

Introduction to classical molecular dynamics simulation

Lee-Ping Wang
OpenMM Workshop
Stanford University
March 2013

Outline

Introduction

- Physical chemistry; the role of theory and computation
- An overview of computer simulations of molecules
- Molecular dynamics within the classical approximation

Methods

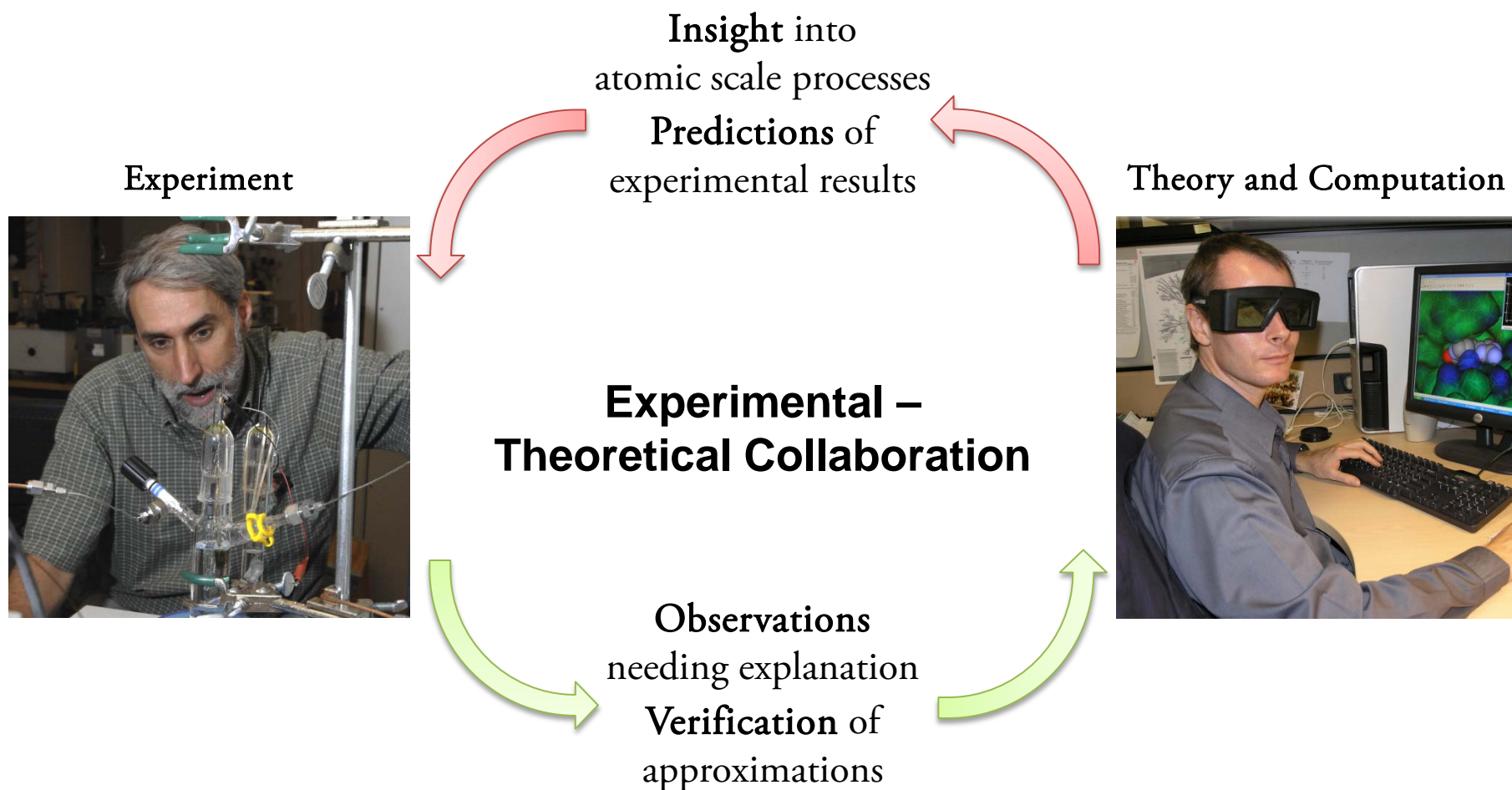
- Force fields
- Integrating the equations of motion
- Sampling from thermodynamic ensembles

Applications

- Protein folding structure and mechanism
- Conformational change and binding free energies
- Properties of condensed phase matter (water)

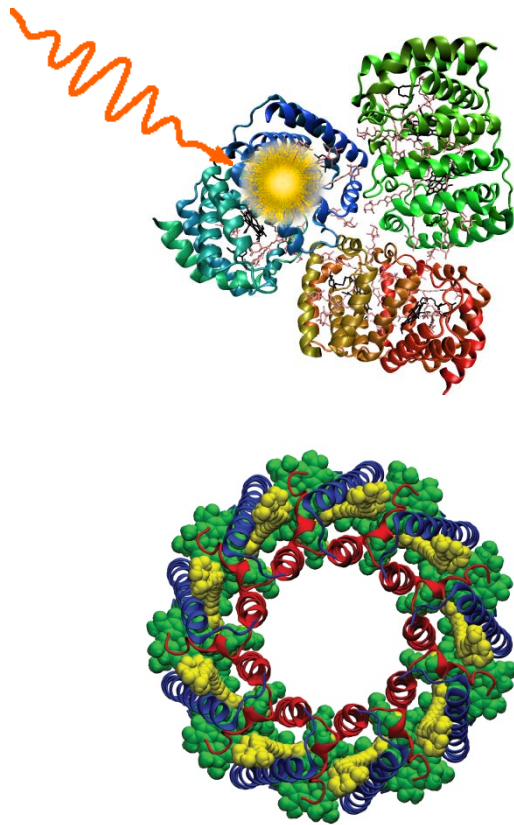
Introduction: Theory and Experiment

Theoretical and computational chemistry can offer *explanations* and *predictions*.



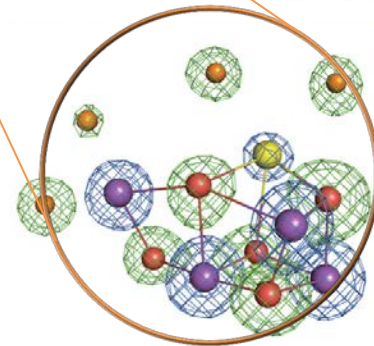
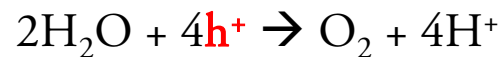
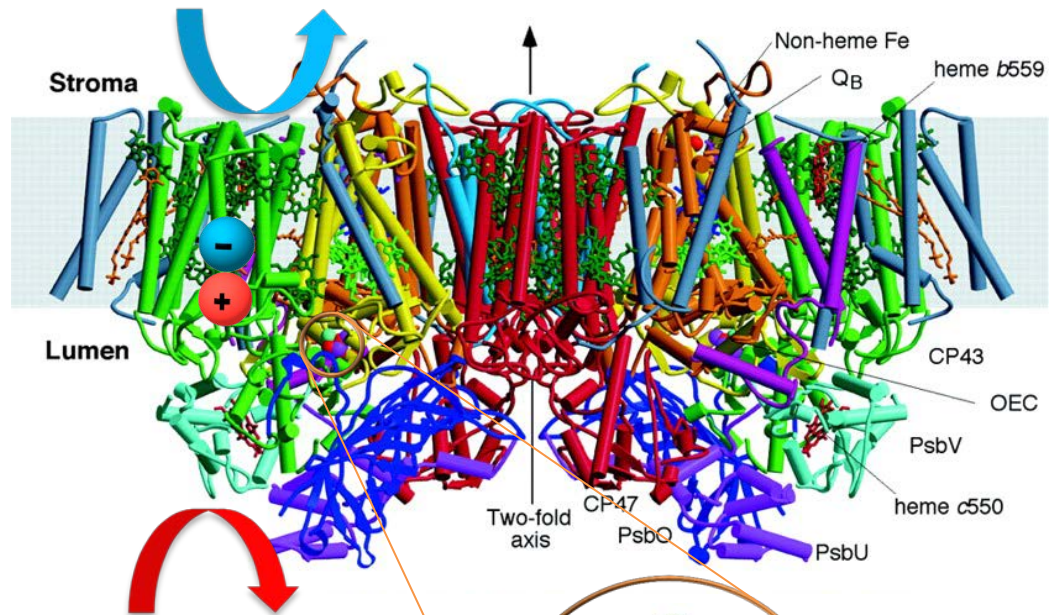
Example of an important process

Understanding the mechanism of photosynthesis requires both experimental and theoretical insights.



Light harvesting complexes in green plants (top) and purple bacteria (bottom)

Photosystem II: Light harvesting and Water splitting protein in green plants
Energy stored: 3.60 eV
 $4\text{H}_2\text{O} \rightarrow 2\text{O}_2 + 4\text{H}^+$



Oxygen evolving center from Umena et al., *Nature*, 2011.

Introduction: Physical Theories

The main tools of theoretical chemistry are quantum mechanics and statistical mechanics.

