Lecture 12

Molecular Dynamics

Required reading: Chapter 6: 6.22 – 6.23
Karplus, M., and Petsko, G. A. (1990) Molecular dynamics simulations in biology. Nature 347: 631-639.

For further reading on the 2013 Nobel Prize, history and current state of computational methods like MD: Smith and Roux. Structure 21: 2102-2105.

Wednesday: Midterm 1

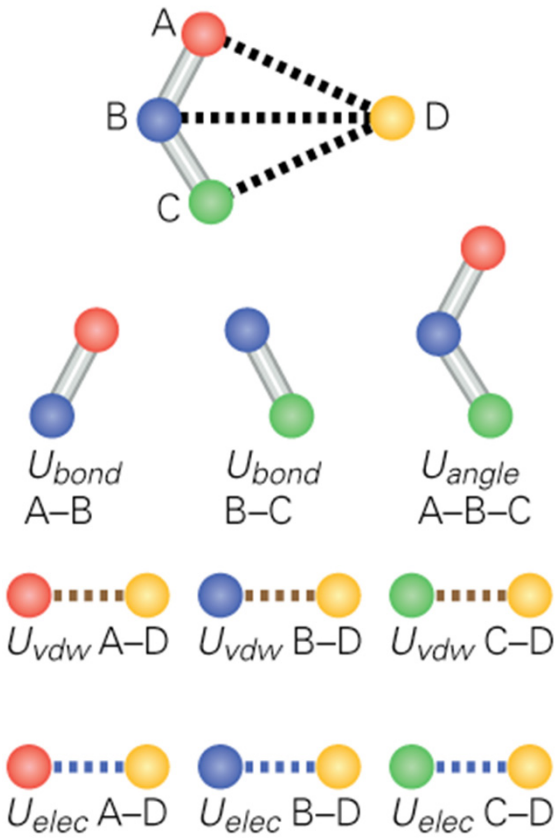
Reading for Friday: Chapter 7, sections 7.1-7.19

Today's goals

- Explain how solvent influences electrostatics
  - Dielectric constant models polarizability of solvent
  - Electrostatics influence interactions of ligands

- Describe the basic principles behind to molecular dynamics (MD)
  - Computational simulation of motions of molecules
  - Challenges and limitations of MD
  - Examples of insights into protein function from MD

Energy of macromolecules



- Component energy terms are assumed to be additive
- parameter values – typically pulled from data on small molecules – are assumed to be transferable

$$U_{total} = \sum U_{bonds} + \sum U_{angles} + \sum U_{dihedrals} + \sum U_{vdw} + \sum U_{elec}$$

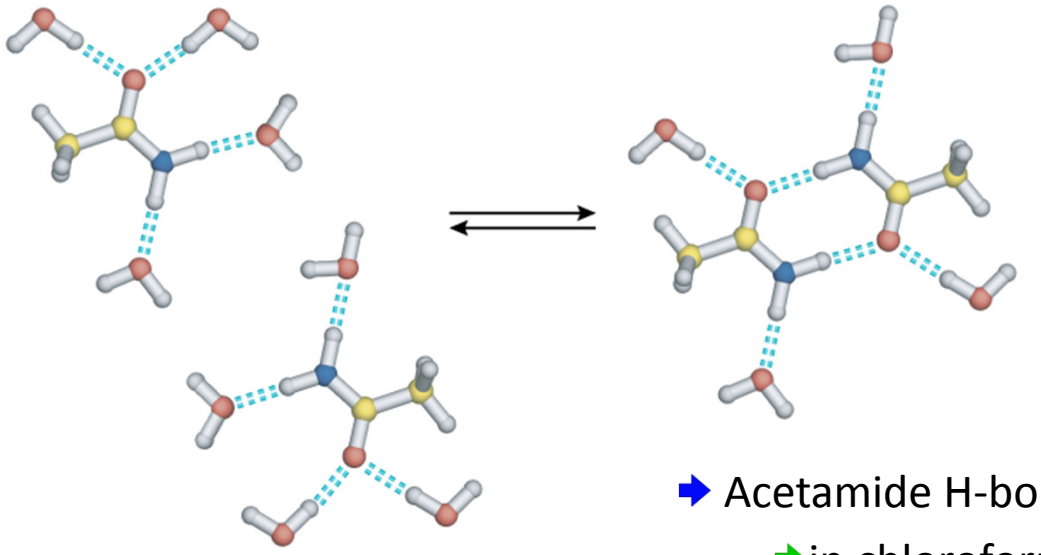
Figure from The Molecules of Life (© Garland Science 2008)

Solvent effects

- Measurements of H-bonds in gases:
  - ~10-20 kJ mol^{-1}
  - ~40 kJ mol^{-1} when one partner is charged
- Calculations for peptide bond to peptide bond H-bond in vacuum:
  - ~20 kJ mol^{-1}
- Measurement of H-bond energy in proteins in aqueous buffer:
  - ~2-4 kJ mol^{-1}
  - ~4-8 kJ mol^{-1} when one partner is charged
- Where does the difference come from?
  - Solvent effect – competition with water

Interactions with water weaken H-bonds

- H-bond energy in solvated proteins:
  - ~2-4 kJ mol^{-1} (~4-8 kJ mol^{-1} when one partner is charged)
  - Energy difference between H-bond with water vs. H-bond with protein group



- Acetamide H-bond (dimer)
  - in chloroform ~17 kJ mol^{-1}
  - in water ~0

Figure from The Molecules of Life (© Garland Science 2008)

Why are H-bonds so important to proteins?

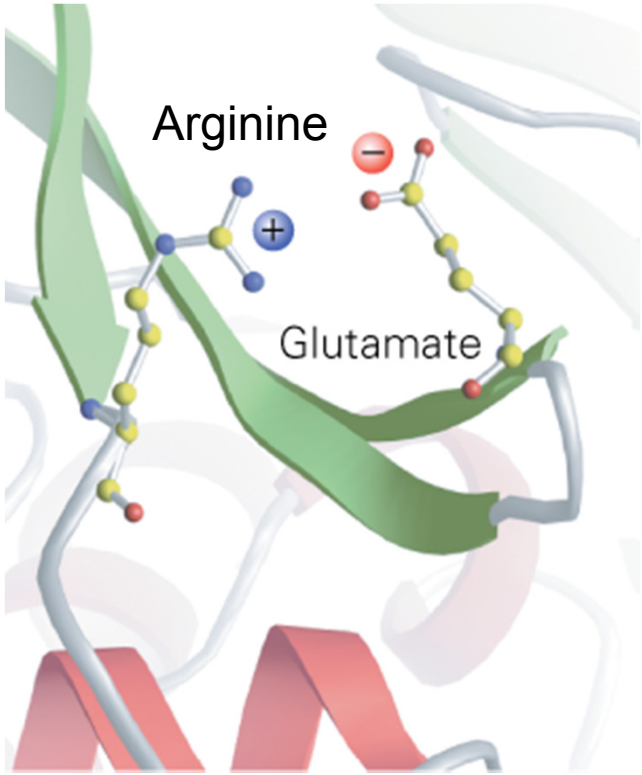
- Solubility provided by the H-bonding groups (before and after folding)

- Specificity
  - van der Waals interactions alone do not lead to a specific, unique structure (e.g. lipid bilayer)
  - H-bonds provide constraints on conformations

- Using a collection of low energy interactions allows for conformational changes in response to external cues

Electrostatics

- Electrostatic interaction between two atoms is described using Coulomb's law:



$$U_{electrostatic} \propto \frac{q_i q_j}{r_{ij}}$$

- q_i and q_j are charge on atoms *i* and *j*
- r_{ij} = distance between the pair
- Used for full or partial charges

Figure from The Molecules of Life (© Garland Science 2008)

Calculating the electrostatics

$$U_{electrostatic} = \left(\frac{1}{4\pi\varepsilon_0}\right)\frac{q_i q_j}{r_{ij}}$$

Coulomb force constant

ε_0 is the vacuum permittivity = 8.854 ×10^{-12} C^2 N^{-1} m^{-2}

- A full elementary charge is 1.602 × 10^{-19} Coulombs
- To express the potential energy in kJ/mol, using elementary charge for q_i and q_j, and the distance r_{ij} in Å:

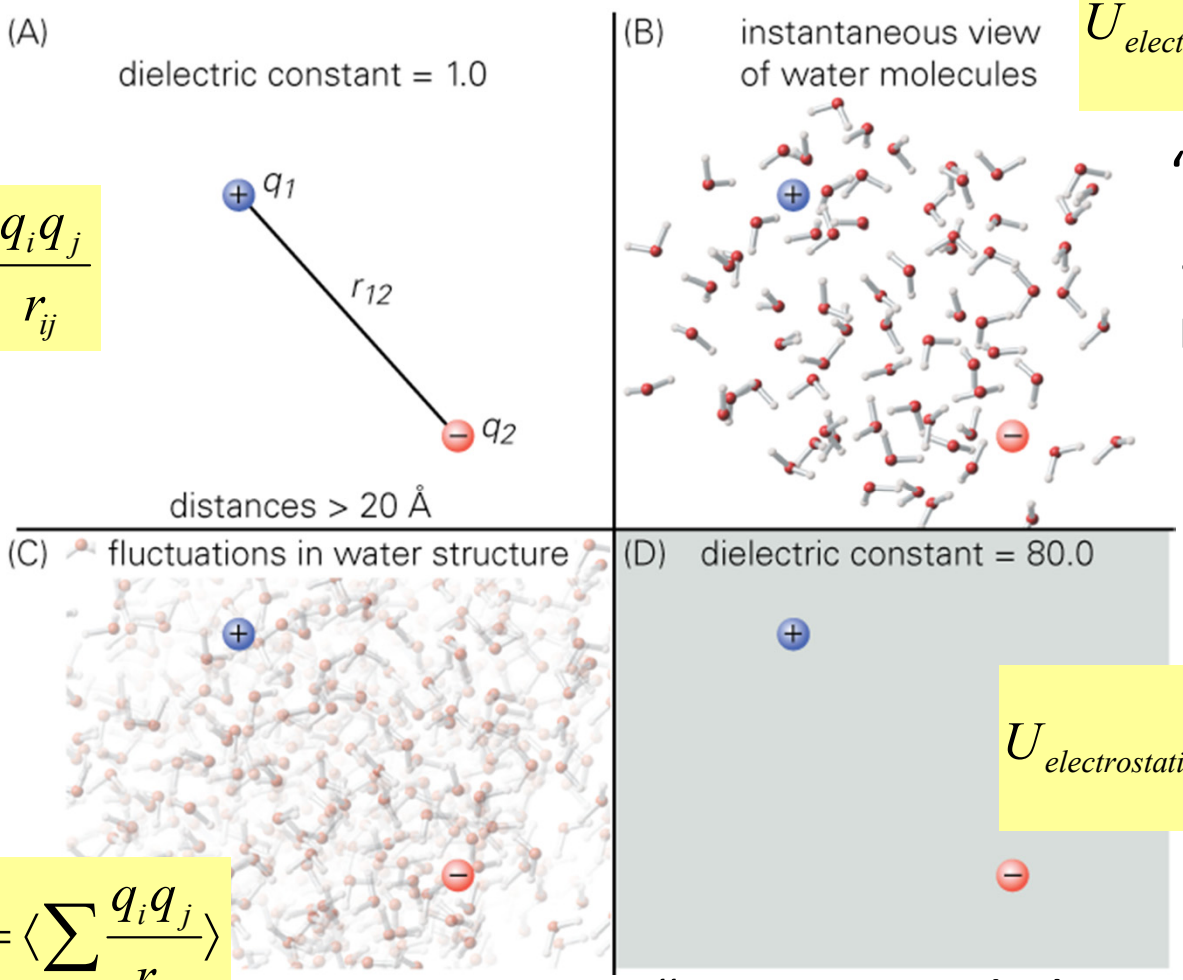
$$U_{electrostatic} = \frac{q_i q_j}{r_{ij}} \times 1390\ \text{kJ/mol}$$

- Two elementary charges 4 Å apart = 347.5 kJ/mol!

Quantifying the shielding effect of water

in vacuum

$$U_{electrostatic} = \sum_{ij} \frac{q_i q_j}{r_{ij}}$$



$$U_{electrostatic} = \frac{q_i q_j}{r_{ij}}$$

"explicit solvent model"

$$U_{electrostatic} = \frac{q_i q_j}{\varepsilon r_{ij}}$$

$$U_{electrostatic} = \langle \sum_{ij} \frac{q_i q_j}{r_{ij}} \rangle$$

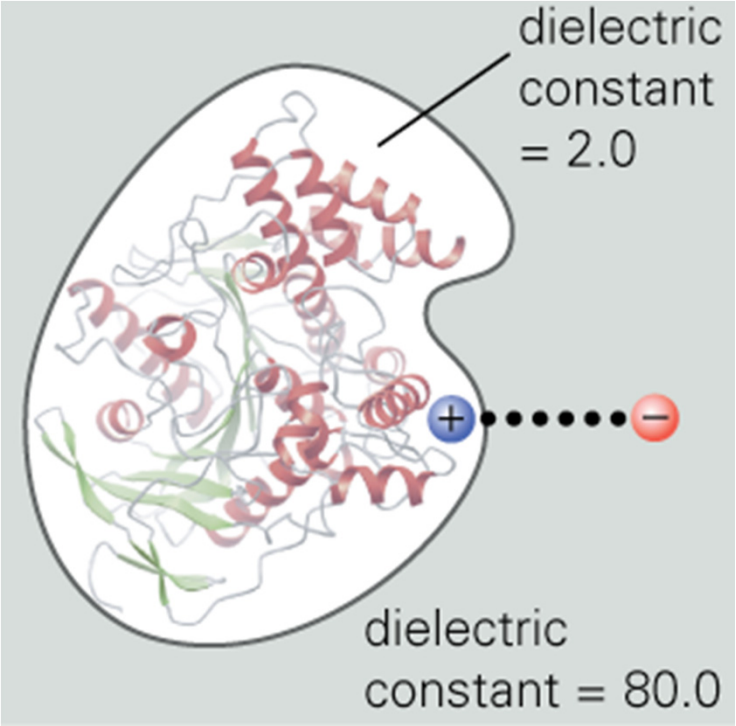
"continuum dielectric model"

Note: all equations above omit the 1390 kJ/mol constant for simplicity

Figure from The Molecules of Life (© Garland Science 2013)

The continuum dielectric model

- The solvent can be modeled using a dielectric constant, $\varepsilon = 80$
  - Two elementary charges 4 Å apart = (347.5/80) = 4.3 kJ/mol
- The inside of the protein, however, is much less *polarizable*, and its dielectric constant is typically set to $\varepsilon \sim 2\text{-}4$



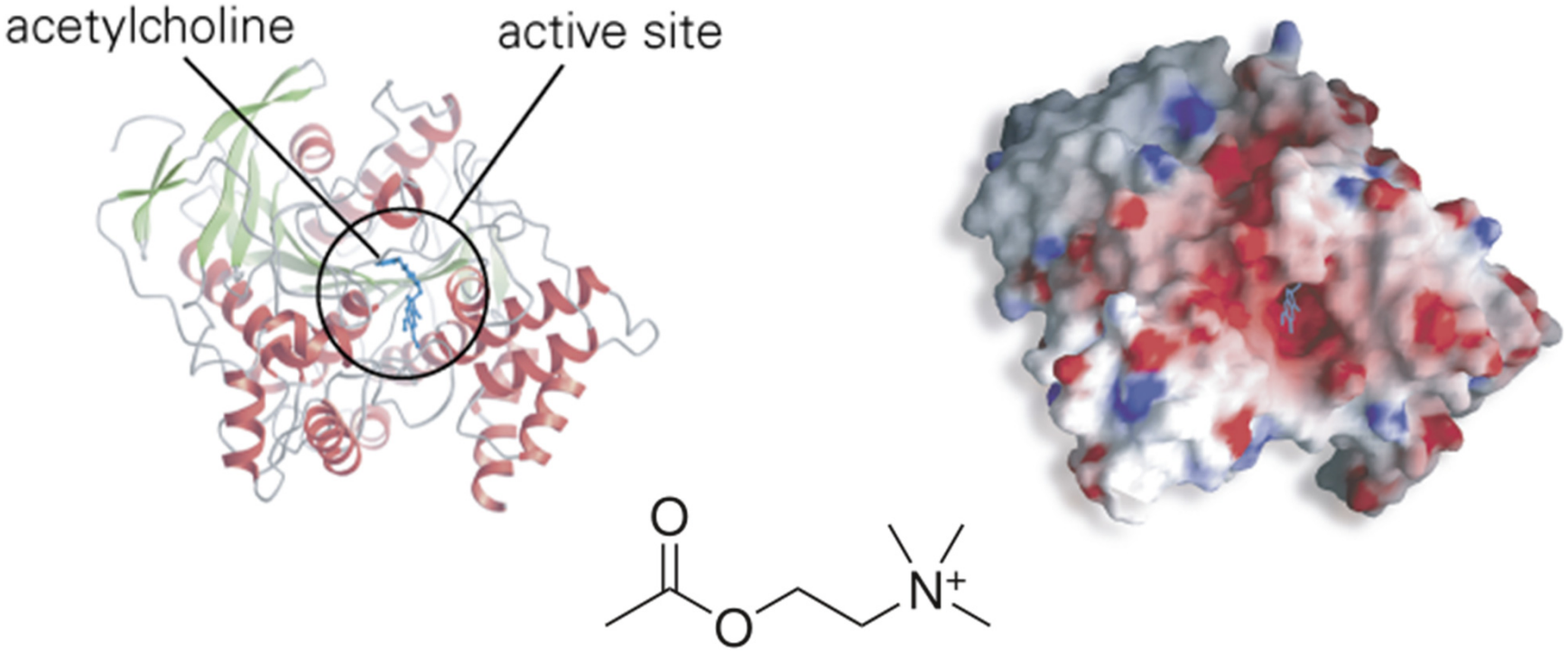
For interactions that bridge the two environments – we need to take into account the non-uniform dielectric environment and the presence of physiological ions and the Poisson-Boltzmann Equation is used.

