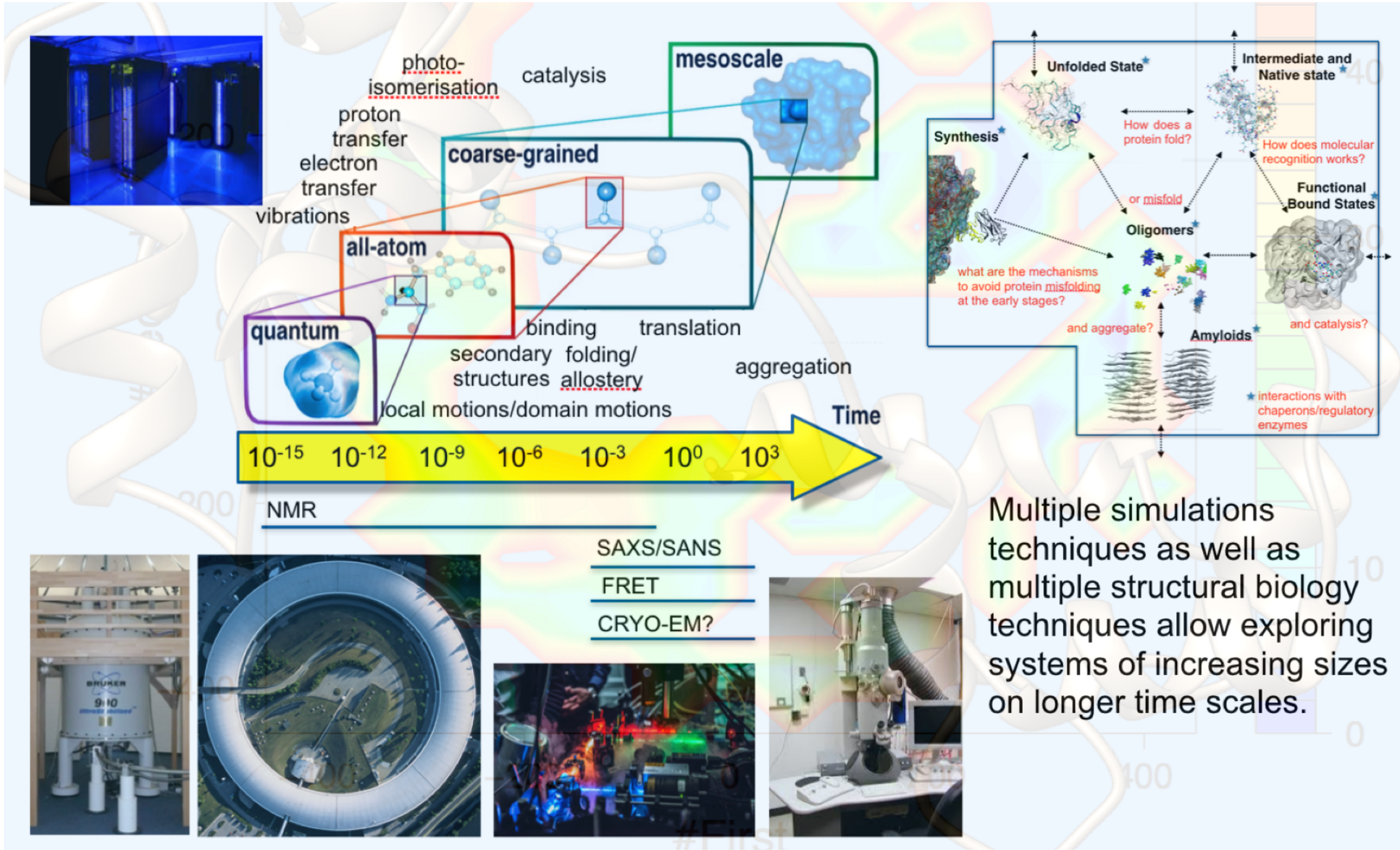


Molecular Dynamics Simulations

Carlo Camilloni

Motivation

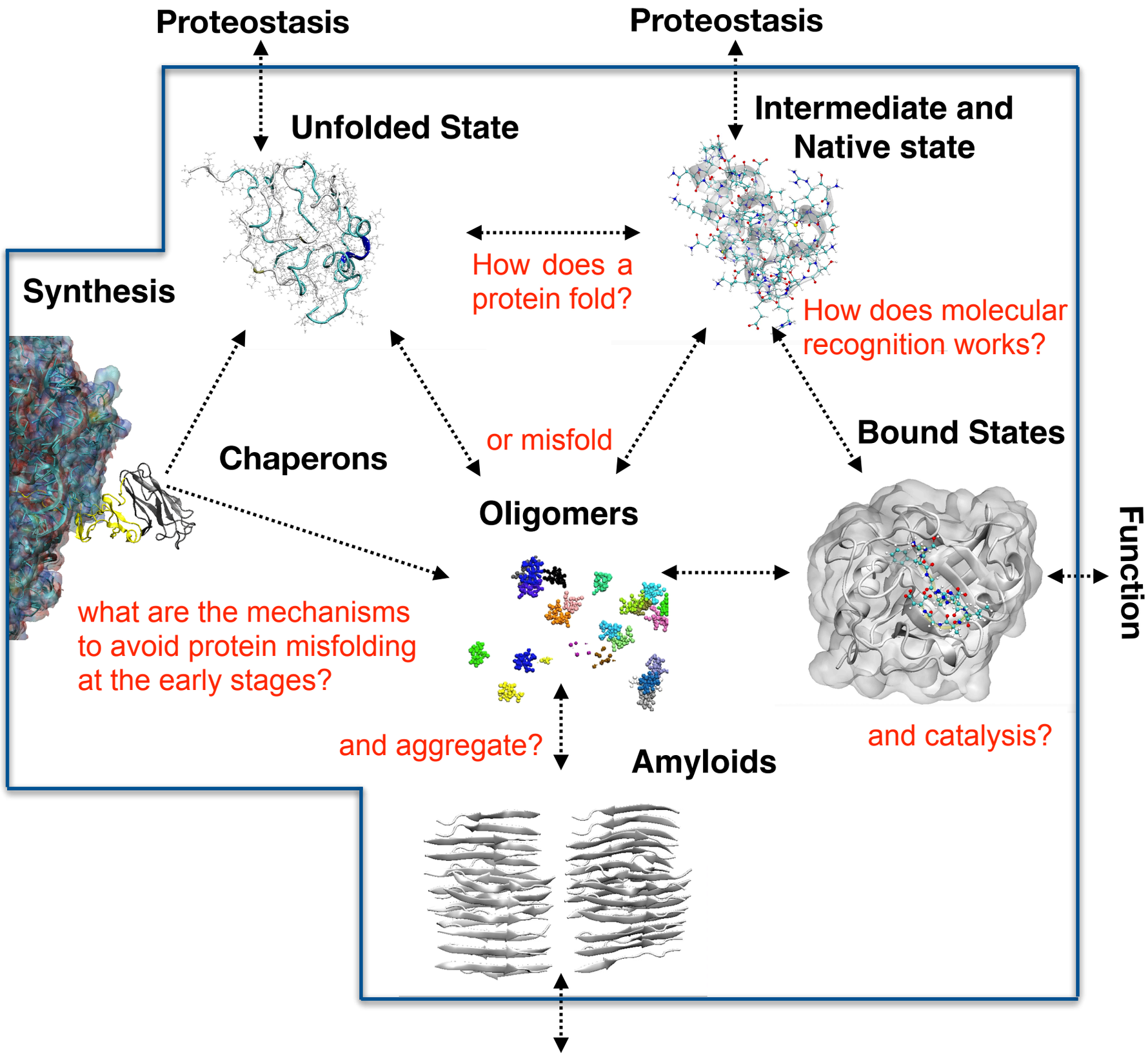
Nowadays there isn't any experiment that allow us seeing how molecules move at atomic resolution. We can observe snapshots from crystals, snapshots of very populated states in cryo-EM or the overall average behaviour of molecules in solution by NMR.

