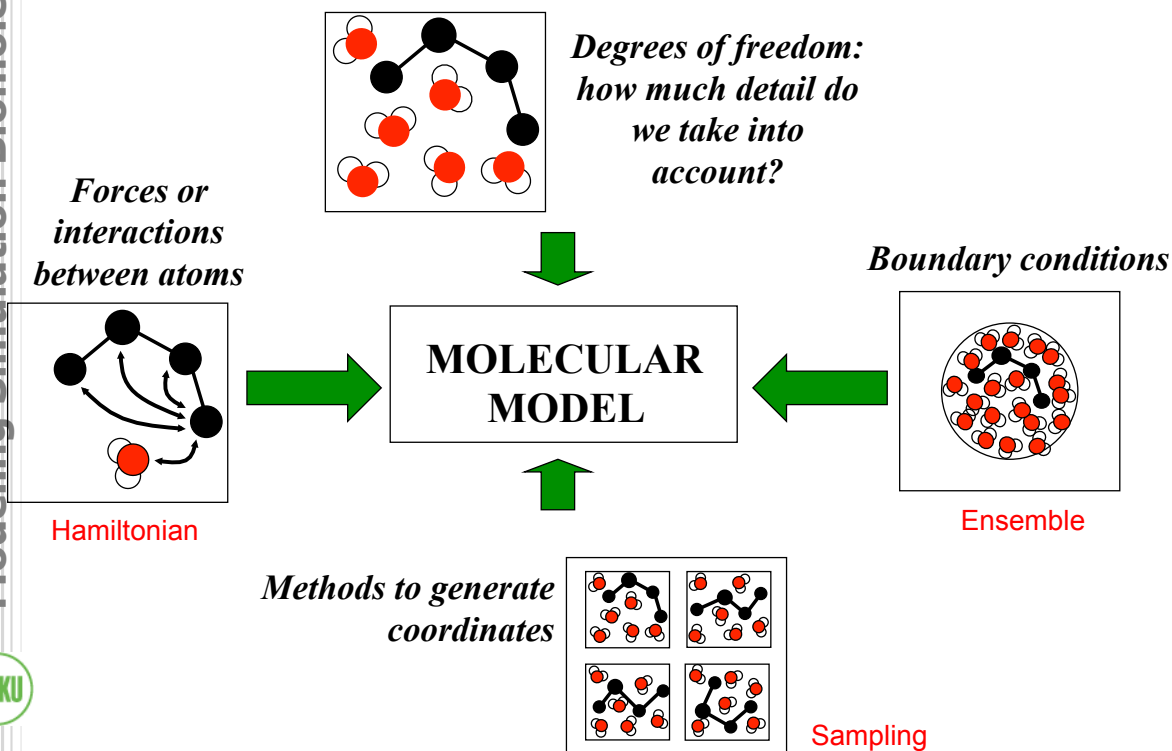# Schedule

Time	Mon. 20/11/17	Tue. 21/11/17	Wed. 22/11/17	Thur. 23/11/17	Fri. 24/11/17	Mon. 27/11/17	Tue. 28/11/17	Wed. 29/11/17	Thur. 30/11/17	Fri. 01/12/17
9:00	Free Time	Welcome JAG	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time
9:30		Lecture 1: Overview and Introduction JAG	Lecture 4: Thermodynamics JAG	Lecture 7: Classical Mechanics I JAG	Lecture 10: Ensembles I CO	Lecture 13: Free energies: reaction coordinates JAG	Lecture 16: Calculating properties from simulations JAG	Lecture 19: Electrostatics WvG	Lecture 22: On the ethics of the academic endeavour: where do we go?	Lecture 24: Left-overs/ questions and future perspectives WvG
10:15		Break	Break	Break	Break	Break	Break	Break		Break
10:30		Lecture 2: Molecular Simulations MD/SD/MC CO	Lecture 5: Force-Field Development CO	Lecture 8: Classical Mechanics II JAG	Lecture 11: Ensembles II JAG	Lecture 14: Boundary Conditions I WvG	Lecture 17: Comparison with Experiments WvG	Lecture 20: Polarization WvG	WvG	Lecture 25: Students plans I
11:15		Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break
11:45		Lecture 3: How to simulate using GROMOS CO	Lecture 6: Structure Refinement CO	Lecture 9: Analizing with GROMOS CO	Lecture 12: Free energies: alchemistry CO	Lecture 15: Boundary Conditions II WvG	Lecture 18: Searching & Enhanced Sampling WvG	Lecture 21: Multi-resolution simulations WvG	Lecture 23: QM/MM WvG	Lecture 26: Students plans II
12:30		Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Tutorial 10: **
14:00	Registration	Tutorial 2: Running MD GROMOS Tutorial	Tutorial 3: Running MD Students Plans	Tutorial 4: Running MD Students Plans	Tutorial 5: Running MD Students Plans	Tutorial 6: Statistical Mechanics exercises	Tutorial 7 Analizing MD: GROMOS Tutorial	Tutorial 8: Analizing MD: Students Plans	Tutorial 9: Analizing MD: Students Plans	How to prepare a barbecue
		Tutorial 1 Linux, NLHPC and OS-dongle installation								
16:00	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	
17:30	End of session	End of session	End of session	End of session	End of session	End of session	End of session	End of session	End of session	
20:30	Free time	Free time	Free time	Free time	Beer, Science & Friendship Good Stock Bar	Free time	Free time	Free time	Free time	Farewell

---

A model for molecular computations

Degrees of freedom: how much detail do we take into account?

Forces or interactions between atoms

Boundary conditions

MOLECULAR MODEL

Hamiltonian

Ensemble

Methods to generate coordinates

Sampling

Conformational analysis

| Method	Surrounding	Information
X-ray diffraction on crystal	neighbour molecules solvent molecules counter-ions	atomic coordinates 1 conformation
NMR in solution	solvent molecules (neighbour molecules)	interatomic distances few (1-50) conformations
Computation (in 'vacuum'?)	solvent molecules? (or none)	many (all?) conformations electronic structure energy levels

---

Optimize structures

- Many computational problems come down to minimising an objective function
- For instance geometry optimisation: minimize the energy
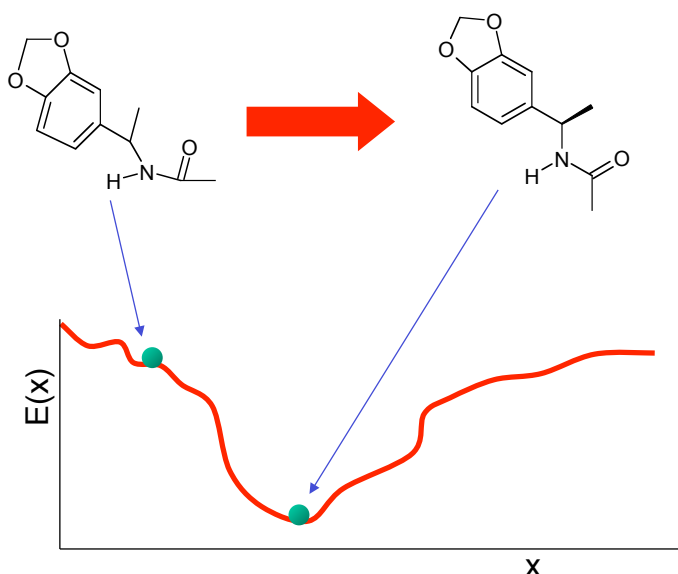
- Compare to a marble rolling down a slope