QUANTUM MECHANICS

$$\hat{H}\Psi = E\Psi$$

- The electronic structure of a molecule is completely described by solving Schrodinger's equation
- For most systems, exact solution not available; *approximations* are computationally intensive

STATISTICAL MECHANICS

$$Z = \int d^{3N}\mathbf{r} \, e^{-\frac{E(\{\mathbf{r}_i\})}{k_b T}}$$

- Given a potential energy surface, the *partition function* provides probabilities of states
- High-dimensional integral is evaluated by *sampling*, requiring computer power and efficient algorithms

The Schrödinger equation

All ground-state quantum chemistry is based on the time-independent Schrödinger's equation.

$$\hat{H}\Psi = E\Psi$$

$$(\hat{T} + \hat{V})\Psi = E\Psi$$

$$\left(-\sum_i \frac{\nabla_i^2}{2} + \sum_{I \neq J} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{i,I} \frac{Z_I}{|\mathbf{R}_I - \hat{\mathbf{r}}_i|} + \sum_{i \neq j} \frac{1}{|\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j|} \right) \Psi(\mathbf{r}_i; \mathbf{R}_I) = E(\mathbf{R}_I) \Psi(\mathbf{r}_i; \mathbf{R}_I)$$

Kinetic
energy
operator

Nuclear
repulsion
(just a “number”)

Electron
nuclear
attraction

Electron
electron
repulsion

Electron
wavefunction

Electronic
energy

Born-Oppenheimer approximation:
when solving for the electronic
wavefunction, treat nuclei as static
external potentials

Approximate solutions to Schrödinger equation

Schrödinger's equation is nearly impossible to solve, so approximate methods are used.

$$\left(-\sum_i \frac{\nabla_i^2}{2} - \sum_{i,I} \frac{Z_I}{|\mathbf{R}_I - \hat{\mathbf{r}}_i|} + \sum_{i \neq j} \frac{1}{|\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j|} \right) \Psi(\mathbf{r}_i) = E \Psi(\mathbf{r}_i)$$

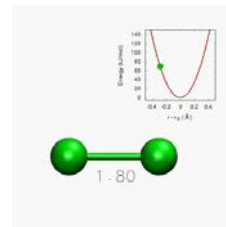
$$|\Psi\rangle_{\text{HF}} = \begin{vmatrix} \chi_1(1) & \chi_1(2) & \cdots & \chi_1(N) \\ \chi_2(1) & \chi_2(2) & \cdots & \chi_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_N(1) & \chi_N(2) & \cdots & \chi_N(N) \end{vmatrix}$$

Slater determinant: Use wavefunction of *noninteracting* electrons for interacting system

$$\rho_0(\mathbf{r}) \leftrightarrow \Psi_0(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, \mathbf{r}_N) \rightarrow E$$

Density functional theory: The ground-state electron density contains all information in the ground-state wavefunction

- **Electron repulsion** makes this problem difficult
- **Approach 1:** Use approximate wavefunction forms (Quantum chem.)
- **Approach 2:** Use the electron density as the main variable
- **Approach 3:** Approximate the energy using empirical functions

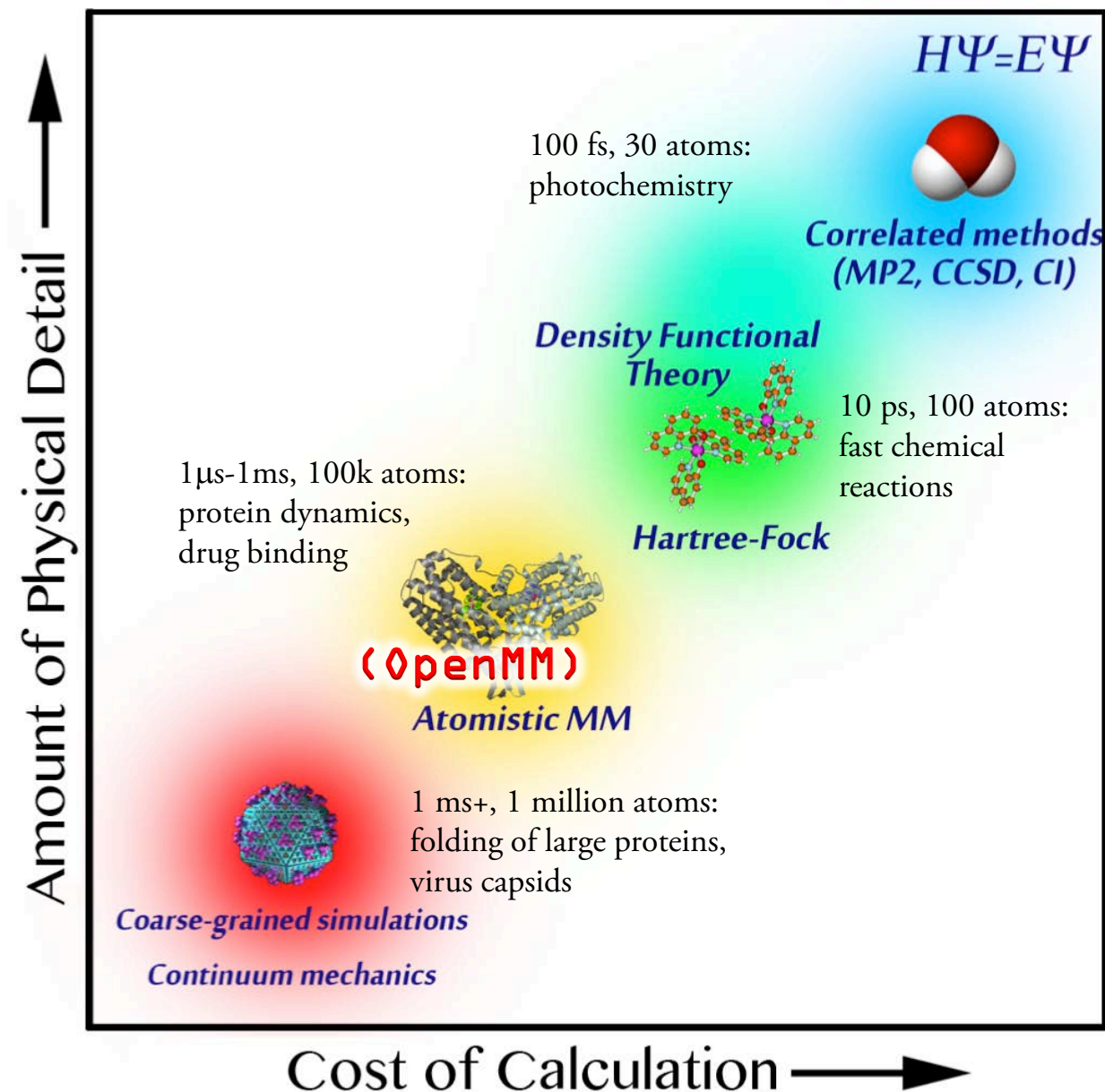


$$E(\mathbf{R}_I) \rightarrow k_{IJ} (R_{IJ} - R_{IJ}^0)^2$$

Molecular mechanics: Use empirical functions and parameters to describe the energy (e.g. harmonic oscillator for chemical bond)

A wide scope of computer simulations

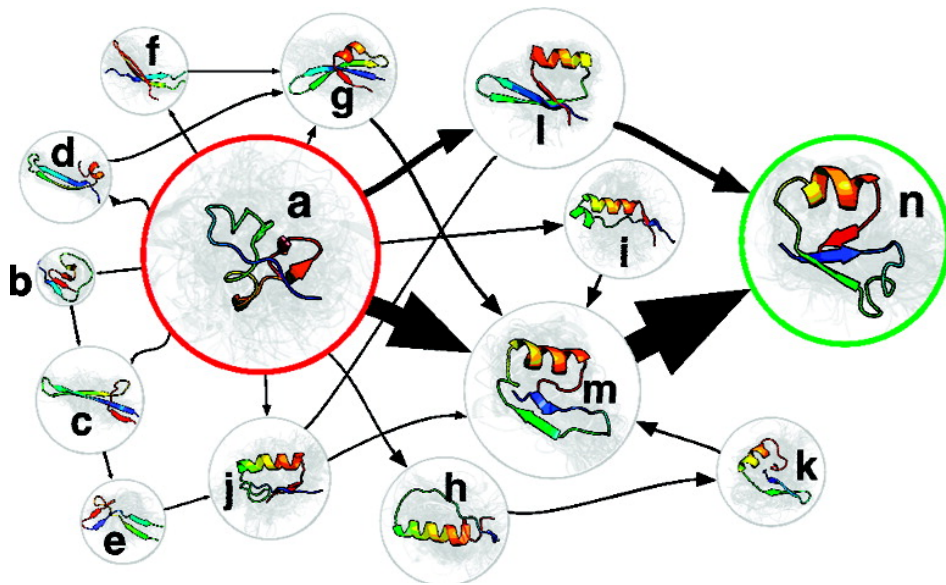
One calculation for 2-3 atoms



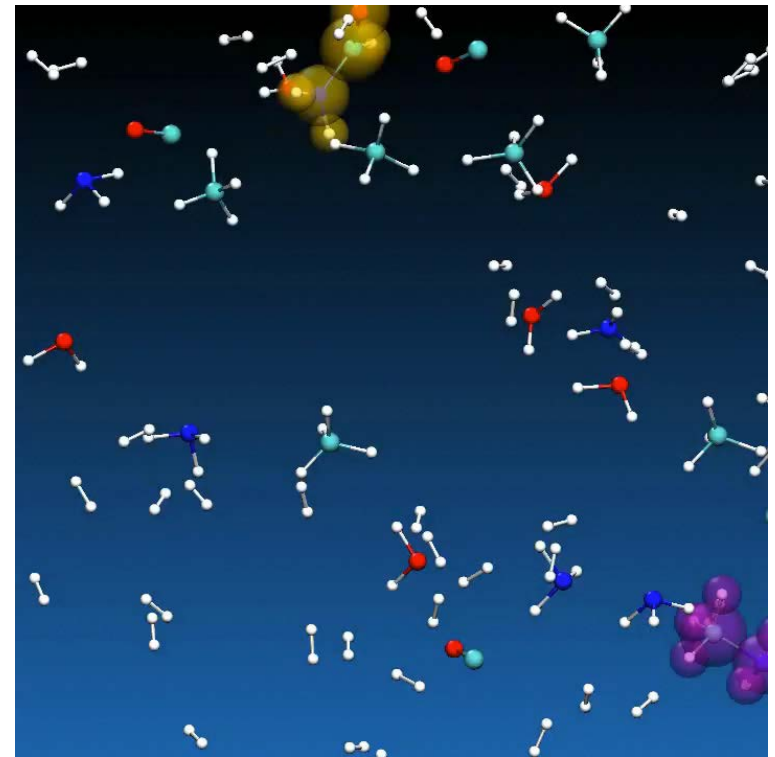
- Computer simulations of molecules span a wide range of resolutions
- More detailed theories can describe complex phenomena and offer higher accuracy
- Less detailed theories allow for simulation of larger systems / longer timescales
- Empirical force fields are the method of choice in the simulation of biomolecules

Introduction: Molecular dynamics simulation

Molecular dynamics (MD) simulation combines concepts from quantum and statistical mechanics.



A Markov state model built using data from several thousand protein folding simulations.



A reactive *ab initio* MD simulation of early Earth atmosphere, with reaction products identified and highlighted

Introduction: Molecular dynamics simulation

The fundamental equation of classical molecular dynamics is Newton's second law.

