Advanced Molecular Dynamics

Introduction

May 10, 2021

Who am I?

- I am a professor at Applied Physics
- Topics I work on:
    - Algorithms for (parallel) molecular simulations including GPU acceleration
    - Methods for (alchemical) free energy calculations
    - Enhanced sampling in protein simulations
    - Wettings (liquids at interfaces)
- Topics I have worked on:
    - Parametrization of atomistic & coarse-grained models, in particular ions in solvent

Who are you?

Tell us:

- Your name
- Your university + department/group
- Your PhD project
- What you expect to get from this course

Molecular Dynamics

Given ***N*** particles, masses and a potential ***V***

$$m_i \frac{d^2 \mathbf{x}_i}{d\, t^2} = -\nabla_i V(\mathbf{x}) \quad , \qquad i = 1, \ldots, N$$

"It's just Newton's equation of motion"

- How to integrate
- Quality of V(**x**) determines (nearly) everthing
- How to choose **x**(0) (and **v**(0))
- Simulating long enough

Aspects of Molecular Dynamics

- How to integrate equations of motion
- How to set up a force field
- How to set up initial conditions
- How long should I simulate?
- What output should I write?
- How should I analyze the results?
- Are my simulations reliable?
- How to improve sampling
- How to develop efficient algorithms

Tentative course schedule

- Monday 10:
  - Introduction to the course, MD and general considerations
  - Force fields & integration
  - MD integrator exercise
- Tuesday 11:
  - Non-bonded interactions
  - Ensembles
  - Water exercise
- Wednesday 12:
  - Complex models
  - Replica exchange (Mark Abraham)
  - Replica exchange exercise
- Monday 17:
  - Sampling
  - Free-energy calculations
  - Free-energy exercise
- Tuesday 18:
  - Systematic course graining
  - Course graining exercise
  - Hybrid CPU/GPU acceleration (Szílard Páll)

Course project

Project for the last course week:

- Decide on a small project of a week
- Should apply an advanced MD technique
- Ideally fits into your PhD project
- You can use any MD code you want
- Mail me a report

Focus on different aspects of MD

- Theoretical Physicist focusses on:
  - focus on sampling methodology

- Biologist focusses on:
  - setting up simulations
  - analysis

- Who makes the tools?
  - efficient simulation & analysis code
  - accurate force fields

When (not) to do simulations

- Simulations are:
  - tedious to set up
  - require a lot of computational power
  - reliability of results?

- Simulations provide:
  - dynamics in full atomistic detail (x,t)
  - information on nearly anything
  - can easily change (initial) conditions

Types of systems studied with MD

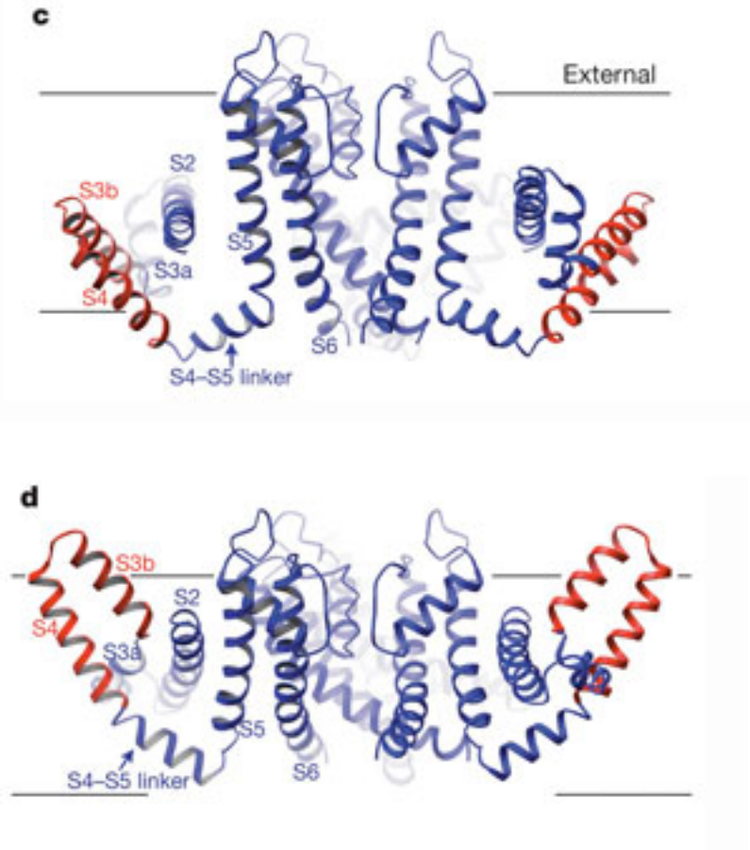
- Quantum mechanics (not in this course):
  - Electronic structure, reactions
  - Metals (QM or embedded atom pot.)
- Classical, mainly pair-wise potentials:
  - Lennard-Jones solid/liquid/gas
  - Water (solid/liquid/gas)
  - Polymers (structure, phase transitions)
  - Bio-molecules: proteins, DNA, RNA
  - ....