Find the minimum

- Minimise a function: set the derivative to zero

$$\frac{\partial E}{\partial x} = 0$$

E(x)

x
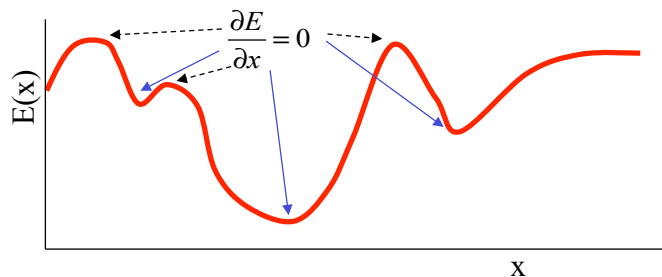
- But we have 3N variables →
  3N derivatives to set to zero
- There are multiple minima →
  different conformations are
  relevant
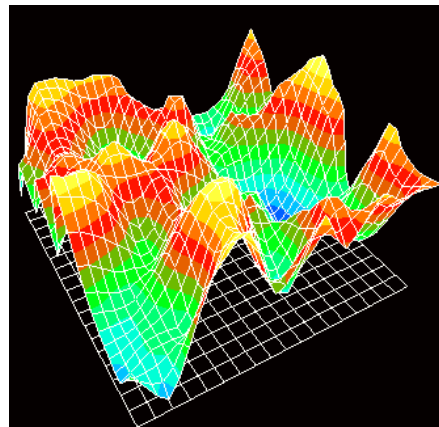
Potential energy surfaces

High Energy

Low Energy
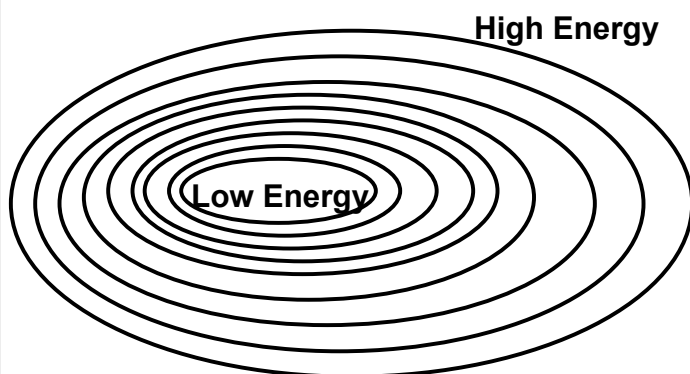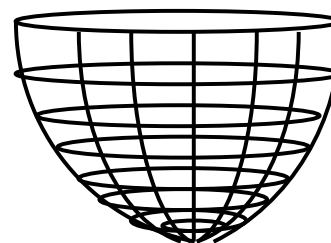
Overhead View   **Side View**

Conformational sampling

Mid-energy   **lower energy**   **lowest energy**   **highest energy**
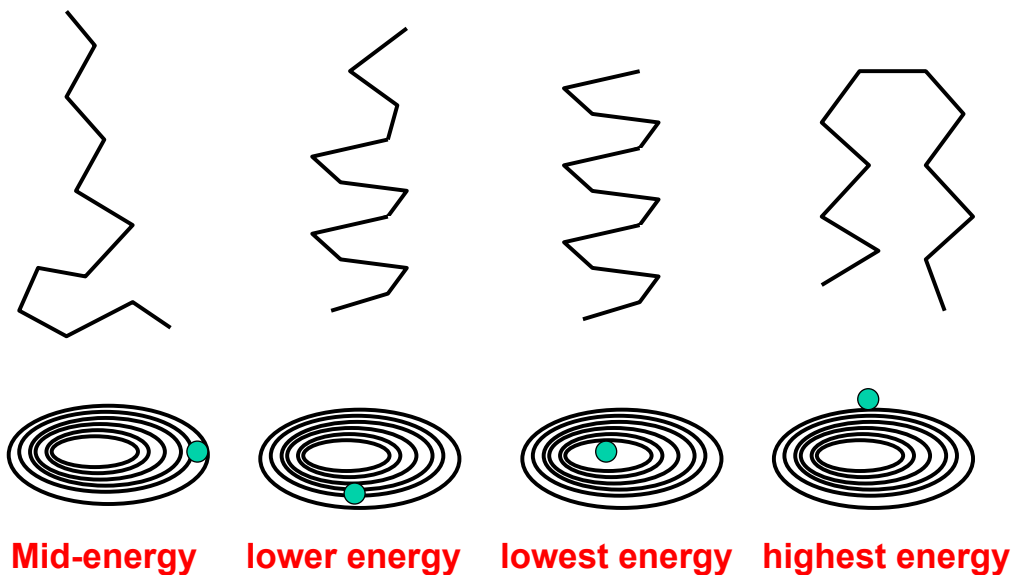
Steepest descent

- Frequently used for energy minimization of large (and small) molecules

- Ideal for calculating minima for complex (I.e. non-linear) surfaces or functions

1. Start at a certain point on the surface
2. Calculate E and all $\partial E/\partial r_x$
3. Take the steepest step down

$$\vec{r}\,' = \vec{r} + \vec{\nabla}E \cdot \lambda \qquad \vec{\nabla} = \left(\partial/\partial r_1, \partial/\partial r_2, \partial/\partial r_3\right)^T$$

4. Repeat 2 and 3 until all $\partial E/\partial r_x \approx 0$ ($\partial E/\partial r_x < e$)

- Goes down to the closest minimum

Steepest descent

High Energy

Low Energy
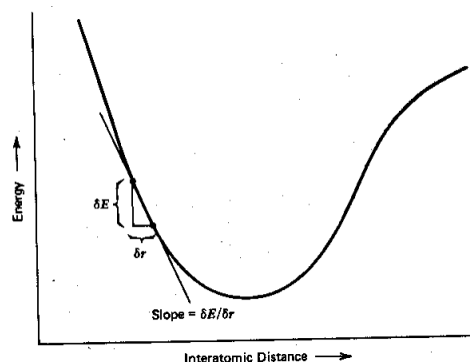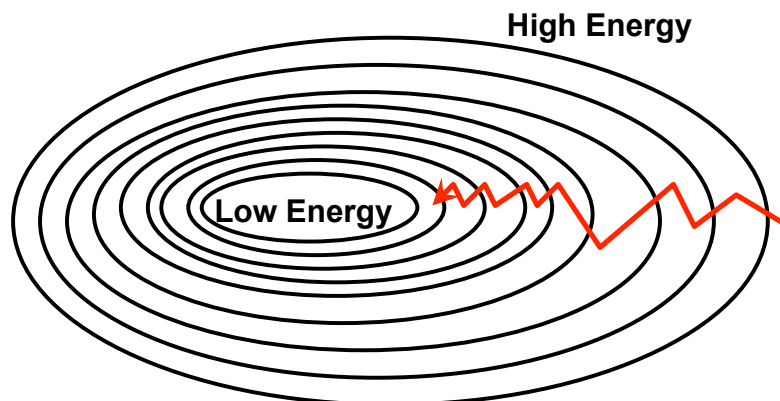
Makes small locally steep moves down gradient