Figure from The Molecules of Life (© Garland Science 2008)

Electrostatic potential can be mapped onto the surface

- The substrate of acetylylcholine esterase, acetylcholine (a neurotransmitter), is positively charged

Blue is + and red is –

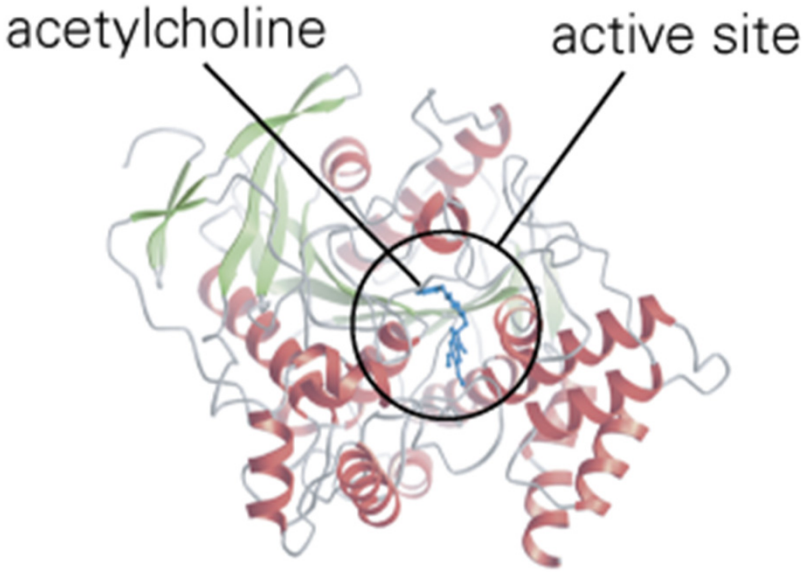


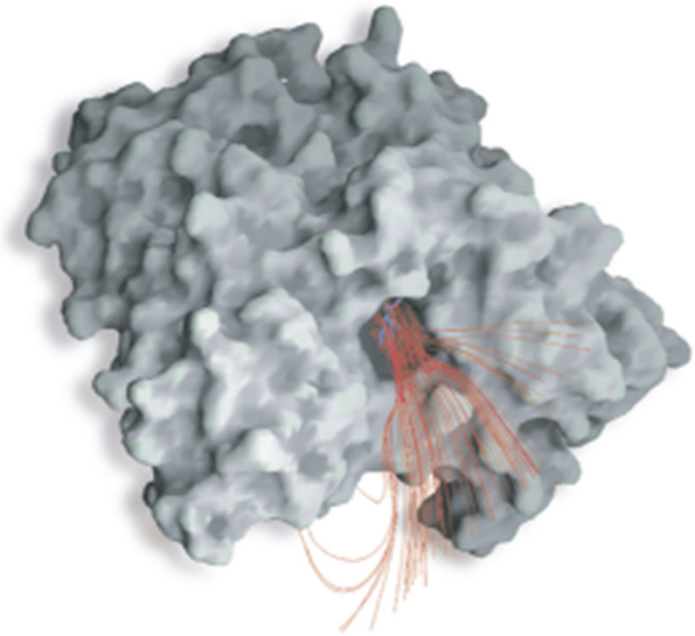
- Electrostatic potential of the enzyme from the distribution of + and – charged side chains

Figure from The Molecules of Life (© Garland Science 2013)

Electrostatic potential can be illustrated by force field lines

- At any point on the map, the electrostatic force on a positive charge is calculated.





- Each red line indicates the path along which a positive charge would move if there were no other force acting on it.

Figure from The Molecules of Life (© Garland Science 2013)

Molecular Dynamics

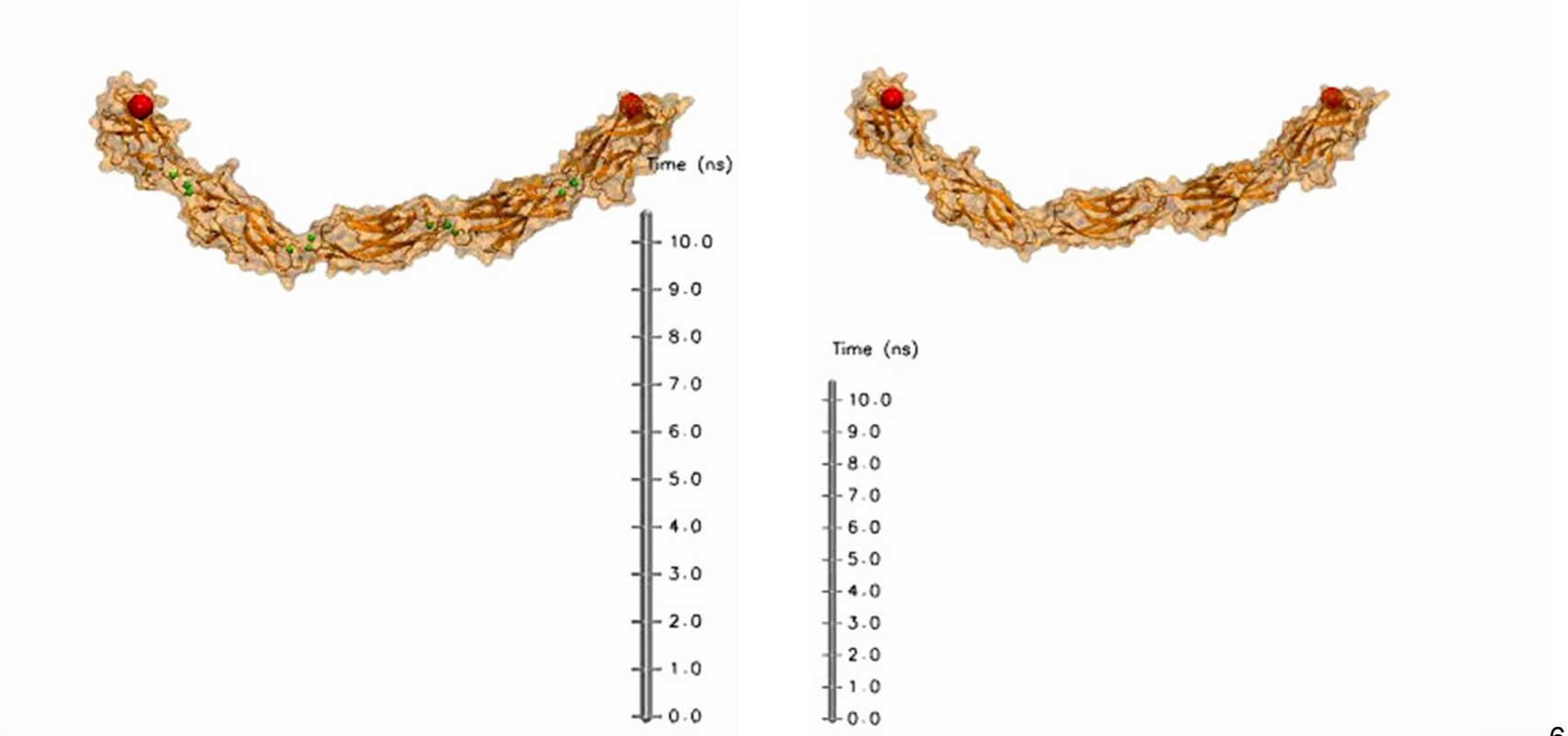
- *Molecular Dynamics* (MD) simulations compute the motions of individual molecules in models of solids, liquids or gases

- *Motion* describes how atom and molecule positions, velocities and orientations change with time

- In principle, the behavior of a given system can be computed if we have a set of initial conditions and forces of interactions.

Molecular Dynamics – the Movie

- Time scale – this represents ~11 ns (shown in ~35 s – slowed down by a factor of 1/3 billion)
- Cadherin cell adhesion protein with and without calcium ions



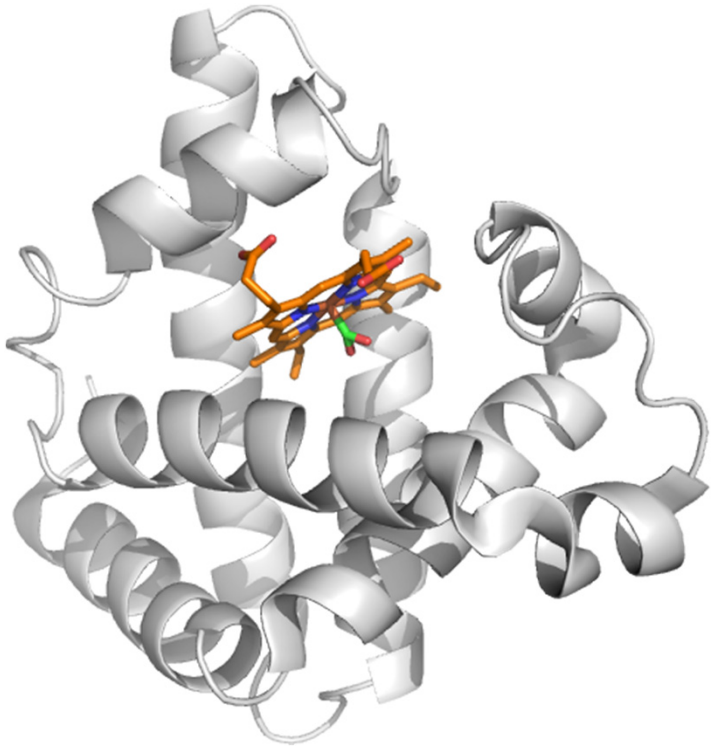
Sotomayor and Schulten, Biophysical Journal (2008) 94: 4621-4633

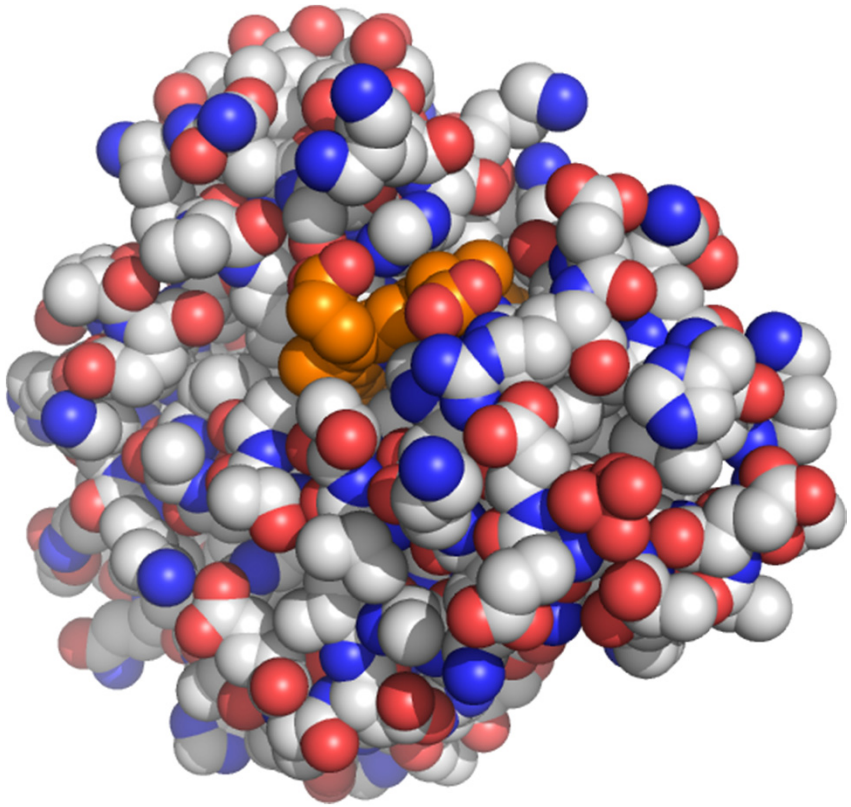
Molecular Dynamics

- "The intrinsic beauty and the remarkable details of the protein structures obtained from x-ray crystallography resulted in the view that proteins are rigid. This created the misconception that atoms in a protein are fixed in position…"
  - M. Karplus and M.A. McCammon: "The Dynamics of Proteins" Scientific American 254 (4): 42 (1986)

Static structures raise questions

- Myoglobin structure leaves no path for the ligand to enter or exit:





- Calculated half-time ~10^{50} years!

