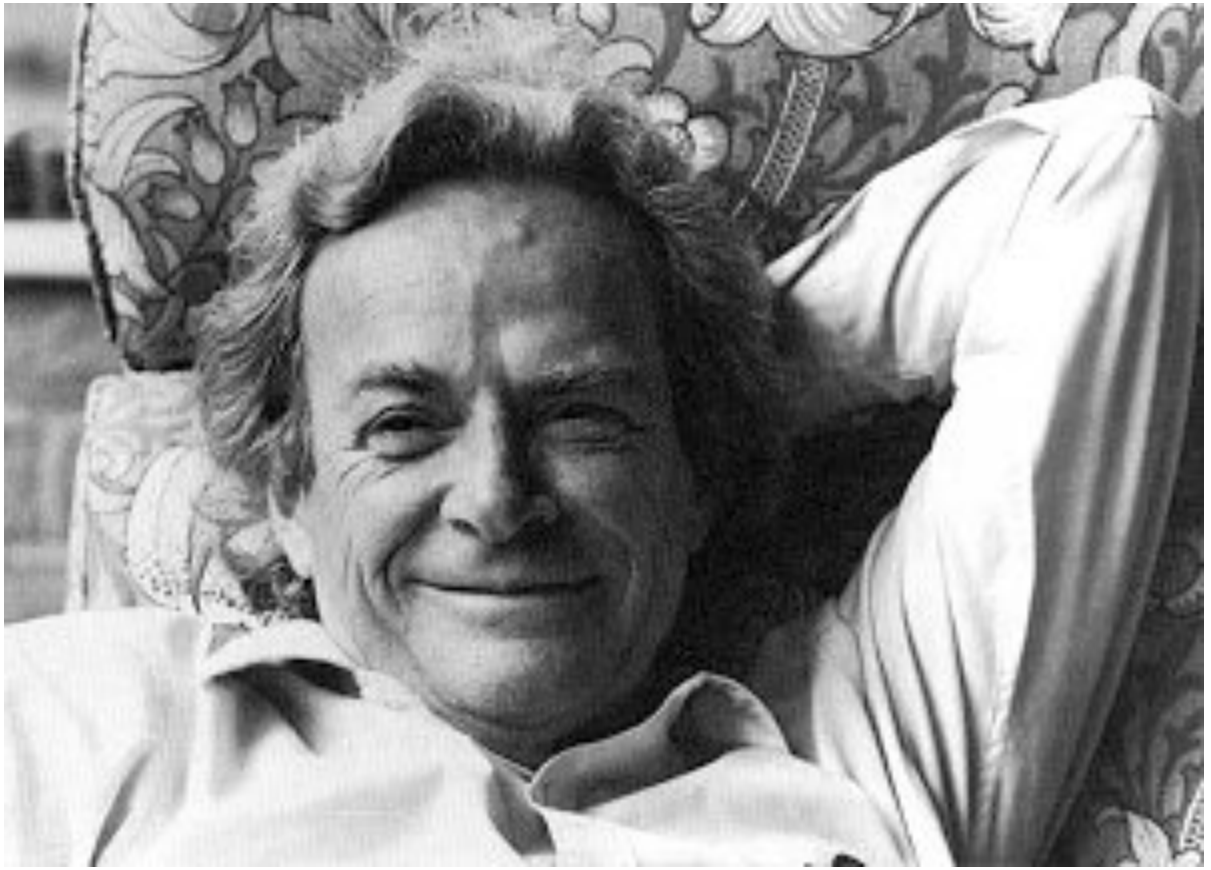
Classical Molecular Dynamics Simulations of Proteins

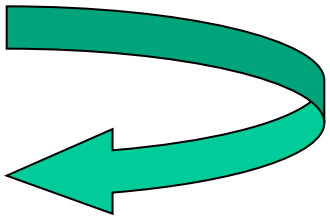


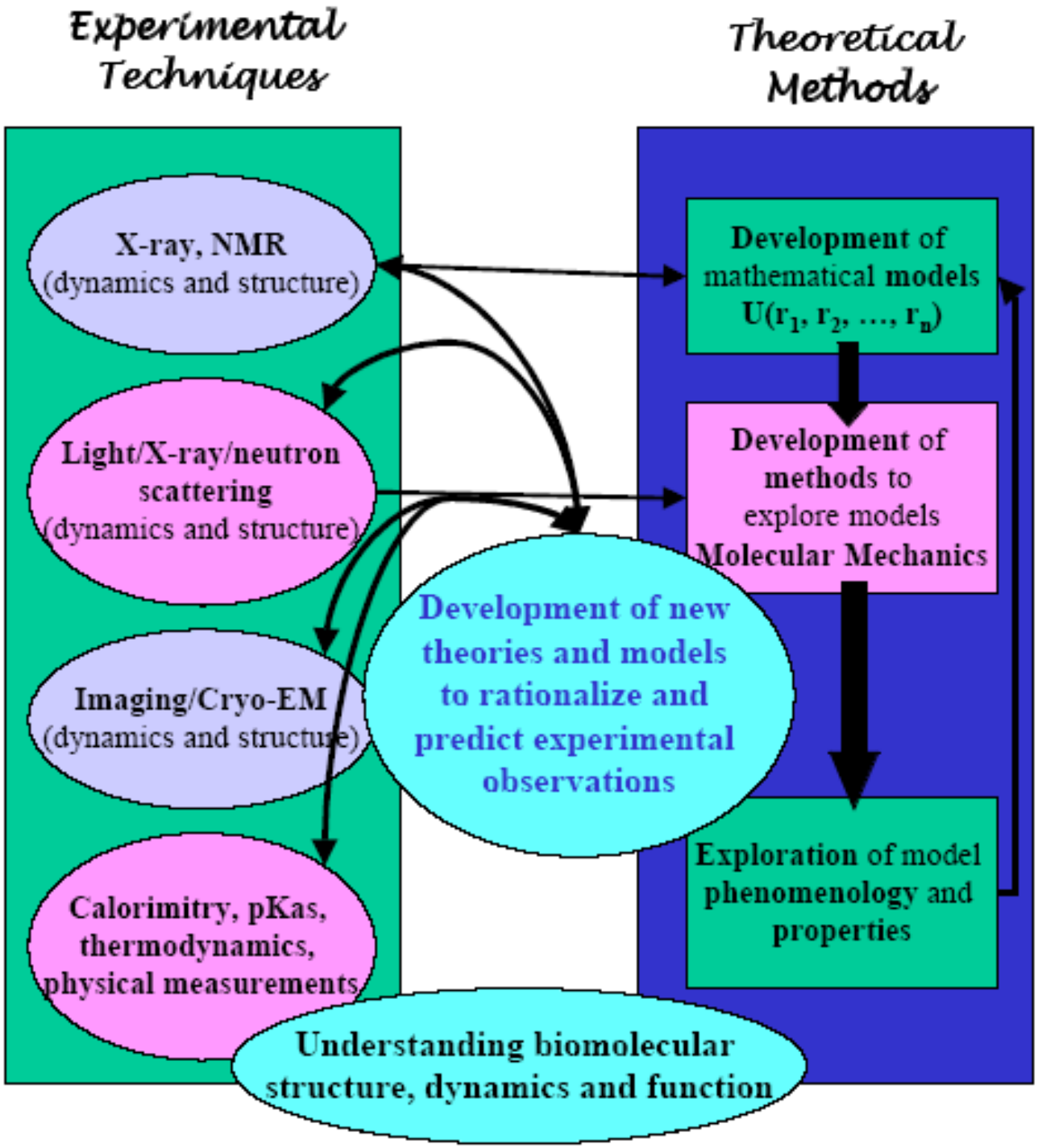
"everything that living things do can be understood in terms of the jigglings and wigglings of atoms."

The Feynman Lectures in Physics vol. 1, 3-6 (1963)

What is Molecular Dynamics?

- "The science of simulating the motions of a system of particles" (Karplus & Petsko)
- From systems
  - As small as an atom
  - As large as a galaxy
- Equations of motion
- Time evolution

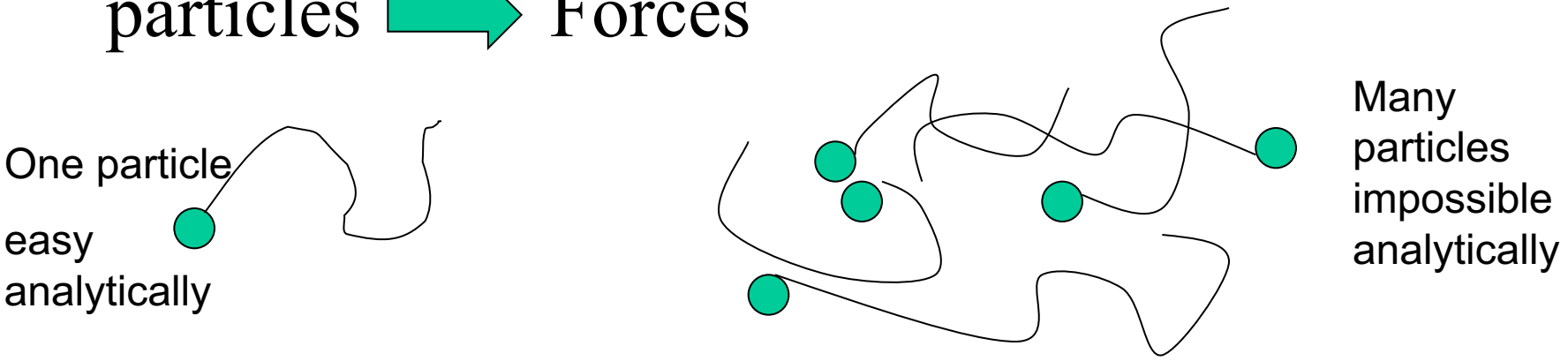



Experimental Techniques
Theoretical Methods
X-ray, NMR (dynamics and structure)
Light/X-ray/neutron scattering (dynamics and structure)
Imaging/Cryo-EM (dynamics and structure)
Calorimitry, pKas, thermodynamics, physical measurements
Development of mathematical models $U(r_1, r_2, \ldots, r_n)$
Development of methods to explore models Molecular Mechanics
Exploration of model phenomenology and properties
Development of new theories and models to rationalize and predict experimental observations
Understanding biomolecular structure, dynamics and function

Why?

