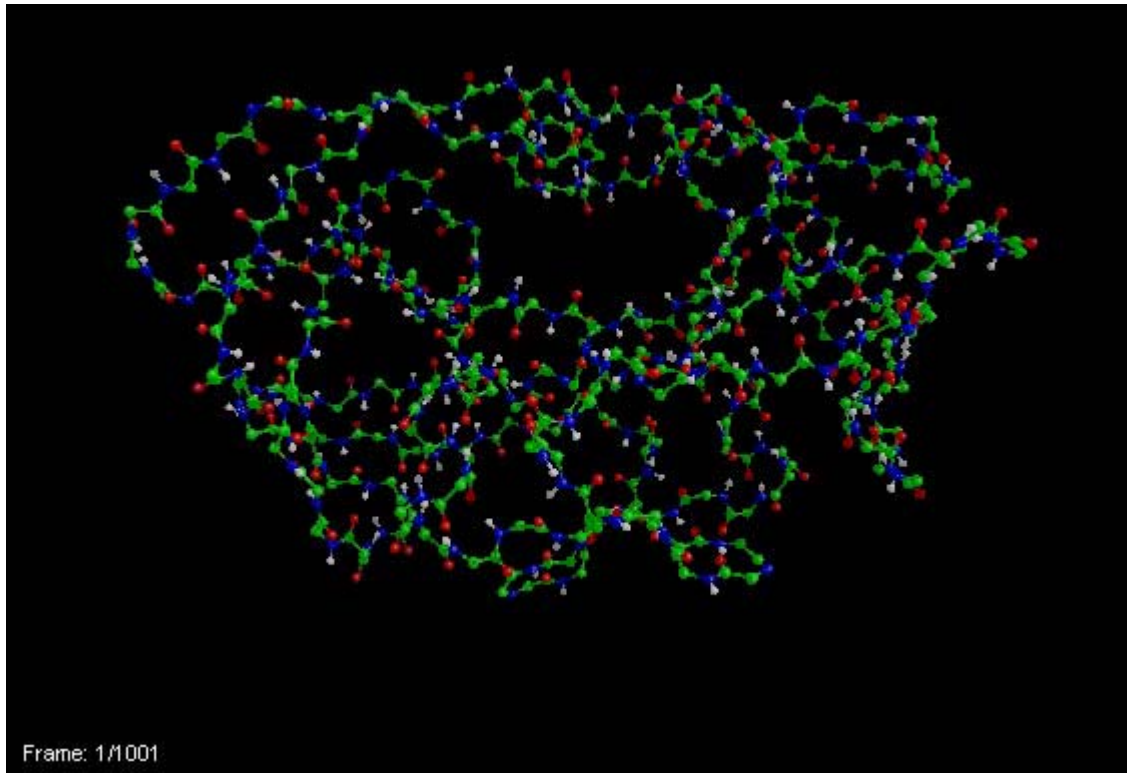
Molecular Dynamics – Sample Run Output

Molecular Dynamics

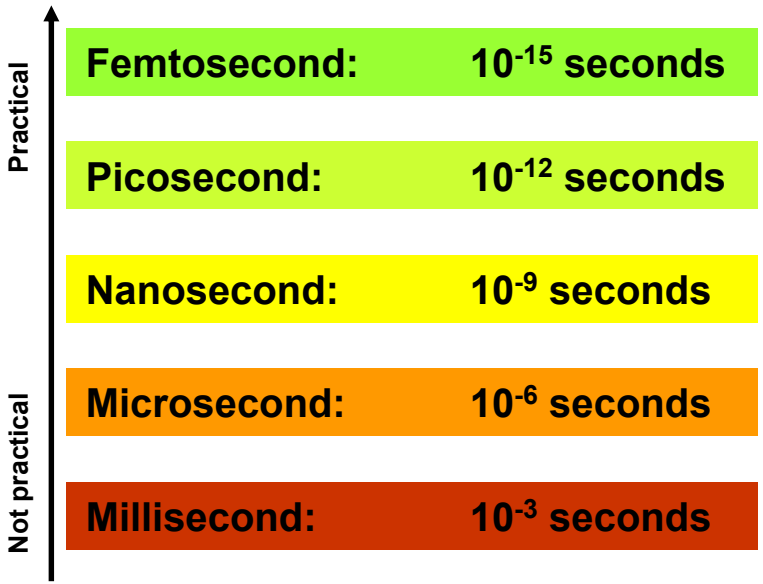
Sample Run
- 100,000 integration steps
- 2.0 femtoseconds
- 200 picoseconds
- 22,009,885 x 10^6 fp ops
- 3 hours 37 minutes, 1CPU

Flap Curling
- 10 nanoseconds
- 10,000 picoseconds
- ~170 hours

Flap Closing/Opening
- 10 microseconds
- 10,000,000 picoseconds
- ~7,000 days



Molecular Dynamics – Is it good? Bad? Feasible?