Can we use computers to build a microscope to see (build models) how biological molecules move and how processes happens?

- we want to **see** things that we cannot see with experiments.
- we want to **understand** what an experiment means.
- we want to **anticipate** what will happen in some conditions.

- The steps of a **chemical reaction** in an enzyme
- The steps of a **self assembly process** (folding, aggregation)
- The **probability distribution of structures** (ensemble) that represents the outcome of an experiment
- The reaction of a system to some change in the experimental conditions, **out-of-equilibrium**.

Structure and Dynamics



Mechanisms: how does this work? (What forces are at play? What is the free energy of the process? …)

Energetics of ion conduction through the K^+ channel

Simon Bernèche & Benoît Roux

Department of Biochemistry, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021, USA; Département de Physique, Université de Montréal, Montréal, Québec H3C 3J7, Canada

K^+ channels are transmembrane proteins that are essential for the transmission of nerve impulses. The ability of these proteins to conduct K^+ ions at levels near the limit of diffusion is traditionally described in terms of concerted mechanisms in which ion-channel attraction and ion–ion repulsion have compensating effects, as several ions are moving simultaneously in single file through the narrow pore[1–4]. The efficiency of such a mechanism, however, relies on a delicate energy balance—the strong ion-channel attraction must be perfectly counterbalanced by the electrostatic ion–ion repulsion. To elucidate the mechanism of ion conduction at the atomic level, we performed molecular dynamics free energy simulations on the basis of the X-ray structure of the KcsA K^+ channel[4]. Here we find that ion conduction involves transitions between two main states, with two and three K^+ ions occupying the selectivity filter, respectively; this process is reminiscent of the 'knock-on' mechanism proposed by Hodgkin and Keynes in 1955[1]. The largest free energy barrier is on the order of 2–3 kcal mol^{-1}, implying that the process of ion conduction is limited by diffusion. Ion–ion repulsion, although essential for