Essential Elements

- Knowledge of the interaction potential for the particles ⇒ Forces



- Classical Newtonian equations of motion
- Many particle systems ⇒ simulation
- Maxwell-Boltzmann averaging process for thermodynamic properties: time averaging

Basis: Molecular Mechanics

- Theoretical foundation
- Potential energy functions
- Energy minimization
- Molecular dynamics

Uses of simulation & modelling

- Conformational searching with MD and minimization
- Exploration of biopolymer fluctuations and dynamics & kinetics
- MD as an ensemble sampler

Free energy simulations

Example applications

- Energy minimization as an estimator of binding free energies
- Protein stability
- Approximate association free energy of molecular assemblies
- Approximate pKa calculations

Theoretical Foundations

1. Force field parameters for families of chemical compounds
2. System modelled using Newton's equations of motion
3. Examples: hard spheres simulations (Alder & Wainwright, 1959); Liquid water (Rahman & Stillinger, 1970); BPTI (McCammon & Karplus, 1976); Villin headpiece (Duan & Kollman, 1998)

Protein Motion

- Protein motions of importance are torsional oscillations about the bonds that link groups together
- Substantial displacements of groups occur over long time intervals
- Collective motions either local (cage structure) or rigid-body (displacement of different regions)
- *What is the importance of these fluctuations for biological function?*

Effect of fluctuations

Thermodynamics: equilibrium behaviour important; e.g., energy of ligand binding

Dynamics: displacements from average structure important; e.g., local sidechain motions that act as conformational gates in oxygen transport myoglobin, enzymes, ion channels

Local Motions

- 0.01-5 Å, 1 fs -0.1s
- Atomic fluctuations
  - Small displacements for substrate binding in enzymes
  - Energy “source” for barrier crossing and other activated processes (e.g., ring flips)
- Sidechain motions
  - Opening pathways for ligand (myoglobin)
  - Closing active site
- Loop motions
  - Disorder-to-order transition as part of virus formation

Rigid-Body Motions

- 1-10 Å, 1 ns – 1 s
- Helix motions
  - Transitions between substates (myoglobin)
- Hinge-bending motions
  - Gating of active-site region (liver alcohol dehydrogenase)
  - Increasing binding range of antigens (antibodies)

Large Scale Motion

- > 5 Å, 1 microsecond – 10000 s
- Helix-coil transition
  - Activation of hormones
  - Protein folding transition
- Dissociation
  - Formation of viruses
- Folding and unfolding transition
  - Synthesis and degradation of proteins

Role of motions sometimes only inferred from two or more conformations in structural studies

Typical Time Scales ....

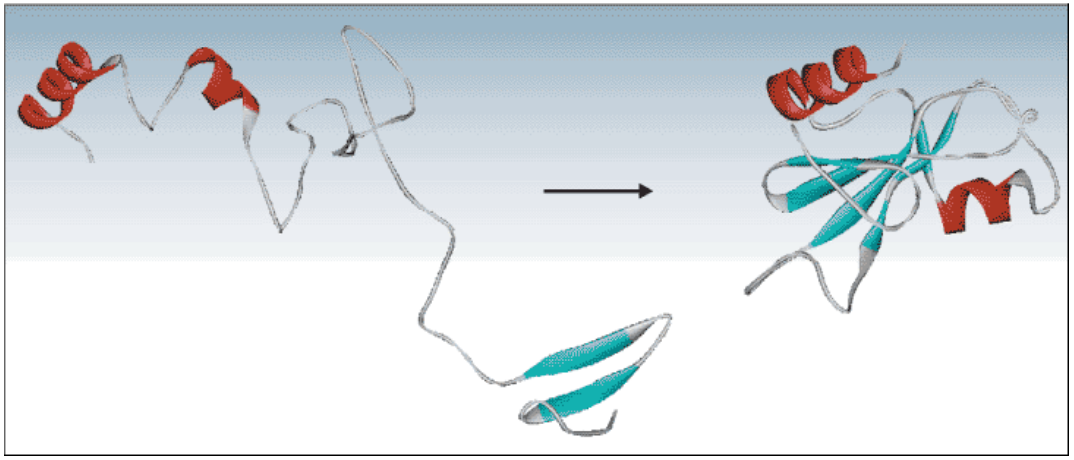
- Bond stretching: 10^{-14} - 10^{-13} sec.
- Elastic vibrations: 10^{-12} - 10^{-11} sec.
- Rotations of surface sidechains: 10^{-11} - 10^{-10} sec.
- Hinge bending: 10^{-11} - 10^{-7} sec.
- Rotation of buried side chains: 10^{-4} - 1 sec.
- Protein folding: 10^{-6} - 10^{2} sec.

Timescale in MD:

- A Typical timestep in MD is 1 fs (10^{-15} sec)

(ideally 1/10 of the highest frequency vibration)

Ab initio protein folding simulation

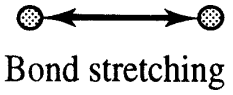


Physical time for simulation	10^{-4} seconds
Typical time-step size	10^{-15} seconds
Number of MD time steps	10^{11}
Atoms in a typical protein and water simulation	32,000
Approximate number of interactions in force calculation	10^{9}
Machine instructions per force calculation	1000
Total number of machine instructions	10^{23}
BlueGene capacity (floating point operations per second)	1 petaflop (10^{15})

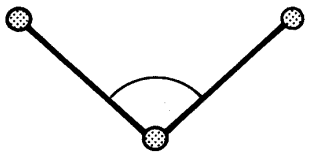
➔ **Blue Gene will need 3 years to simulate 100 μsec.**

[http://www.research.ibm.com/bluegene/]

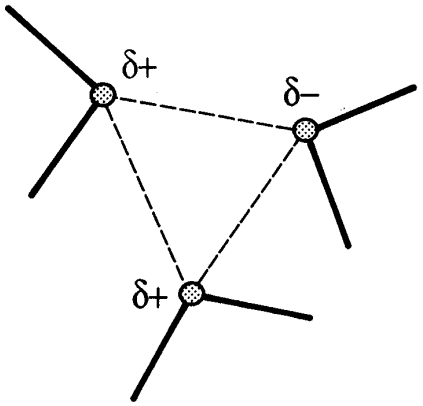
Empirical Force Fields and Molecular Mechanics



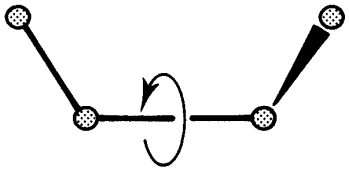
Bond stretching



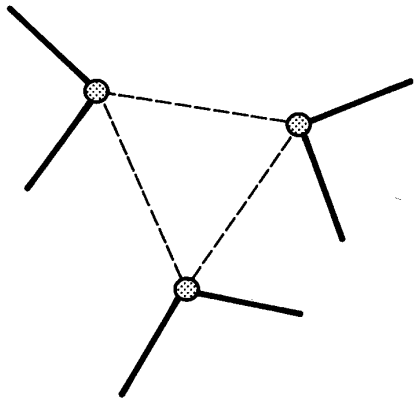
Angle bending



Non-bonded interactions
(electrostatic)

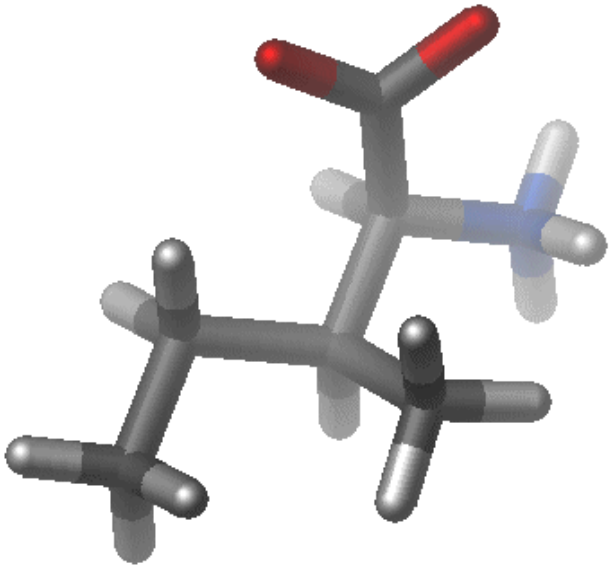


Bond rotation
(torsion)



Non-bonded interactions
(van der Waals)

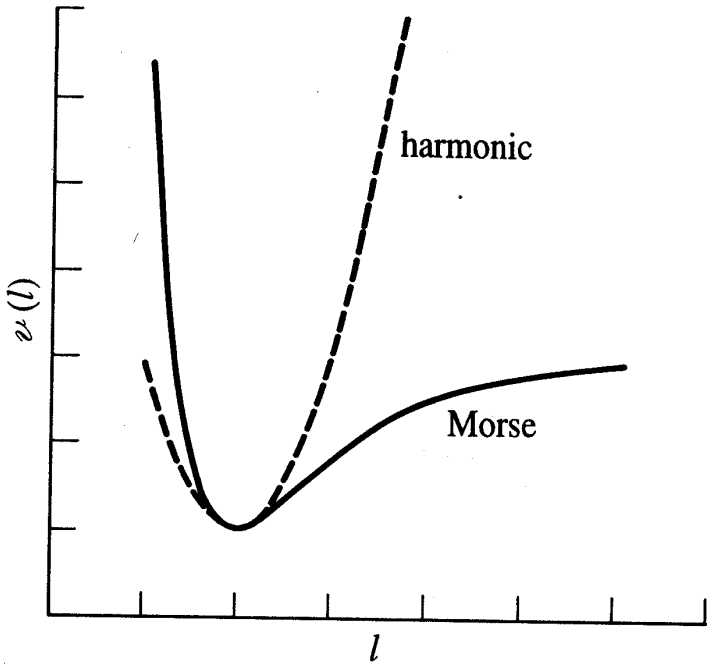
- describe interaction of atoms or groups
- the parameters are "empirical", i.e. they are dependent on others and have no direct intrinsic meaning