Local Motions (0.01 to 5 Å, 10^{-15} to 10^{-1} s)

- Atomic Fluctuations
- Sidechain Motions
- Loop Motions

Rigid Body Motions (1 to 10Å, 10^{-9} to 1s)

- Helix Motions
- Domain Motions
- Subunit Motions

Large-Scale Motions (> 5Å, 10^{-7} to 10^{4} s)

- Helix coil transitions
- Dissociation/Associations
- Folding and unfolding

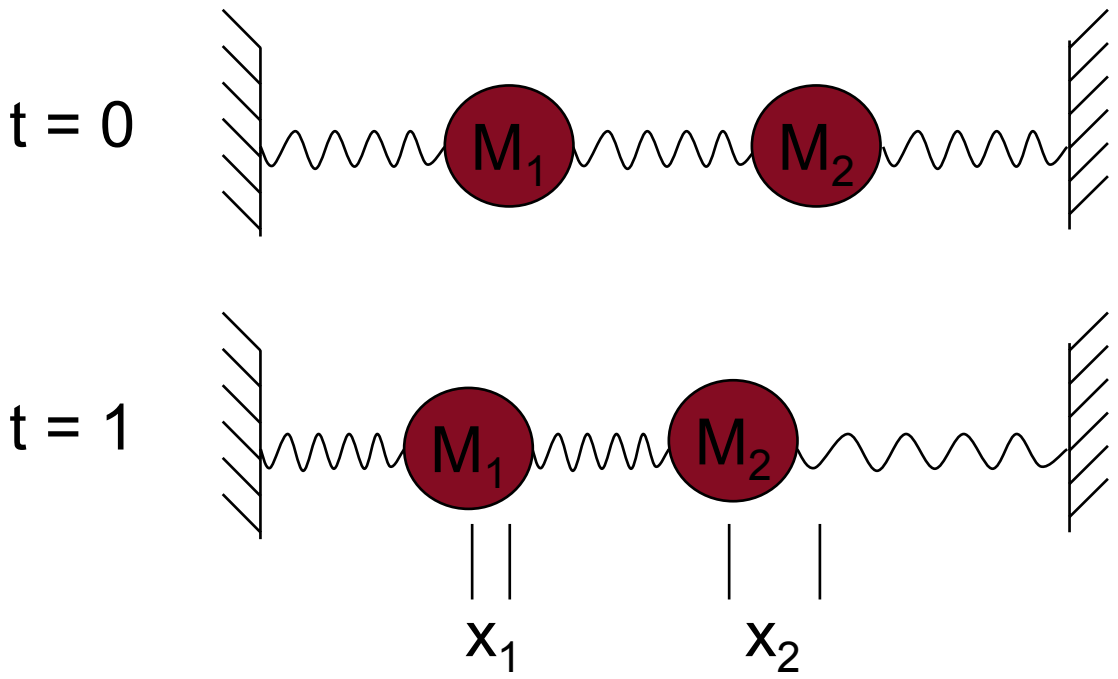
Normal Modes – What are they?

- A **normal mode** in an oscillating system is the frequency at which a deformable structure will oscillate when disturbed.

- A set of these frequencies are unique to each structure.

- Why is this relevant to proteins?