Conformational searches

- We don't want to find *a* minimum energy conformation, but
  - *The* absolute energy minimum, or sometimes
  - *All* important conformations
- We have to cross barriers

- Systematic search: generate all possible conformations and pick the one(s) with the lowest energy
  very inefficient
- Sample conformational space
  Monte Carlo
  Molecular Dynamics
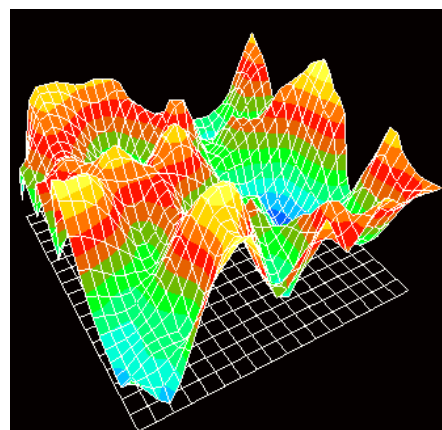
Systematic search

- Generate all possible conformations and calculate their energy
- Pick the best / lowest one
- Very inefficient: take for example a small protein
  - 50 amino acids with 2 rotatable torsional angles each (in the backbone)
  - 3 possibilities per torsion (trans, gauche+, gauche -)
  - Total 3^{50*2} = 10^{47} possibilities to calculate!

high energy

low
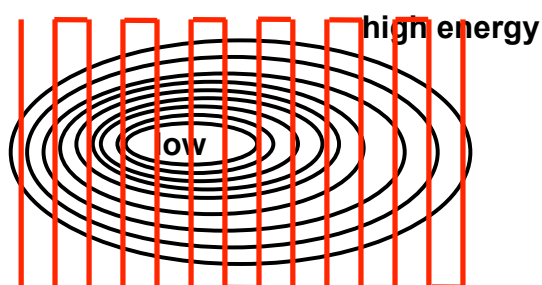
Conformations

- Only those conformations that have a low energy occur in reality
- How often does a molecule show a given conformation?

- The probability of a certain conformation depends on the Boltzmann factor

$$P(\mathbf{r}) = Q^{-1} e^{-U(\mathbf{r})/k_B T}$$

Probability for conformation defined by **r**

Energy of the conformation

(Absolute) temperature

"Normalisation constant"

Boltzmann constant (8.314 J mol^{-1}K^{-1})

Phase space

- The state of a system is completely determined by the coordinates and momenta (p= m·v) of the constituting particles

  X = (**q**, **p**)   **q** = $(x_1, y_1, z_1, x_2, y_2, z_2, \ldots)$ (or we use **r**)

      p = $(p_{x1}, p_{y1}, p_{z1}, p_{x2}, p_{y2}, p_{z1}, \ldots)$ (or we use **v**)

- The probability of a certain state or configuration is given by

$$P(\mathbf{q}, \mathbf{p}) = Q^{-1} e^{-E(\mathbf{q}, \mathbf{p})/k_B T}$$

Total energy,
including kinetic energy

- The probability also determines the distribution of different states in an ensemble of systems

---

Partition function

- The normalisation constant, Q, is called the partition function

$$Q_{NVT} = \iint e^{-E(\mathbf{q}, \mathbf{p})/k_B T} d\mathbf{q} d\mathbf{p}$$

- An ensemble is defined by various constants

  N   number of particles
  V   volume of the system
  E   energy
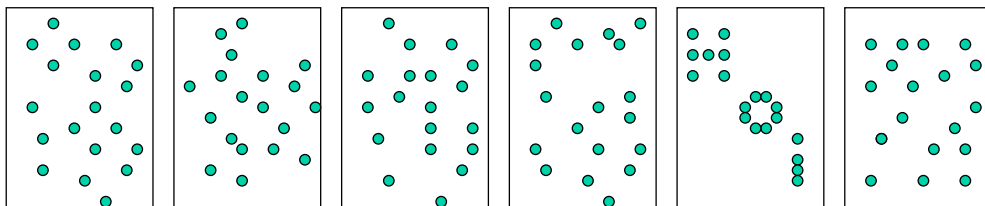
- Ensemble of systems with various configurations

Boltzmann

- You can visit him at the Zentralfriedhof

- For the NVE ensemble, the entropy is given by

$S = k_B \log Q$

S = k. log W

LVDWIG
BOLTZMANN
1844-1906
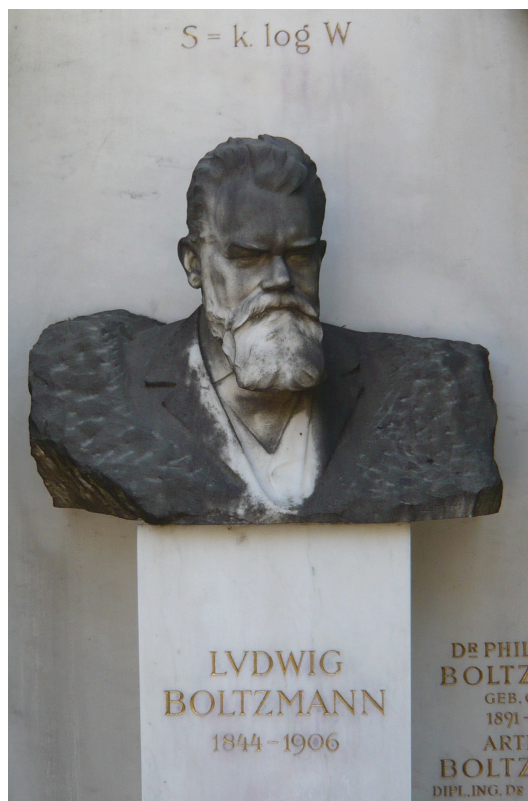
Experimental quantities

- Experimentally determined quantities are usually the average over very many molecules or systems
- Weighted average of a quantity A may be calculated as

$$\langle A \rangle = \iint A(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) d\mathbf{q} d\mathbf{p}$$

- For instance, the dipole moment of HCl ($\mu(\mathbf{r}) = q \cdot r$)
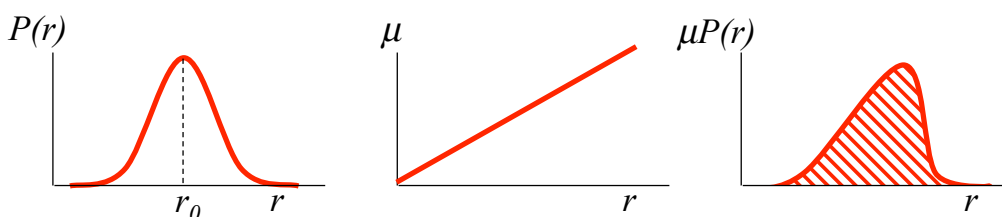
Ergodicity

- Ergodicity theorem:
  The average of a large number of systems is the same as the average over time of a single system

$$\langle A \rangle = \lim_{t \to \infty} \frac{1}{t} \int_0^t A(\tau) \, d\tau$$