Experiments

- Thermodynamic measurements:
  - Phase diagrams, density, heat of vaporization, ...
- X-ray crystallography: atomistic detail, but static structures, very resource intensive
- Nuclear Magnetic Resonance: averaged atomistic detail, some averaged dynamics, very resource intensive, size limits
- Single molecule experiments: single molecule, but very few observables
- Cryo-electron microscopy: atomistic detail of large, static structures, small size limits

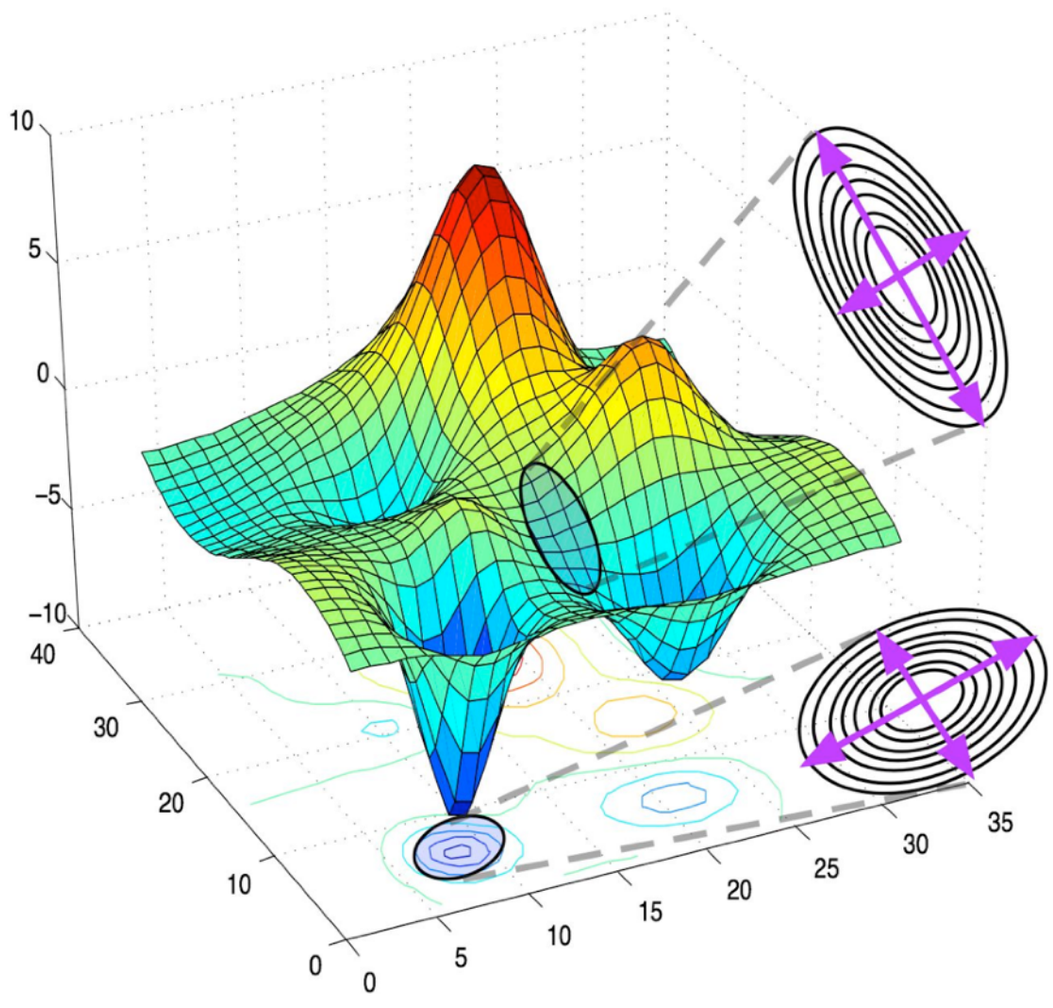
Measuring averages - an example

- Voltage-gated ion channel:
  - Voltage sensing through movement of charged ‘paddle’.
- How to investigate this with simulation?
  - First need to check stability: is there a force on the open structure?
    - Calculate *average* force
    - What is the average?



Sampling a landscape

- 3N-dimensional space
- Native structure is the free energy minimum
- Ideally, we would sample all of phase space exhaustively
- In practice we have to make do with the most populated parts



Chaotic systems

- Molecular dynamics is chaotic
- Is it even possible to simulate?
- Try to track a single particle in the air:
  - Extremely small differences in initial position/velocity will lead to large deviations later