Normal Mode Analysis – Harmonic Oscillations



$F = -kx$

$F = -dV/dx = -kx$

$V = 0.5k\ x^2$

$k = d^2V/dx^2$

$F = ma$

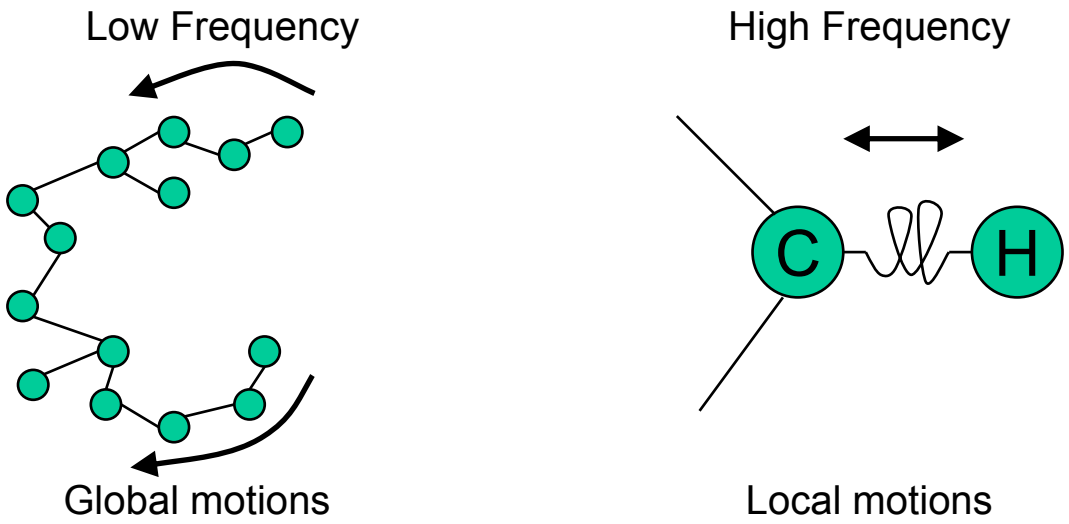
$md^2x/dt^2 = -kx$

$x(t) = A \sin(2\pi\nu t)$

$\mathbf{-4\pi^2\nu^2mx = -kx}$

ν = Frequency of vibration

Normal Modes – for Proteins?



1. Each normal mode acts like a simple harmonic oscillator.
2. A normal mode is a concerted motion of many atoms.
3. Normal modes are independent; they don't interact.
4. Normal mode analysis is less expensive than MD simulation, but requires much more memory.

Approximate the potential energy and describe the motion of the molecules as collection of independent harmonic oscillators.

Normal Mode Analysis

$$k = \frac{d^2 V}{dx^2}$$

Molecular orbital calculations are used to find the potential energy of the molecules as a function of the position of each atom.

The second derivative of the potential energy can then be used to calculate the force constants.

$$\frac{\partial^2 V}{\partial x_1^2} = k_{xx}^{11}$$

The change of the force on atom 1 in the x-direction when you move atom 1 in the x-direction.

$$\frac{\partial^2 V}{\partial x_1 \partial y_2} = k_{xy}^{12}$$

The change of the force on atom 1 in the x-direction when you move atom 2 in the y-direction.

Normal Mode Analysis – derivation 3

$$-4\pi^2\nu^2 m_1x_1 = -k_{xx}^{11}x_1 - k_{xy}^{11}y_1 - k_{xz}^{11}z_1 - k_{xx}^{12}x_2 - k_{xy}^{12}y_2 - \ldots - k_{xz}^{1N}z_N$$

$$-4\pi^2\nu^2 m_1y_1 = -k_{yx}^{11}x_1 - k_{yy}^{11}y_1 - k_{yz}^{11}z_1 - k_{yx}^{12}x_2 - k_{yy}^{12}y_2 - \ldots - k_{yz}^{1N}z_N$$

$$\vdots$$

$$-4\pi^2\nu^2 m_2x_2 = -k_{xx}^{21}x_1 - k_{xy}^{21}y_1 - k_{xz}^{21}z_1 - k_{xx}^{22}x_2 - k_{xy}^{22}y_2 - \ldots - k_{xz}^{2N}z_N$$

$$\vdots$$

$$-4\pi^2\nu^2 m_Nz_N = -k_{zx}^{N1}x_1 - k_{zy}^{N1}y_1 - k_{zx}^{N1}z_1 - k_{zx}^{N2}x_2 - k_{zy}^{N2}y_2 - \ldots - k_{zz}^{NN}z_N$$