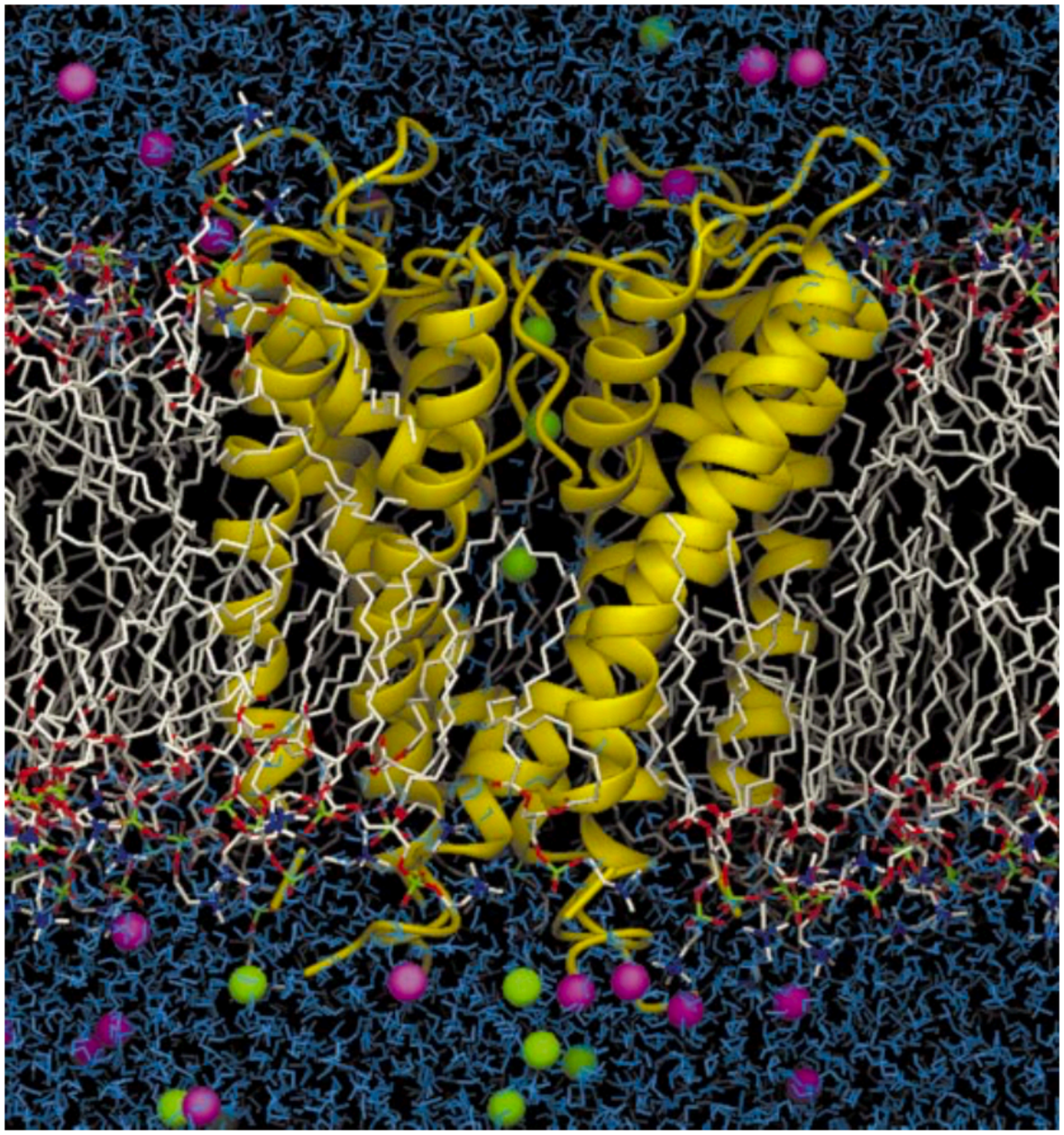


Figure 1 Molecular representation of the atomic model of the KcsA K^+ channel embedded in an explicit DPPC phosphilipid membrane bathed by a 150 mM KCl aqueous salt solution[11].

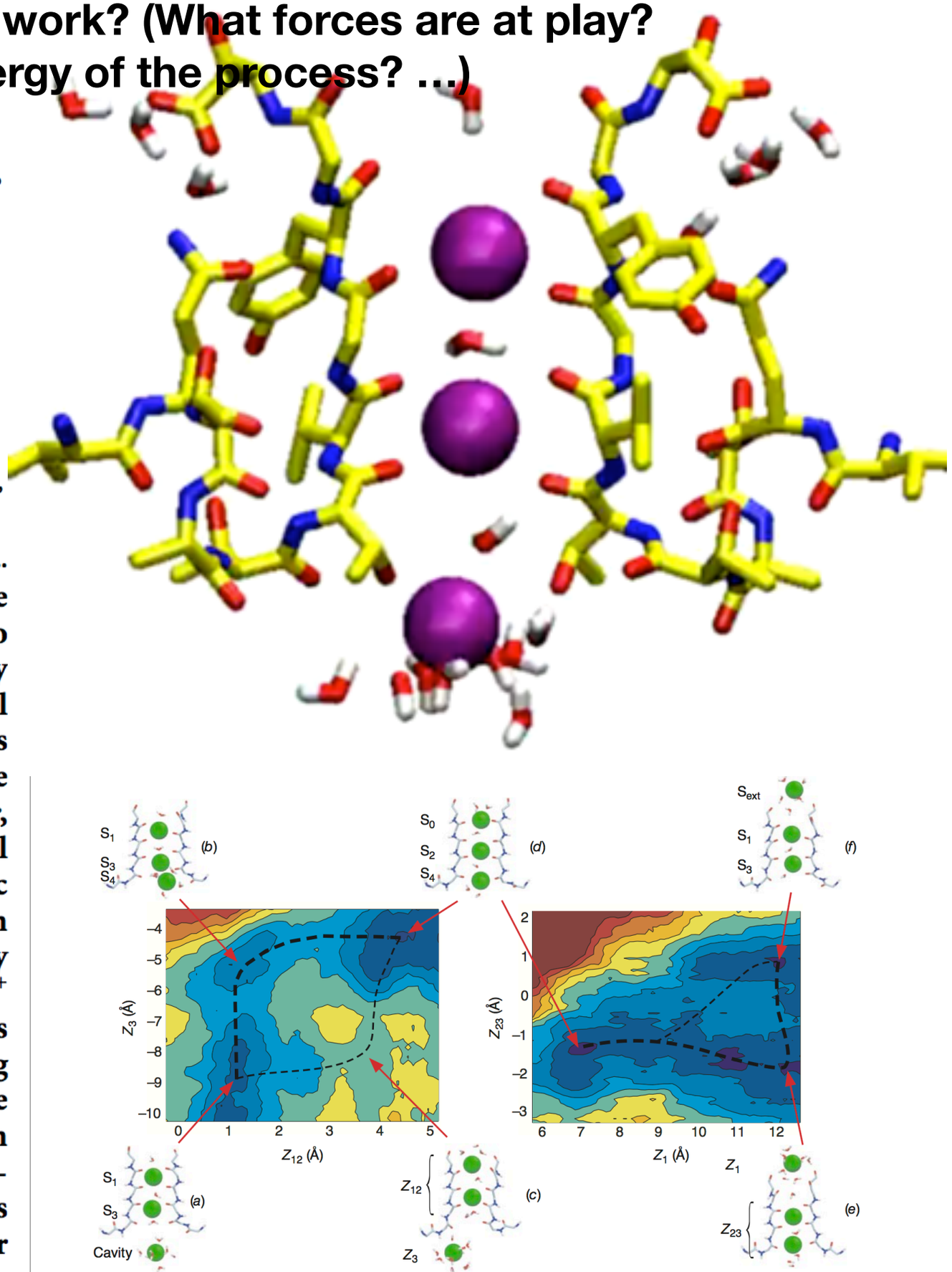
Mechanisms: how does this work? (What forces are at play? What is the free energy of the process? …)