The force is given by the negative gradient of the potential energy.

$$\mathbf{F} = -\nabla V(\mathbf{r})$$

Knowing the force allows us to accelerate the atoms in the direction of the force.

$$\mathbf{F} = m\mathbf{a}$$

Key approximations:

- Born-Oppenheimer approximation; no transitions between electronic states
- Approximate potential energy surface, either from quantum chemistry or from the force field
- Classical mechanics! Nuclei are quantum particles in reality, but this is ignored.

Outline

Introduction

- Physical chemistry; the role of theory and computation
- An overview of computer simulations of molecules
- Molecular dynamics within the classical approximation

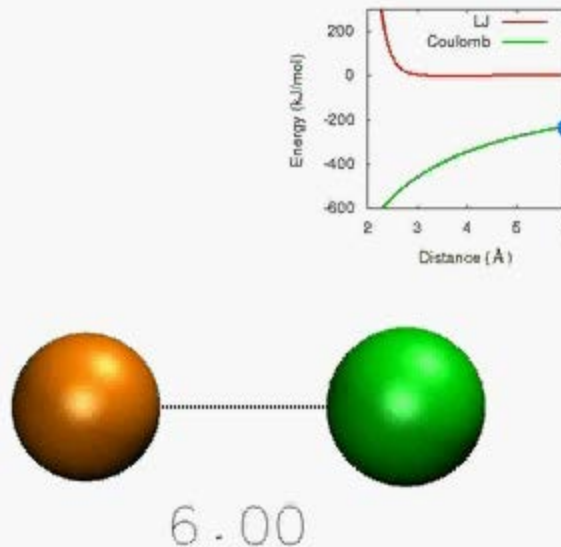
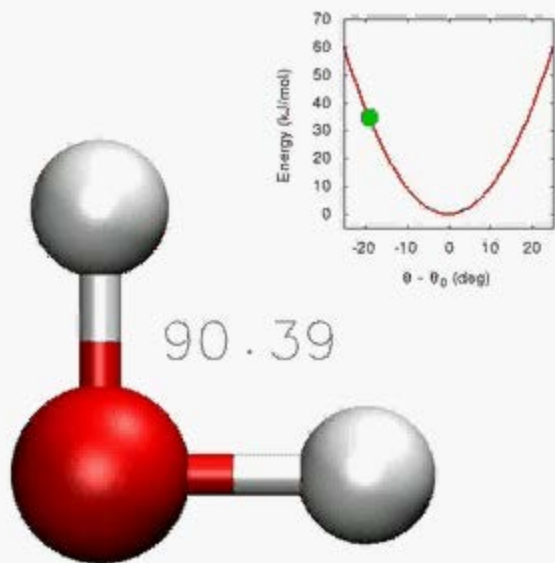
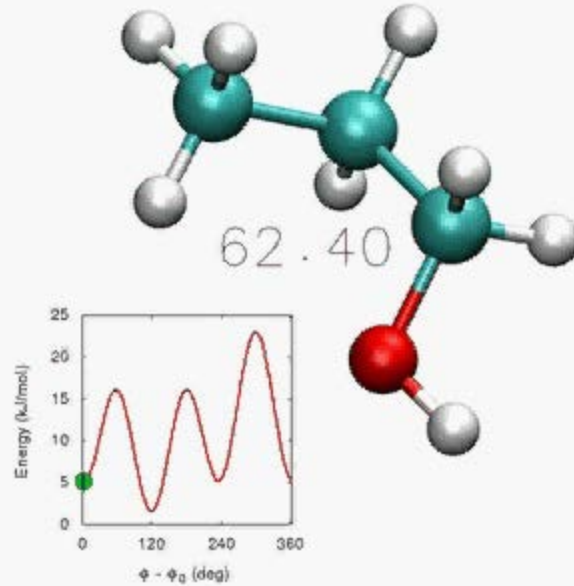
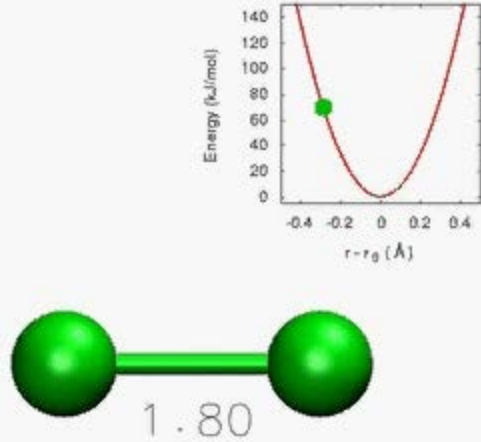
Methods

- Force fields
- Integrating the equations of motion
- Sampling from thermodynamic ensembles

Applications

- Protein folding structure and mechanism
- Conformational change and binding free energies
- Properties of condensed phase matter (water)

Force Fields



- **Force fields** are built from *functional forms* and empirical *parameters*
- Interactions include bonded pairwise, 3-body, and 4-body interactions...
- ... as well as non-bonded pairwise interactions
- Simulation accuracy depends critically on choice of parameters

Force Field Parameterization

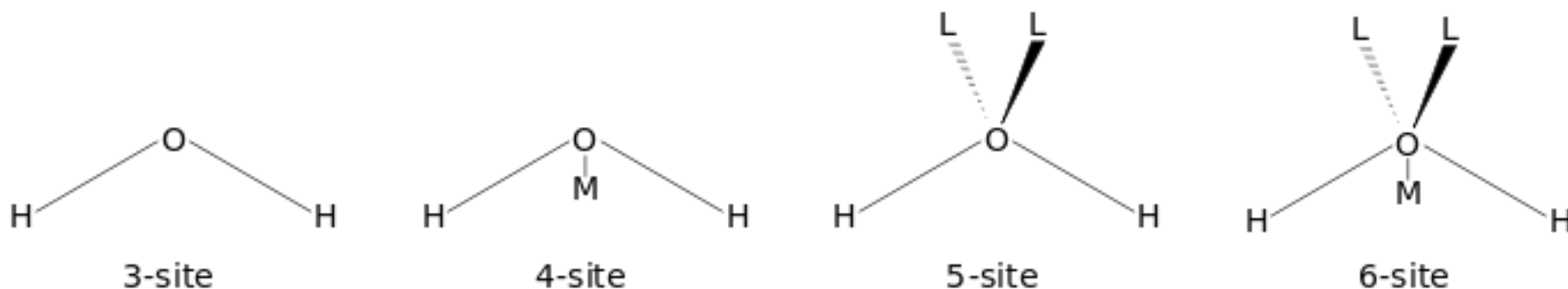
Force fields are parameterized to compensate for their simplified description of reality.

Most models have incomplete physics:

- Fixed point charges (no electronic polarization)
- Classical mechanics (no isotope effects)
- Fixed bond topology (no chemistry)

However, much can be recovered through parameterization:

- Increase the partial charges to recover polarization effects
- Tune vdW parameters to recover the experimental density
- In many cases, force fields exceed the accuracy of quantum methods!



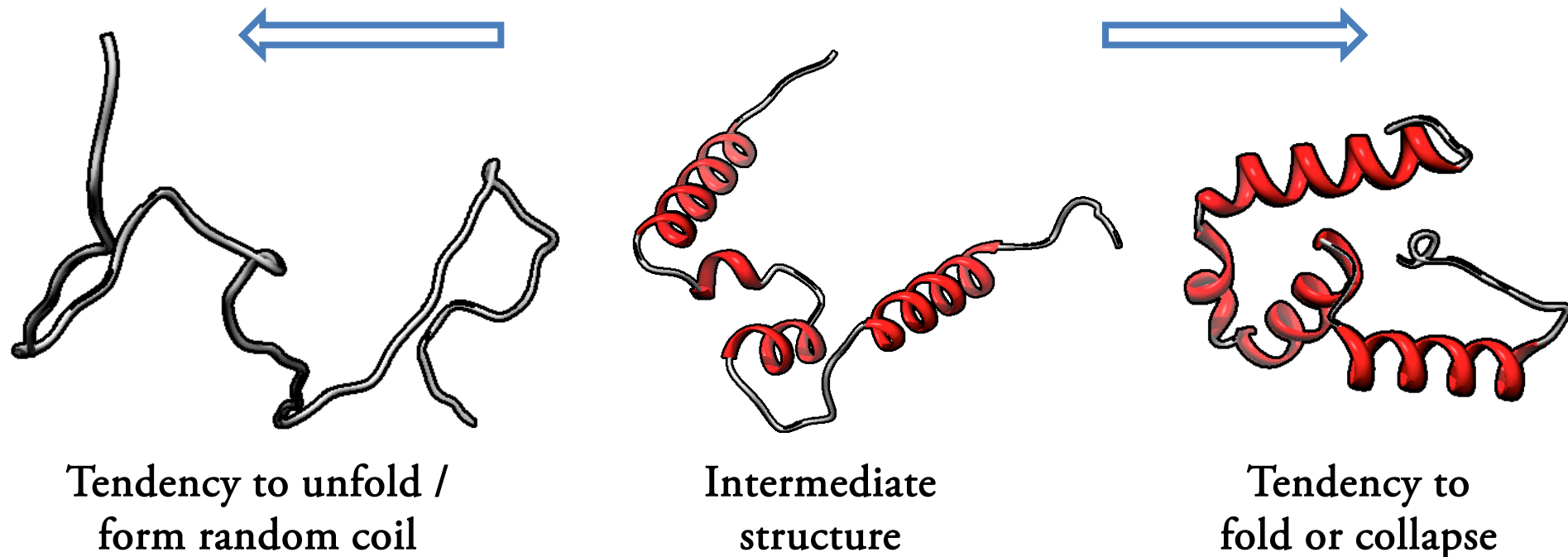
A careful balance of solute and solvent

The interactions in a force field must be carefully balanced to ensure correct behavior.

One must be careful to avoid force field bias in the simulation results...