The complete list of force constants

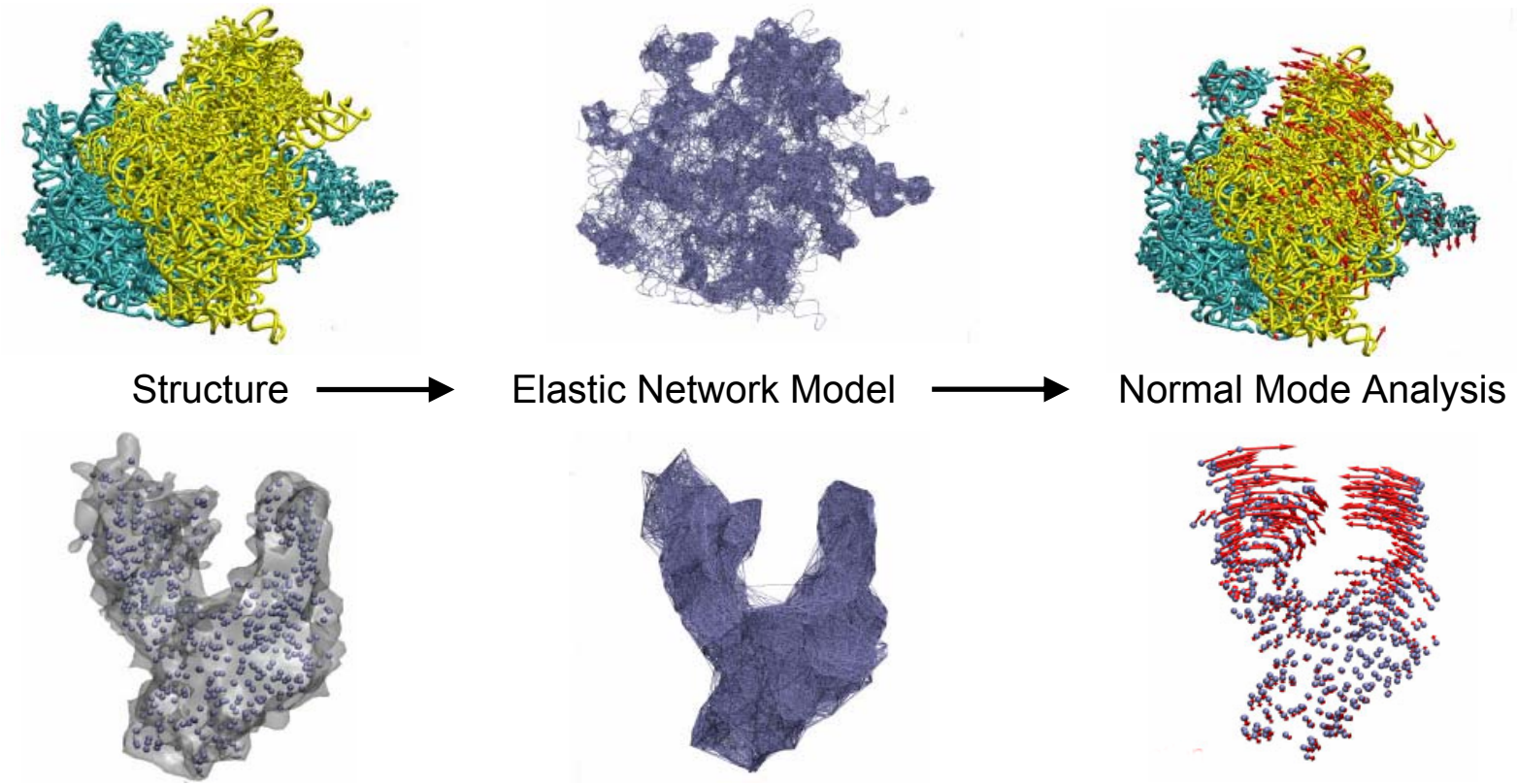
symmetric matrix

$$-\begin{pmatrix} \frac{k_{xx}^{11}}{\sqrt{m_1}\sqrt{m_1}} & \frac{k_{xx}^{12}}{\sqrt{m_1}\sqrt{m_2}} & \frac{k_{xx}^{13}}{\sqrt{m_1}\sqrt{m_3}} \\ \frac{k_{xx}^{21}}{\sqrt{m_2}\sqrt{m_1}} & \frac{k_{xx}^{22}}{\sqrt{m_2}\sqrt{m_2}} & \frac{k_{xx}^{23}}{\sqrt{m_2}\sqrt{m_3}} \\ \frac{k_{xx}^{31}}{\sqrt{m_3}\sqrt{m_1}} & \frac{k_{xx}^{32}}{\sqrt{m_3}\sqrt{m_2}} & \frac{k_{xx}^{33}}{\sqrt{m_3}\sqrt{m_3}} \end{pmatrix} \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{pmatrix} = -4\pi^2\nu^2 \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{pmatrix}$$

eigenvalue-eigenvector equation

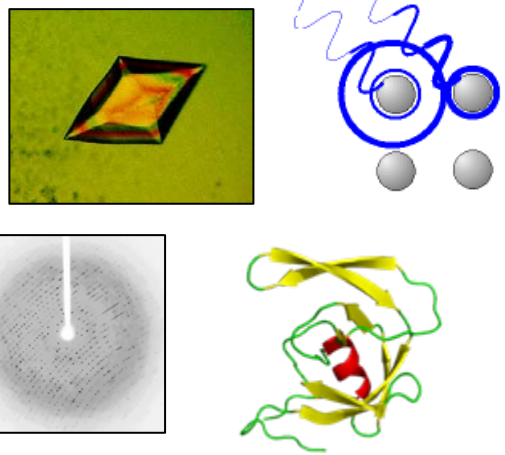
Eigenvalues = normal mode frequencies
Eigenvectors = coordinate displacements