Energetics of ion conduction through the K^+ channel

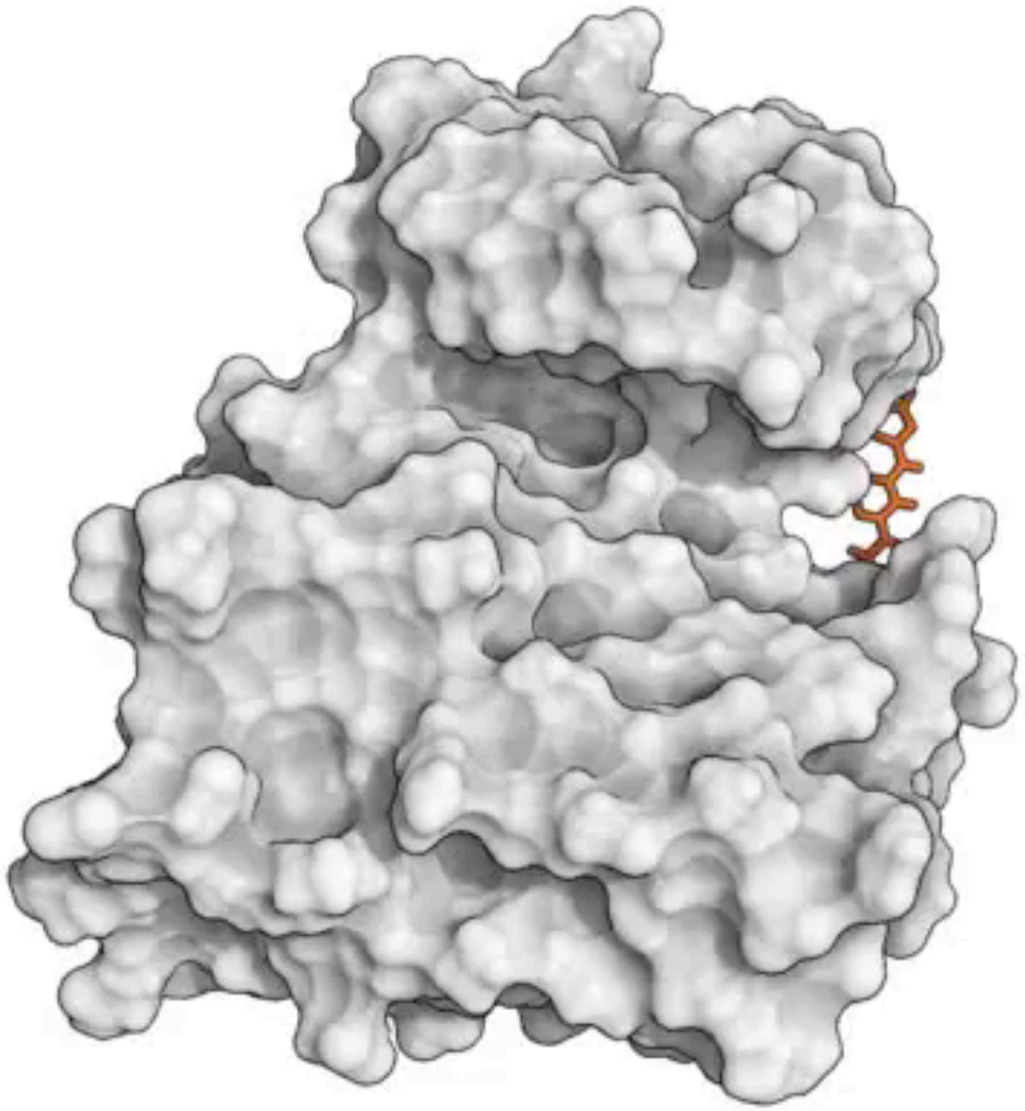
Simon Bernèche & Benoît Roux

Department of Biochemistry, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021, USA; Département de Physique, Université de Montréal, Montréal, Québec H3C 3J7, Canada

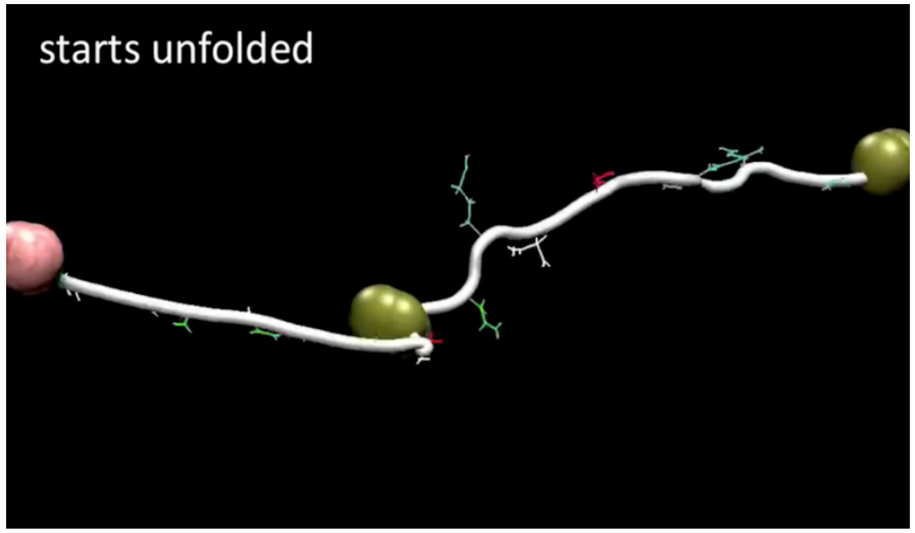
K^+ channels are transmembrane proteins that are essential for the transmission of nerve impulses. The ability of these proteins to conduct K^+ ions at levels near the limit of diffusion is traditionally described in terms of concerted mechanisms in which ion-channel attraction and ion–ion repulsion have compensating effects, as several ions are moving simultaneously in single file through the narrow pore[1–4]. The efficiency of such a mechanism, however, relies on a delicate energy balance—the strong ion-channel attraction must be perfectly counterbalanced by the electrostatic ion–ion repulsion. To elucidate the mechanism of ion conduction at the atomic level, we performed molecular dynamics free energy simulations on the basis of the X-ray structure of the KcsA K^+ channel[4]. Here we find that ion conduction involves transitions between two main states, with two and three K^+ ions occupying the selectivity filter, respectively; this process is reminiscent of the 'knock-on' mechanism proposed by Hodgkin and Keynes in 1955[1]. The largest free energy barrier is on the order of 2–3 kcal mol^{-1}, implying that the process of ion conduction is limited by diffusion. Ion–ion repulsion, although essential for