Solvent-protein interactions *strong*
Protein-protein interactions *weak*

Solvent-protein interactions *weak*
Protein-protein interactions *strong*

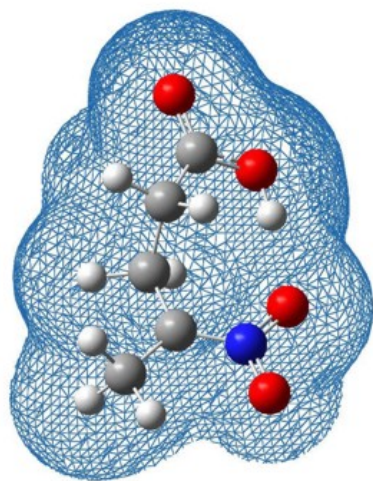


Solvent models

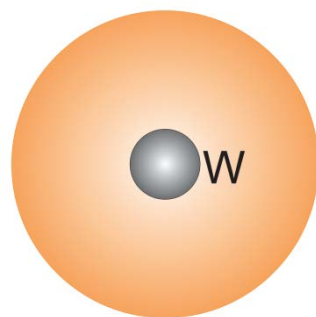
Solvent models span many levels of detail.

Choose a compatible solvent model for your solute force field.

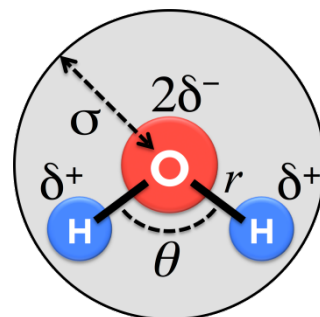
Less physical detail, accuracy
Lower computational cost



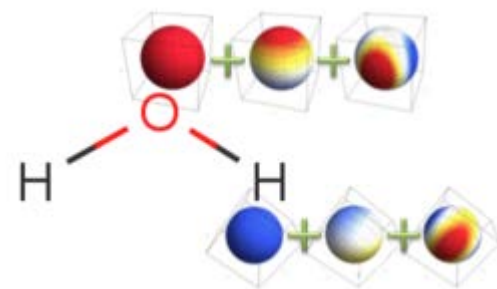
Implicit solvent;
no solvent particles
(GBSA)



Coarse grained
water model
(MARTINI)



All-atom
water model
(TIP3P, SPC/E)



Multipoles and
polarization
(AMOEBA)

Integrating the equations of motion

MD simulation requires accurate numerical integration of Newton's equations.

Euler method (*unstable*)

$$x(t + \Delta t) = x(t) + v(t)\Delta t + O(\Delta t^2)$$

$$v(t + \Delta t) = v(t) + a(t)\Delta t + O(\Delta t^2)$$

Velocity Verlet method

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2 + O(\Delta t^4)$$

$$v(t + \Delta t) = v(t) + \frac{1}{2}(a(t) + a(t + \Delta t))\Delta t + O(\Delta t^2)$$

Beeman predictor-corrector method

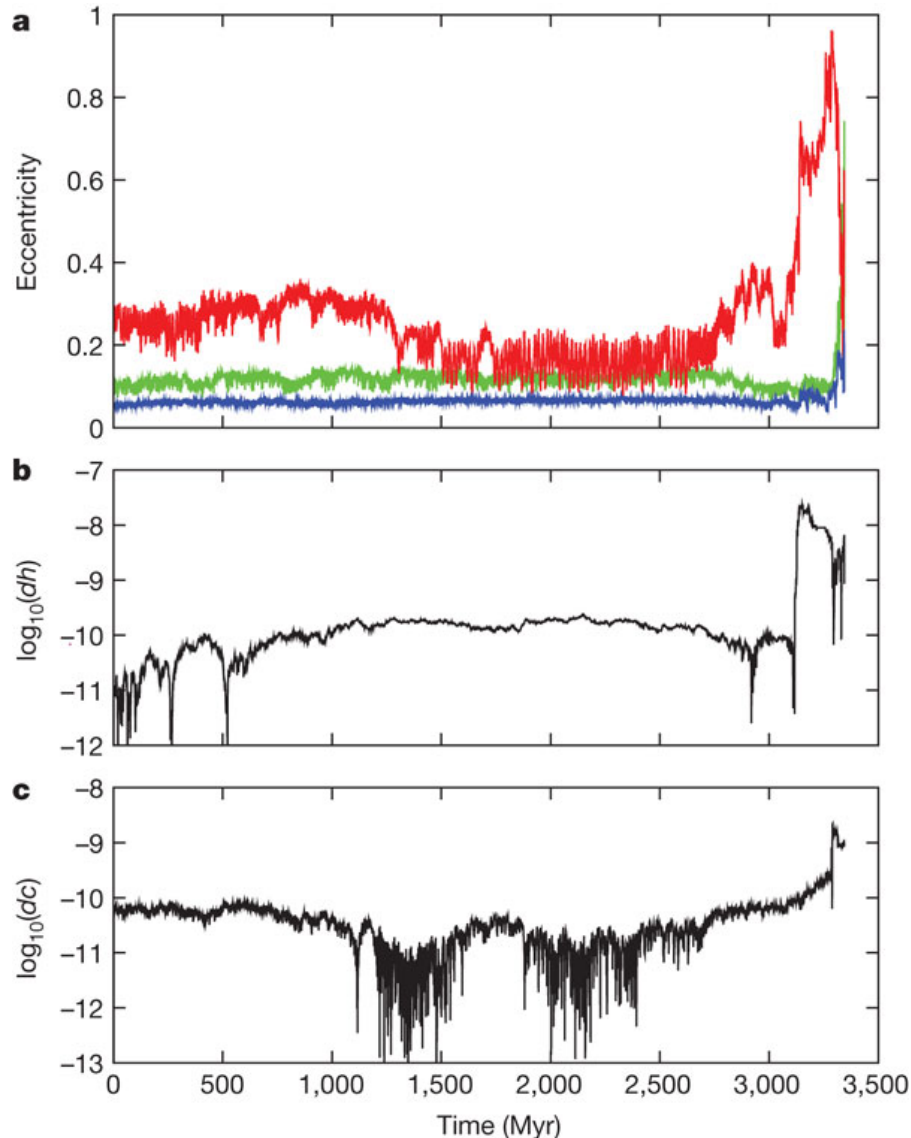
$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{6}(4a(t) - a(t - \Delta t))\Delta t^2 + O(\Delta t^4)$$

$$v(t + \Delta t) = v(t) + \frac{1}{6}(2a(t + \Delta t) + 5a(t) - a(t - \Delta t))\Delta t + O(\Delta t^3)$$

- **Integrators** are algorithms that accelerate the atoms in the direction of the force.
- More sophisticated algorithms include higher order terms for better accuracy
- Time step is limited by fast degrees of freedom (bond vibrations)
- MD simulations are *chaotic*; small differences in the initial conditions quickly lead to very different trajectories

A brief distraction...

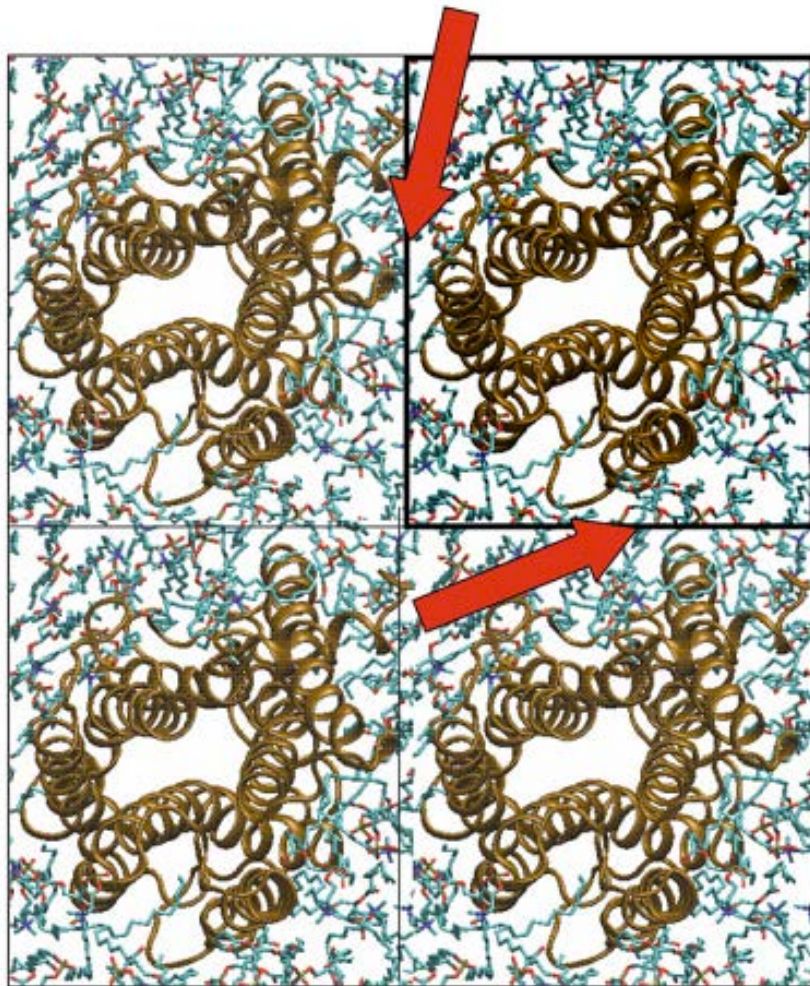
Example of collisional trajectory
for Mars and the Earth.



- Numerical algorithms for integrating Newton's equations are applied across many fields, including gravitational simulations in astrophysics
- Chaos exists for simple systems with only a few interacting bodies
- A small change in the initial conditions (0.15 mm) results in Venus and/or Mars eventually colliding with the Earth in 3 billion years

Boundary conditions

Boundary conditions allow us to simulate condensed phase systems with a finite number of particles.



- van der Waals interactions are typically treated using finite distance cutoffs.
- Ewald summation treats long range electrostatics accurately and efficiently using real space and reciprocal space summations

Important considerations!

- Choose large enough simulation cell to avoid close contact between periodic images
- Make sure the simulation cell is neutralized using counter-ions (otherwise Ewald is unphysical)

Statistical mechanical ensembles

Statistical mechanical ensembles allow our simulation to exchange energy with an external environment.

Ensemble menu: Choose one from each row	
Particle number N	Chemical potential μ
Volume V	Pressure P
Energy E	Temperature T

$$P_{NVT}(\mathbf{r}) \propto e^{-\frac{E(\mathbf{r})}{k_b T}}$$

Probability of a microstate
in the *canonical* (NVT) ensemble.