(Thermo)dynamics

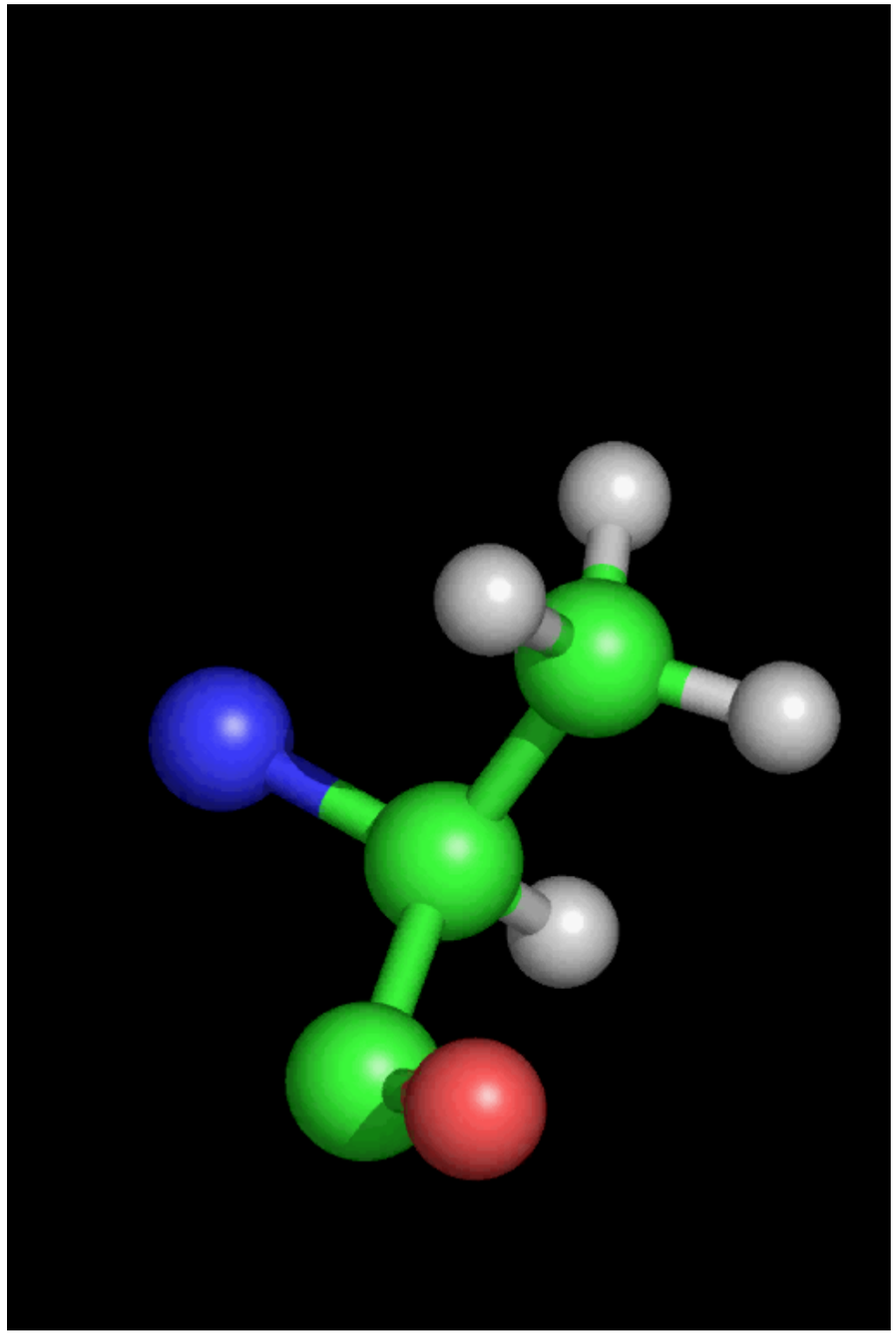
$$m_i \frac{d^2 \mathbf{x}_i}{d\, t^2} = -\nabla_i V(\mathbf{x}) \quad , \qquad i = 1, \ldots, N$$

- Molecular Dynamics
- But we are often not interested in dynamics, but rather thermodynamics or statistical mechanics

$$\rho(\mathbf{x}) \sim \exp(-V(\mathbf{x})/k_B T)$$

Simulations

- Monte Carlo simulation (later)
  - No time dependence
  - Requires intelligent moves
- Newton's equations of motion
  - Time dependence
  - Dynamical events
- Stochastic / Brownian dynamics
  - Add noise to forces



Molecular dynamics

- Try to calculate paths / trajectories of individual particles
- We know there are errors in these
- But if the errors do not systematically add up, the averages should still be OK (note: this is a very complex issue)
- How do we calculate motion?

Newton's equations

Leap-Frog integrator

- From the acceleration we can calculate how the velocity of the atom is changing:

$$v_i(t + \frac{\Delta t}{2}) = v_i(t - \frac{\Delta t}{2}) + a_i(t)\Delta t$$

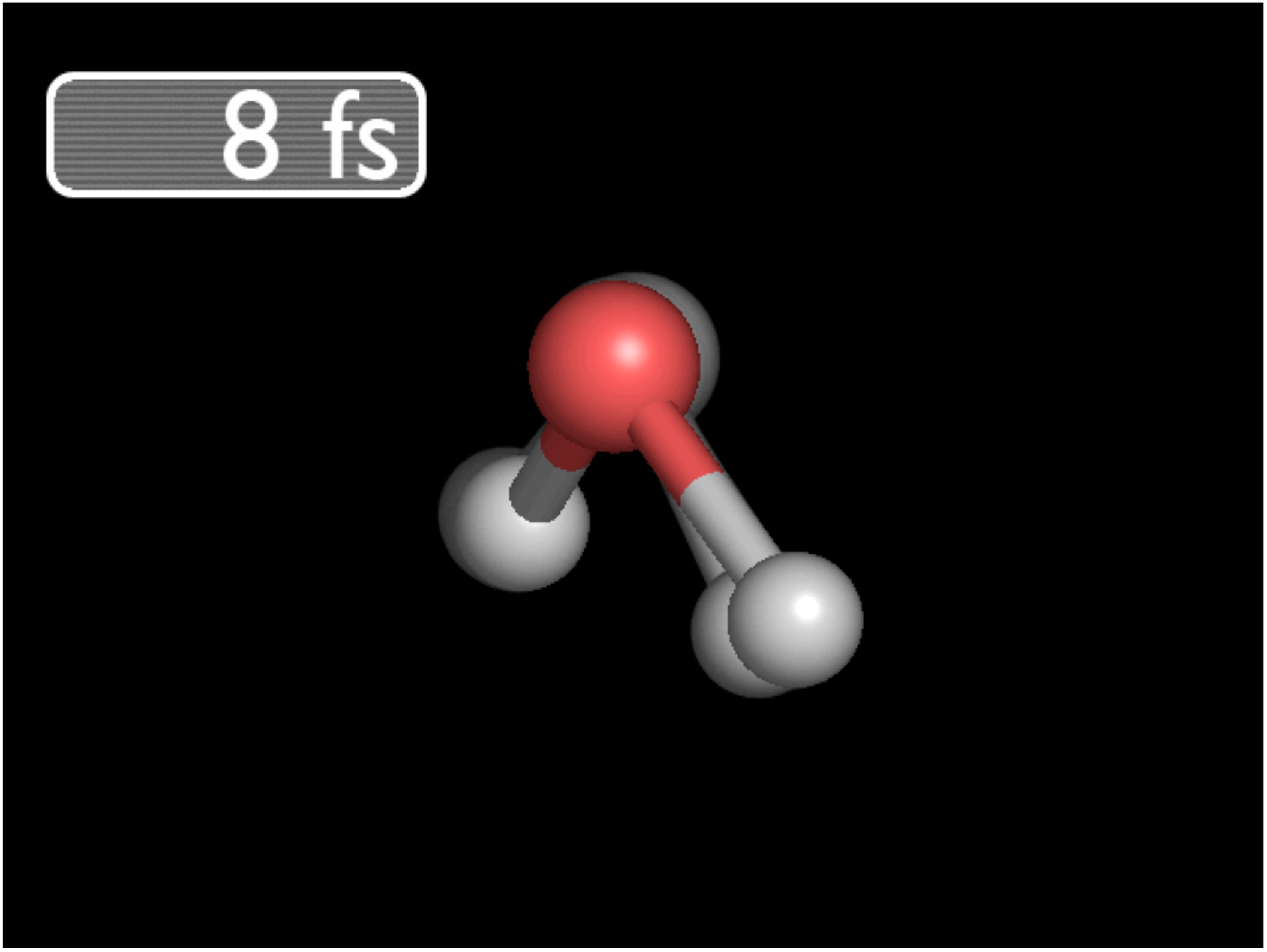
- From the velocity, we can calculate new coordinates for the atom:

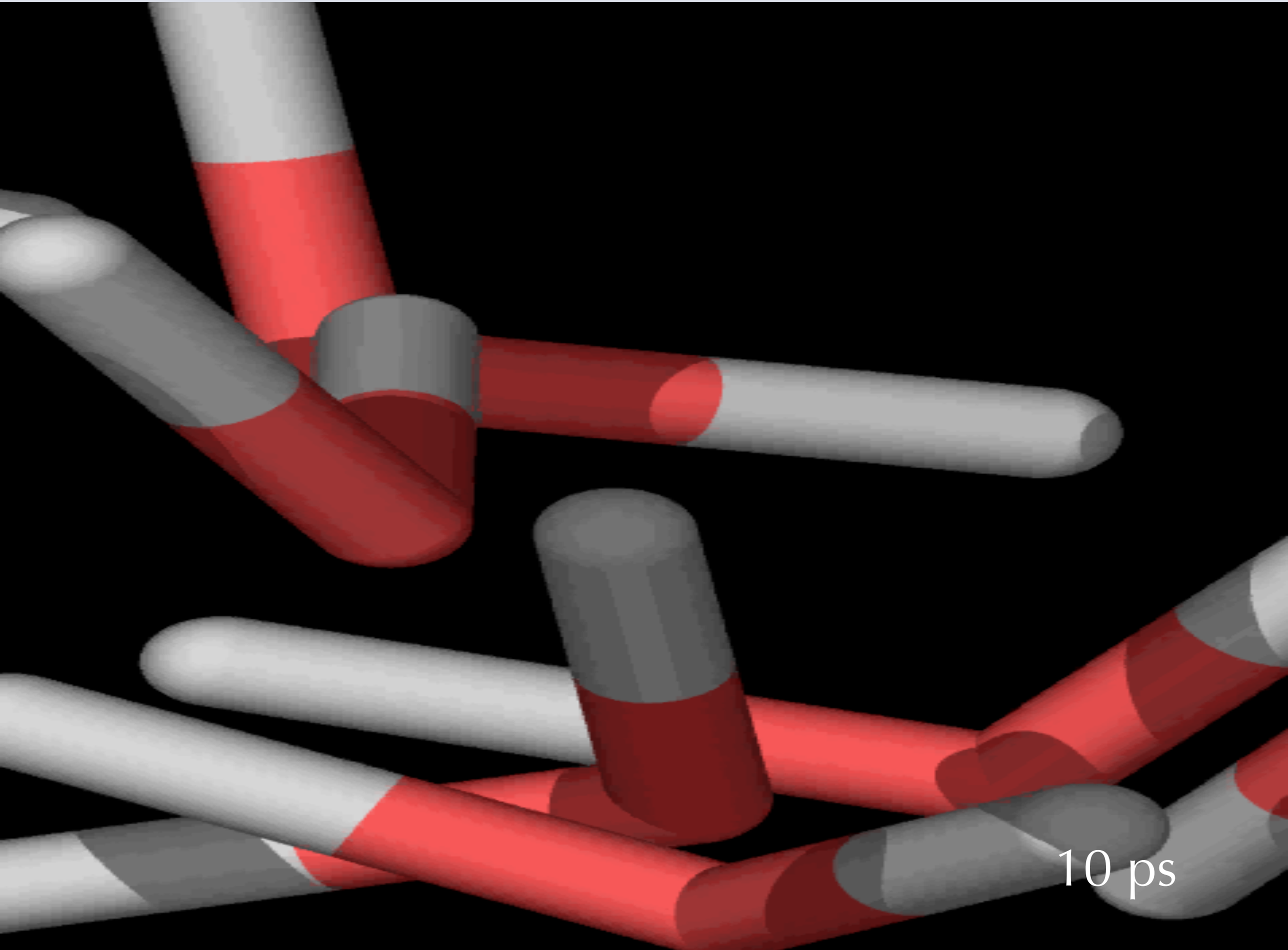
$$x_i(t + \Delta t) = x(t) + v_i(t + \frac{\Delta t}{2})\,\Delta t$$