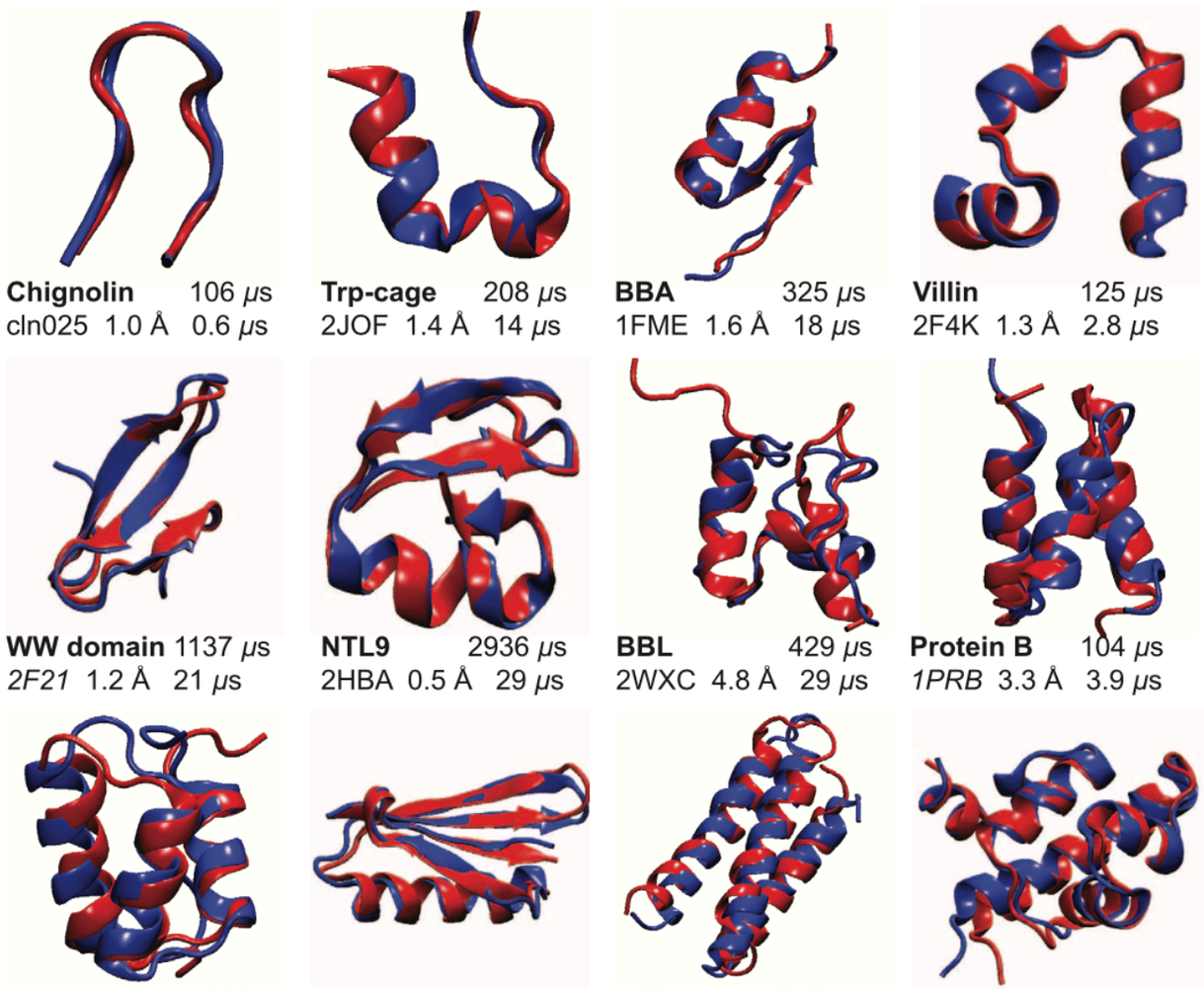


Intermolecular interactions: how does a ligand bind?
Is there a first encounter? How many possible binding sites?
Ecc.



Protein Folding: how does a protein fold?
Are there many possible pathways?
How do mutations affect it?





Chignolin 106 *μs*
cln025 1.0 Å 0.6 *μs*

Trp-cage 208 *μs*
2JOF 1.4 Å 14 *μs*

BBA 325 *μs*
1FME 1.6 Å 18 *μs*

Villin 125 *μs*
2F4K 1.3 Å 2.8 *μs*

WW domain 1137 *μs*
2F21 1.2 Å 21 *μs*

NTL9 2936 *μs*
2HBA 0.5 Å 29 *μs*

BBL 429 *μs*
2WXC 4.8 Å 29 *μs*

Protein B 104 *μs*
1PRB 3.3 Å 3.9 *μs*

Homeodomain 327 *μs*
2P6J 3.6 Å 3.1 *μs*

Protein G 1154 *μs*
1MIO 1.2 Å 65 *μs*

***α*3D** 707 *μs*
2A3D 3.1 Å 27 *μs*

***λ*-repressor** 643 *μs*
1LMB 1.8 Å 49 *μs*

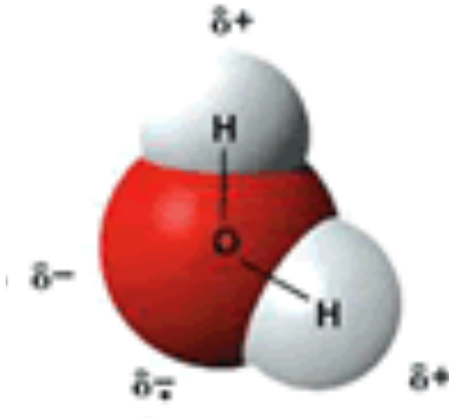
What is a good representation of a molecule?