Bond stretching

- Approximation of the Morse potential by an "elastic spring" – model

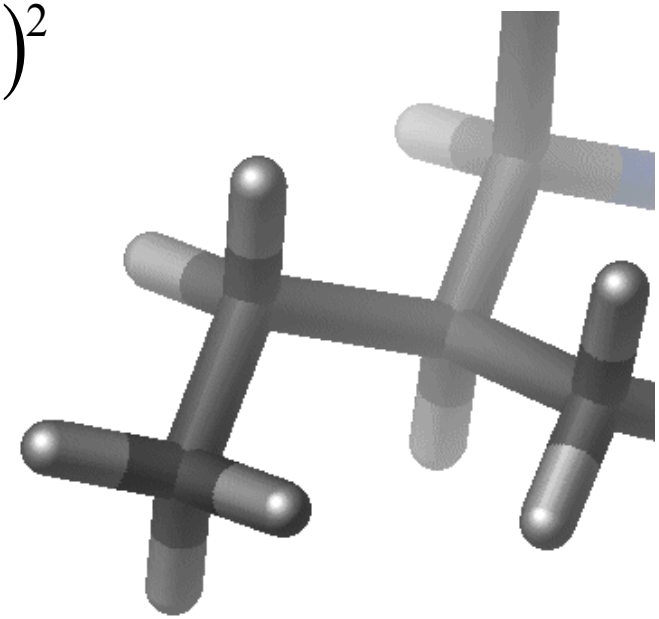
- Hooke's law as reasonable approximation close to reference bond length l_0



$$V(l) = \frac{k}{2}\left(l - l_0\right)^2$$

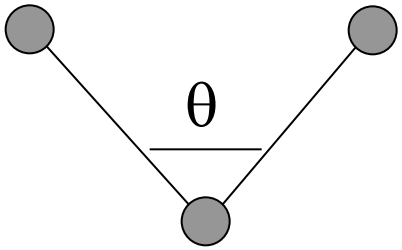
k : Force constant

l : distance



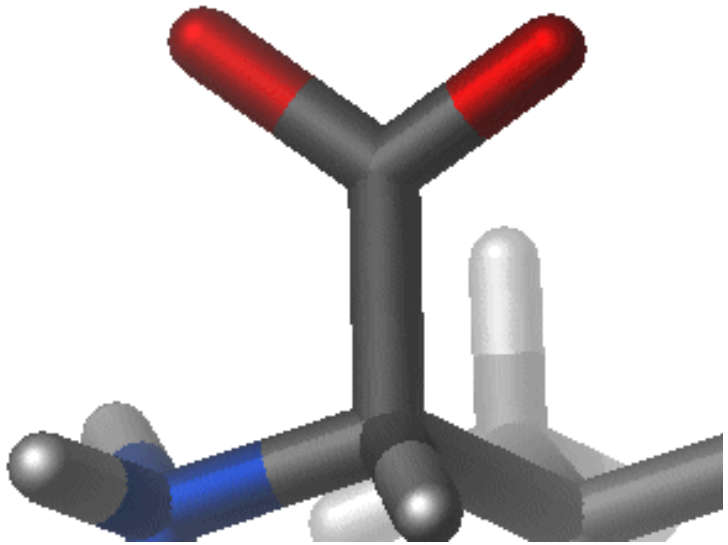
Angle Bending

- Deviation from angles from their reference angle θ_0 often described by Hooke's law:



$$V(\theta) = \frac{k}{2}(\theta - \theta_0)^2$$

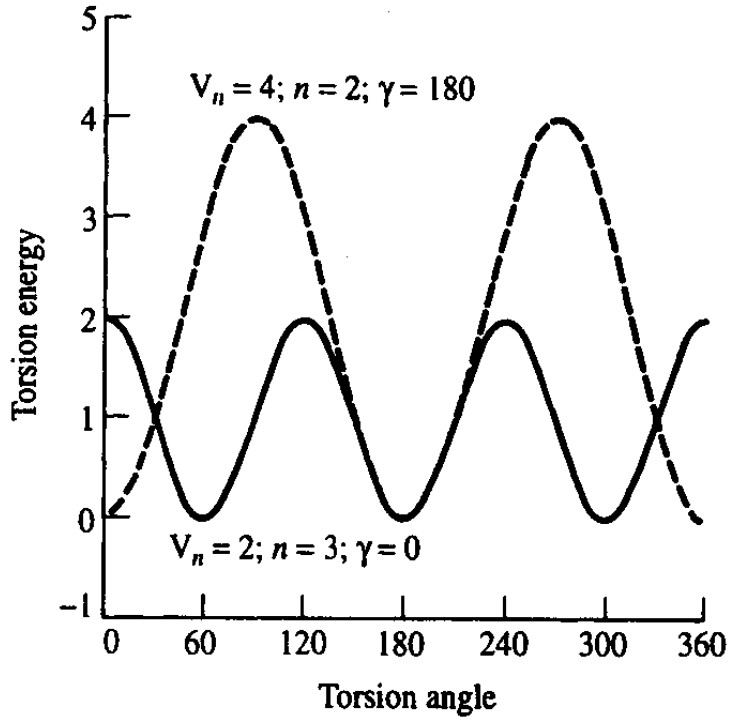
k : Force constant
θ : bond angle



- Force constants are much smaller than those for bond stretching

Torsional Terms

- Hypothetical potential function for rotation around a chemical bond:

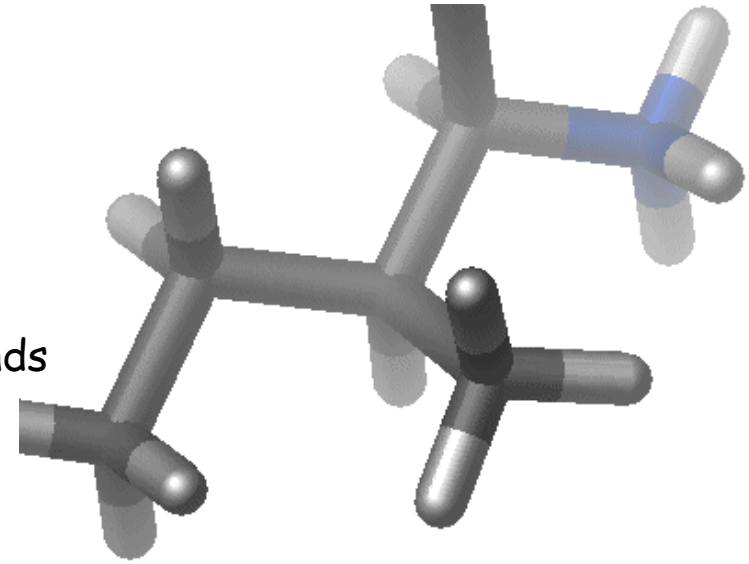


$$V(\omega) = \frac{V_n}{2}\left[1 + \cos\left(n\omega - \gamma\right)\right]$$

V_n : 'barrier' height
n : multiplicity (e.g. n=3)
ω : torsion angle
γ : phase factor

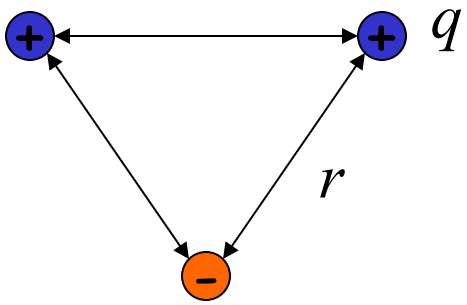
- Need to include higher terms for non-symmetric bonds

(i.e. to distinguish trans, gauche conformations)



Electrostatic interactions

- Electronegative elements attract electrons more than less electronegative elements
- Unequal charge distribution is expressed by fractional charges
- Electrostatic interaction often calculated by Coulomb's law:



$$V = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Example for a (very) simple Force Field:

$$V = \sum_{bonds} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2$$

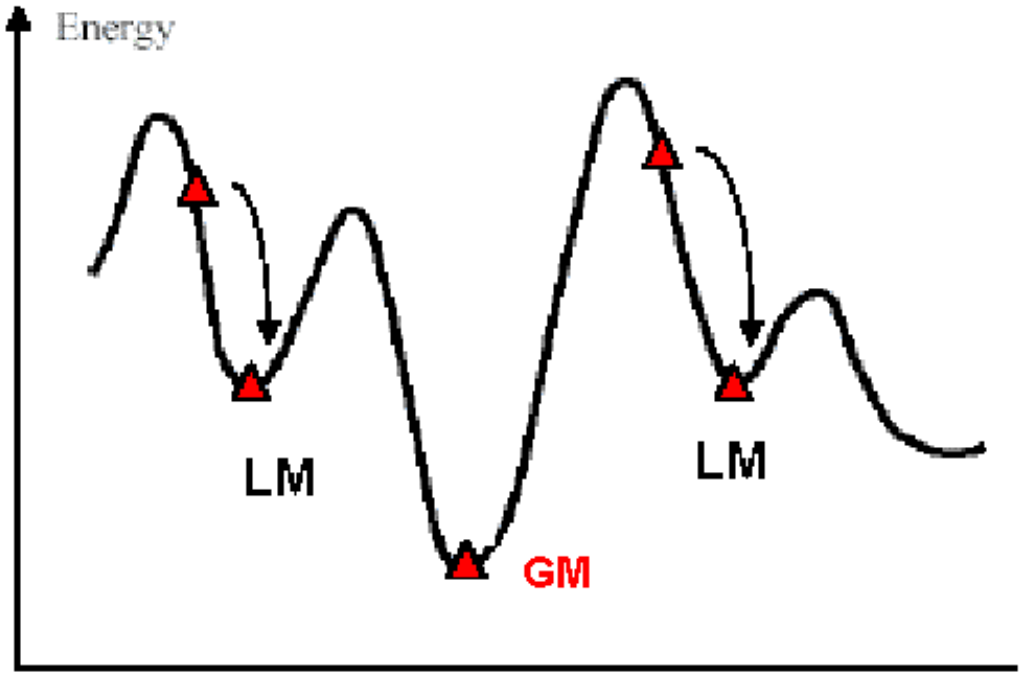
$$+ \sum_{angles} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2$$

$$+ \sum_{torsions} \frac{V_N}{2}\left(1 + \cos\left(n\omega - \gamma\right)\right)$$

$$+ \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\pi\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