Normal Mode Analysis – RNAP Example

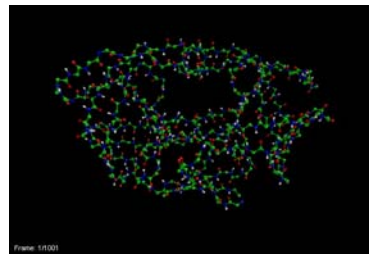


Modeling and Analysis of Proteins

Crystallography

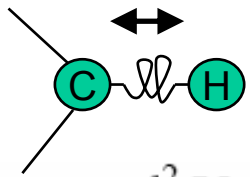


Molecular Dynamics

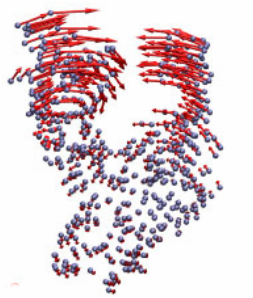


$$U(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

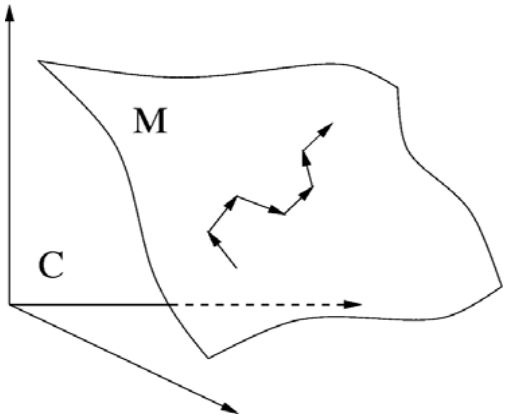
Normal Mode Analysis



$$k = \frac{d^2 V}{dx^2}$$



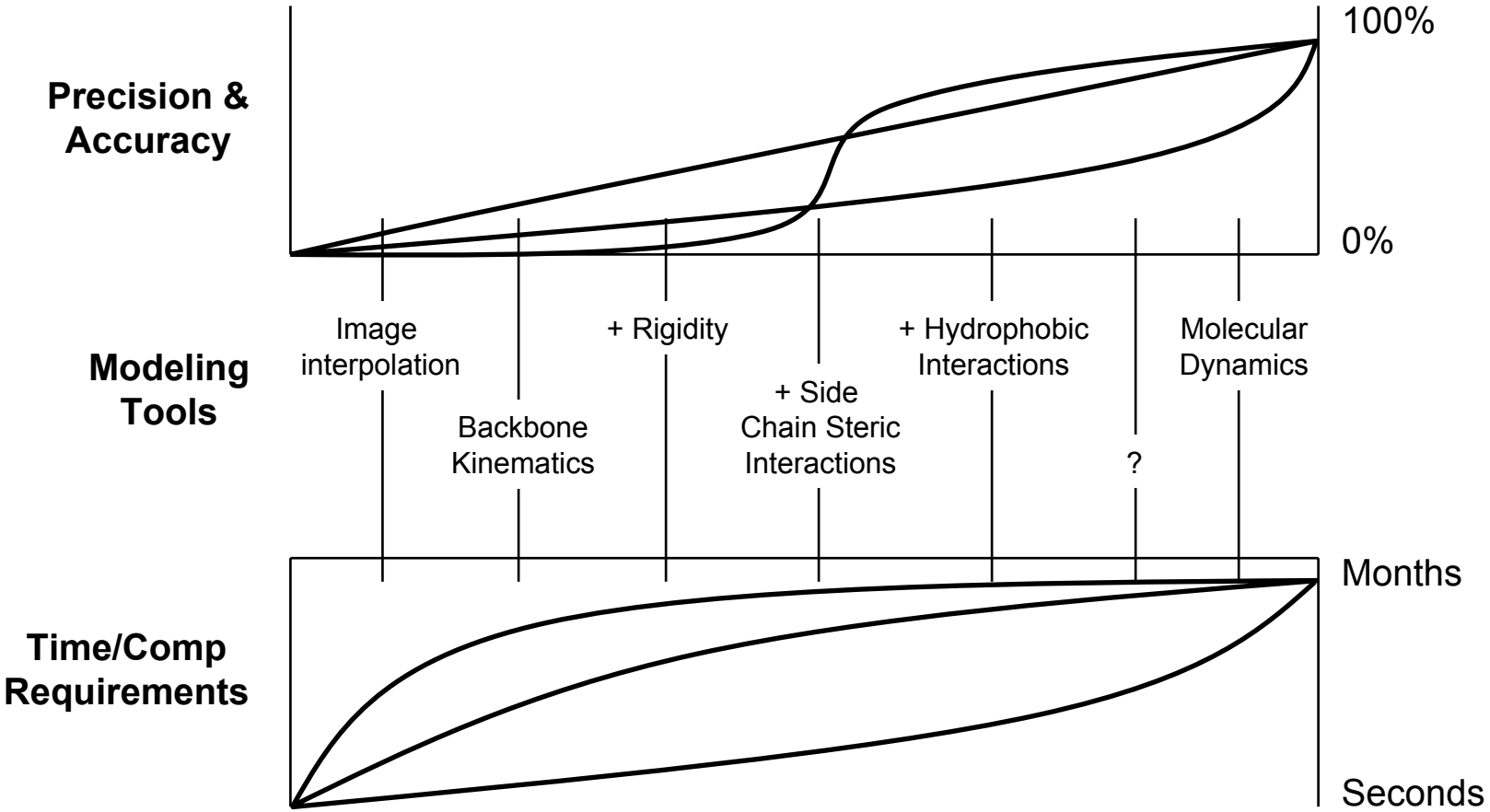
Part III: Opportunities for Improvement



- How can we improve upon MD and NMA
- Are we asking the right questions?
- What features of protein motion should we be looking at?
- Utilize ideas from operational space control; kinematics from mechanics.

Choose the simplest representation that will illustrate the property of interest satisfactorily.
-Tsjerk Wassenaar, Alex de Vries, Dr. Alan E. Mark

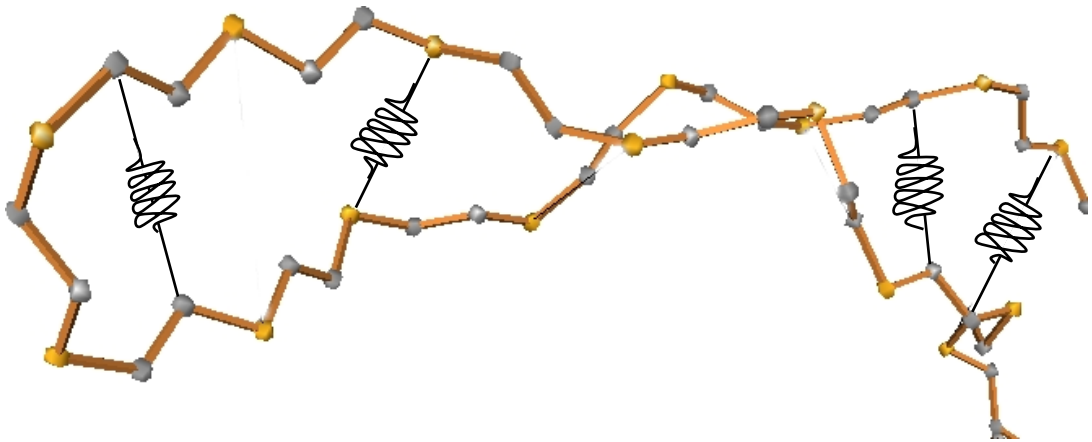