Molecules are made of atoms that are made of nuclei and electrons, nuclei are made of protons and neutrons, that are made of quarks. Structural biology experiments allow us to see molecules at atomic resolution and in a fixed chemical configuration, this is a molecular perspective (i.e. a molecule is given)

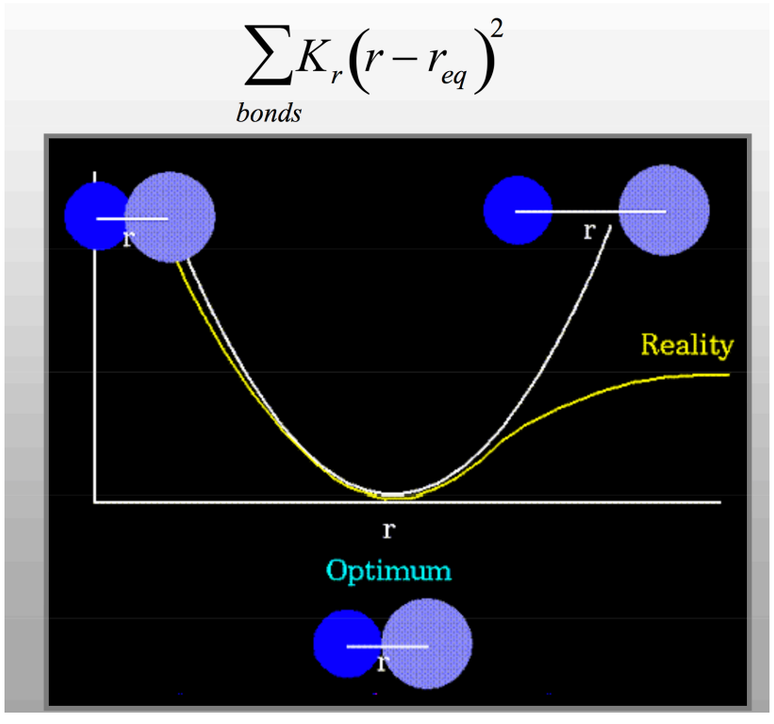
Molecules have covalent properties and non-covalent interactions

Covalent/Geometry

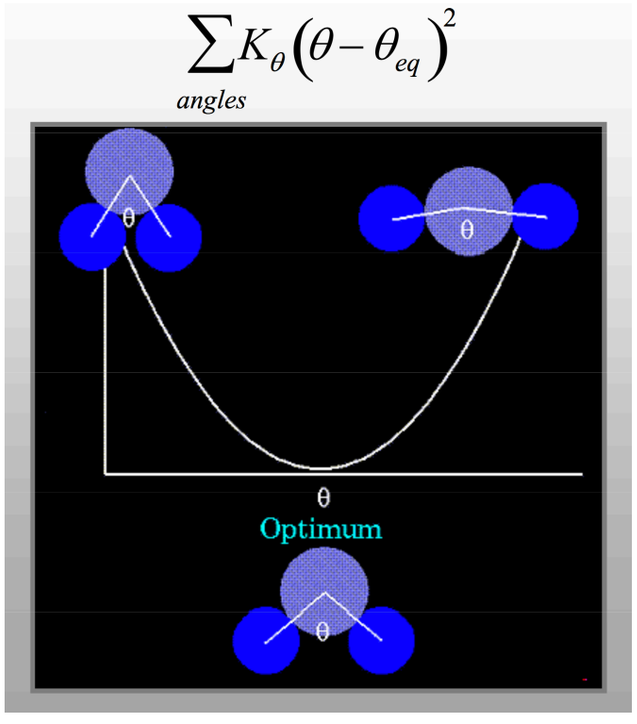
O-H distance: 0.9572 Å
H-O-H angle: 104.52
With fluctuations around these average values



=



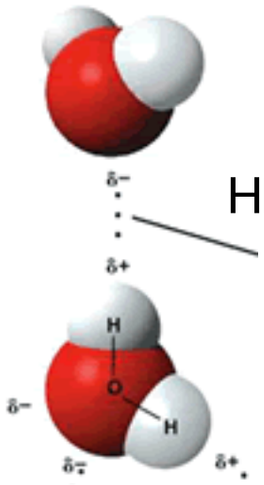
+



What is a good representation of a molecule?

Molecules are made of atoms that are made of nuclei and electrons, nuclei are made of protons and neutrons, that are made of quarks. Structural experiments allow us to see molecules at atomic resolution and in a fixed chemical configuration, this is a molecular perspective (i.e. a molecule is given)

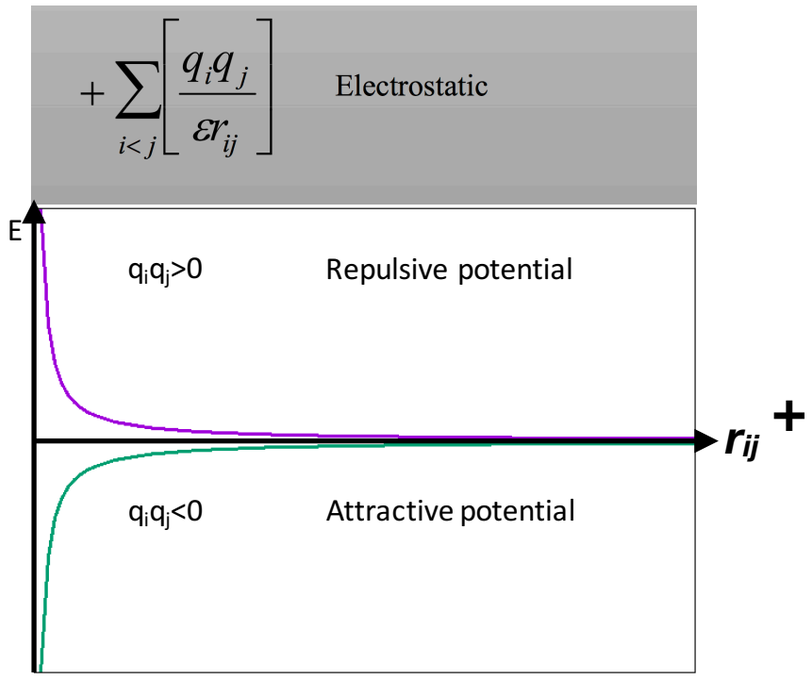
Molecules have covalent properties and non-covalent interactions