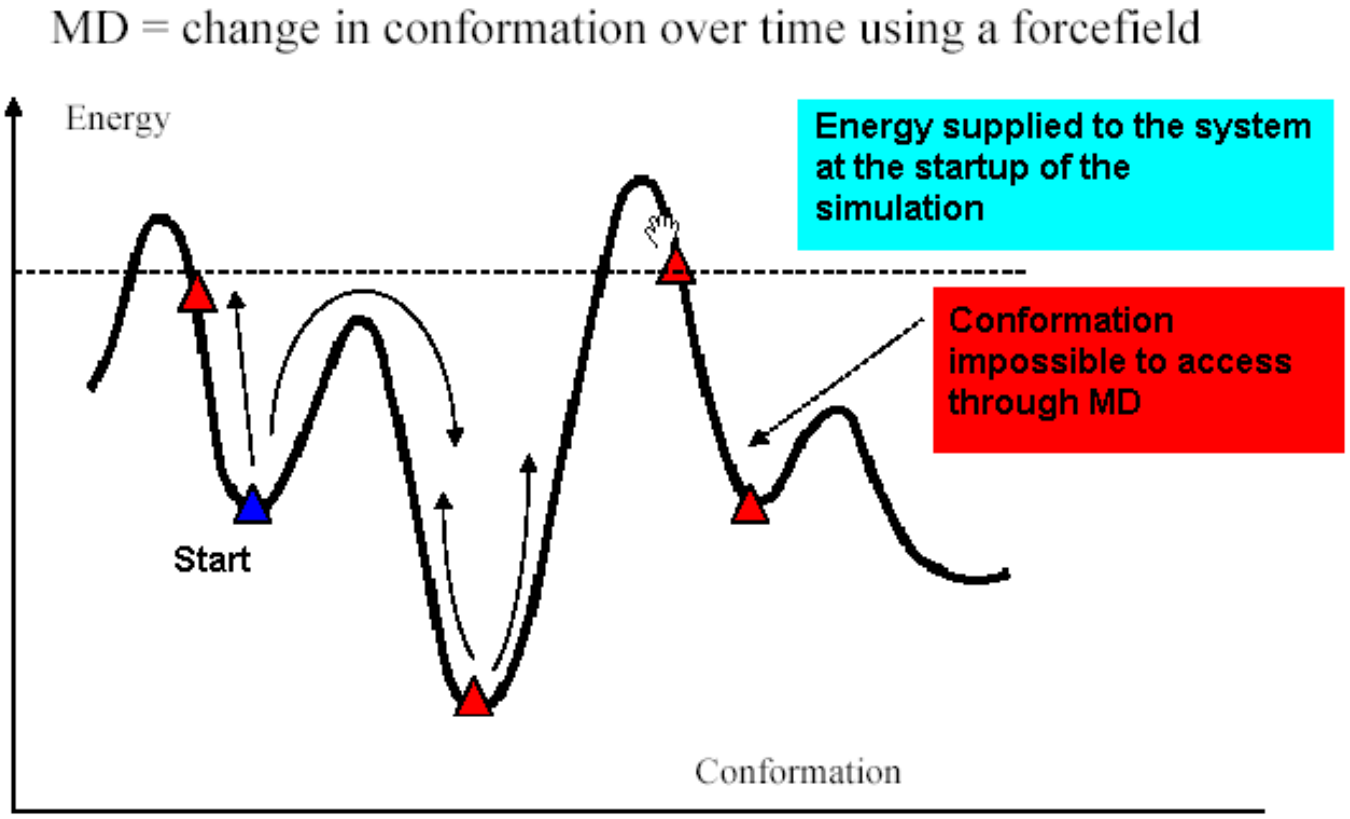
Molecular Mechanics - Energy Minimization

- The energy of the system is minimized. The system tries to relax
- Typically, the system relaxes to a local minimum (LM).



Molecular Dynamics (MD)

In molecular dynamics, energy is supplied to the system, typically using a constant temperature (i.e. constant average constant kinetic energy).



Newton's Laws of Motion

1. A body maintains its state of rest or of uniform motion in a straight line, unless acted upon by a force.
2. The applied force is equal to the rate of change of momentum.
3. Two isolated bodies acting upon each other experience equal and opposite forces.

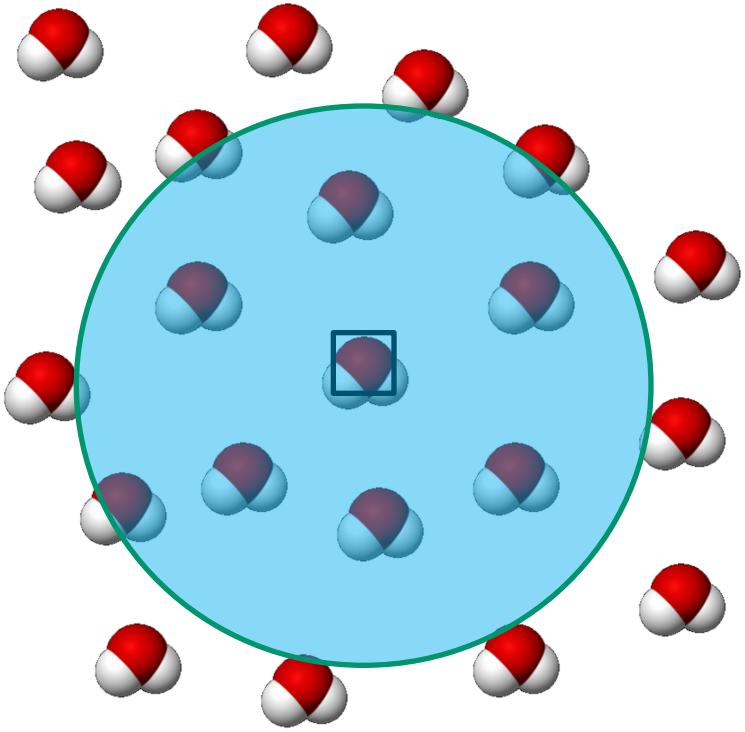
Molecular Dynamics (MD)

- Use Newtonian mechanics to calculate the net force and acceleration experienced by each atom.
- Each atom i is treated as a point with mass m_i and fixed charge q_i
- Determine the force F_i on each atom:

$$\vec{F}_i = m_i a_i = -\frac{\mathrm{d}V}{\mathrm{d}r_i}$$

- Use positions and accelerations at time t (and positions from $t - \delta t$) to calculate new positions at time $t + \delta t$

Cutoffs

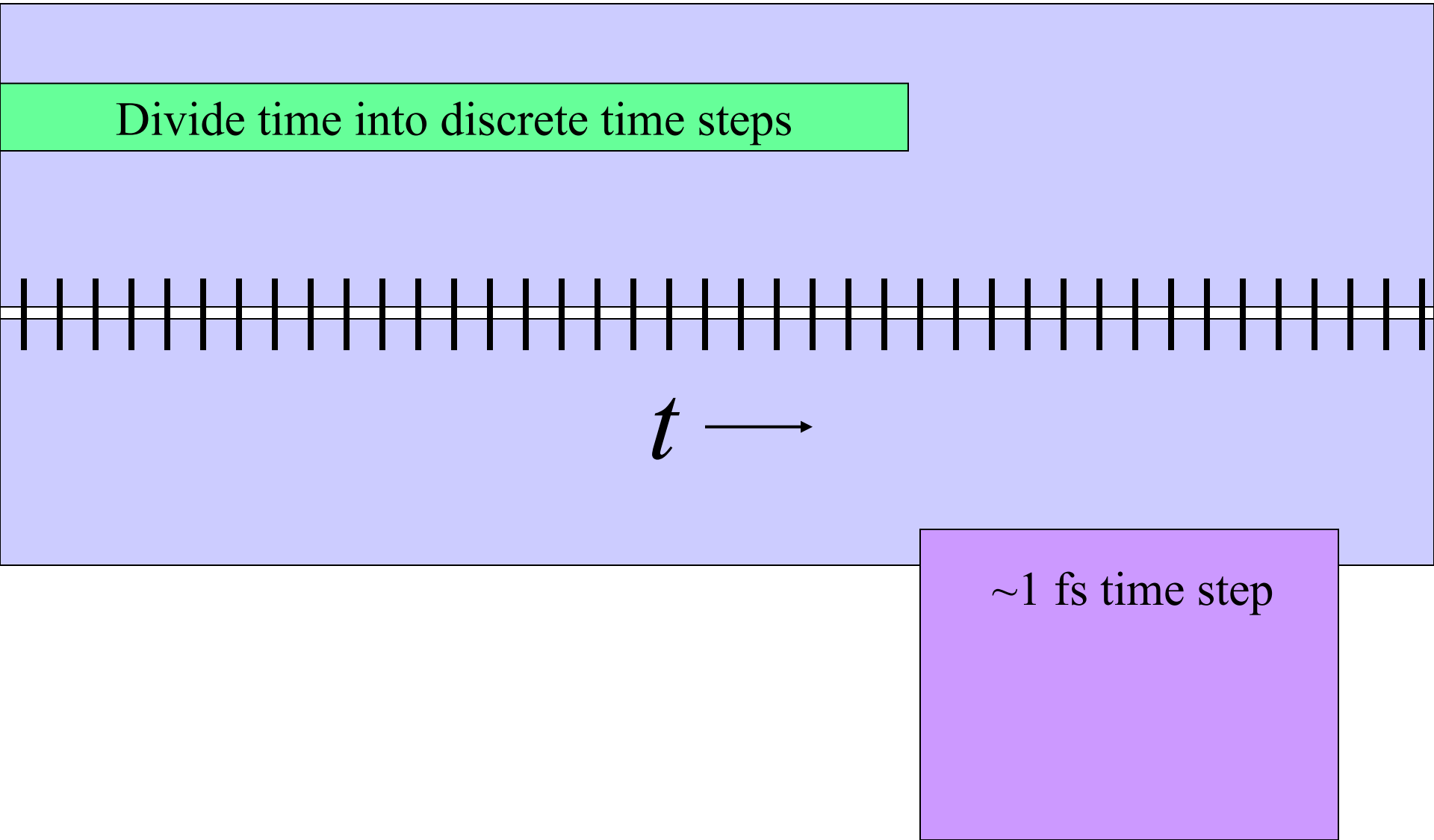


(a) Estimate the total number of possible structures of a polypeptide consisting of 10 amino acid residues. State and justify any assumptions that you make.