- Has been proven for very simple systems, but is generally accepted

- So, if we follow a system in time, we generate an ensemble of states (**q**,**p**) each with the correct probability $P(\mathbf{q},\mathbf{p})$
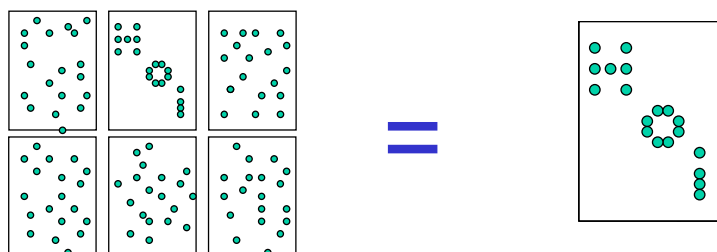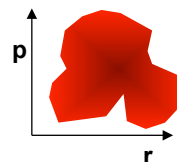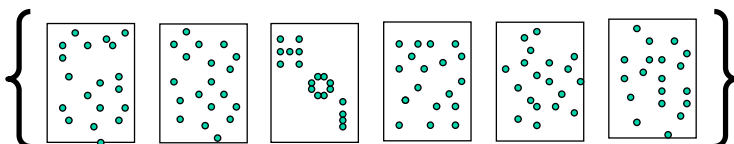

Phase space summary

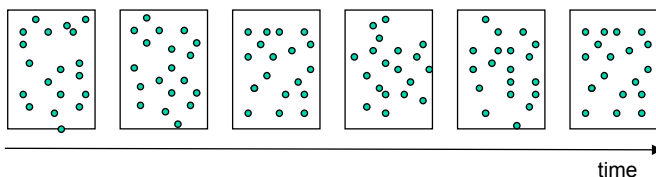A single configuration is given by the position and momenta of all particles

p
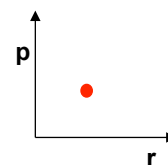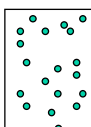
A point in phase space

An ensemble is the collection of all possible configurations, with their probabilities

p

A probability distribution
in phase space

A trajectory follows the configuration through time

time

p
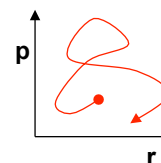
A continuous trajectory
in phase space

Monte Carlo simulation

1. Generate a conformation
2. Calculate *E*
3. a. If the new energy is less than the previous energy accept that conformation and go back to step 1

   b. If the new energy is greater than the previous energy accept it if a randomly chosen number is less than $e^{-\Delta E/kT}$ otherwise reject it

4 Go back to step 1 and repeat until done

- The accepted conformations represent an NVT ensemble

Metropolis criterion

1 $(=e^0)$

$P = e^{-\Delta E/kt}$

0 $(=e^{-\infty})$
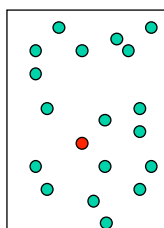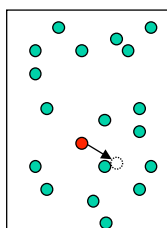
High Energy

Low
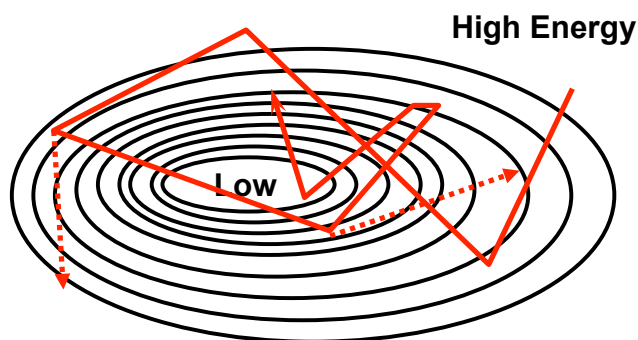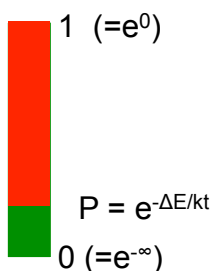
Monte Carlo in liquids

- Pick a random particle, move it randomly
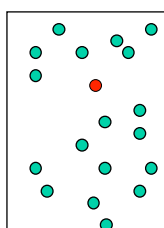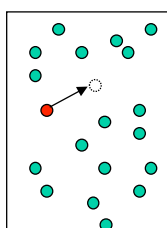
The energy of the new state is favourable, this configuration is accepted
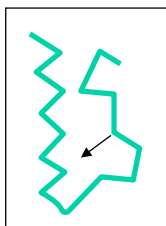
The energy of the new state is unfavourable, this configuration is rejected
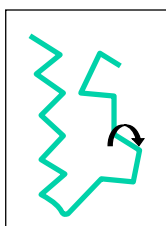
Large movements lead to worse acceptance ratios
Small movements lead to slow sampling
--> optimize to get acceptance ratio ~ 0.5

Monte Carlo for macromolecules

- For macromolecules, you have to design your moves cleverly
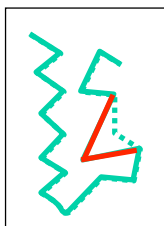
Approach 1: move the atoms
But: bond lengths and angles are very stiff, so only very small moves are accepted
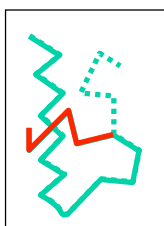
Approach 2: rotate a torsional angle
But: this may lead to large motion further in the sequence, and then to clashes

Some success with double torsional moves to keep the modifications local

---

Trajectory in time: classical dynamics
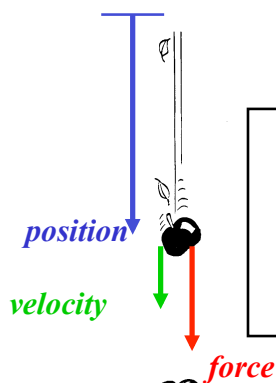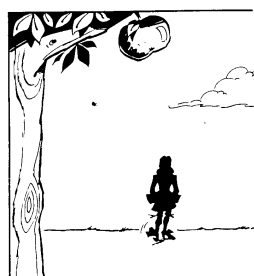
Situation at time t

Force is determined by relative positions

acceleration = force / mass

Δ velocity = acceleration × Δ t

Δ position = velocity × Δ t

Situation at time t+Δt

position

velocity

force

Sir Isaac Newton
1642 -1727

Determinism …

Forces from a force field

- The force on an atom is defined as the negative derivative of the potential energy with respect to the coordinates

$$F_{x,1} = -\frac{\partial U(\mathbf{r})}{\partial x_1}$$

$$F_{y,1} = -\frac{\partial U(\mathbf{r})}{\partial y_1}$$

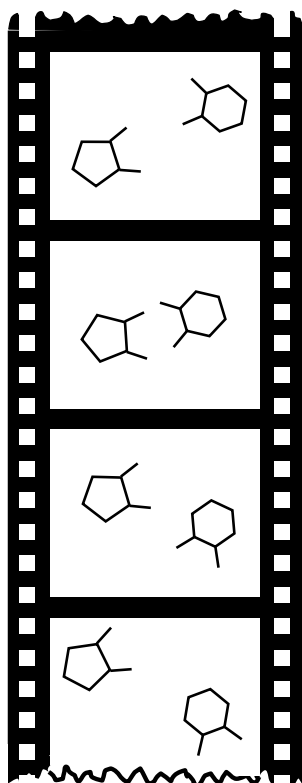$$F_{z,1} = -\frac{\partial U(\mathbf{r})}{\partial z_1}$$

$$\mathbf{a_1} = \frac{\partial^2 \mathbf{x_1}}{\partial t^2} = \frac{\mathbf{F_1}}{m_1}$$