Molecular Dynamics

- "The intrinsic beauty and the remarkable details of the protein structures obtained from x-ray crystallography resulted in the view that proteins are rigid. This created the misconception that atoms in a protein are fixed in position…"
  - M. Karplus and M.A. McCammon: "The Dynamics of Proteins" Scientific American 254 (4): 42 (1986)

- "The most powerful assumption of all ... (is) that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."
  - R. R. P. Feynman, R. B. Leighton, M. L. Sands: "The Feynman Lectures on Physics". Addison-Wesley, Reading, MA, 1963

Why is MD useful?

- Most structural information is a *static*, spatial and/or temporal average of thousands (EM) to millions (crystallography) of molecules
  - Note: NMR and other spectroscopy techniques do allow us to measure some dynamic processes
- Biochemical reactions (enzyme catalysis, protein complex assembly, protein or RNA folding, chromatin maintenance and assembly, etc) are *dynamic* processes
- MD allows us to describe some of these processes

MD – the basics

- To run a molecular dynamics simulation we need:
  - The interaction potential for the particles in the system, from which we can calculate the forces acting on atoms:
    - For molecules, the force is the derivative of the potential energy with respect to the atom position

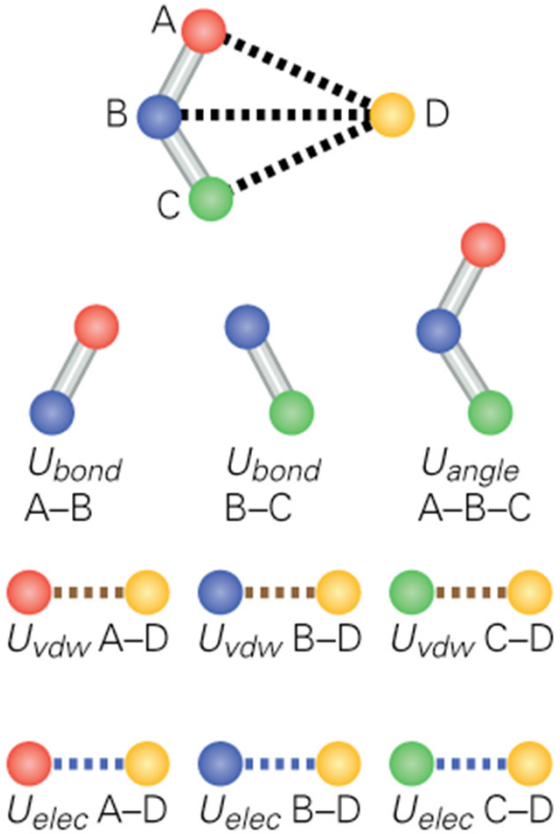
$$F(r) = -dU/dr$$

  - The equation of motion governing the dynamics of the particles
    - Classical Newtonian equations are adequate for biomolecules

$$F_i = m_i a_i$$

Energy of macromolecules

$$U_{total} = \sum U_{bonds} + \sum U_{angles} + \sum U_{dihedrals} + \sum U_{vdw} + \sum U_{elec}$$



- Component energy terms are assumed to be additive
- parameter values – typically pulled from data on small molecules – are assumed to be transferable
  - Assumptions are likely reasonable for van der Waals and bonded energy terms, but less so for electrostatics

Figure from The Molecules of Life (© Garland Science 2013)

MD: the Force Field Equation

$$U(R) = \sum_{\text{bonds}} \frac{K_b}{2}(r-r_0)^2 + \sum_{\text{angles}} \frac{K_\theta}{2}(\theta-\theta_0)^2 + \sum_{\text{dihedral}} \frac{K_\chi}{2}[1+\cos(n\chi-\delta)] + \sum_i \sum_{i \neq j} \varepsilon_{ij}\left[\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}}$$

Bonding terms

Non-bonded interactions

Lennard-Jones

electrostatic

- Where R = atomic coordinates

MD: the Force Field Equation

$$U(R) = \sum_{\text{bonds}} \frac{K_b}{2}(r-r_0)^2 + \sum_{\text{angles}} \frac{K_\theta}{2}(\theta-\theta_0)^2$$

$$+ \sum_{\text{dihedral}} \frac{K_\chi}{2}[1+\cos(n\chi-\delta)]$$

$$+\sum_i \sum_{i\neq j} \varepsilon_{ij}\left(\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min_{ij}}}{r_{ij}}\right)^{6}\right) + \frac{q_i q_j}{r_{ij}}$$

- In red: depends on the atomic position coordinates
- In blue: predetermined force field parameters

Energy scales

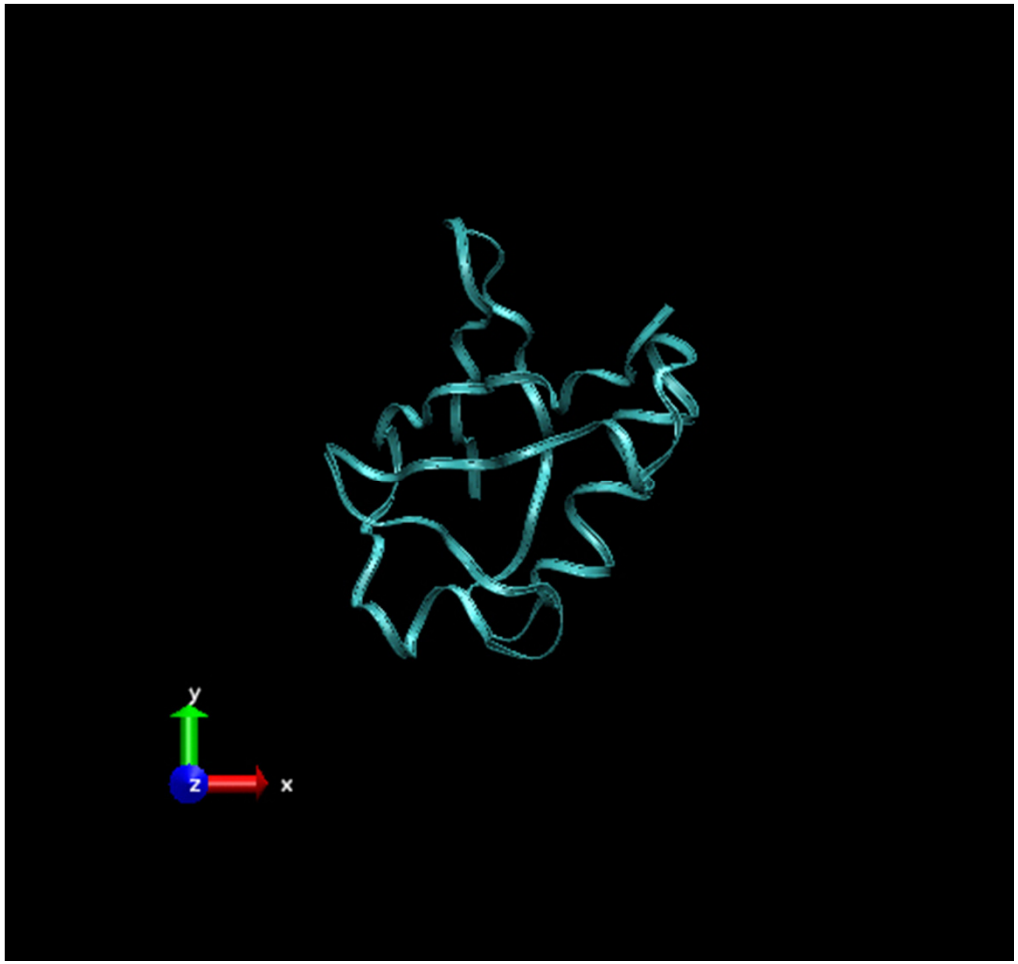
- Room temperature $k_B T$ ~ 2.5 kJ/mol

- Bond vibrations – stiff (K ~ 400-2000 kJ mol^{-1} A^{-2})
- bond angle bending – less stiff (K ~ 50-200 J mol^{-1} deg^{-2})
- Dihedral rotations – soft (K ~ 0-10 kJ mol^{-1})

- van der Waals – ~ 2 kJ mol^{-1} ~ $k_B T$
- Hydrogen bonds – ~2-4 kJ mol^{-1} in aqueous solution
  - (uncharged partners)
- Salt bridges – ~5-10 kJ mol^{-1} in aqueous solution