(b) Calculate the number of pairwise interactions which need to be evaluated to calculate the energy of a 10-residue peptide, stating any assumptions you make. If a computer capable of calculating one million pairwise interactions per second is used, and the time to perform a systematic search of all conformations is one structure per 10^{-13} seconds, estimate both the simulation time required to fold the peptide and the time it would take to calculate the energy of all the conformers.

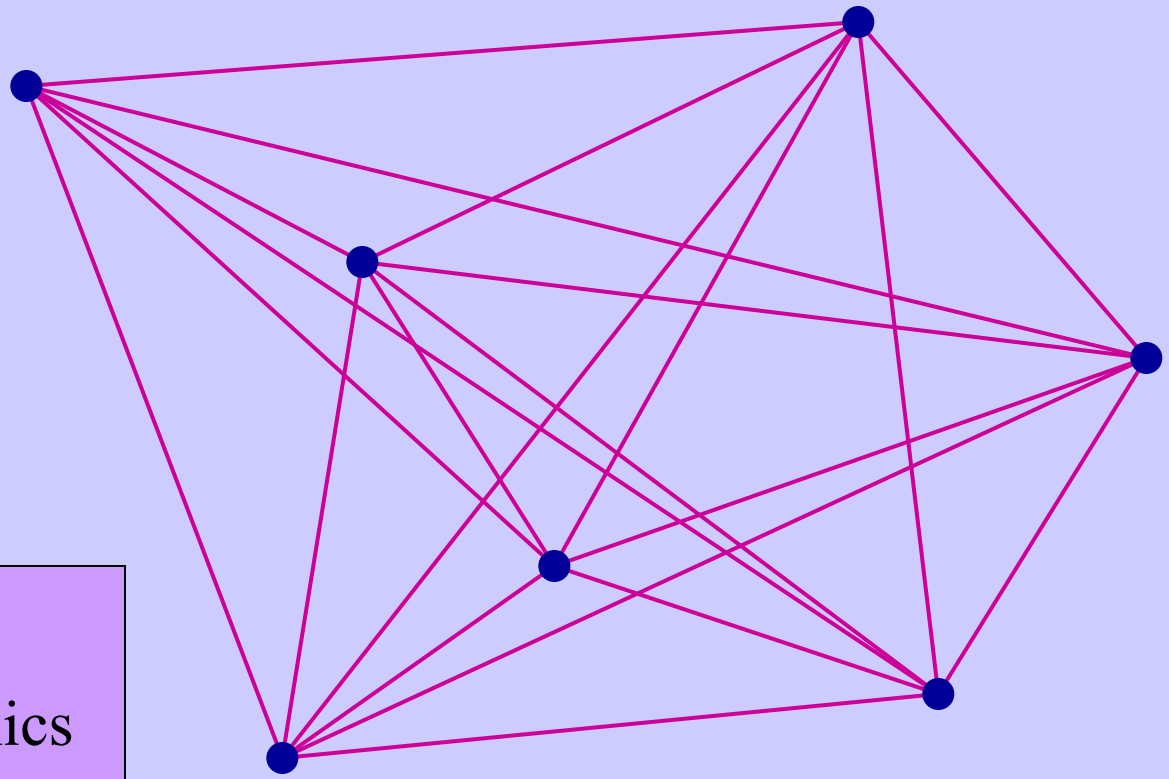
Molecular Dynamics

Molecular Dynamics



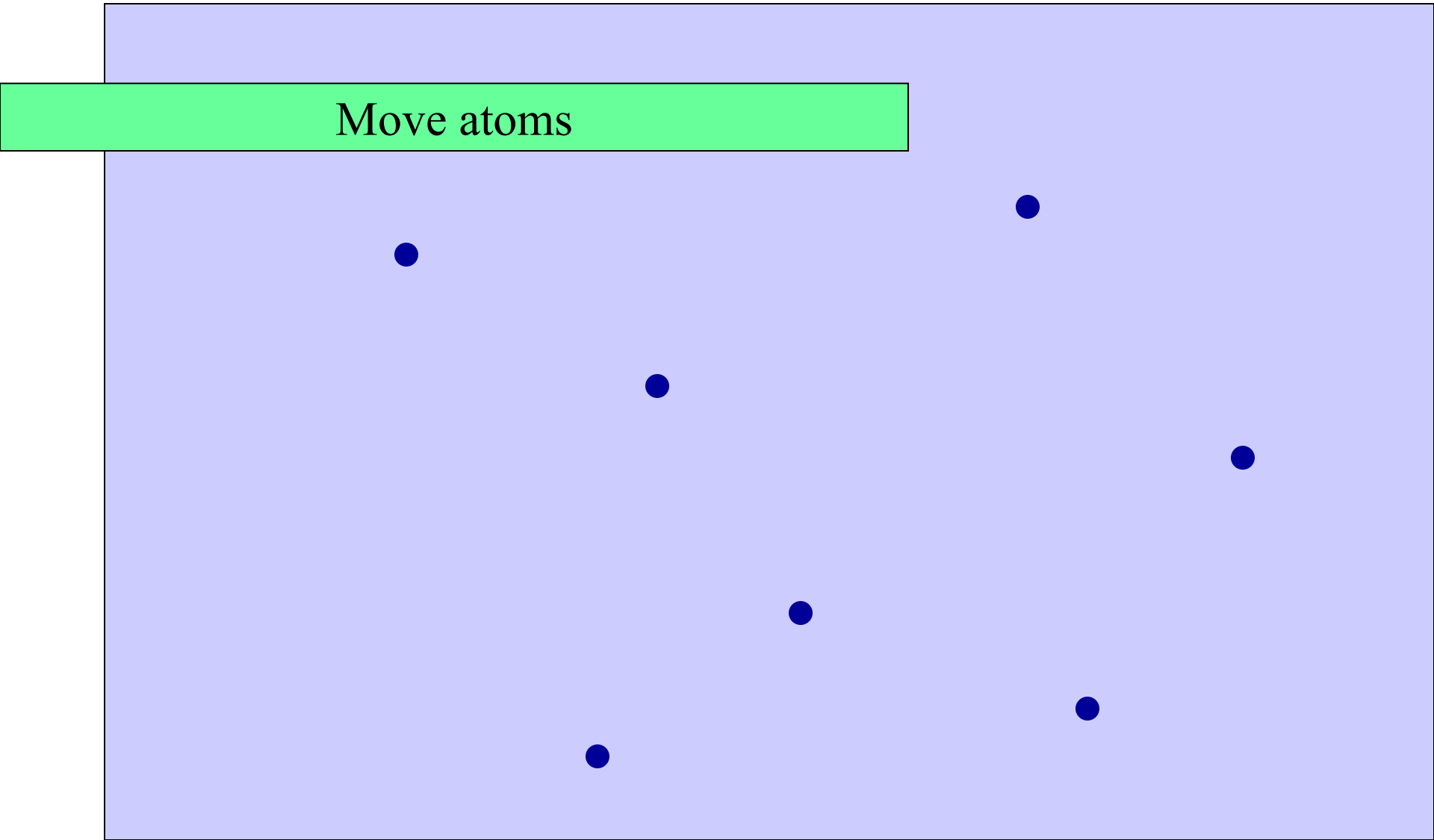
Molecular Dynamics

Calculate forces

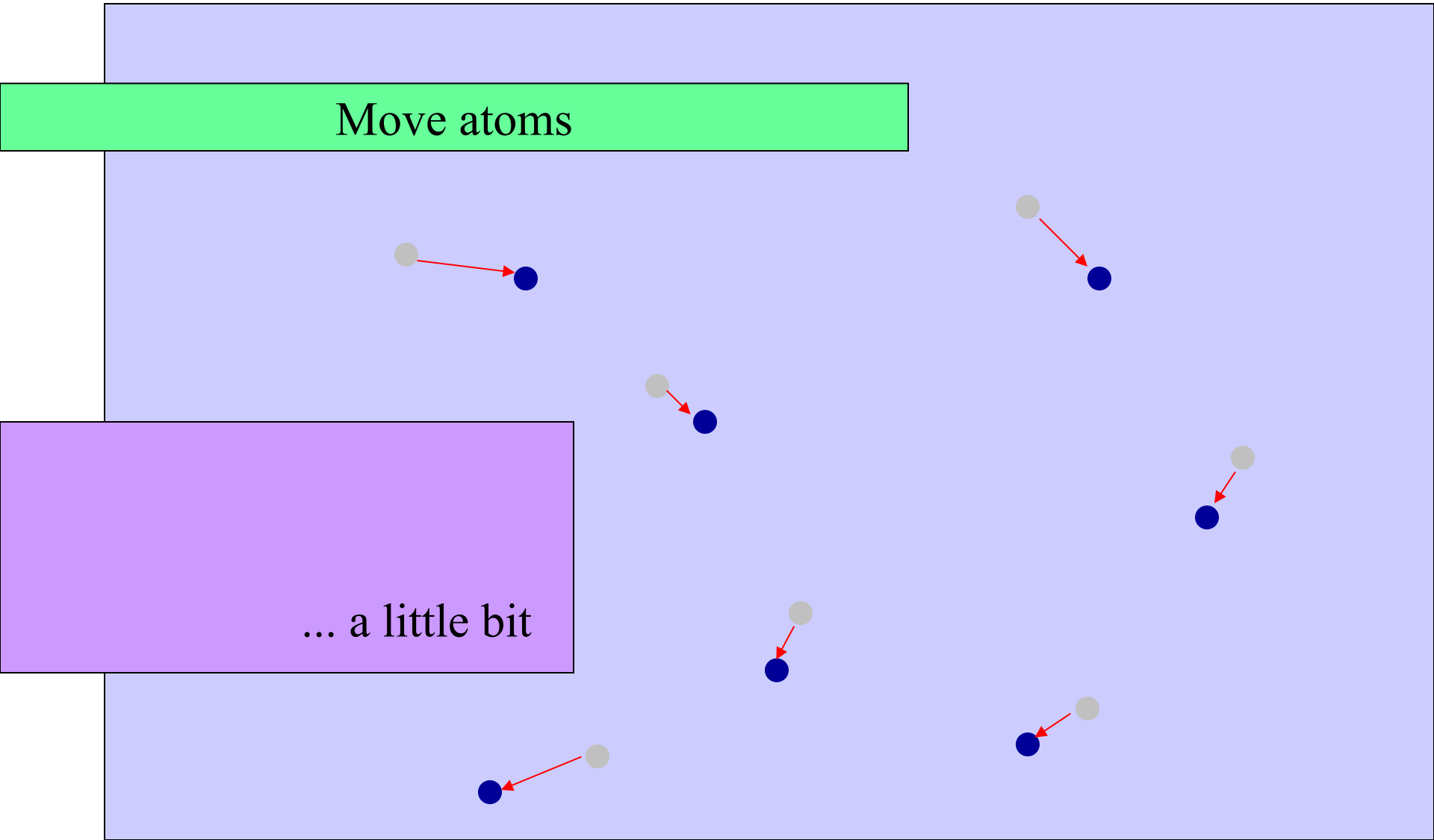


Molecular mechanics force field

Molecular Dynamics



Molecular Dynamics



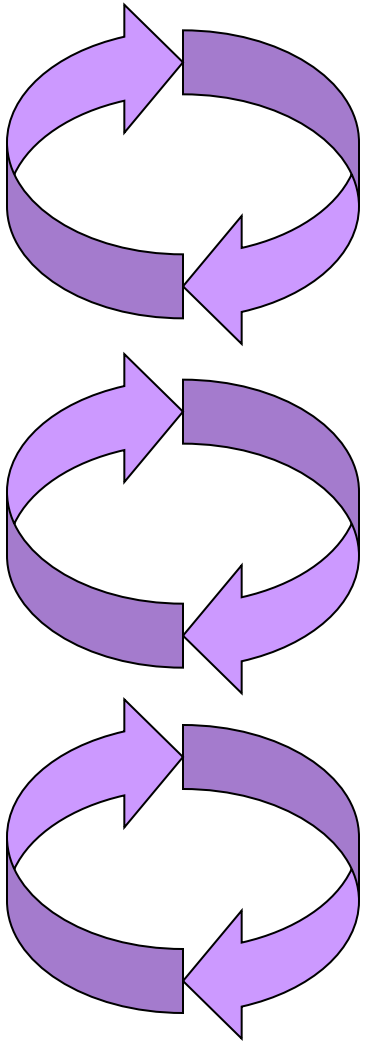
Molecular Dynamics

Iterate

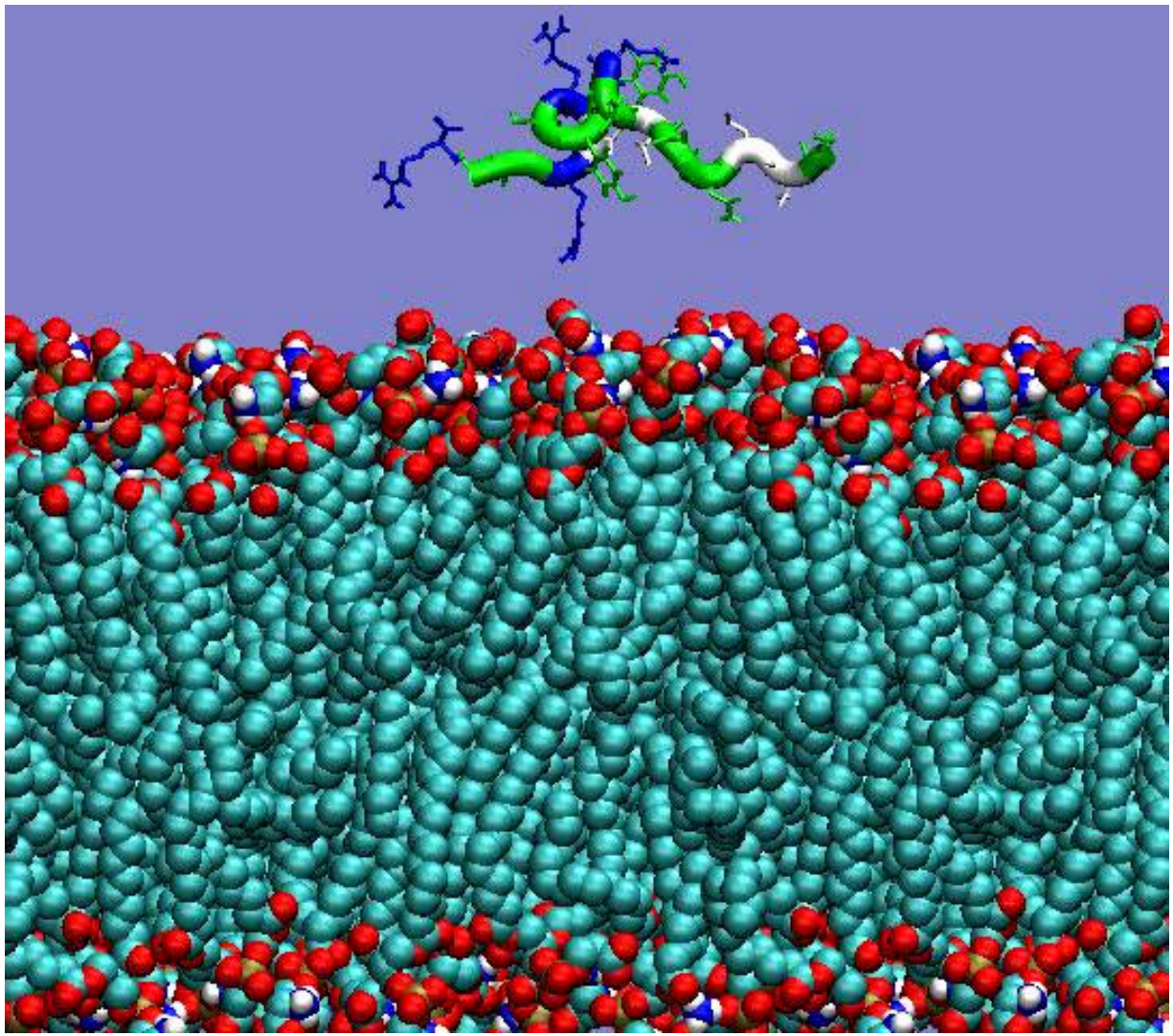
... and iterate

... and iterate

Integrate Newton's laws of motion



Example of an MD Simulation



Main Problem With MD

Too slow!

Example I just showed:

- 2 ns simulated time
- 3.4 CPU-days to simulate

Goals and Strategy

Thought Experiment

- What if MD were
  - Perfectly accurate?
  - Infinitely fast?

- Would be easy to perform **arbitrary computational experiments**
  - Determine structures by watching them form
  - Figure out what happens by watching it happen
  - Transform measurement into data mining

Two Distinct Problems

Problem 1: Simulate many short trajectories

Problem 2: Simulate one long trajectory

Simulating Many Short Trajectories

- Can answer surprising number of interesting questions

- Can be done using
  - Many slow computers
  - Distributed processing approach
  - Little inter-processor communication

- E.g., Pande's *Folding at Home* project

Simulating One Long Trajectory

- Harder problem
- Essential to elucidate many biologically interesting processes
- Requires a single machine with
  - Extremely high performance
  - Truly massive parallelism
  - Lots of inter-processor communication

DESRES Goal

- **Single, millisecond-scale MD simulations (long trajectories)**
  - Protein with 64K or more atoms
  - Explicit water molecules

- Why?
  - **That's the time scale at which many biologically interesting things start to happen**