- An *ensemble* represents all of the microstates (i.e. geometries) that are accessible to the simulation, and provides the probability of each microstate.
- An ideal MD simulation conserves the total energy and entropy, and samples the *microcanonical* (NVE) ensemble.
- More realistic systems may exchange energy, volume or particles with external reservoirs
- However, this could make the algorithms more **difficult**

Temperature and pressure control

Thermostats and barostats allow MD simulations to sample different thermodynamic ensembles.

Kinetic energy / temperature relationship

$$T = \frac{2}{3Nk_B} \times KE$$

Berendsen thermostat (uses velocity scaling)

$$\frac{dT}{dt} = \frac{T - T_0}{\tau}$$

Andersen thermostat (velocity randomization)

$$P(\mathbf{p}) = \left(\frac{1}{2\pi m k_B T} \right)^{3/2} e^{-\frac{p^2}{2m k_B T}}$$

- Thermostat algorithms ensure that the temperature of our system (derived from kinetic energy) fluctuates around a *target* temperature that we set.
- The Berendsen thermostat achieves the target temperature but does not produce the correct canonical ensemble
- The Andersen thermostat produces the canonical ensemble by randomly resetting the momenta of particles (imitating random collisions)

Other integrators and algorithms

Langevin dynamics represents a different physical process; Monte Carlo is a pure sampling strategy.

Langevin equation

$$\mathbf{F} = m\mathbf{a} = -\nabla V(\mathbf{r}) - \gamma m\mathbf{v} + \sqrt{2\gamma M k_B T} \mathbf{R}(t)$$

Metropolis criterion for canonical ensemble

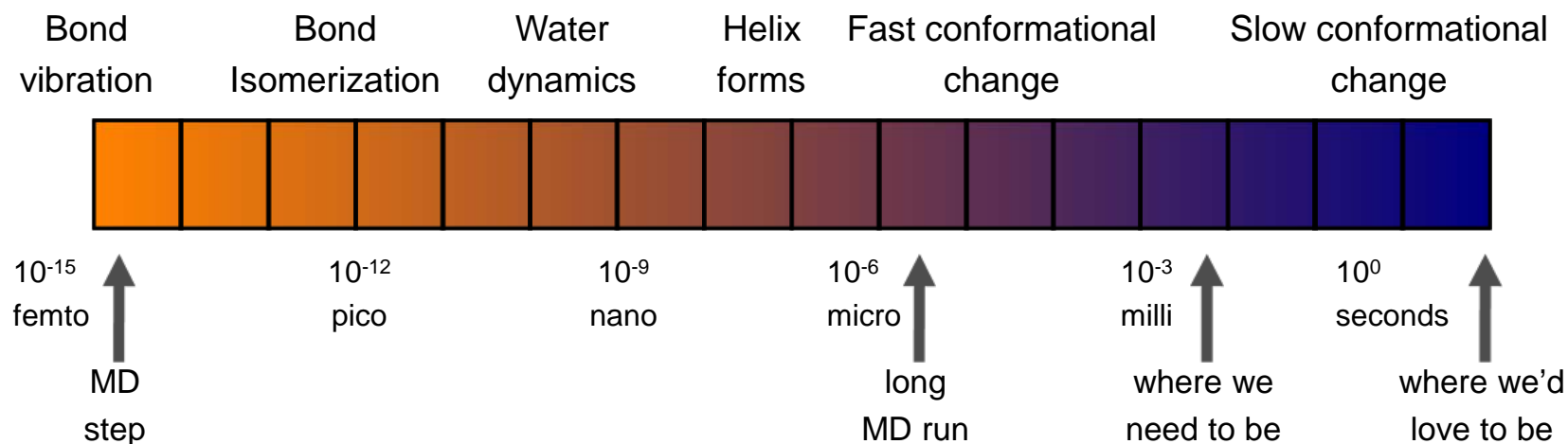
$$P(\mathbf{r}_n \rightarrow \mathbf{r}_{n+1}) = \begin{cases} \exp\left(\frac{E_n - E_{n+1}}{k_B T}\right), & \text{when } E_{n+1} > E_n \\ 1, & \text{otherwise.} \end{cases}$$

- Langevin dynamics (a.k.a. stochastic dynamics): Particles experience a *friction* force as well as a *random* force from particles “outside” the simulation

- Metropolis Monte Carlo method (a.k.a. Metropolis-Hastings): Generate any trial move you like, and accept / reject the move with a probability given by the *Metropolis criterion*.

Methodological challenges

The main challenges in molecular dynamics methods are simulation timescale and accuracy.



Statistical Mechanics: Efficiently sample the correct thermodynamic ensemble

Algorithms and Computer Science: Make our simulations run faster by designing faster algorithms and taking maximum advantage of current hardware

Force Fields: Achieve higher accuracy without unduly increasing the complexity of the calculation, using better parameterization methods (my work!)

Data Analysis: Draw scientific conclusions from huge volumes of data

Outline

Introduction

- Physical chemistry; the role of theory and computation
- An overview of computer simulations of molecules
- Molecular dynamics within the classical approximation

Methods

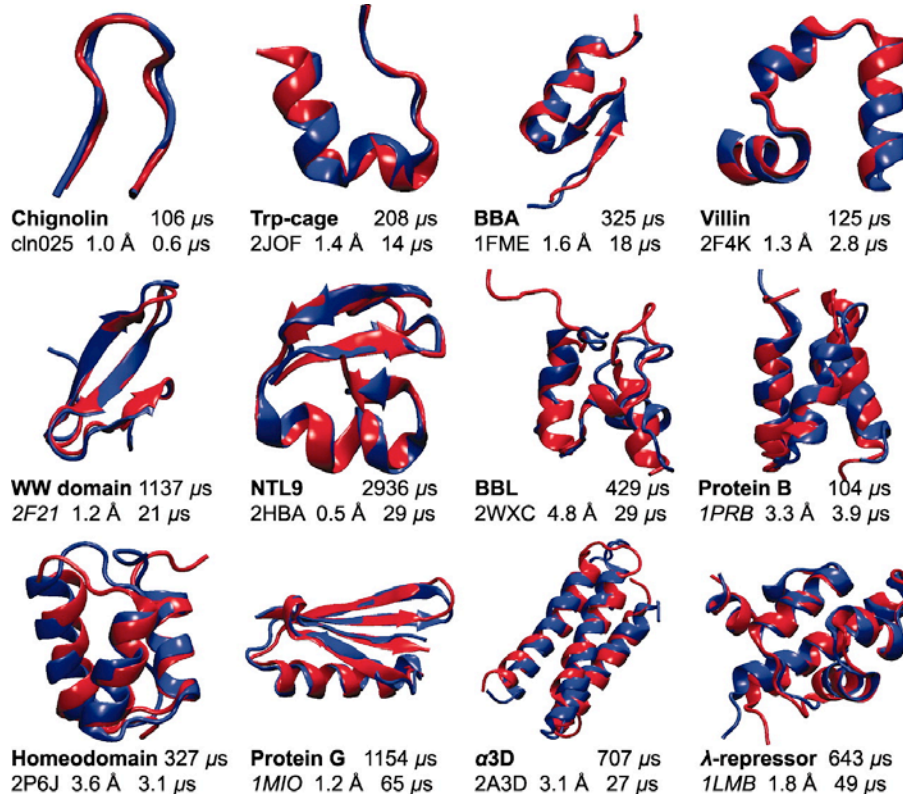
- Force fields
- Integrating the equations of motion
- Sampling from thermodynamic ensembles

Applications

- Protein folding structure and mechanism
- Conformational change and binding free energies
- Properties of condensed phase matter (water)

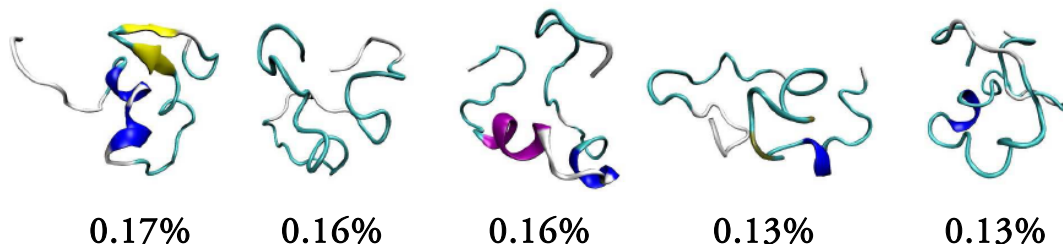
Protein Folding: Structure Prediction

A collection of fast-folding proteins



- Can we accurately reproduce / predict the structures of experimentally crystallized proteins?
- In many cases, protein folding simulations work amazingly well
- Folding studies help us understand intrinsically disordered proteins, which may be relevant in neurodegenerative disease (e.g. Alzheimer's, Parkinson's, Huntington's)