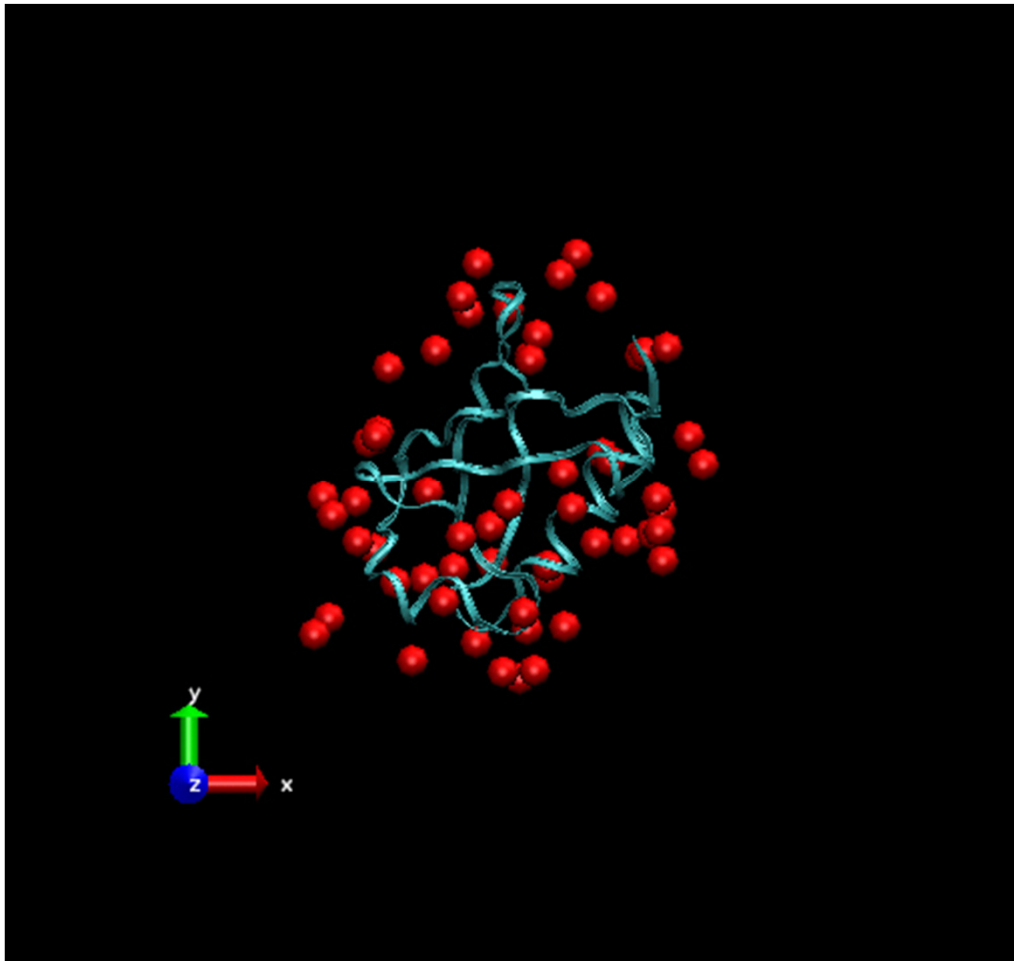
Building the system

- Start with a crystal structure (~600 atoms)
- We need some solvent too



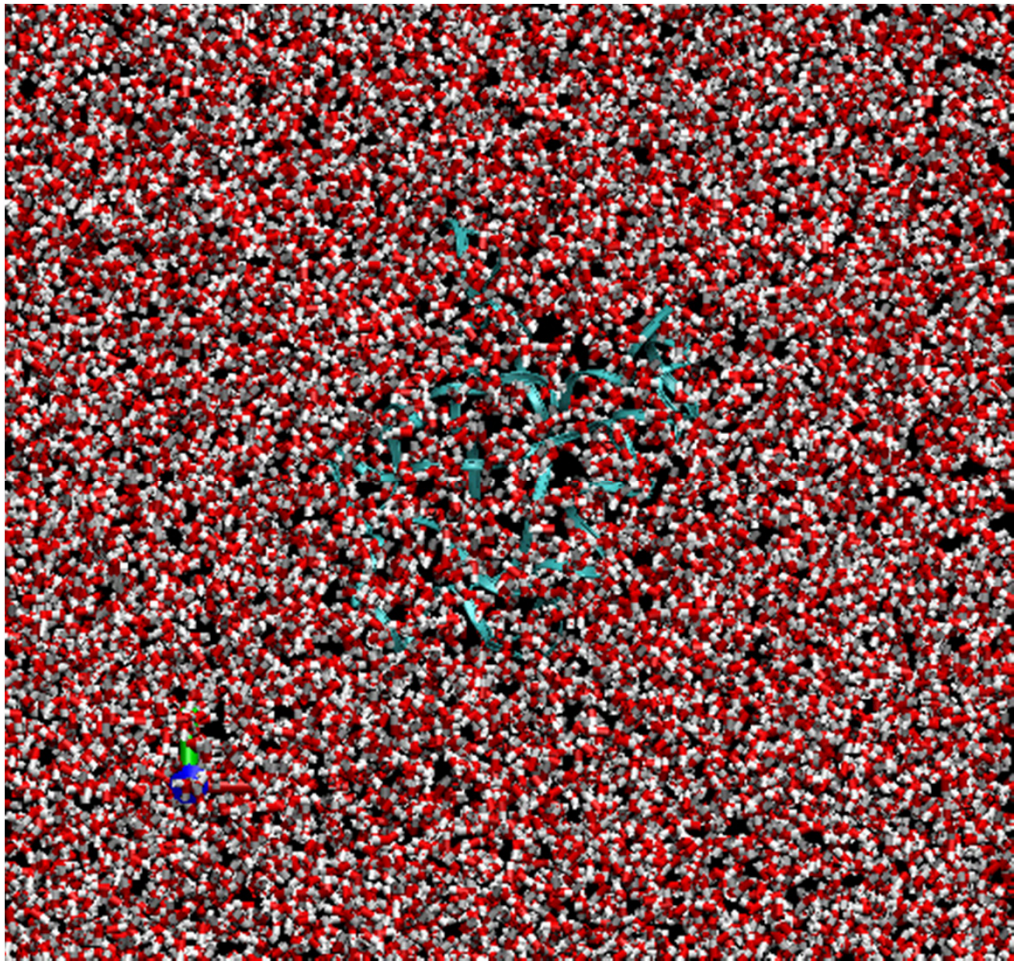
Building the system

- Start with a crystal structure (~600 atoms)
- We need some solvent too
- The crystal structure has some solvent
  - These water molecules are bound to the protein, form a solvent shell



Building the system

- Start with a crystal structure (~600 atoms)
- We need some solvent too
- The crystal structure has some solvent
  - These water molecules are bound to the protein, form a solvent shell
- Add more solvent with hydrogens (~26,000 atoms)



MD: the simulation process

- Given a potential energy function, solve Newton's equations of motion for all atoms in the system:

$$F_i = m_i a_i$$

  - Where *F* is the force on the atom, calculated from the potential energy
  - *m* is its mass
  - *a* is the acceleration

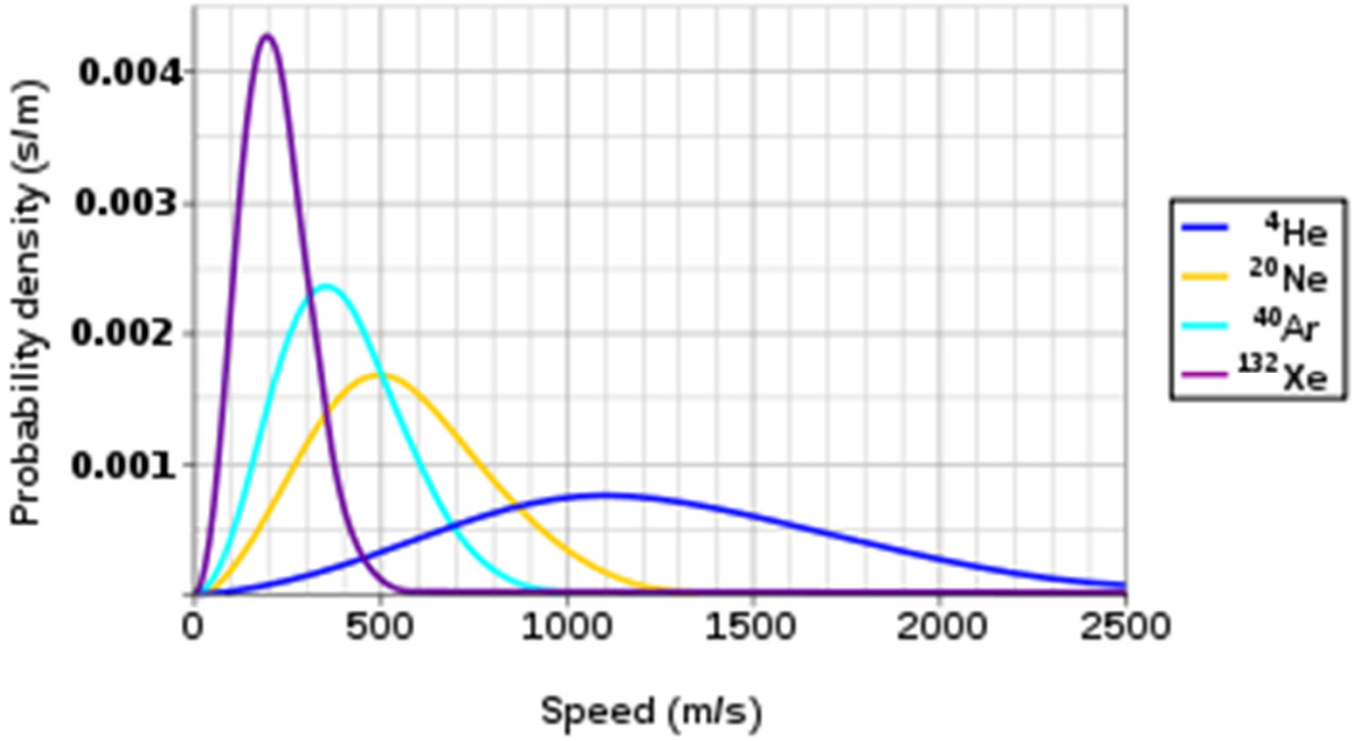
- Atoms are assigned *initial velocities* at random from a Maxwell-Boltzmann distribution of kinetic energy.
  - The velocities reflect the *temperature* in the system (kinetic energy)