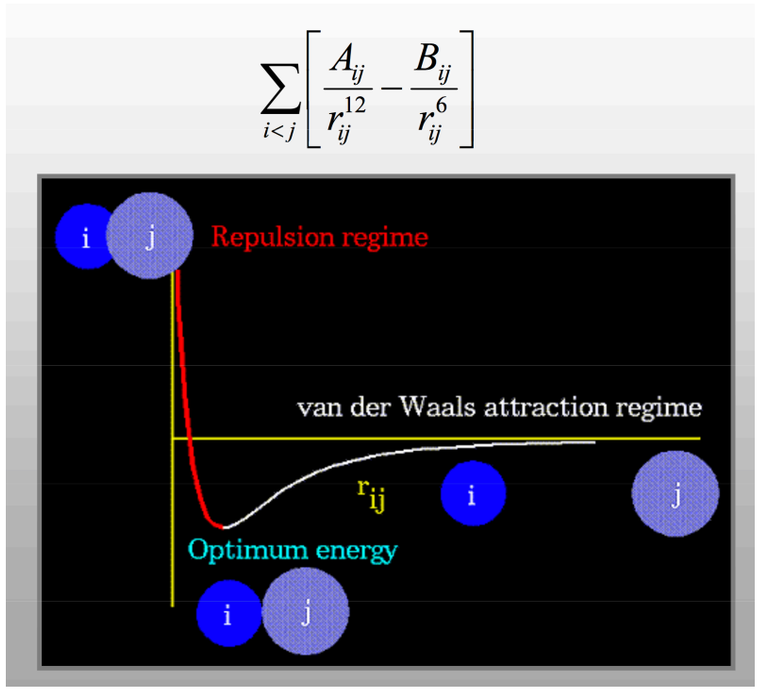
=

Non-covalent interactions

Lone pair electrons that can form hydrogen bonds
The oxygen is negatively charged while hydrogens are positively charged



+



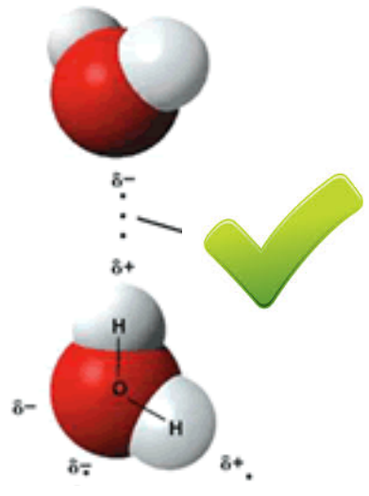
Otherwise an O can overlap with an H

What is a good representation of a molecule?

Are Coulomb and Lennard-Jones interactions enough for non-covalent interactions?

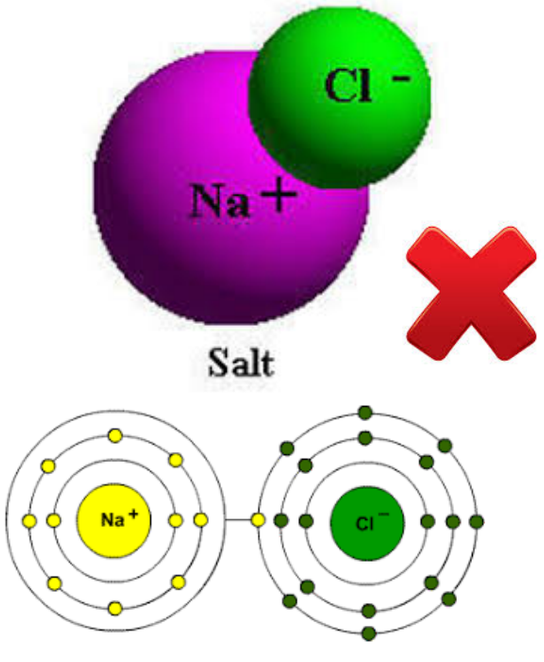
Neutral molecules

Yes, interaction energies between neutral molecules can be reproduced rather accurately (dipolar interaction)



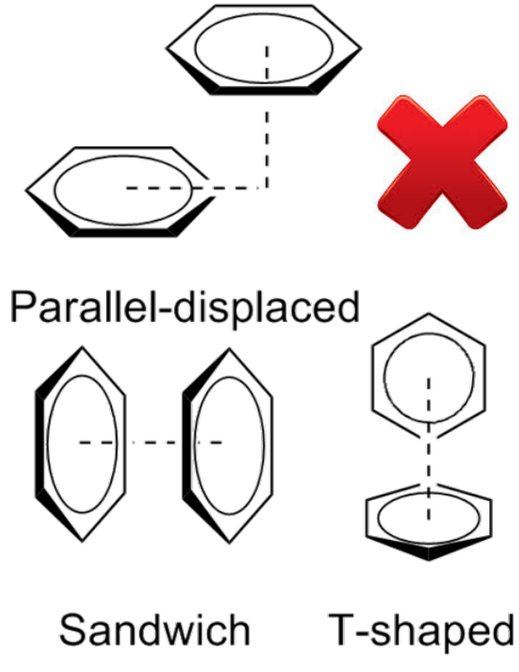
Ionic interactions

No because the charge of an atom is fixed, while in reality it reacts to the environment (polarisation)



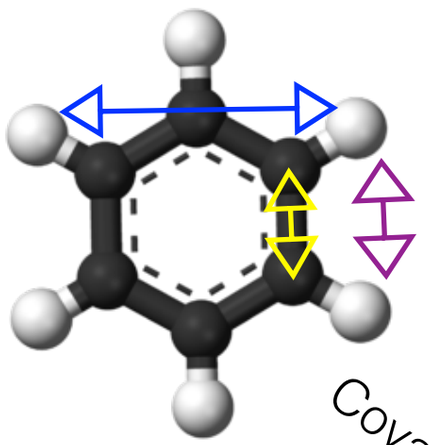
Pi-stacking

Weakly, because the charge of a ring is distributed over and below the ring (quadrupole)