U is defined by the force field with relatively simple equations
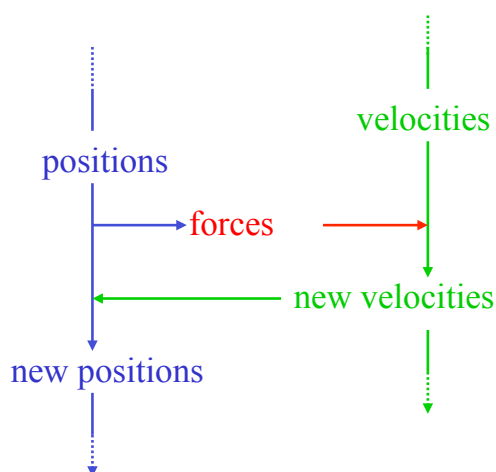
The derivatives may be calculated analytically

For simple systems (harmonic oscillator) we can solve the equations of motion exactly

Modeling Simulation Biomolecules

Molecular dynamics

... Comparable to shooting a movie of molecular motion...

time t

positions → forces

velocities

time (t+Δt)

new positions ← new velocities

Leap frog algorithm

Binding equilibrium

Complex :

Hydrogen bonds

Cyclohexane-diamine

?
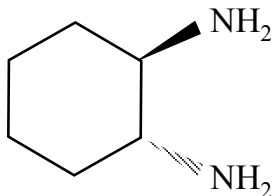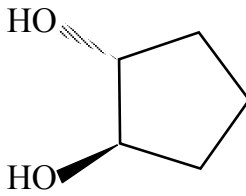
Many different binding modes

Cyclopentane-diol
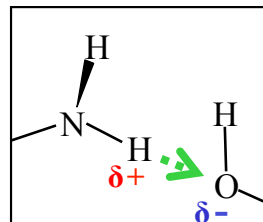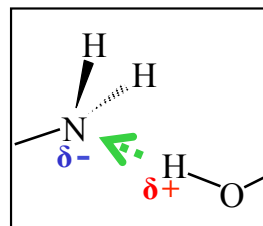
Average binding strength (free enthalpy) :

	Experimental		MD simulation
	Benzene	*CCl_4*	
ΔG_b [kJ/mol]	**-9.3**	**-11.5**	**-10.4**

Diol + Diamine + 252 CCl_4 Molecules
$2.1 - 2.2 \cdot 10^{-9}$ seconds

Diol + Diamine + 252 CCl₄ Molecules
3.2 – 4.0·10⁻⁹ seconds

Hydrogen bonds

▮▮▮▮▮ **O → N**

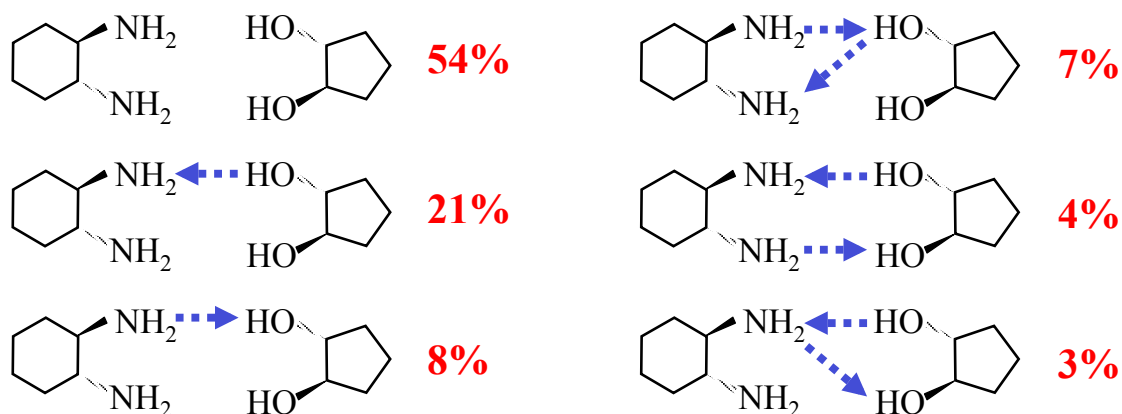▮▮▮▮▮ **N → O**

•

the molecules are free again…

Results of the simulation (100 ns)

Experimentally hardly (or not) possible !

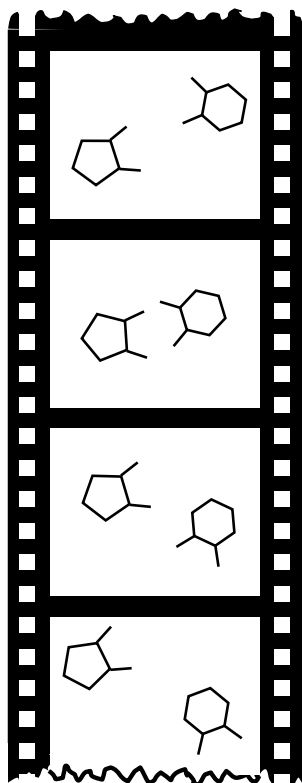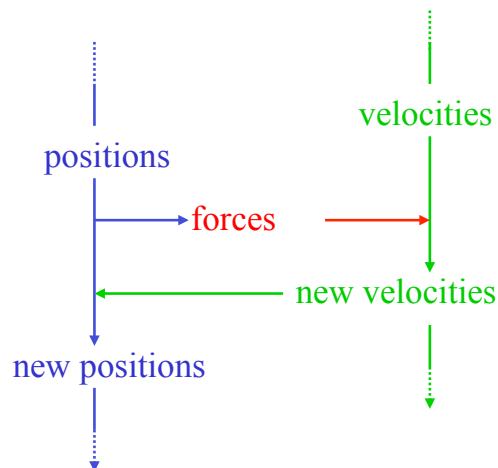*Occurrence of different binding modes :*

54%

7%

21%

4%

8%

3%

Life time :

- Average life time of the complex: $2 \cdot 10^{-10}$ sec (max. $3 \cdot 10^{-9}$ sec)
- Average life time of a hydrogen bond: $5 \cdot 10^{-12}$ sec

Molecular dynamics

... Comparable to shooting a movie of molecular motion...

time t

positions

velocities

forces

time (t+Δt)

new velocities

new positions

Leap frog algorithm

Leap-frog algorithm

- Taylor expansions around t_0 with ½Δt:

$$\mathbf{q}(t_0 + \tfrac{1}{2}\Delta t) = \mathbf{q}(t_0) + \tfrac{1}{2}\Delta t \frac{\partial \mathbf{q}(t_0)}{\partial t} + \frac{1}{2!}\left(\tfrac{1}{2}\Delta t\right)^2 \frac{\partial^2 \mathbf{q}(t_0)}{\partial t^2} + O(3)$$

$$\mathbf{q}(t_0 - \tfrac{1}{2}\Delta t) = \mathbf{q}(t_0) - \tfrac{1}{2}\Delta t \frac{\partial \mathbf{q}(t_0)}{\partial t} + \frac{1}{2!}\left(\tfrac{1}{2}\Delta t\right)^2 \frac{\partial^2 \mathbf{q}(t_0)}{\partial t^2} + O(3)$$