- Each *time step* should be small enough so that the potential energy does not change too much during the time step (1 fs)

Maxwell-Boltzmann Distribution

- A normalized form of the Boltzmann distribution to describe the velocity:

$$P(v) = 4\pi v^2 \left(\frac{m}{2\pi k_B T} \right)^{\frac{3}{2}} e^{\frac{-mv^2}{2k_B T}} dv$$



The process of calculating an MD time step

- Solve for acceleration (a_i) at time t: $-\frac{dU_i}{dr} = F_i = m_i a_i$

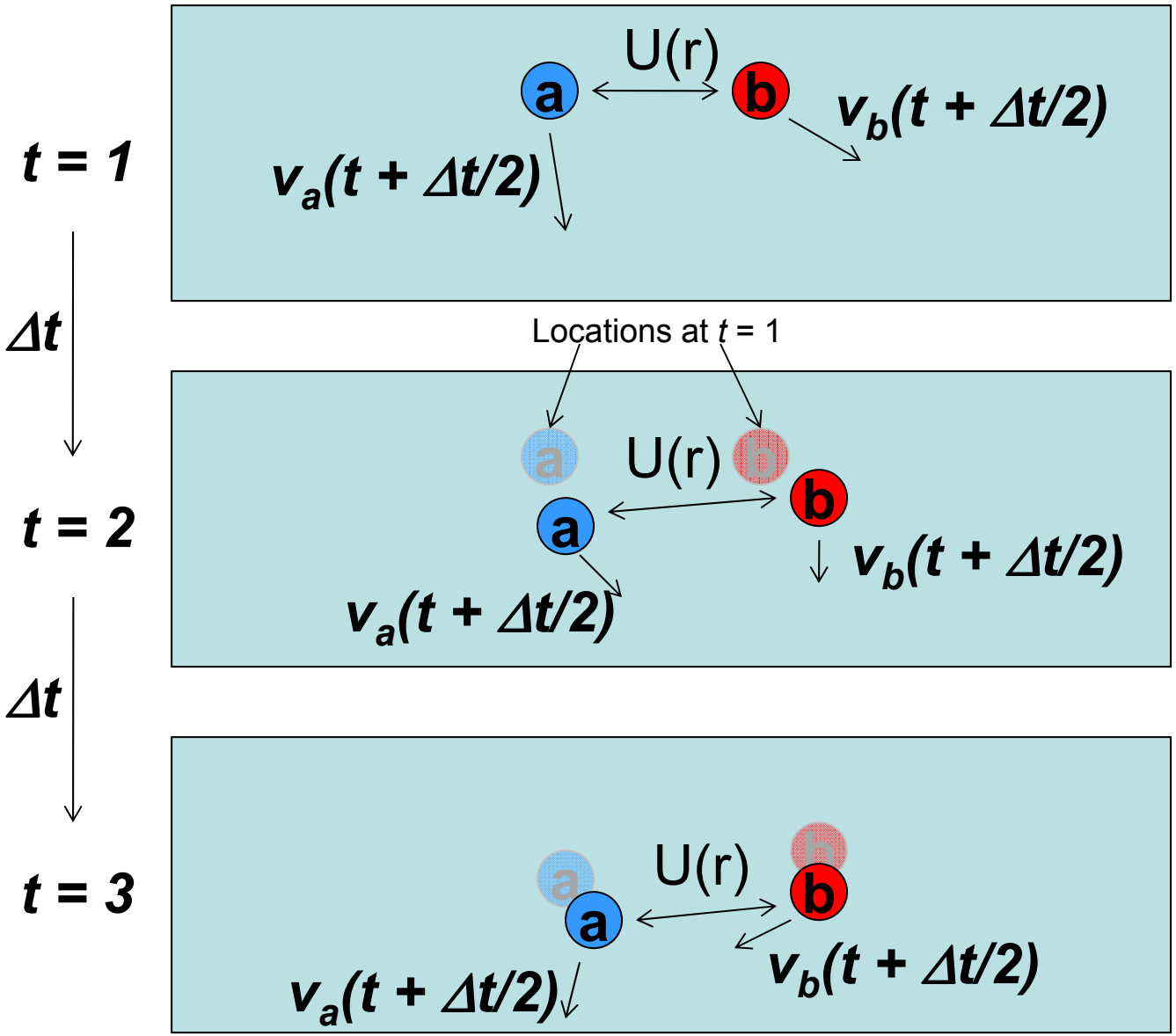
- Update velocity (v_i at $t + \Delta t/2$, middle of the next time step):

$$v_i(t + \Delta t/2) = v_i(t - \Delta t/2) + a_i \cdot \Delta t$$

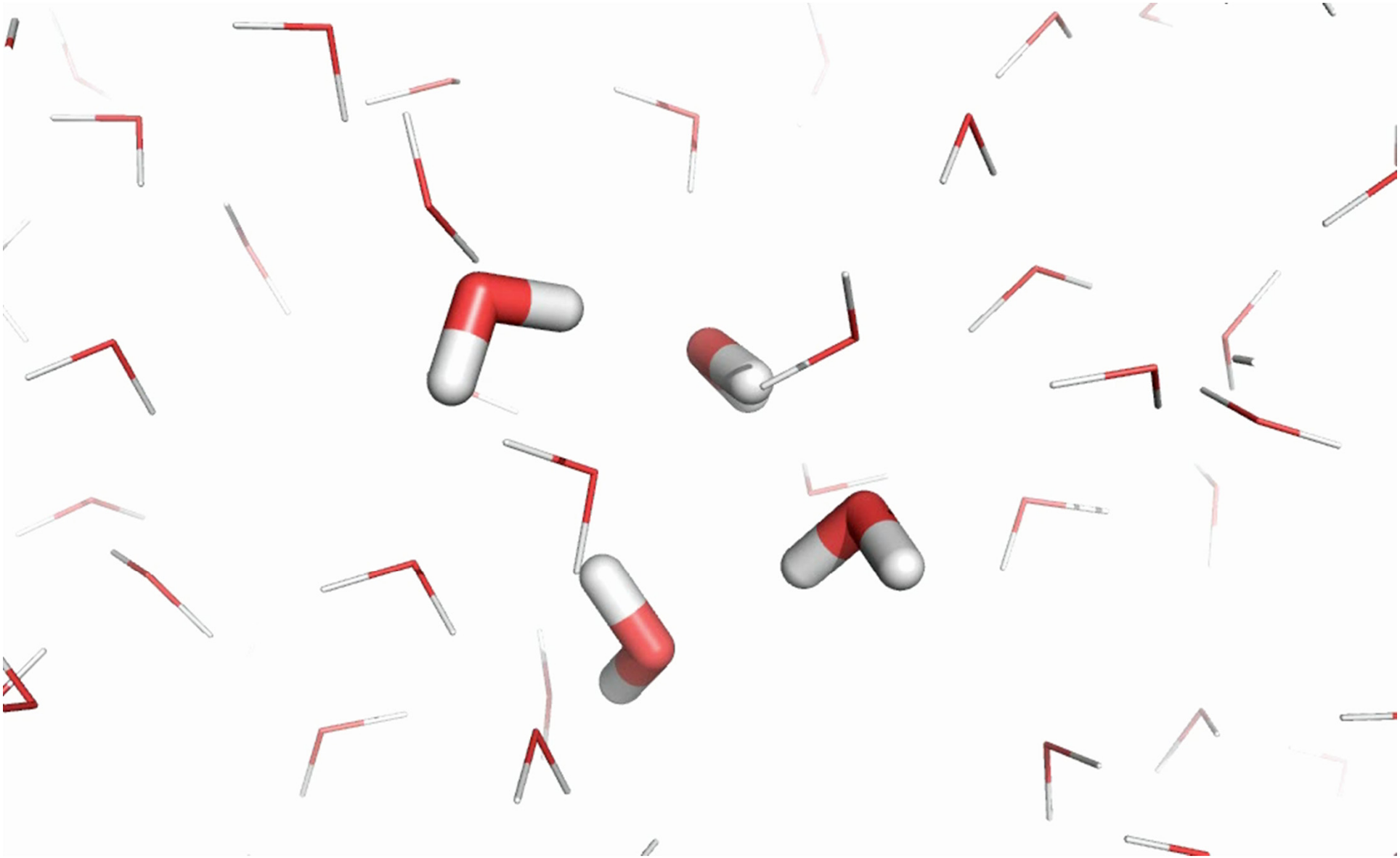
- Update atomic positions (r_i at $t + \Delta t$):