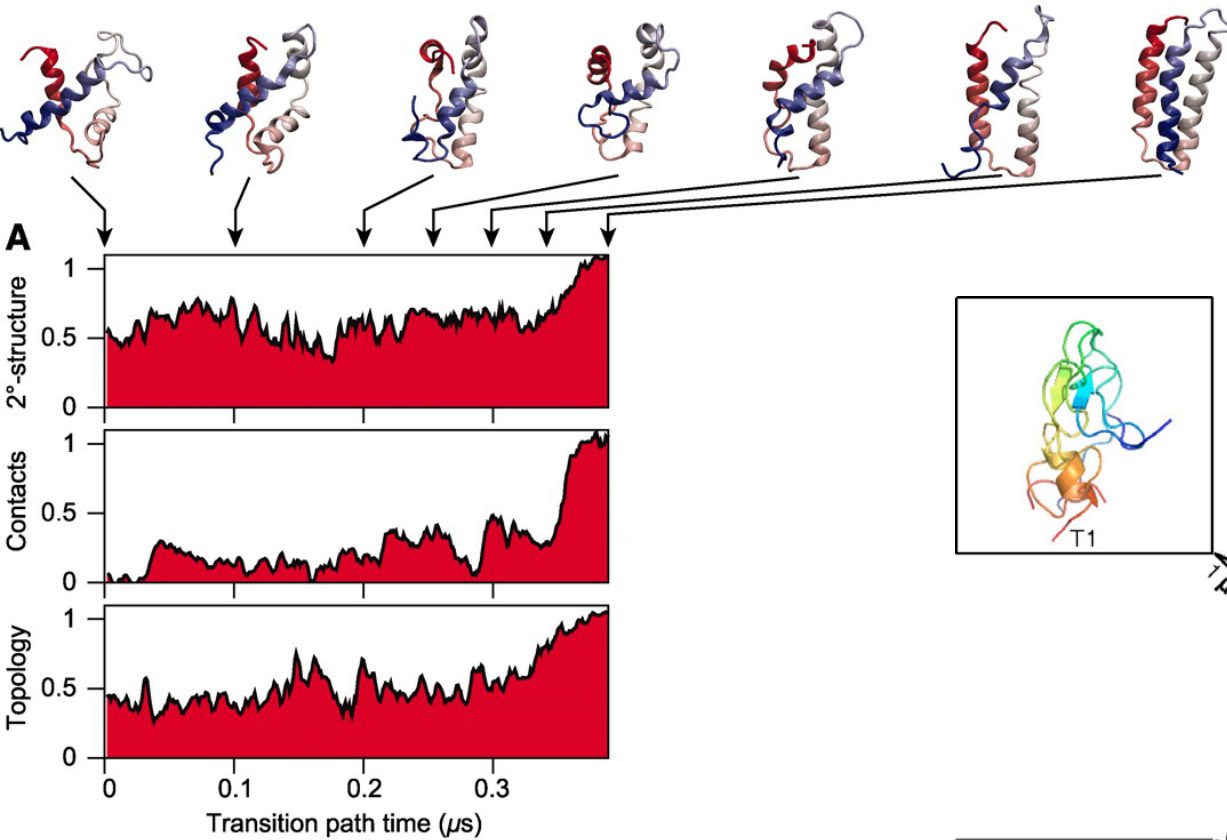
Amyloid beta peptide (intrinsically disordered)



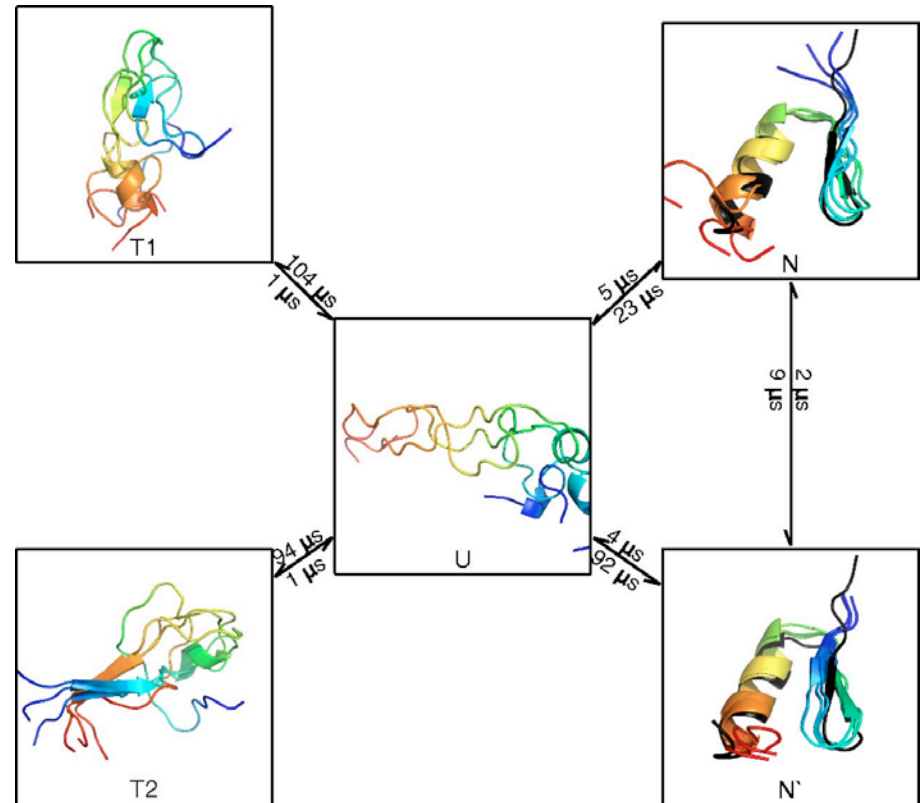
Lindorff-Larsen et al., *Science* 2011

Y-S. Lin and V. S. Pande, *Biophys. J.* 2012

Protein Folding Mechanism and Kinetics



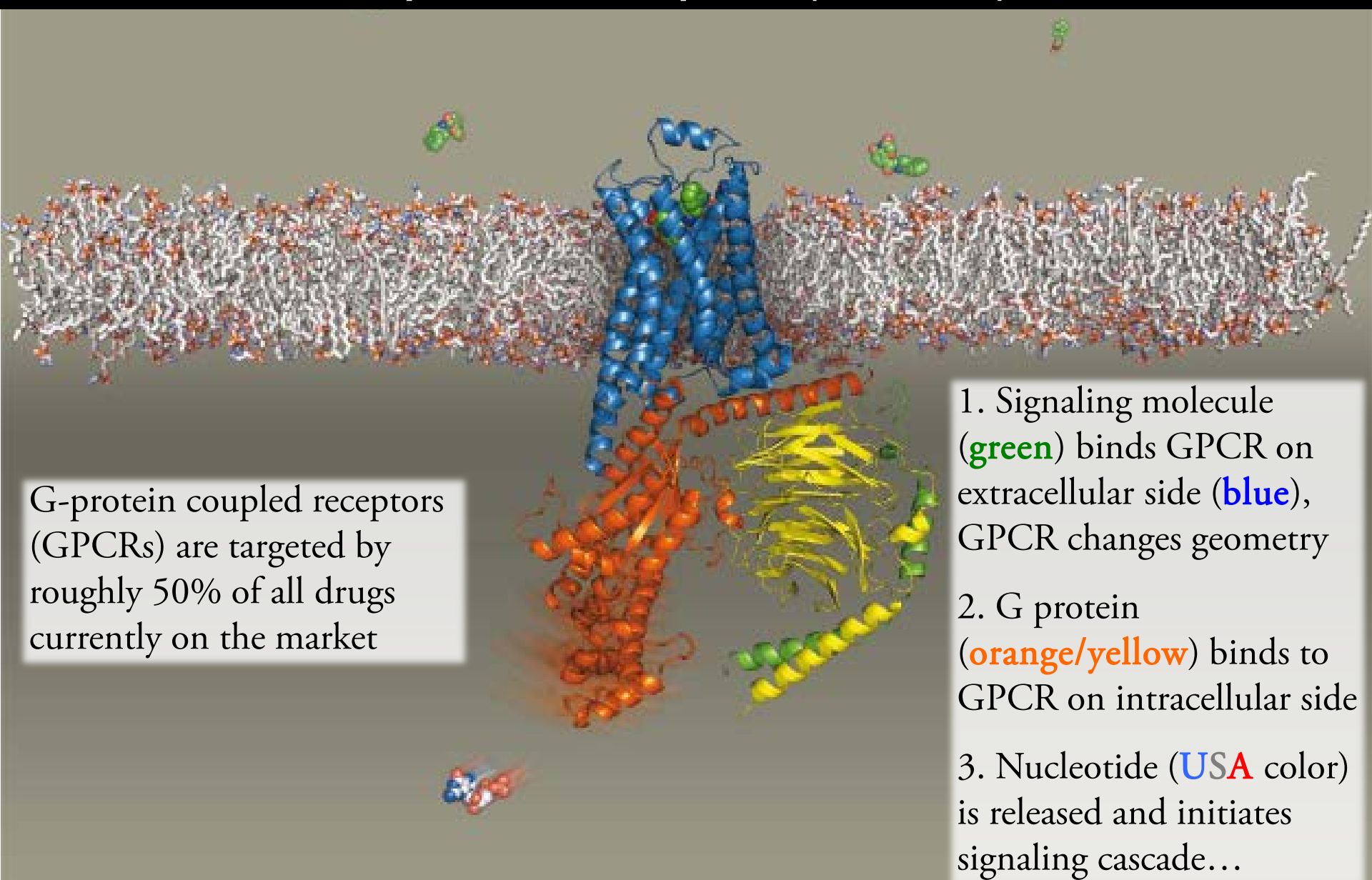
What are the pathways of protein folding? Is there a single preferred pathway, or are there many pathways?



- DE Shaw approach (above): Very long simulation trajectories ($> 10 \text{ ms}$) using special *Anton* supercomputer

- Pande Group approach (right): Synthesize many short simulations (e.g. from F@H) into a model with discrete states and rates (MSM)

G-Protein Coupled Receptor (GPCR)