Remaining terms are of the order 3 and higher ~(½Δt)³

- Subtract the first from the second:

$$\mathbf{q}(t_0 + \tfrac{1}{2}\Delta t) - \mathbf{q}(t_0 - \tfrac{1}{2}\Delta t) = \Delta t \frac{\partial \mathbf{q}(t_0)}{\partial t} + O(3)$$

$$\mathbf{q}(t_0 + \tfrac{1}{2}\Delta t) = \mathbf{q}(t_0 - \tfrac{1}{2}\Delta t) + \Delta t \frac{\partial \mathbf{q}(t_0)}{\partial t}$$

- Use $t_0 = t + \tfrac{1}{2}\Delta t$ and $v = \partial q / \partial t$, do the same for v
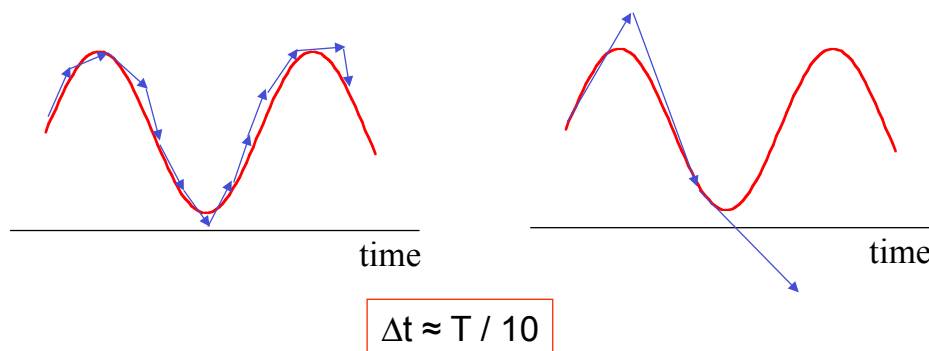
$$\mathbf{q}(t + \Delta t) = \mathbf{q}(t) + \mathbf{v}(t + \tfrac{1}{2}\Delta t) \cdot \Delta t$$

$$\mathbf{v}(t + \tfrac{1}{2}\Delta t) = \mathbf{v}(t - \tfrac{1}{2}\Delta t) + \mathbf{a}(t) \cdot \Delta t$$

BOKU

Leap-frog algorithm

- There are many integration algorithms
    - Verlet, Beeman and Leap-frog give identical coordinate trajectories

- The integration time step Δt should be sufficiently small, such that the fastest motion is correctly described

time   time

$\Delta t \approx T / 10$

Time step

- Fastest motion is the bond vibration: $\Delta t \sim 0.5$ fs
- Interesting simulations: ps / ns / ms   So very many steps!

- Bond vibrattions do not really influence the overall dynamics
- At these frequencies, one should treat bonds using QM / relativistically

- Better approximation is a fixed bond length: bond constraints (SHAKE or Lincs algorithm)
    - Integrate the equations of motion under the condition that the bonds remain at the same length

- Using SHAKE or Lincs we can use time steps of 2 fs
   - Now the angles or the water libration is the fastest motion

Molecular dynamics

1. Start at a certain conformation with initial velocities

2. Calculate the energy and the force on every atom i:

$$\vec{F}_i = -\vec{\nabla}_i E^{pot}(\vec{r}_1, \vec{r}_2, ..., \vec{r}_N)$$

3. From the force (acceleration) update velocity for every atom

4. From the velocity update the position

5. Propagate through time

- Total energy $E^{tot} = E^{pot} + E^{kin}$ is conserved (class. mech.)
- Kinetic energy allows us to go over barriers
- If we simulate infinitely long, we get the NVE ensemble

Energy conservation

- Simulation of liquid argon (256 atoms)
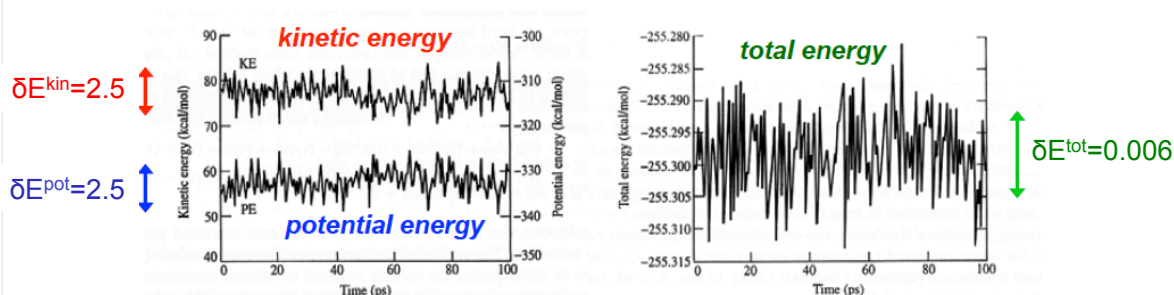
- The kinetic and potential energy fluctuates considerably
- The total energy is conserved
  - Remaining noise comes from the integration accuracy (Δt)

- This explains why we can find new minima:

E^{pot} is high E^{kin} is low

E^{pot} is low E^{kin} is high
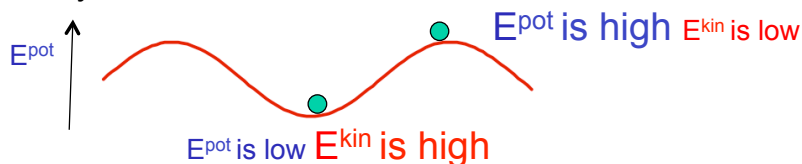
Stochastic dynamics

Vacuum

- Surface effects (surface tension)
- No dielectric screening

Implicit solvation

- Surrounding is a continuum with dielectric constant ε > 1
- Different ε within and outside the solute: solving Poisson-Boltzmann equation
- Nonpolar interaction proportional to solvent accessible surface area

Stochastic dynamics
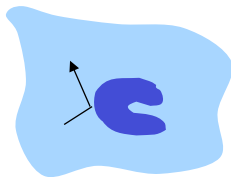
- Change of the equations of motion
- Additional frictional coefficient represents viscosity of solvent
- Random (stochastic) force represents collision with solvent

$$\mathbf{F}(t) = -\frac{\partial \mathcal{V}}{\partial \mathbf{r}} - \zeta \mathbf{p}(t) + \mathbf{F}_{stoch}(t)$$

History

| Year	molecular system: type, size	length of the simulation in seconds
1957	first molecular dynamics simulation (hard discs, two dimensions)	
1964	atomic liquid (argon)	10^{-11}
1971	molecular liquid (water)	$5 \cdot 10^{-12}$
1976	protein (no solvent)	$2 \cdot 10^{-11}$
1983	protein in water	$2 \cdot 10^{-11}$
1989	protein-DNA complex in water	10^{-10}
1997	polypeptide folding in solvent	10^{-7}
2001	micelle formation	10^{-7}
2010	folding of a small protein	10^{-6}

Summary

- Conformation of a molecule defines a (force field) energy
  - We can optimize the conformation, by minimizing the energy
- The energy determines the probability that a conformation occurs
- All possible conformations form an ensemble of structures
- Experiments give us the average over an ensemble
- A trajectory in time approximates the ensemble