$$r_i(t + \Delta t) = r_i(t) + v_i \cdot \Delta t$$

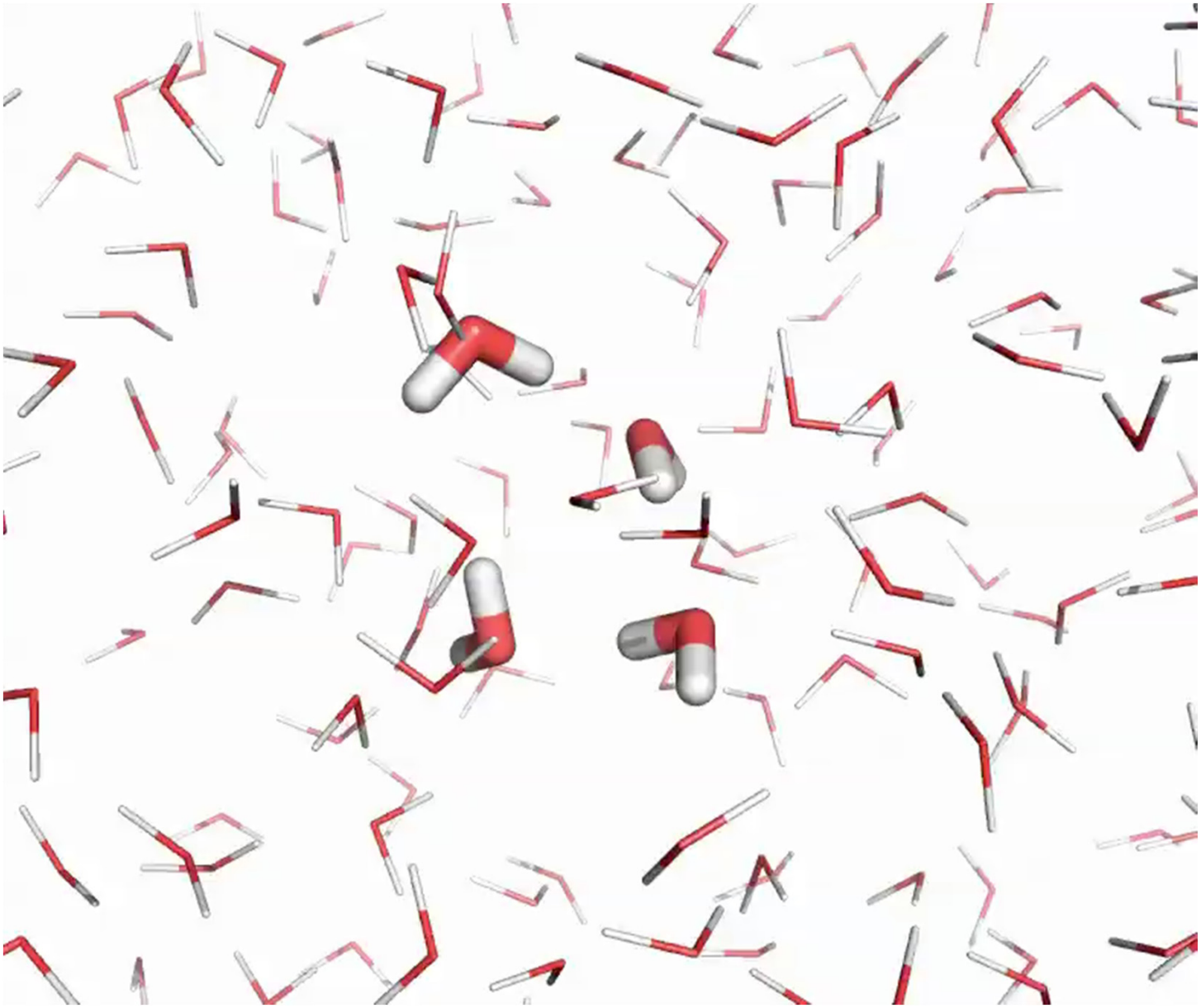
- Recalculate the potential energy (U_i) based on the new atom positions


t = 1
Δt
t = 2
Δt
t = 3
U(r)
a
b
$v_b(t + \Delta t/2)$
$v_a(t + \Delta t/2)$
Locations at t = 1
U(r)
a
b
$v_b(t + \Delta t/2)$
$v_a(t + \Delta t/2)$
U(r)
a
b
$v_b(t + \Delta t/2)$
$v_a(t + \Delta t/2)$