What is a trajectory?

- NOT an accurate prediction of the motion of an individual particle
- Molecular dynamics is a chaotic process - differences grow exponentially
- But trajectories aren't random either?
- Shadow trajectory: Simulated path stays close to (some) real path for shorter times

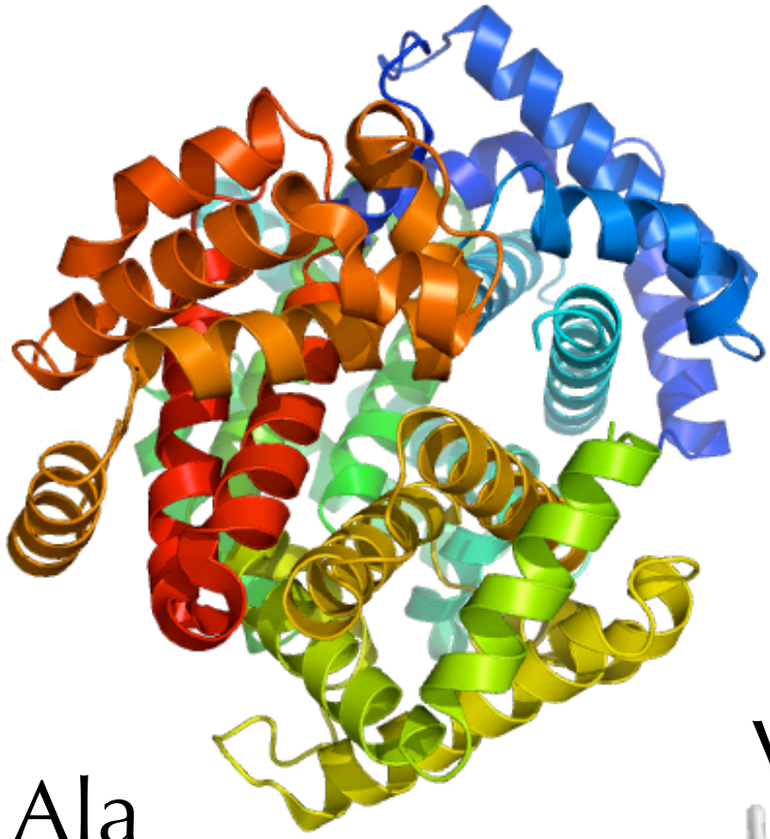
Taking a time step

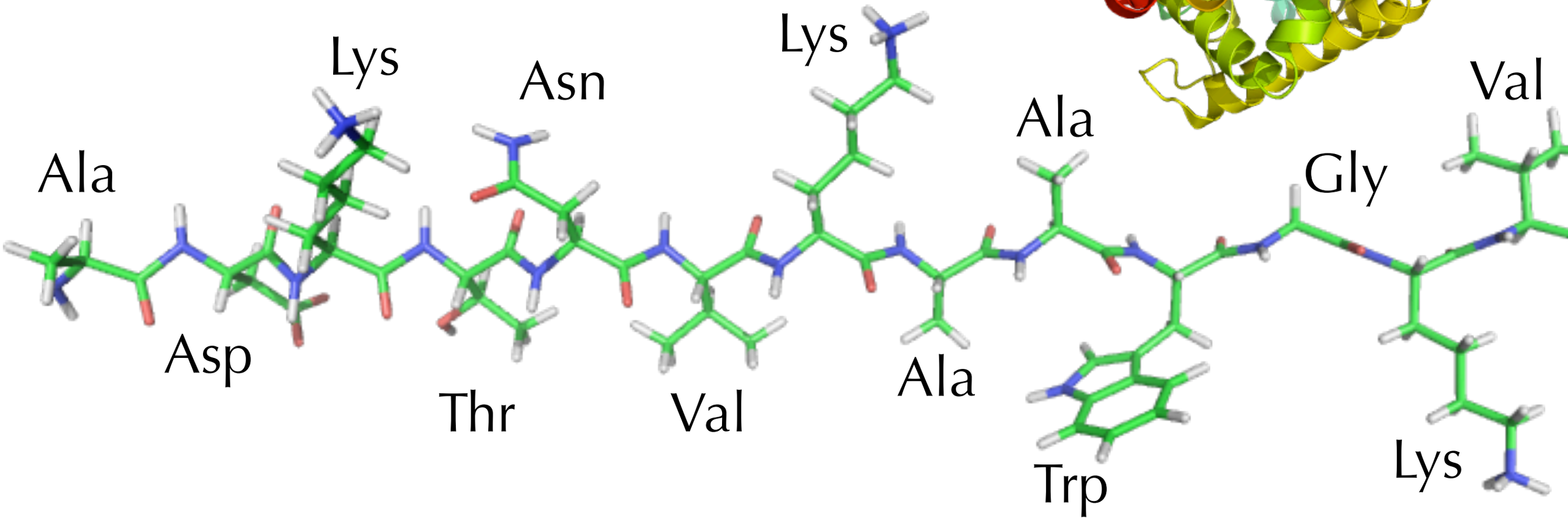



10 ps

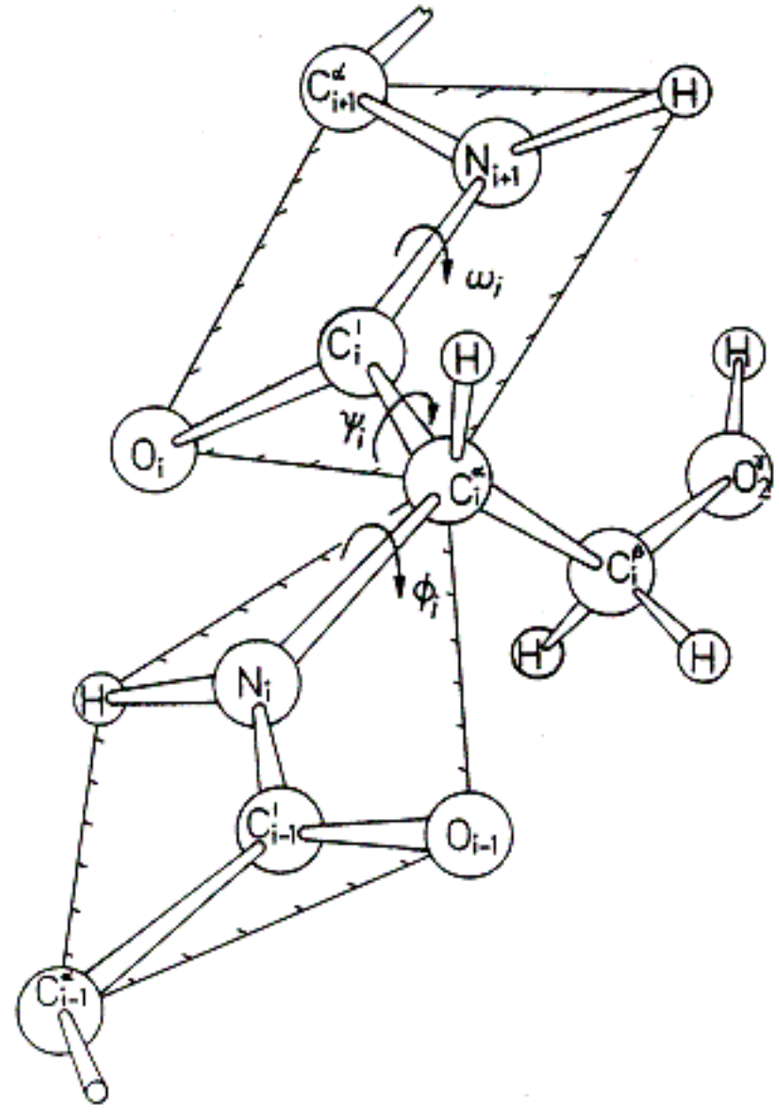
Proteins

Start of the amino acid sequence for chain A of Hemoglobin

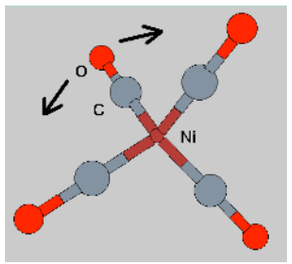


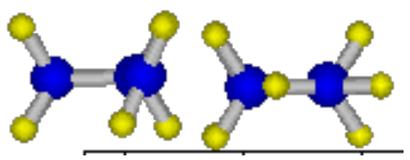


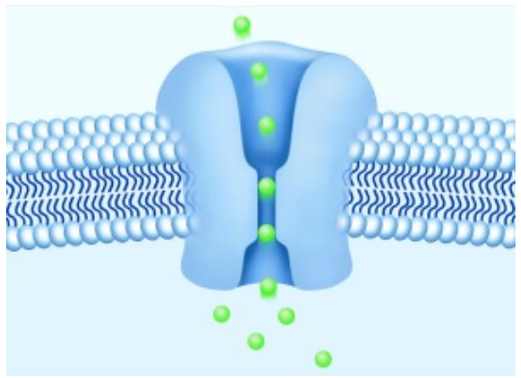
Example of motion in a chain

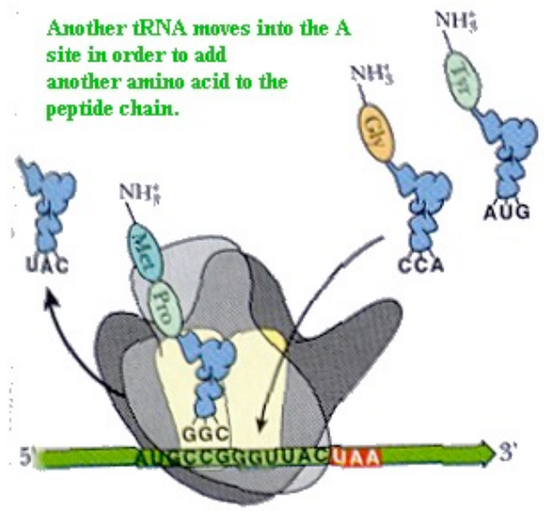


Time scales, again