What should we study? What model should we use?



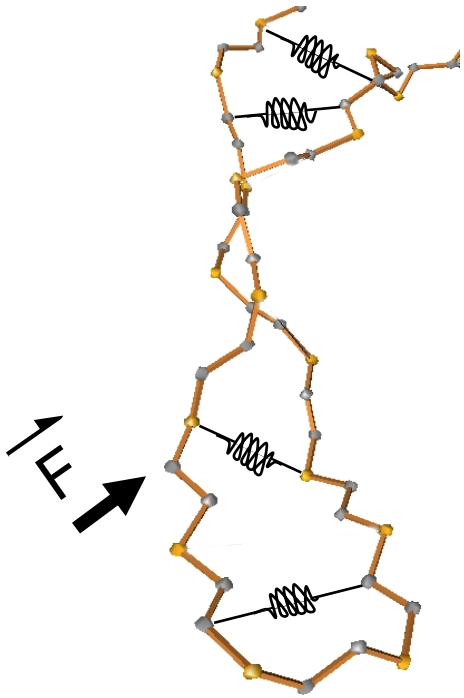
Which model best maximizes accuracy while minimizing computational requirements?

Operational Space Control, Dynamics

Observation:	Length of covalent bonds of the protein backbone are relatively fixed; only in the most extreme circumstances do the lengths of these bonds change.
Result:	Fix the length of the covalent bonds and have them assume the links in a kinematic chain.
Observation:	Hydrogen bonds have a higher degree of flexibility than do covalent bonds.
Result:	Model hydrogen bonds as springs; any deviation of the springs from the stable state results in an appropriate restoring force

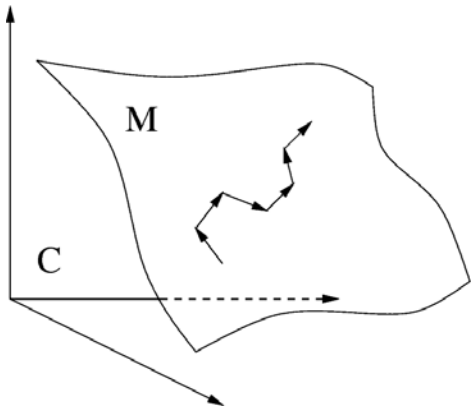


Motion on the Motion manifold



Mechanics: transform information about the interactions of individual atoms into valid conformational changes of the kinematic structure.