Water molecules rotate, bend in 100 fs

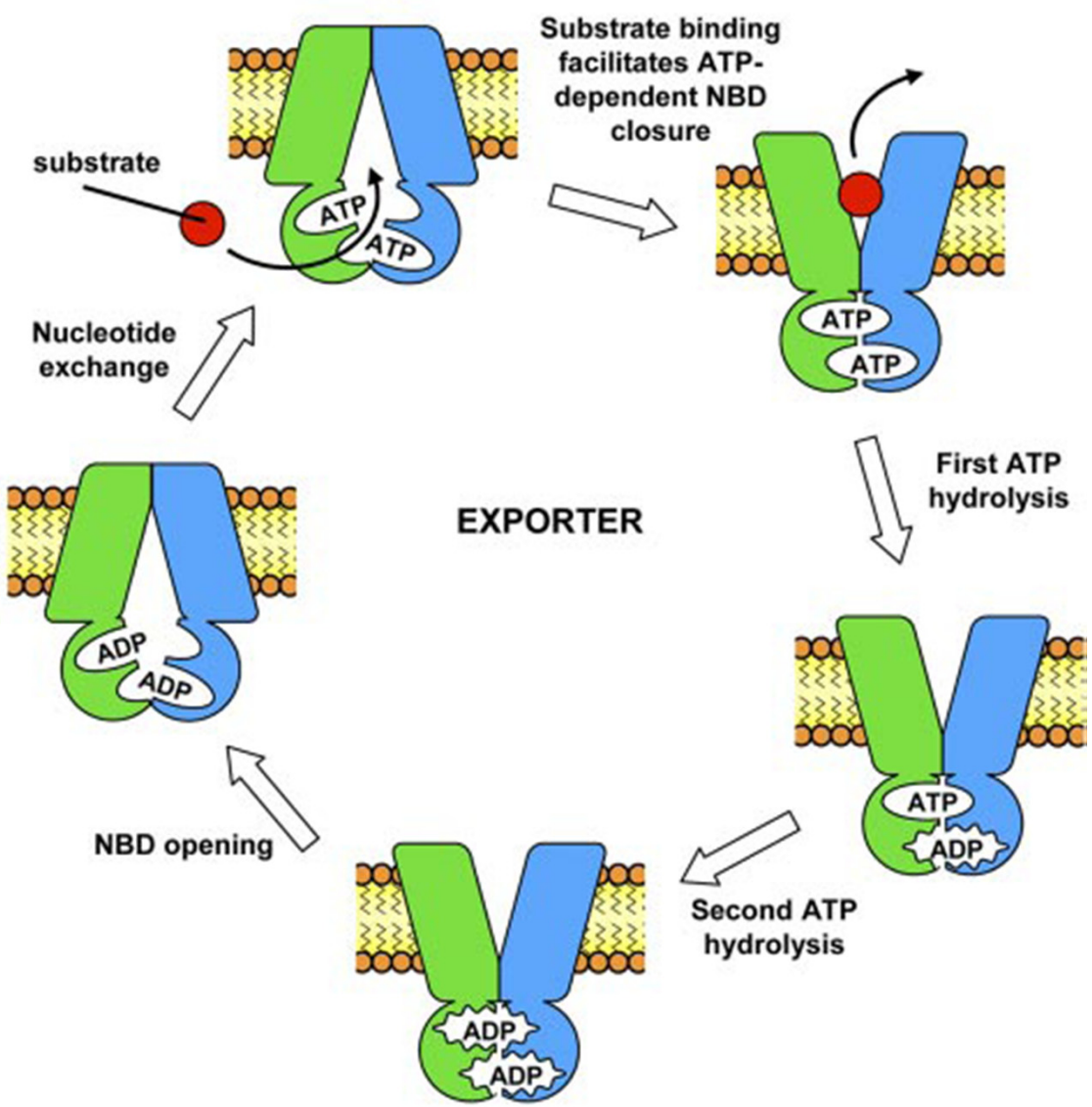


Water molecules diffuse a little over 1 ps



Example: nucleotide binding domains

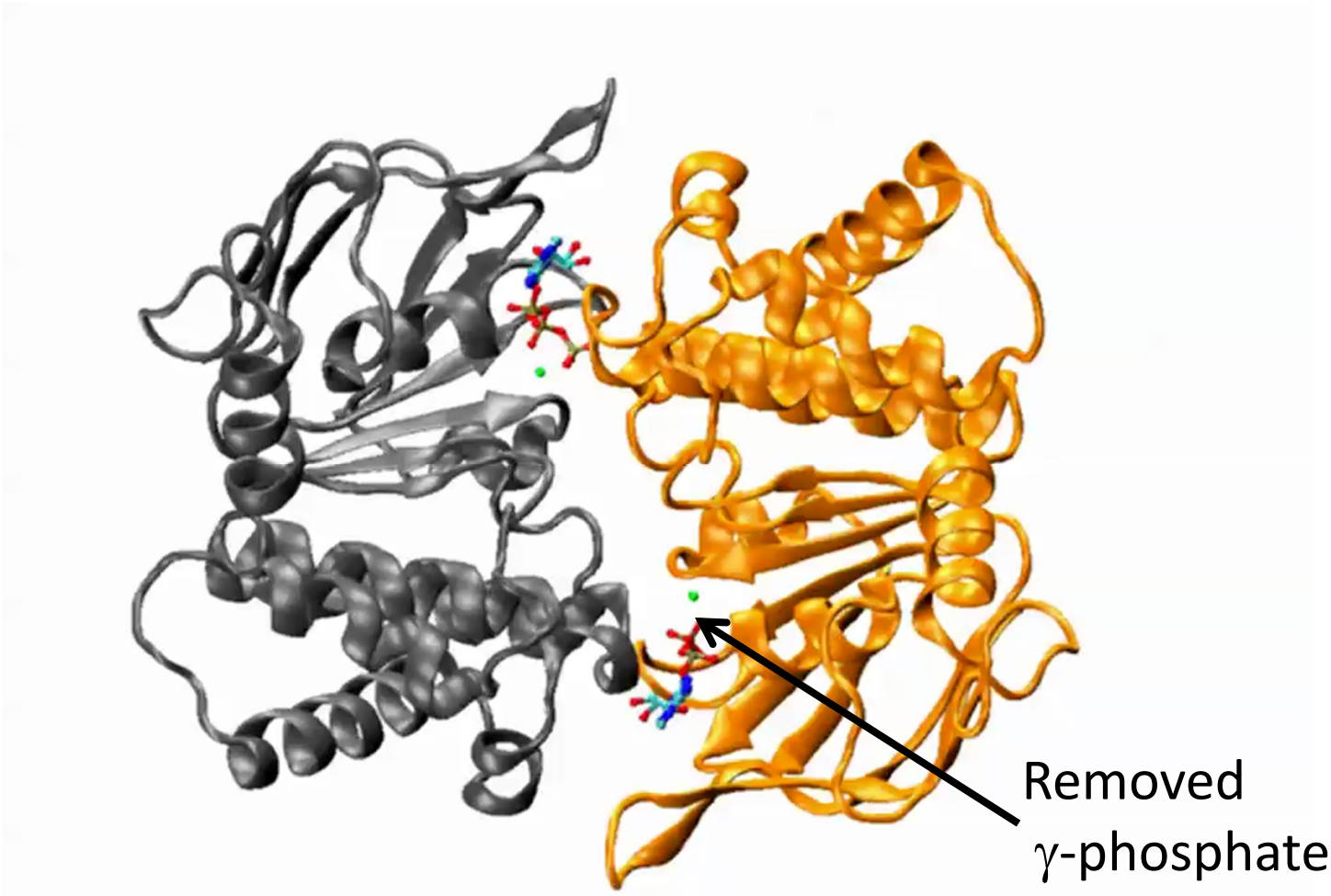
- ATP-binding cassette transporters
- Two nucleotide binding domains that cycle between closed and open
- ATPase rates are often ~ a few ATPs per second
- Current computers can do: ~100 K atoms on 200 processors at 15-20 ns/day



Procko and Gaudet FASEB Journal (2009)

Example: nucleotide binding domains

- Take a structure with 2 ATPs bound. Remove the γ-phosphate from one of the ATPs, then run MD – that side of the dimer opens in ~10 ns: (Simulation is 19 ns.)



Jones and George Proteins (2008) 75:387-396.

Timescale of macromolecular motions

- Local motions (femtoseconds to milliseconds):
  - Atomic fluctuations
  - Sidechain motions
  - Loop motions
- Rigid-body motions (nanoseconds to seconds):
  - Helix motion
  - Domain motions (hinge-bending)
  - Subunit motions (allostery)

- Larger-scale motions (microseconds to days):
  - Macromolecular folding/unfolding
  - Dissociation/association of proteins/nucleic acids

Challenges in MD

- Approximations of the force field lead to *systematic errors* (hence, calculations of free energy differences is still very difficult)

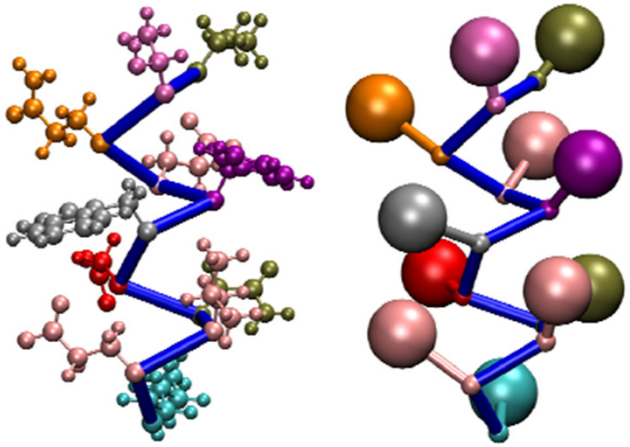
- Computing cost of sampling trajectories at the appropriate timescale leads to *statistical errors*
    - E.g. a typical timestep is 1 fs; for a 1 ns simulation, there are 10^6 integrations

- Reduce the problem to minimize the computing time (e.g. rigid bodies for domain motions, restrict motions to the active site) – "Coarse grain"

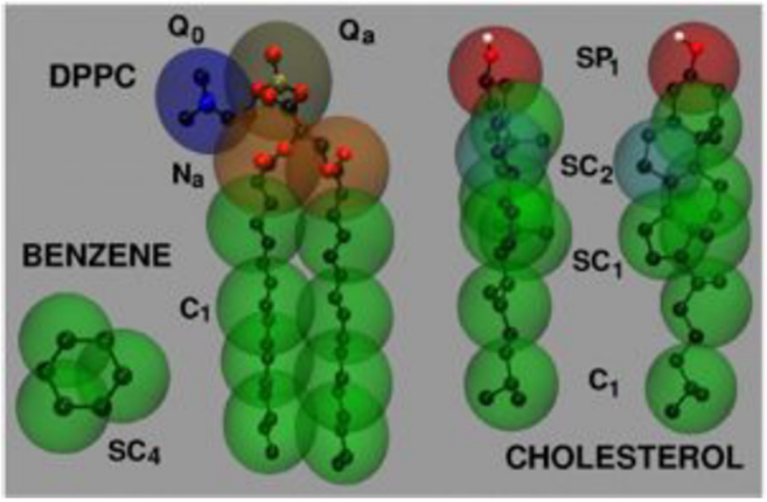
Coarse grain example – lipoprotein assembly

- Coarse grain (CG bead) models reduces the number of "atoms"

  - Protein model uses two CG beads per residue
  - One CG bead per side chain another for backbone



All-atom peptide CG peptide

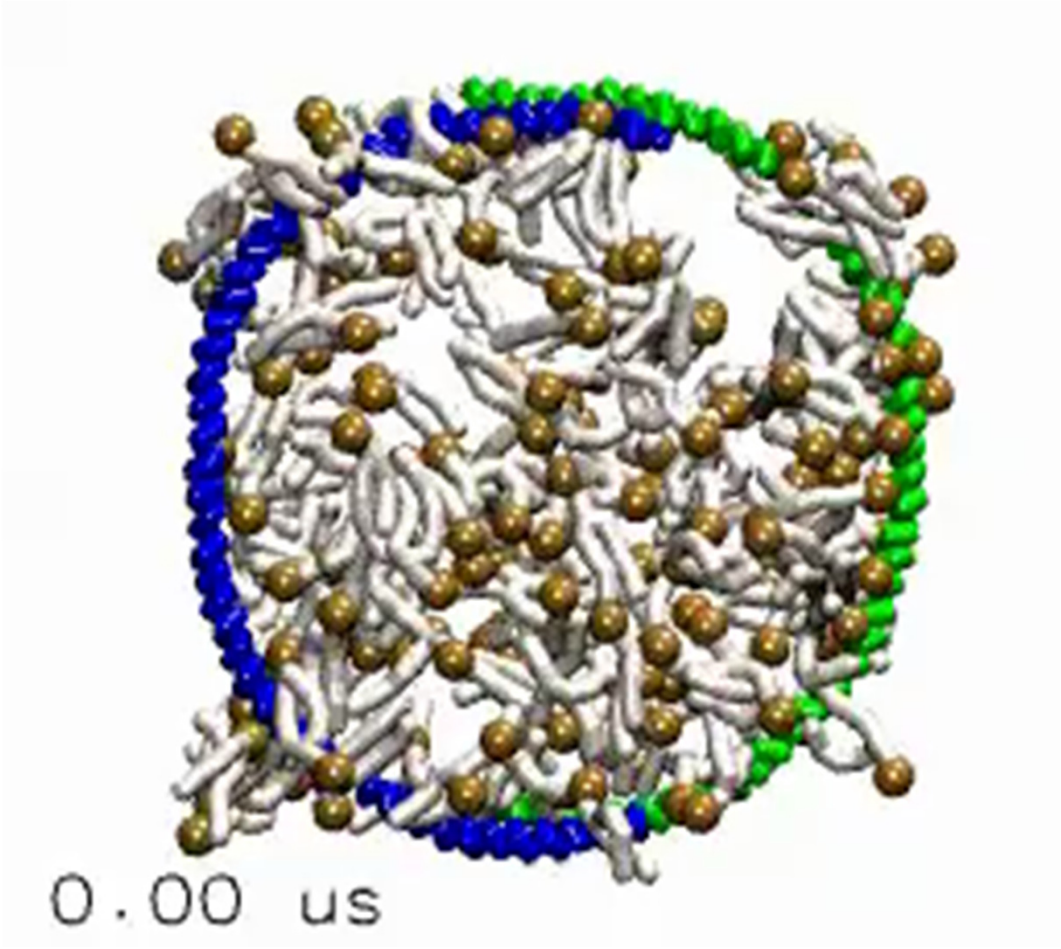


  - Level of coarse-graining: ~4 heavy atoms per CG bead

Application of residue-based and shape-based coarse graining to biomolecular simulations. P. L. Freddolino, A. Arkhipov, A. Y. Shih, Y. Yin, Z. Chen, and K. Schulten. In G. A. Voth, editor, Coarse-Graining of Condensed Phase and Biomolecular Systems, chapter 20, pp. 299-315. .CRC Press, 2008.

Coarse grain example – lipoprotein assembly

- 10 μs simulation (25 fs time steps)



- What drives assembly?

http://www.ks.uiuc.edu/Research/Lipoproteins/

Some concepts to remember

- Molecular dynamics studies the *motions* of molecules (internal and external)
- Some parameters that can affect the quality of the simulation:
  - Force field description
  - Time step
  - Starting model (macromolecular structure and surrounding solvent, substrates, etc)
  - Length and reproducibility of the simulation