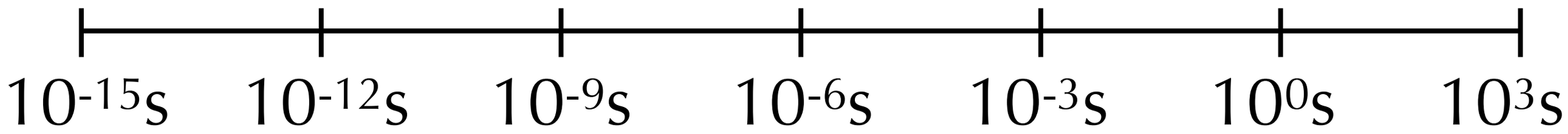








Biological Experiments

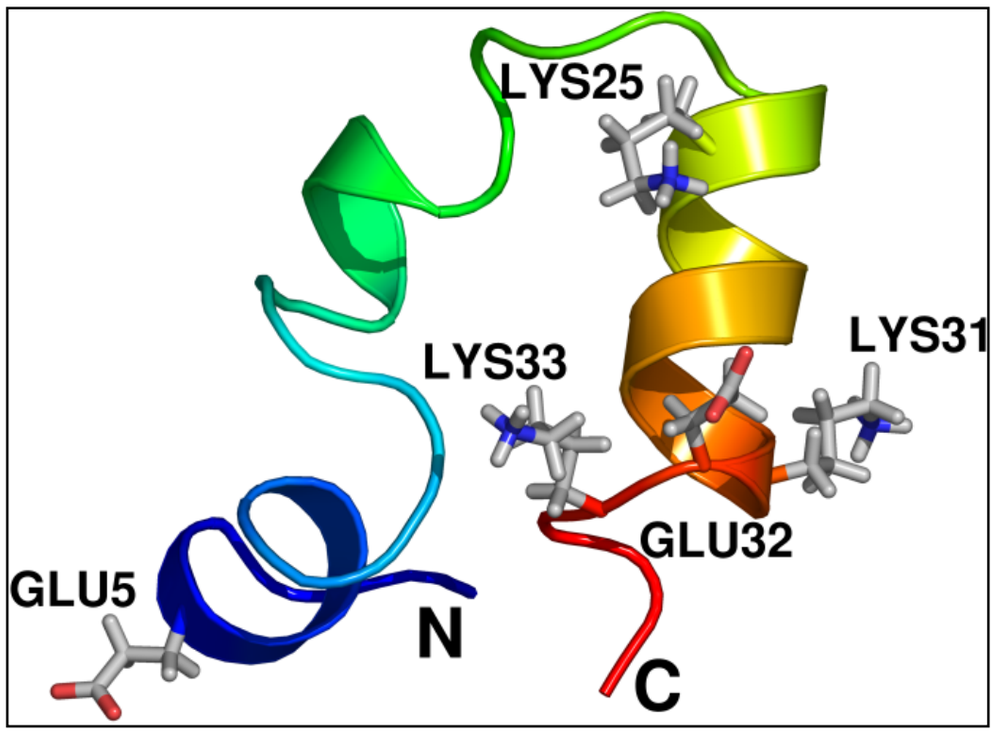


where we are

where we need to be

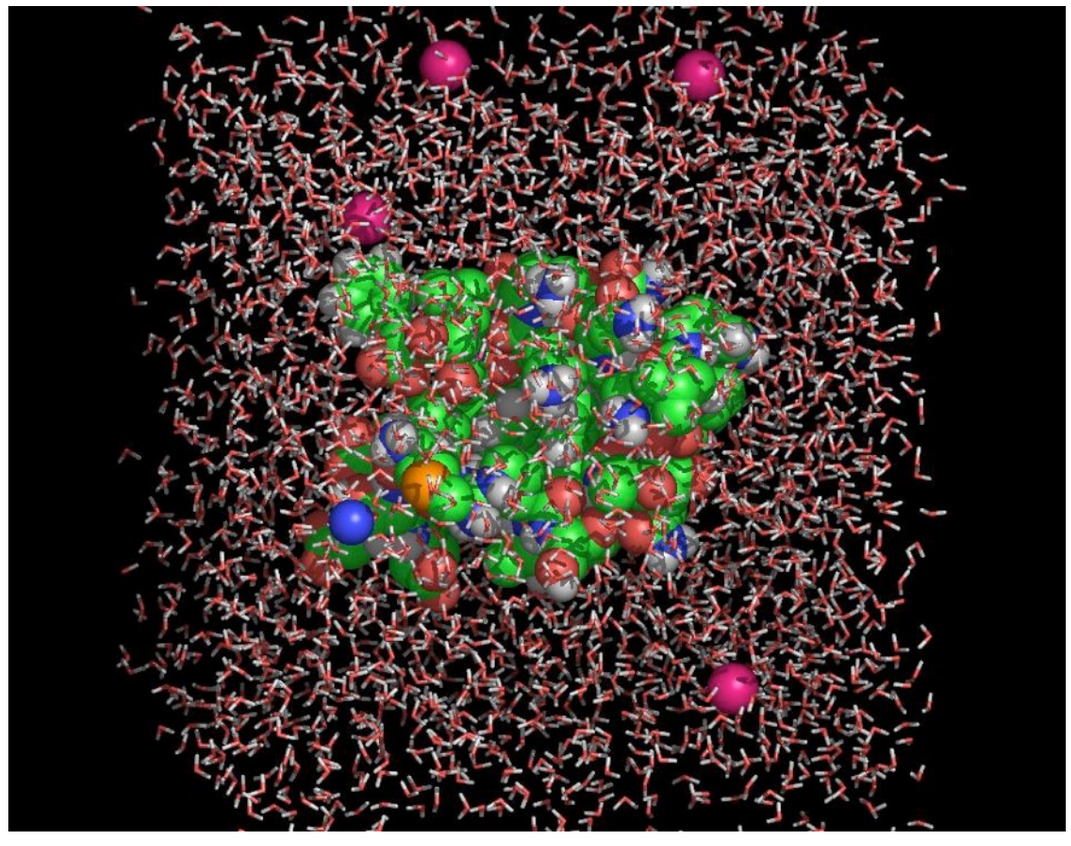
where we want to be

Example system



33 residues
~750 atoms

The Villin headpiece



Add water & ions:
~10000 atoms

MD programs

- GROMACS, Amber, Charmm, NAMD, LAMMPS, AceMD, ...
- Programs used to be intimately tied to a force field (Amber, Charmm, OPLS)
- Most of these are freely available to academia
- GROMACS is written by us

A typical protein simulation

- Get a structure (usually from PDB)
- Fix missing segments, sidechains, determine protonation states, etc.
- Prepare a topology (parameters)
- Add solvent water