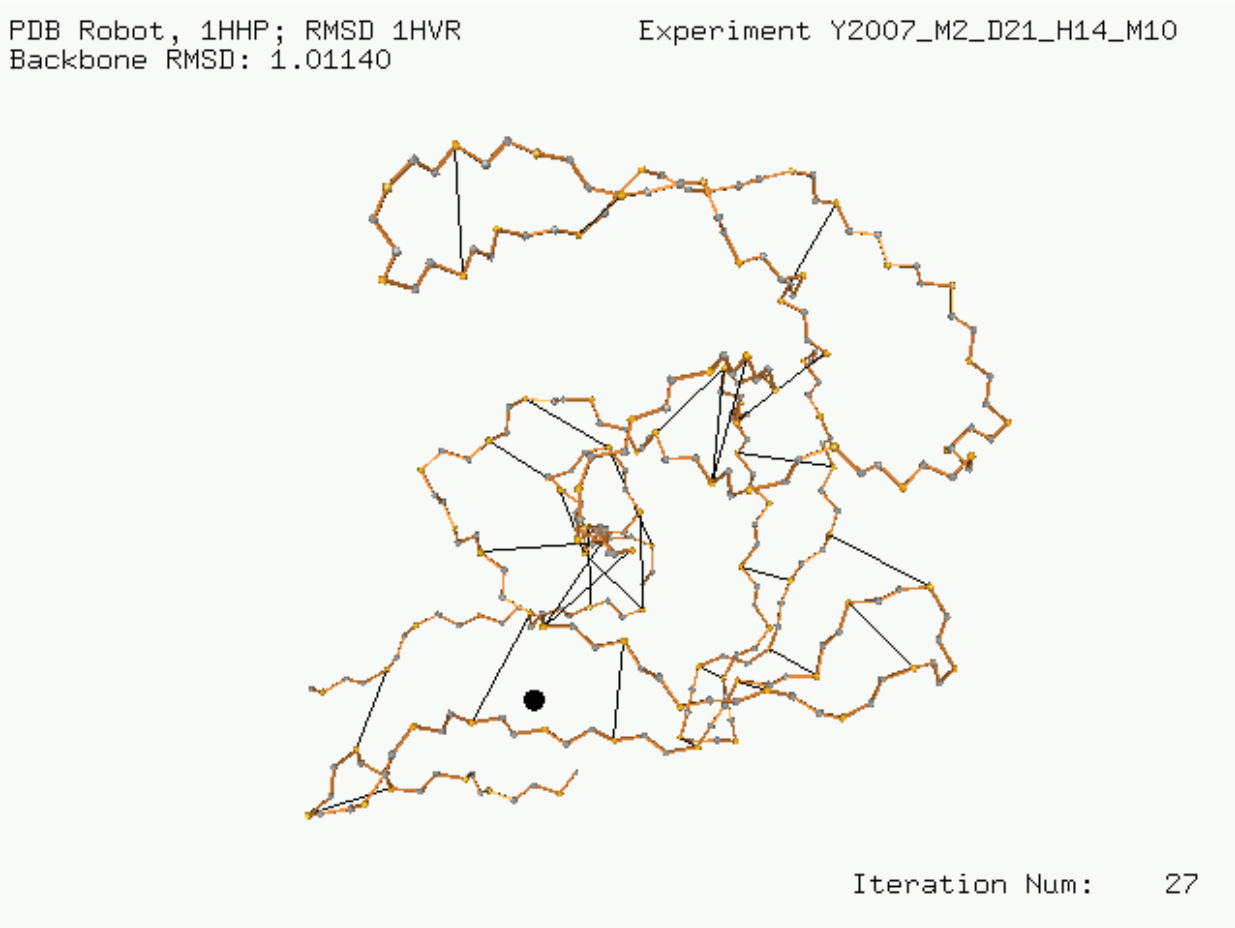
Use the Jacobian Matrix, *J*, to transform a force applied in Cartesian space to an equivalent torque in joing space