Protein Folding

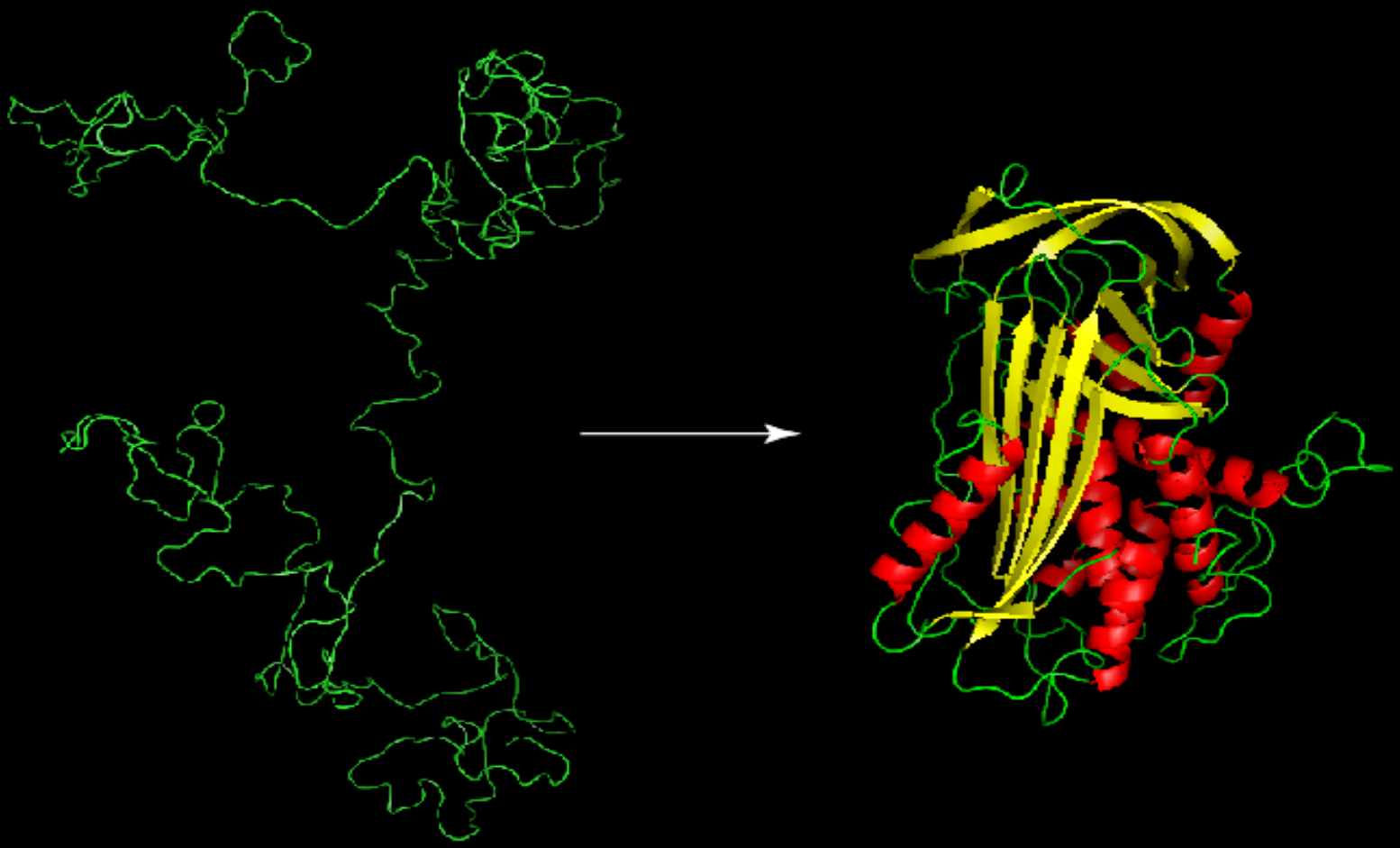


Image: Istvan Kolossvary & Annabel Todd, D. E. Shaw Research

Interactions Between Proteins

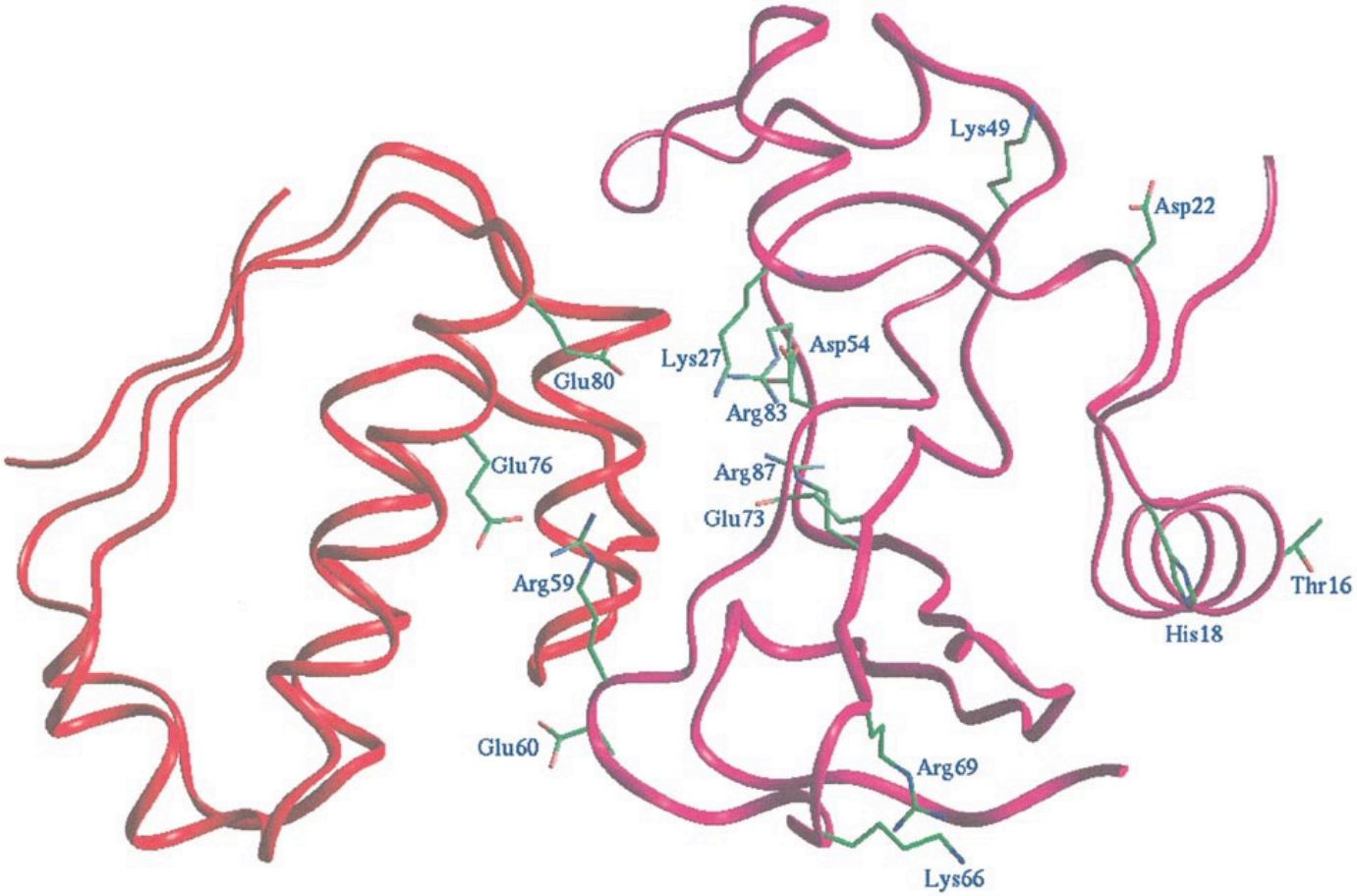
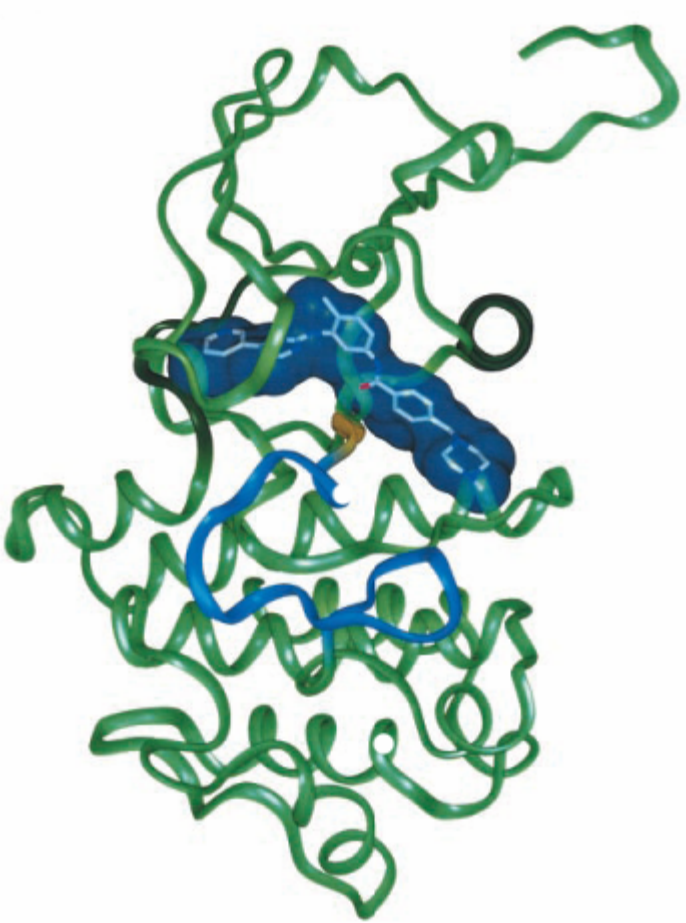


Image: Vijayakumar, et al., *J. Mol. Biol.* 278, 1015 (1998)