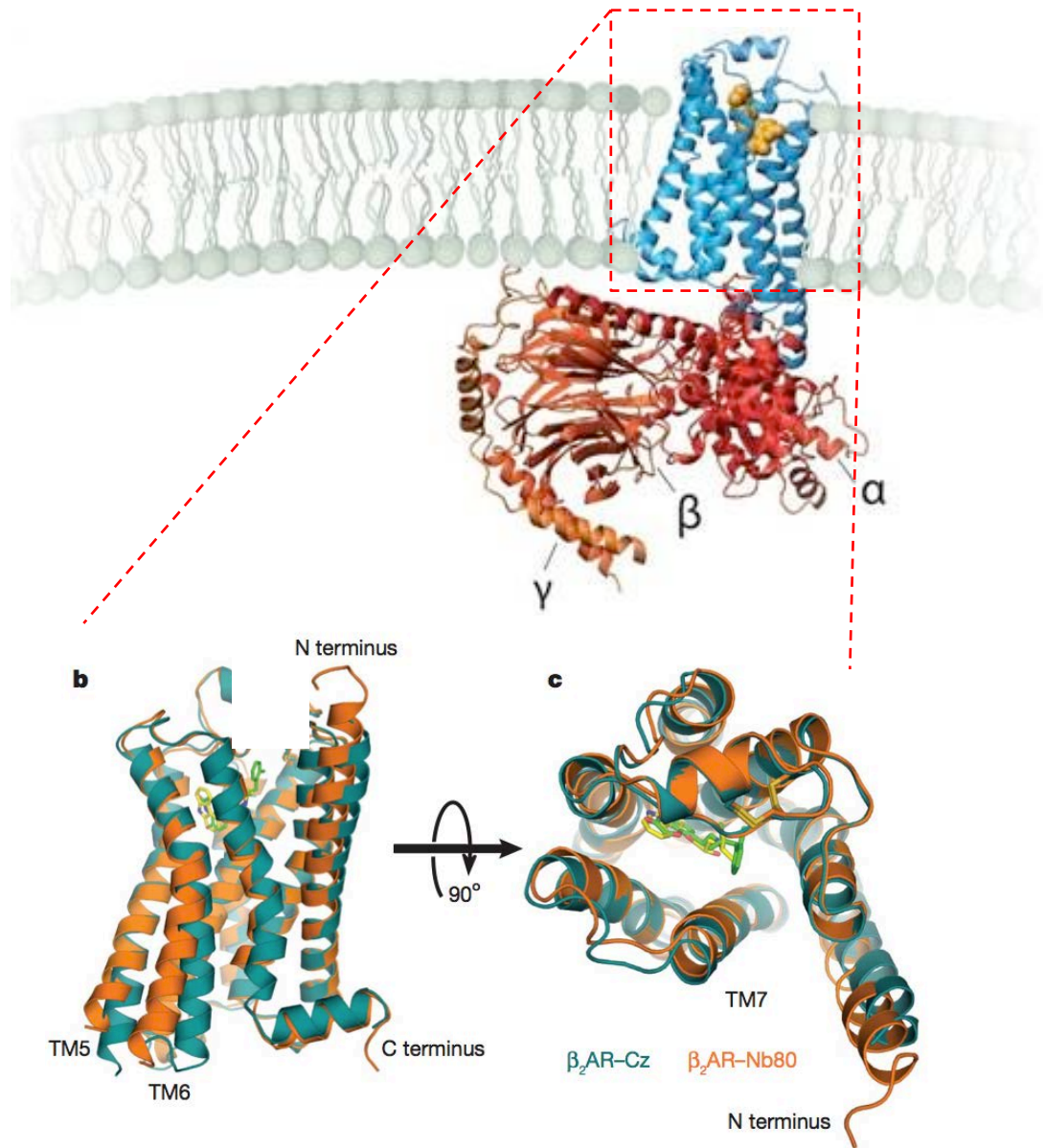


G-Protein Coupled Receptor (GPCR)

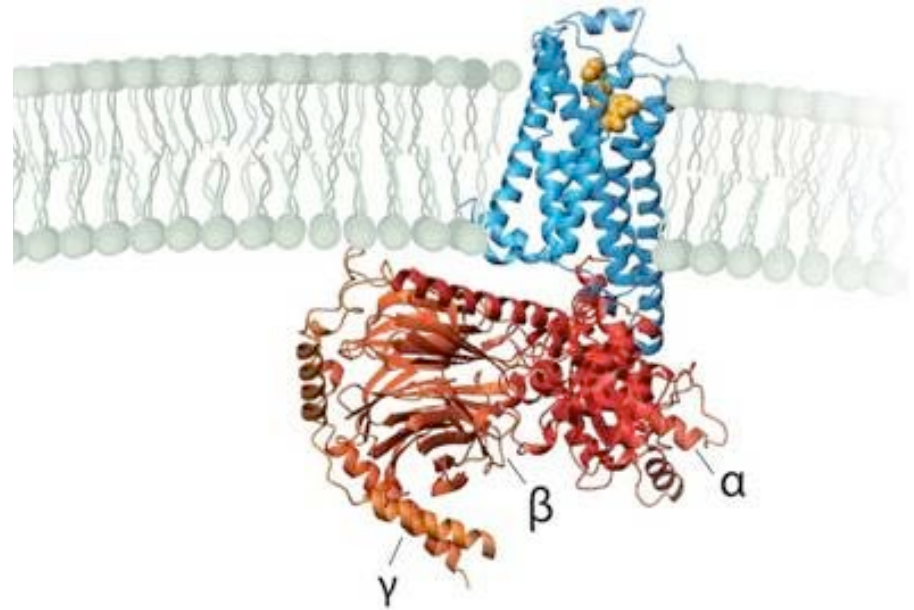
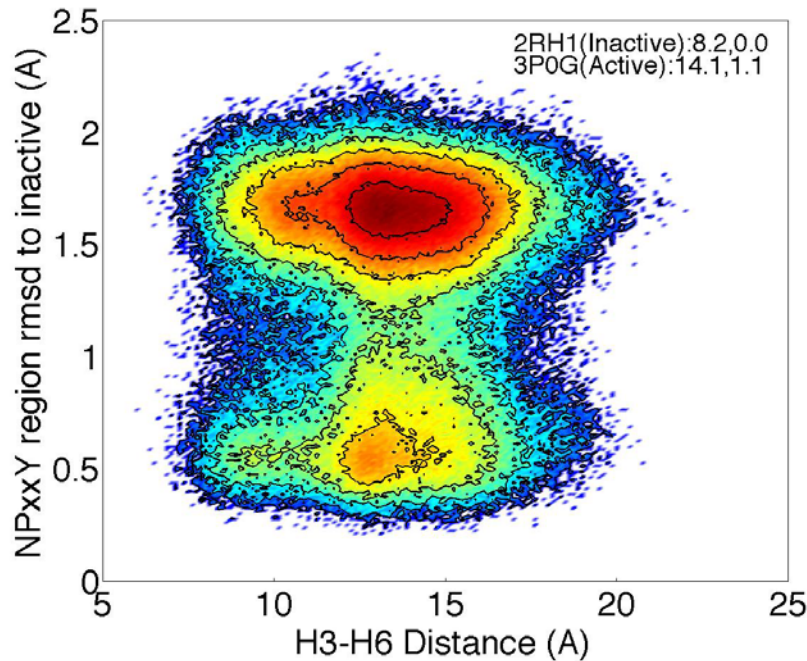


2012 Nobel Prize in Chemistry First Crystal Structure of a GPCR

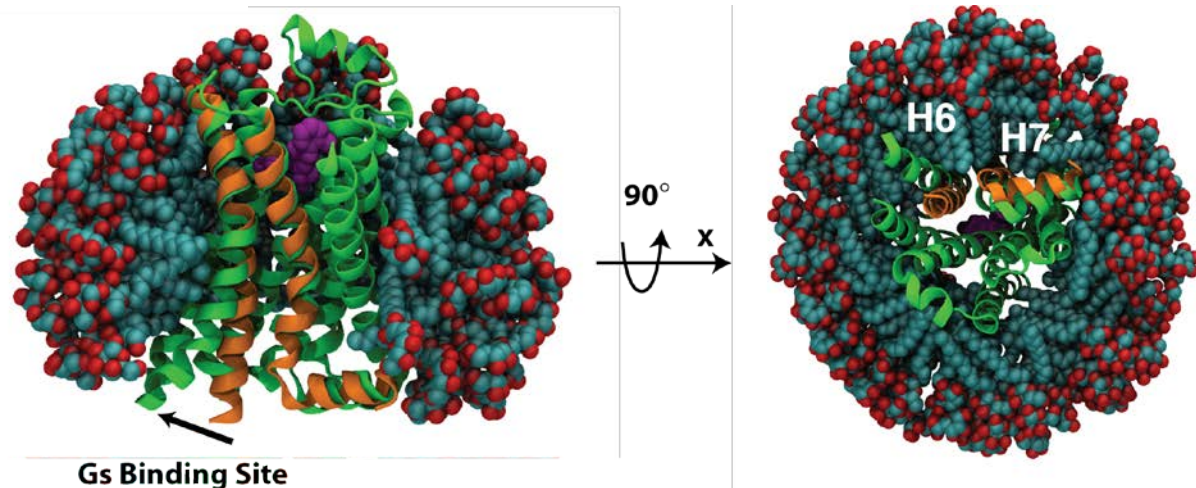
- For some GPCRs (e.g. β_2 AR) both the agonist-bound active and inactive structures have been crystallized
- The conformational change of the receptor is very subtle!



GPCR Conformational change



- What is the activation mechanism of GPCRs?
- Investigate using large-scale MD simulations and Markov state models for analysis