- Monte Carlo simulations: get the ensemble directly
- Molecular Dynamics simulations: get the trajectory in time
  - Leap-frog algorithm

- Bond constraints
- Temperature and pressure coupling

Look at the movie – loop dynamics
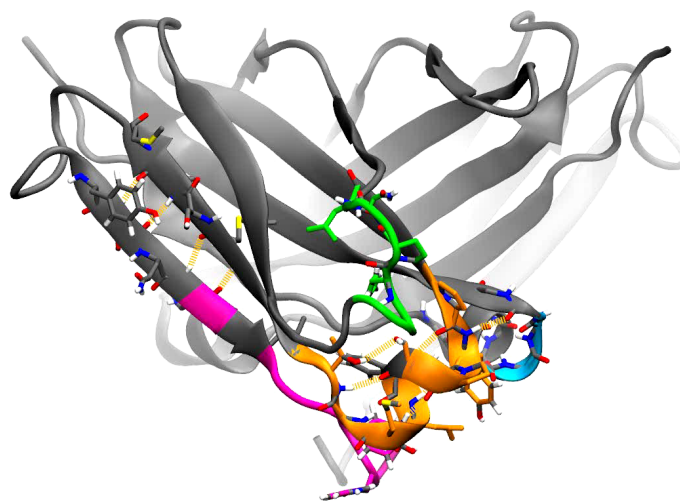
Backup slides

PREVIEW OTHER ENSEMBLES

Other ensembles

- Extensive properties are additive
  Increase with system size: volume, mass, entropy
- Intensive properties are not additive
  Stay constant with system size: temperature, pressure

- There are pairs of extensive and intensive properties that cannot both be constant:
  - Volume (extensive) and Pressure (intensive)
  - Energy (extensive) and Temperature (intensive)
  - Number of particles (extensive) and Chemical potential (μ; intensive)

- Choice of which properties stay constant define the ensemble
  - Microcanonical ensemble: NVE
  - Canonical ensemble: NVT
  - Isothermal-isobaric ensemble: NpT
  - Grand-canonical ensemble: μVT

Other ensembles

- Standard MD simulations: NVE-ensemble
- Standard MC simulations: NVT-ensemble

But experiments are usually at constant temperature and constant pressure

- MD in NVT-ensemble:
  - Scale the velocities of the particles in every step in such a way that the average temperature is constant:

$$E_{kin}(t) = \sum_{i=1}^{N} \tfrac{1}{2} m_i \mathbf{v_i}^2(t) = \tfrac{1}{2} N_{df} k_B T(t) \qquad T(t) = \frac{1}{N_{df} k_B} \sum_{i=1}^{N} m_i \mathbf{v_i}^2(t)$$

- MD / MC in NpT-ensemble
  - Scale the positions of the particles such that the average pressure is constant:

$$p(t) = \frac{1}{V(t)} \left[N k_B T(t) + \tfrac{1}{3} \sum_{i}^{N} \sum_{j>i}^{N} F_{ij} r_{ij} \right]$$

# MD algorithm

1. Apply periodic boundary conditions: put atoms in the box
2. Calculate the kinetic energy and the temperature
3. Calculate the potential energy and the force, **F**(t)
4. Calculate the pressure in the box (depends on **F** and E^{kin})
5. Calculate the velocities, **v**(t+½Δt) , from the force, **F**(t) (~acceleration, **a**(t))
6. Scale the velocities to keep <T> constant
7. Calculate the positions, **q**(t+Δt), from the velocities **v**(t+½Δt)
8. Satisfy the bond constraints (SHAKE) corrections for **q**(t+Dt) and **v**(t+½Δt)
9. Scale the positions to keep <p> constant
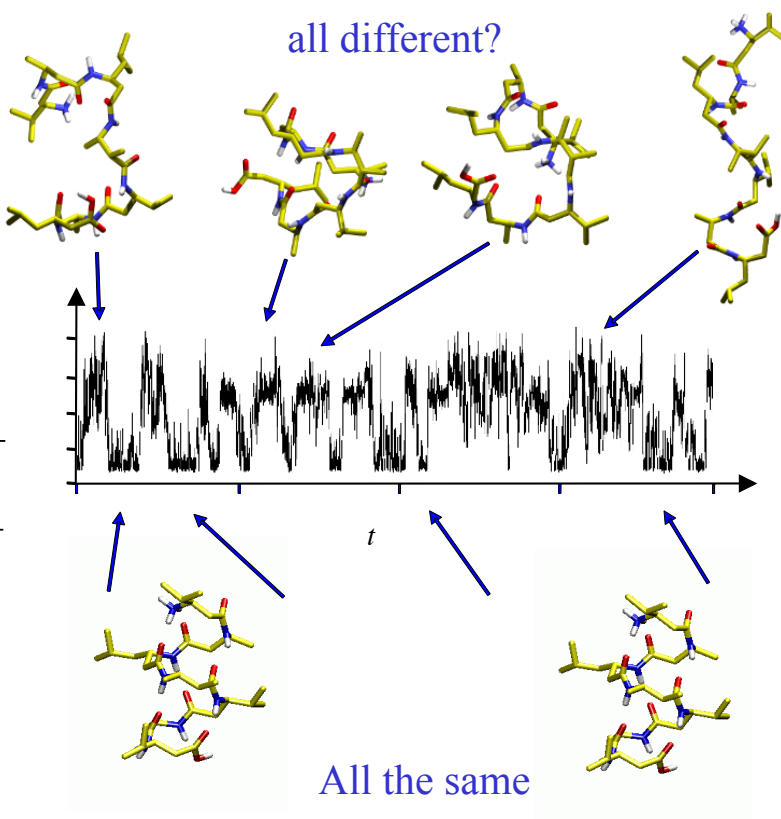10. Set t = t + Δt and return to step 1.
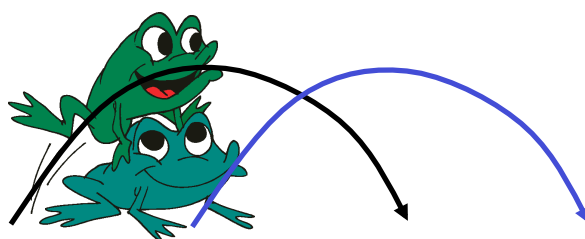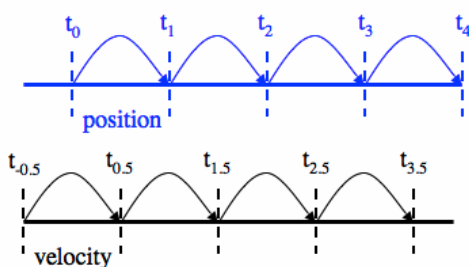
Molecular dynamics