- Energy minimization
- Equilibration simulation
- Run production simulation
- Analyze the output trajectory data

How to accelerate MD

- Take bigger time steps
- Use less particles
- Simulate in parallel
- Use accelerator hardware
- Coarse-graining

Competences involved

- Physics
- Chemistry
- Biology
- Mathematics (numerical analysis)
- Computer science

Exercise 1: integrators

- Euler integrator:
  - $x_{i+1} = x_i + dt * v_i$
  - $v_{i+1} = v_i + dt/m * f(x_i)$
- Velocity Verlet integrator:
  - $x_{i+1} = x_i + dt * v_i + dt*dt/(2m) * f(x_i)$
  - $v_{i+1} = v_i + dt/(2m) * (f(x_i) + f(x_{i+1}))$
- Leap frog
  - $v_{i+1/2} = v_{i-1/2} + dt/m * f(x_i)$
  - $x_{i+1} = x_i + dt * v(i+1/2)$

Exercise 1 ctnd

- Take a 1-dimensional system
- Use a harmonic potential: $V(x) = k/2\ x^2$
- Choose parameters m and k
- Implement the 3 integrators
- Choose initial conditions
- Try a range of values for dt, how does this affect the stability of the 3 integrators, when do things go wrong?
- Determine the average potential and kinetic energy ($1/2\ m\ v^2$) and compare Epot and Ekin

Exercise 1 cntd

- Determine the average potential and kinetic energy (1/2 m v^2) for the velocity Verlet and Leap-Frog integrators and compare Epot with Ekin.
- The comparison will give different results for VV and LF, especially for larger time steps. Can you explain the difference, knowing that v_i = (v_(i-1/2) + v_(i+1/2))/2, as can be found in Frenkel&Smit?