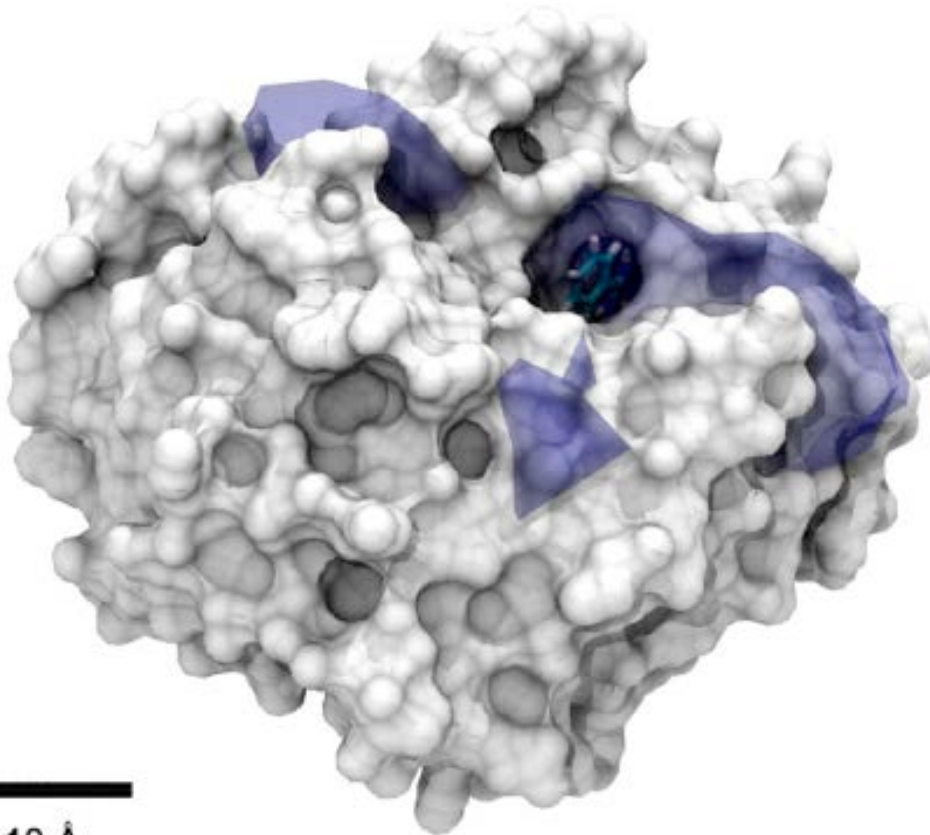


Protein-Ligand Binding Free Energies

Trypsin with benzamidine

187/495 100 ns simulations achieved binding pose within 2 Å of the crystal pose

Compute binding free energy within 1 kcal/mol of experiment

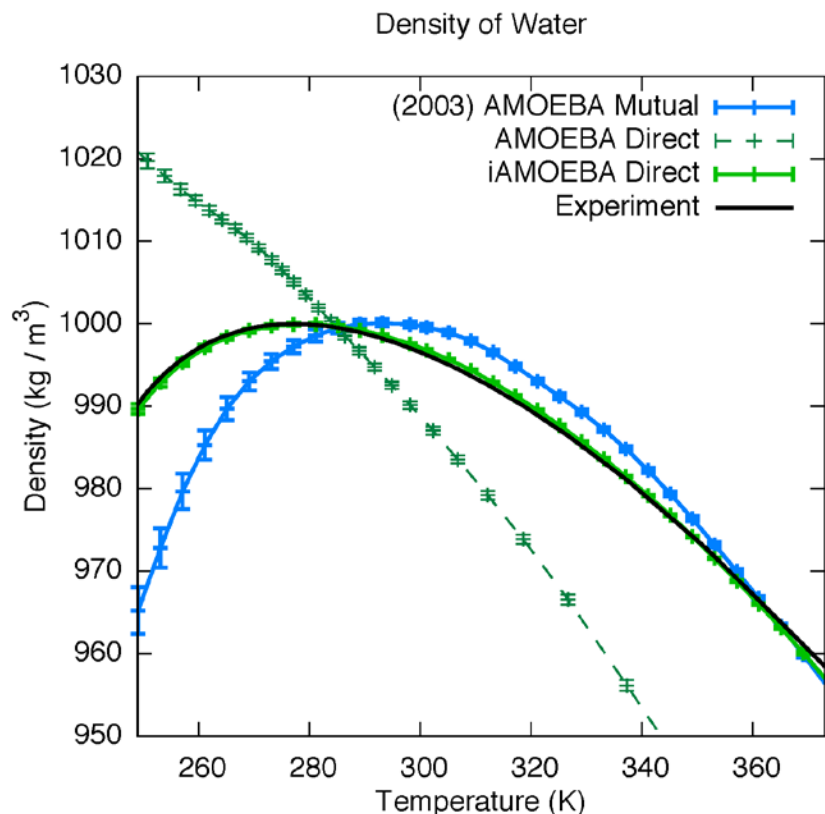


- How important is the role of dynamics in the process of protein / ligand binding?
- Can MD simulations improve the drug discovery process?
- Simulations involve the ligand molecule repeatedly visiting the binding pocket
- Lots of sampling required! Much lower throughput than “docking” but incorporates many more physical effects

Force field development for water

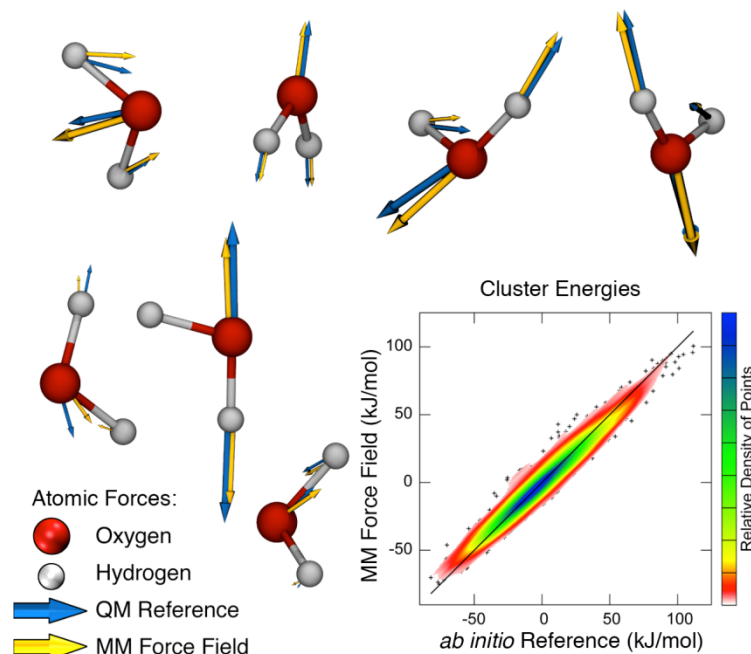
Simulated density of water vs. temperature.

Our model is shown in **light green**



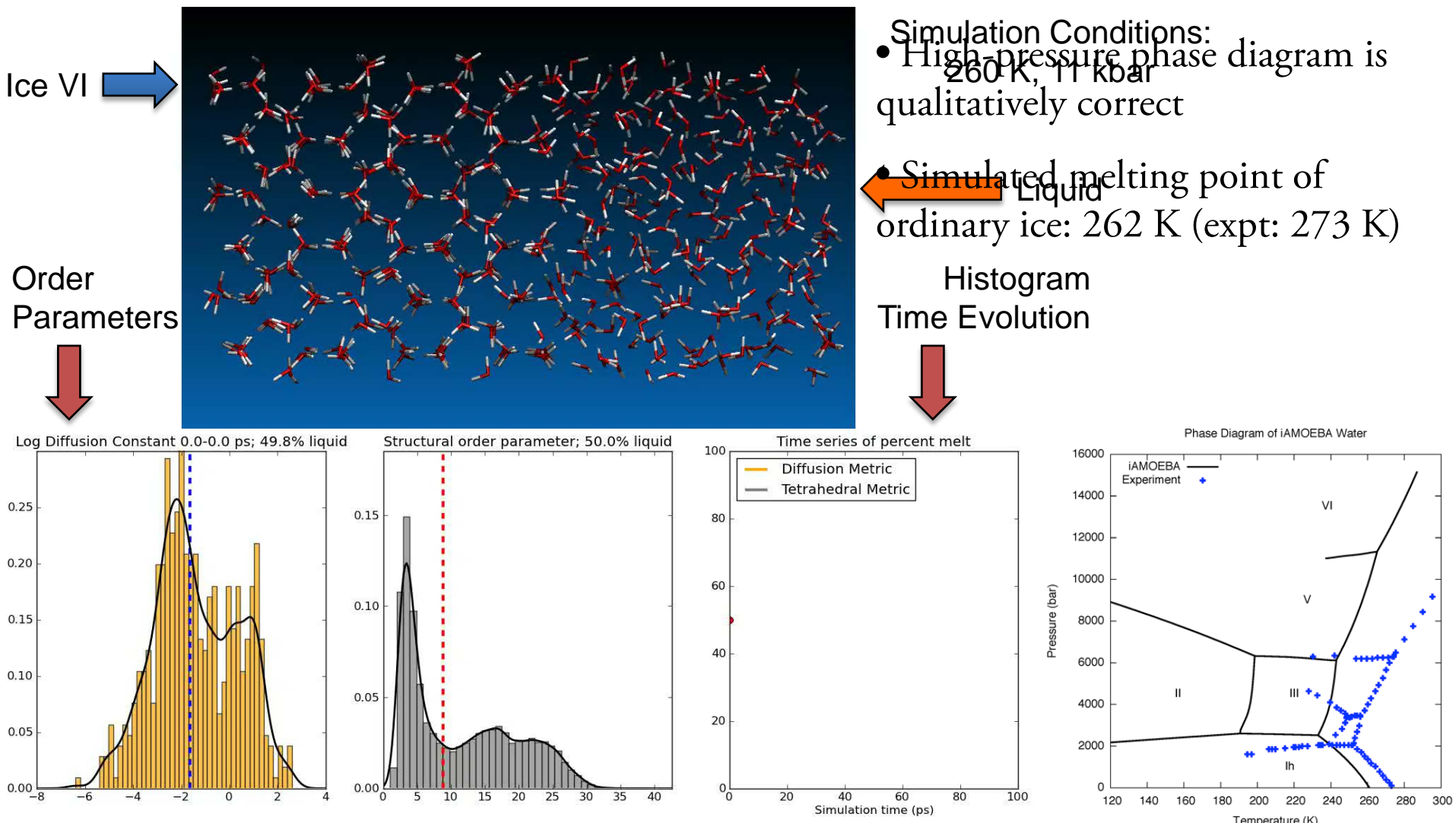
Right: Comparison of our model to high-level QM (MP2/aug-cc-pVTZ) energy and force calculations.

- An inexpensive polarizable water force field for biomolecular simulations, largely based on AMOEBA (2003)
- Model is fitted to combined experimental data and QM calculations
- Simpler and faster to calculate than AMOEBA, and much more accurate.



Force field development for water

Our model was validated by calculating the phase diagram and comparing to experiment.



Some classical molecular dynamics programs

Program	“Best feature”	License
AMBER	Large community, GPU acceleration Great modeling tools	Not free
CHARMM	Force fields for nucleic acids, carbohydrates and lipids	Not free
GROMACS	Fast, also humorous	Free, Open Source
TINKER	Easy to edit, AMOEBA force field	Free
OpenMM	Multi-platform compatibility (including GPUs), highly flexible	Free, Open Source
NAMD	Scales to 100,000+ cores, interface to VMD	Free
LAMMPS	Force fields for condensed matter physics / materials science	Free, Open Source

Recommended reading

S. Adcock and A. McCammon. “Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins.” *Chem. Rev.* 2006, **106**, 1589.

S. Rasmussen and B. Kobilka. “Crystal structure of the human $\beta 2$ adrenergic G-protein-coupled receptor.” *Nature* 2007, **454**, 383.

B. Cooke and S. Schmidler. “Preserving the Boltzmann ensemble in replica-exchange molecular dynamics.” *J. Chem. Phys.* 2008, **129**, 164112. (Contains good introductions to integrators, thermostats etc.)

C. Vega and J. L. F. Abascal. “Simulating water with rigid non-polarizable models: a general perspective.” *Phys. Chem. Chem. Phys.* 2011, **13**, 19663. (All about water models.)

Two classic texts:

D. Frenkel and B. Smit, Understanding Molecular Simulation.

M. P. Allen and D. J. Tildesley, Computer Simulations of Liquids.

For the ambitious:

L. D. Landau and E. M. Lifshitz, Classical Mechanics 3rd ed.