- Extract mostly populated conformations from simulation
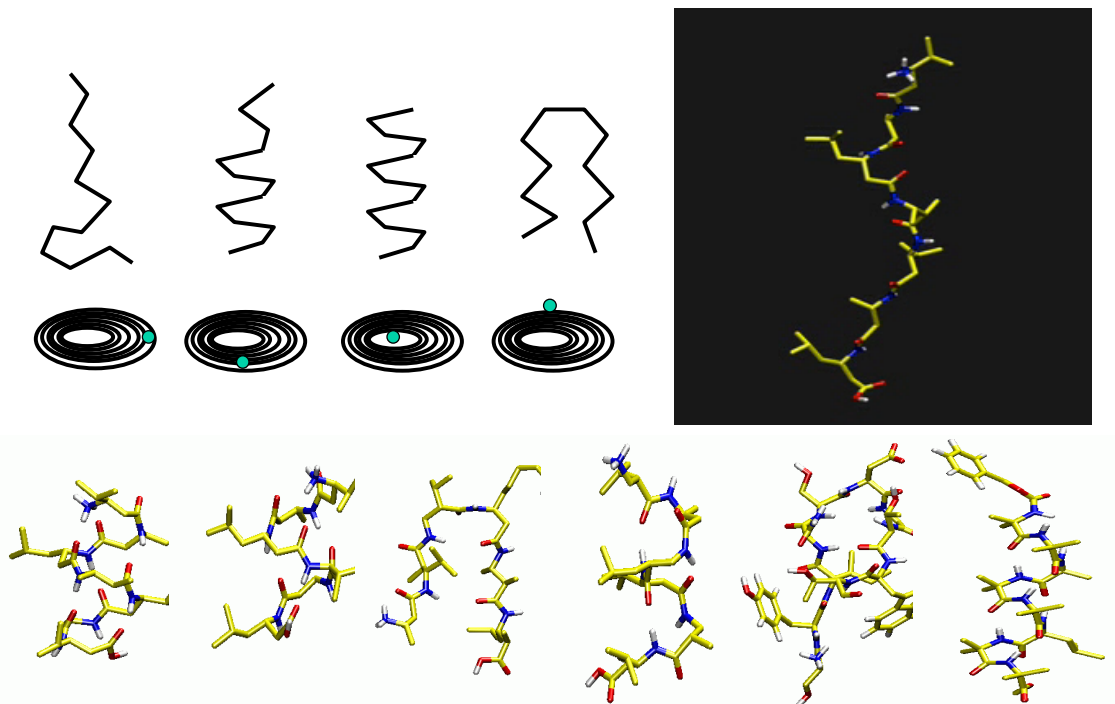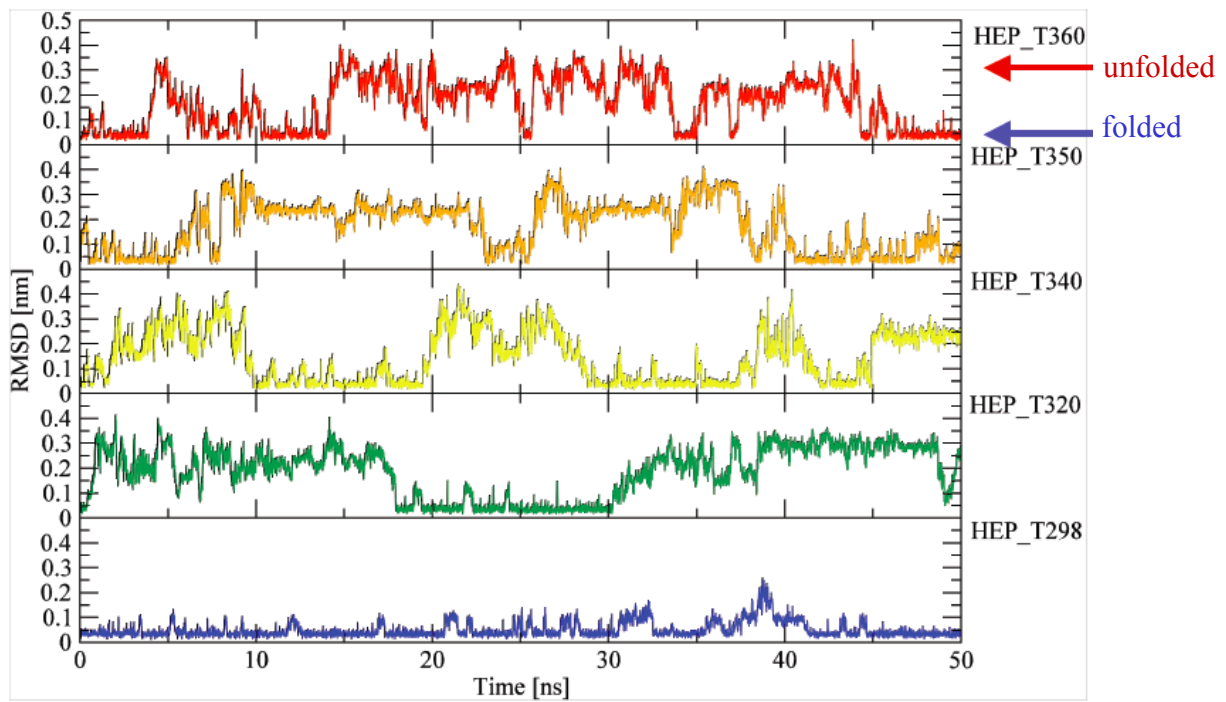

RMSD: Root mean square deviation

- A measure to compare two structures
- Here we compare the structures seen in the simulation to the experimentally determined 'folded' structure

all different?

$$D = \sqrt{\frac{\sum_i \left| \mathbf{q}_i - \mathbf{q}_i^{ref} \right|^2}{N}}$$
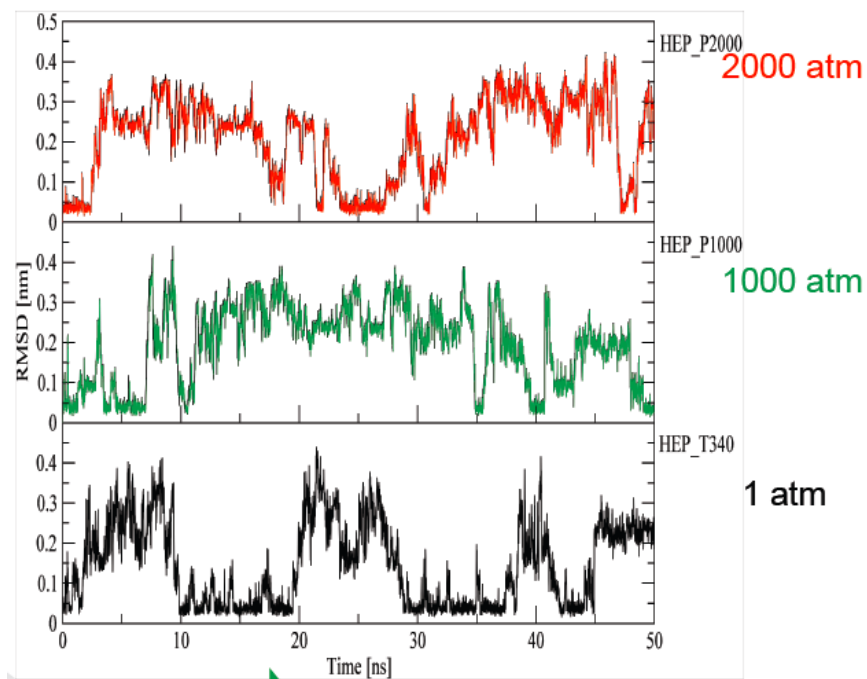

All the same

Temperature dependency

folding equilibrium depends on temperature

Pressure dependency

folding equilibrium depends on pressure