Binding of Drugs to their Molecular Targets



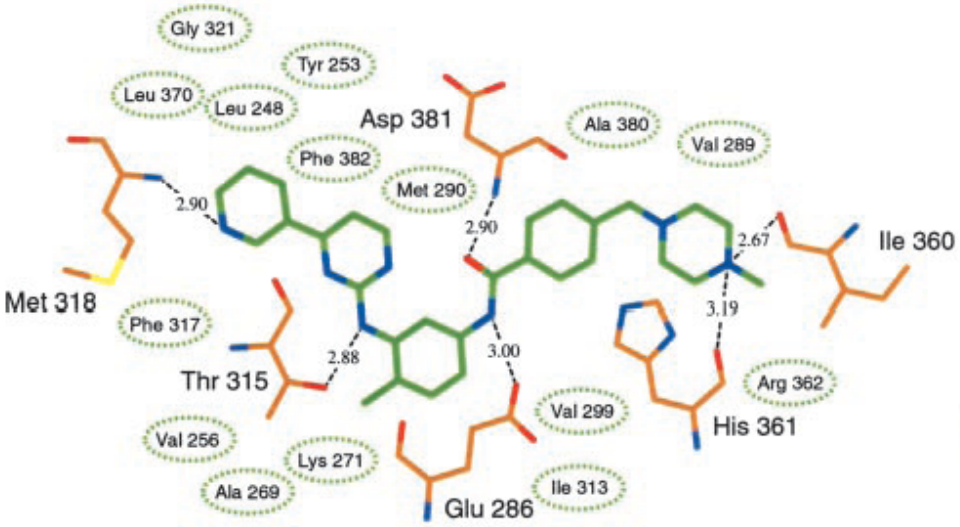
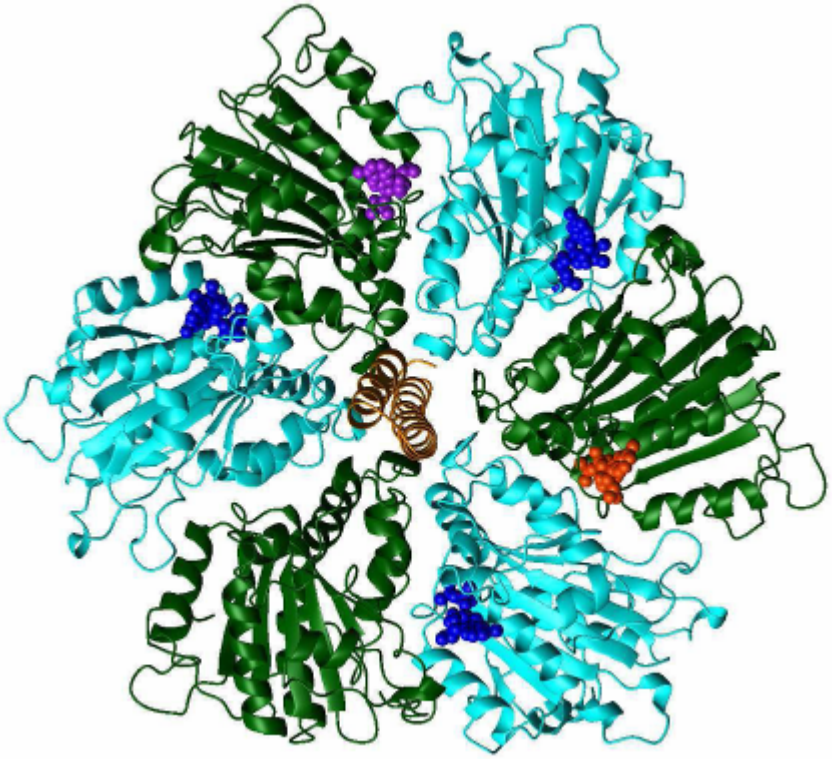


Image: Nagar, et al., *Cancer Res.* 62, 4236 (2002)

Mechanisms of Intracellular Machines



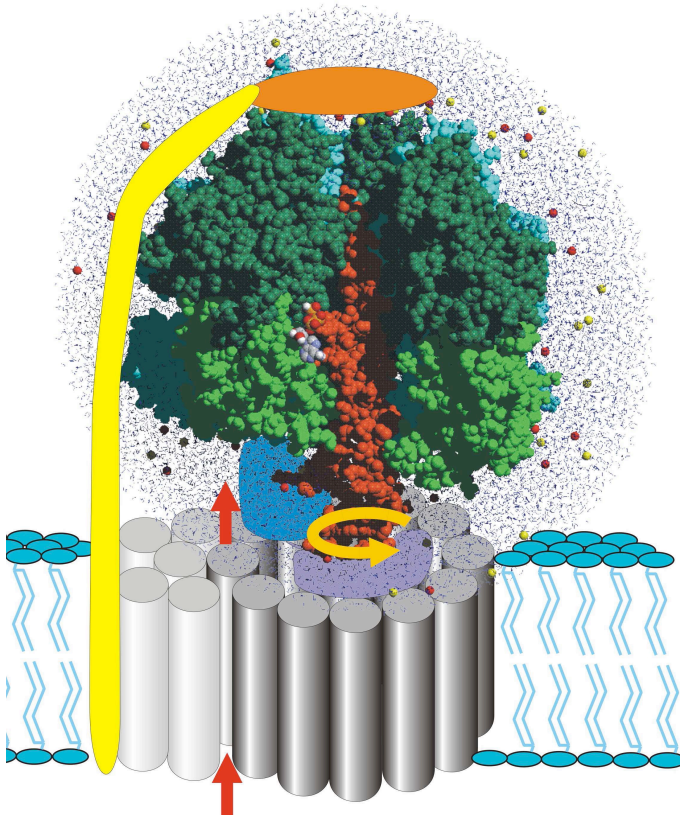


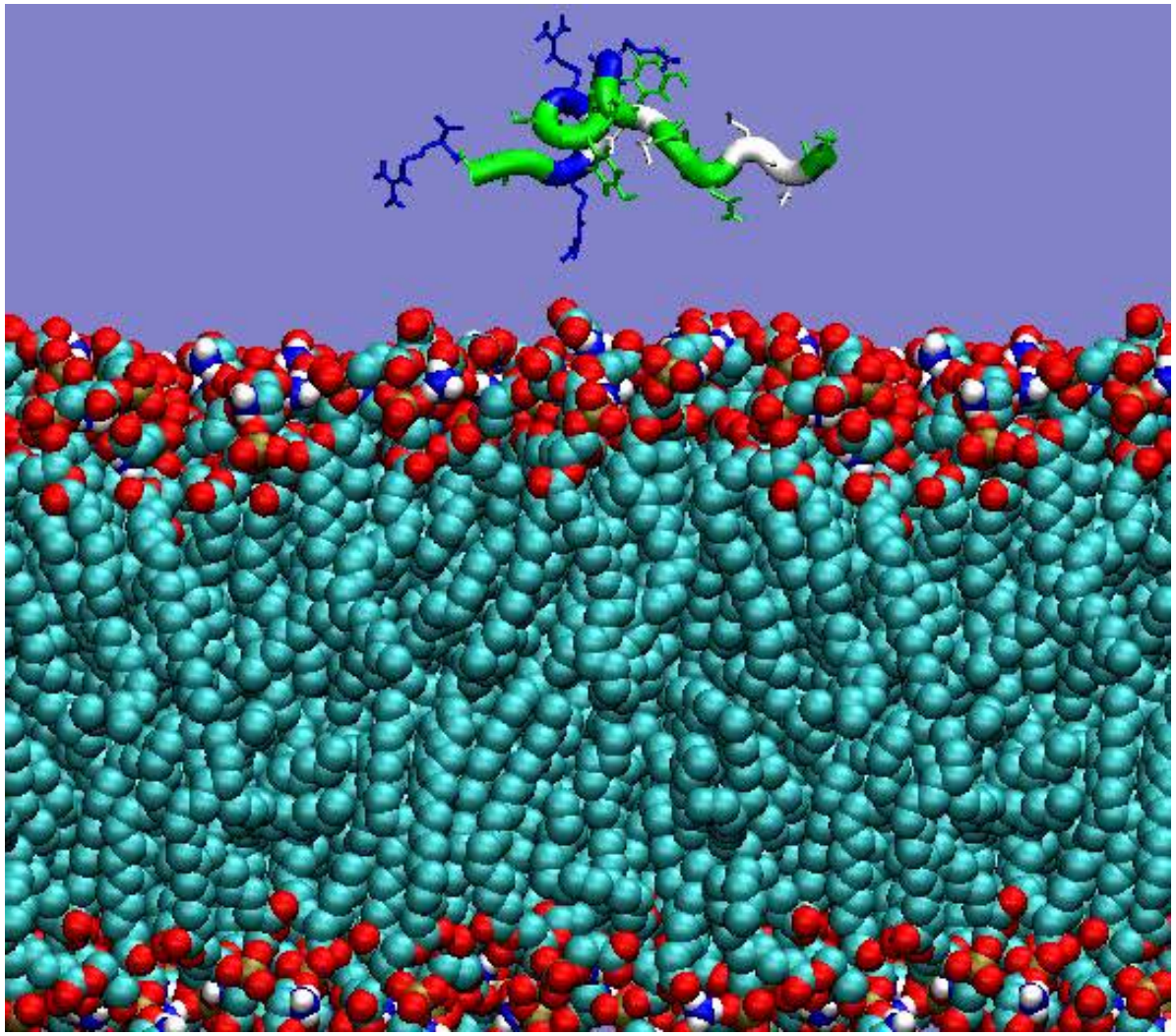
Image: H. Grubmüller, in Attig, et al. (eds.), *Computational Soft Matter* (2004)

What Will It Take to Simulate a Millisecond?

- We need an **enormous** increase in speed
  - Current (single processor): ~ 100 ms / fs
  - Goal will require < 10 ***m***s / fs

- **Required speedup:**
  - **> 10,000 x faster than current single-processor speed**
  - **~ 1,000x faster than current parallel implementations**
- ***Can't accept >10,000x the power (~5 Megawatts)!***

Target Simulation Speed



3.4 days today
(one processor)

~ 13 seconds on
HP machine
(one segment)

Molecular Mechanics Force Field

$$E = \sum_{\text{bonds}} k_b \left(r - r_0\right)^2$$

Stretch

$$+ \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2$$

Bend

$$+ \sum_{\text{torsions}} A[1 + \cos(n\tau - \varphi)]$$

Torsion

Bonded

$$+ \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}}$$

Electrostatic

$$+ \sum_i \sum_{j>i} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}$$

Van der Waals

Non-Bonded

What Takes So Long?

- Inner loop of force field evaluation looks at all ***pairs*** of atoms (within distance *R*)
- On the order of 64K atoms in typical system
- Repeat ~10^{12} times
- Current approaches too slow by several orders of magnitude
- What can be done?

Our Strategy

- **New architectures**
  - Design a specialized machine
  - Enormously parallel architecture
  - Based on special-purpose ASICs
  - Dramatically faster for MD, but less flexible
  - Projected completion: 2008

- **New algorithms**
  - Applicable to
    - Conventional clusters
    - Our own machine
  - Scale to very large # of processing elements

Interdisciplinary Lab