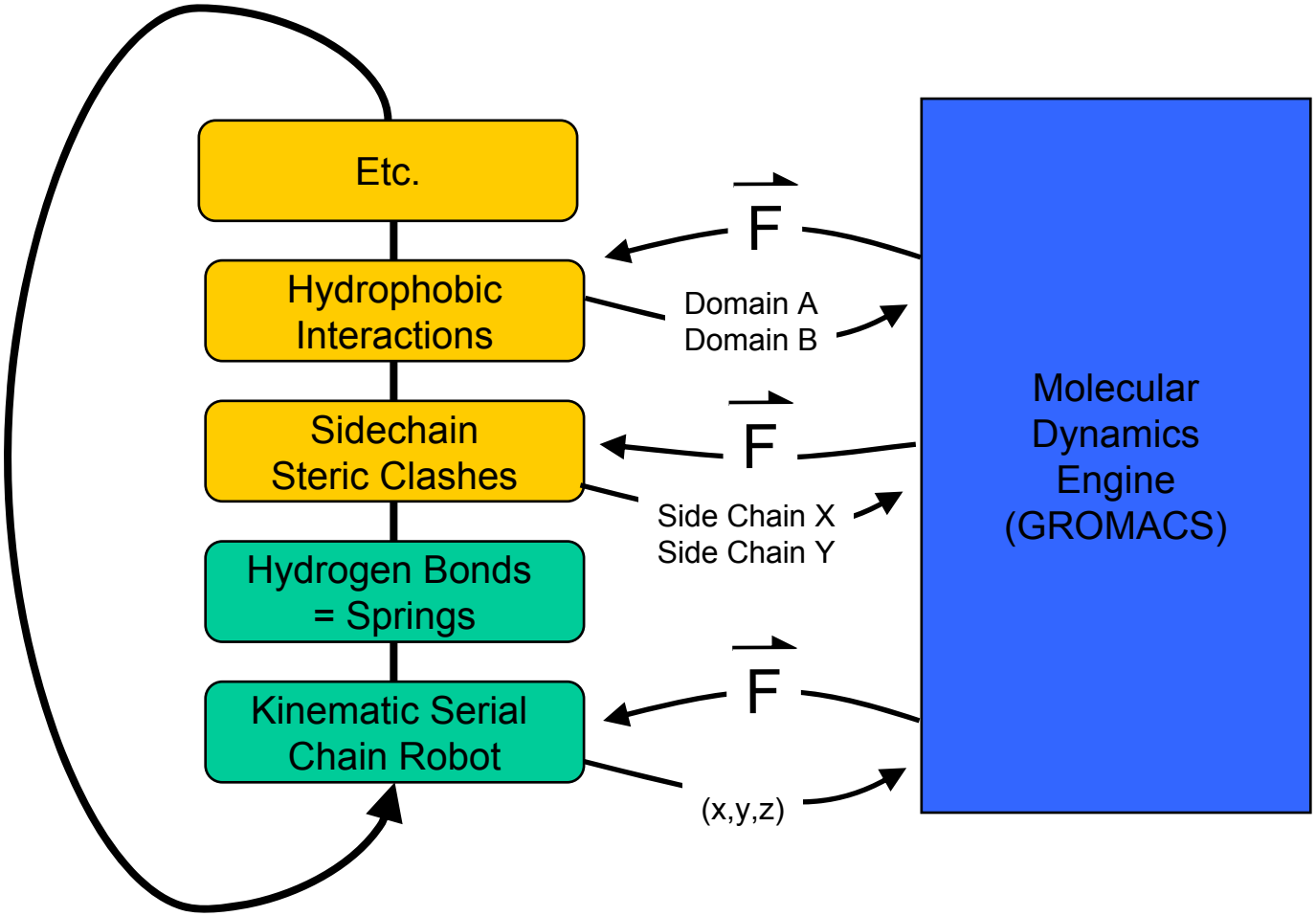
$$\tau = J^T F$$

We do not need to calculate the force between all possible atom-partners
Consider only how a displacement of an atom affects the connected atoms

Protein Motion; Using Kinematic Representation



Future Directions



Modeling Biological Systems
Opportunities for Computer Scientists

Filip Jagodzinski
RBO Tutorial Series
25 June 2007

Statistical Mechanics

Statistical mechanics: the study of the macroscopic system from a molecular point of view.

Mechanical (microscopic) state	**Thermodynamic (macroscopic) state**
defined by the atomic positions, **q**, and momenta, **p** - Configuration	a small set of parameters, including T, P, and number of particles, N - Energy

A single point in mechanical space, denoted by Γ, describes the state of the system. An **ensemble** is a collection of points in mechanical space satisfying the conditions of a particular thermodynamic state.

The dilemma appears to be that one can calculate time averages by molecular dynamics simulation, but the experimental observables are assumed to be ensemble averages.

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time}$$

Ensemble average = Time average

Ergodic Hypothesis

The trajectory of an isolated mechanical system runs through all states compatible with the total energy of the system.