Molecular Mechanics Force Fields

Non-covalent



1-4

Covalent

first approx for vibrations
(anharmonic potential can be used for more accurate vibrations)

geometrical consideration pi-bonds, etc

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1]$$

$$+ \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right]$$

point charge Coulomb

excluded volume

Dispersions (interactions of neutral molecules)

Force Fields

a force field as ~5000-10000 parameters
there are multiple possible solutions

AMBER ✓ CHARMM

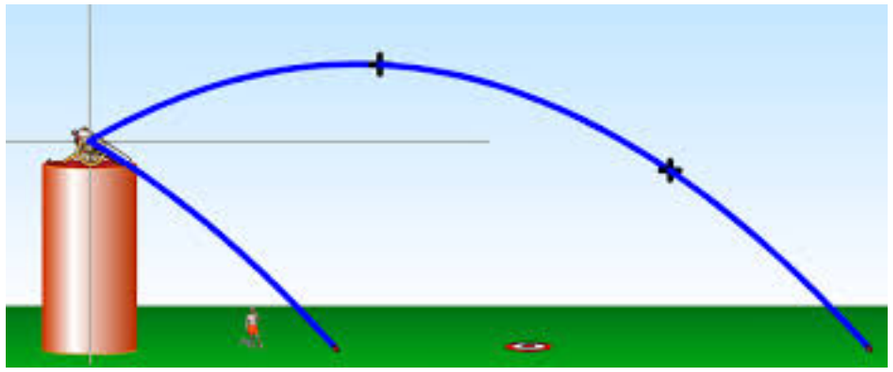
Gromos ✗ OPLS

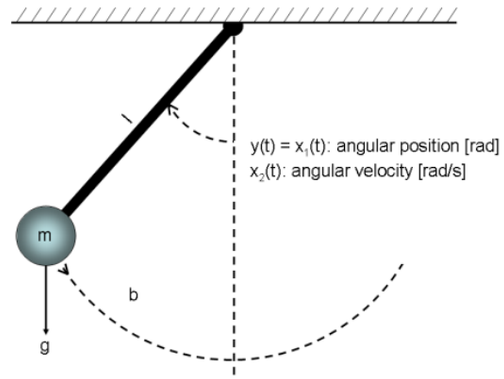
For each family there are multiple generations and variants.
Some of them should be used with a specified water models other can be used with any.

Generally speaking newer force fields are better than the older, but some variants are yet not enough tested… so read and follow the literature

MD, How?

If we want to study how things move, we need to go to physics…





The action of an external force **F** on a body produces an acceleration **a** equal to the force divided by he mass *m* of the body

$$\mathbf{F}(\mathbf{x}) = -\nabla \mathbf{E}(\mathbf{x})$$

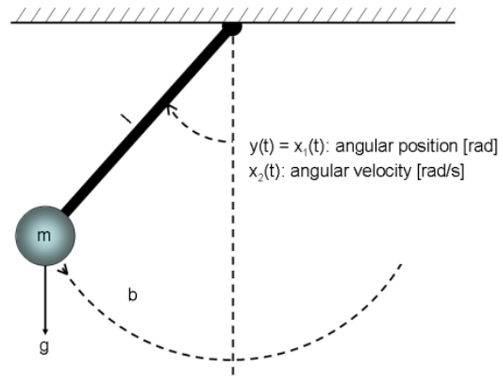
$$\frac{d^2}{dt^2}\mathbf{x} = \frac{\mathbf{F}(\mathbf{x})}{\mathrm{m}}$$

Important Properties

1. If a force does not depend explicitly on time (is conservative) then the Total Energy of the system is conserved.

$$E=U(x)+E_K(v)$$

Total energy is the sum of the potential Energy and the Kinetic energy.

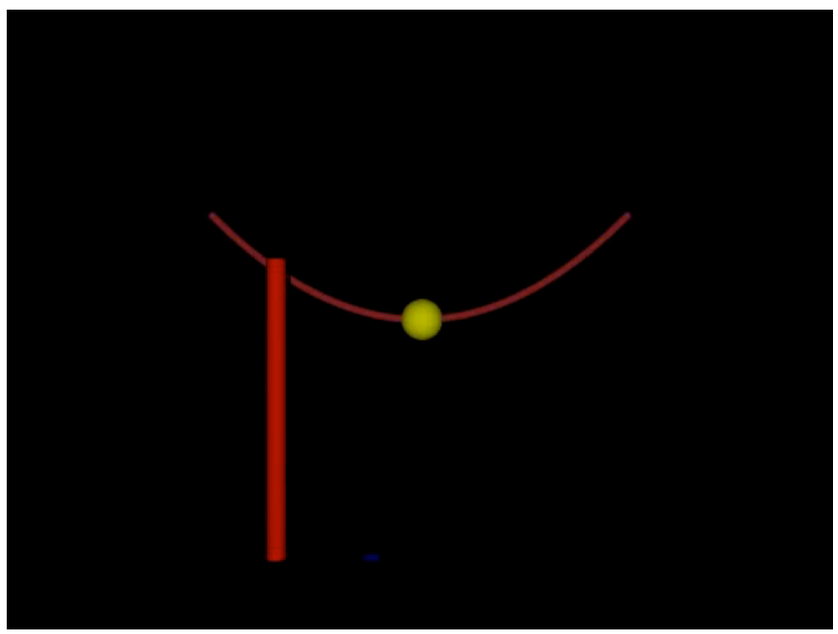


$$E_k = \frac{1}{2}m\boldsymbol{v}^2$$

Kinetic Energy (work needed to accelerate up to speed **v**)

$$\boldsymbol{F}(\boldsymbol{x}) = -\nabla U(\boldsymbol{x})$$

The force is the negative of the gradient of the energy.



2. If a force does not depend explicitly on time (is conservative) then there is a symmetry for time inversion, i.e. going back in time is equivalent to invert the velocities.

The Euler algorithm

Acceleration

$$\boldsymbol{v}(t+\Delta t) = \boldsymbol{v}(t) + \boxed{\frac{d\boldsymbol{v}(t)}{dt}}\Delta t$$

Velocity

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \boxed{\frac{d\boldsymbol{r}(t)}{dt}}\Delta t$$

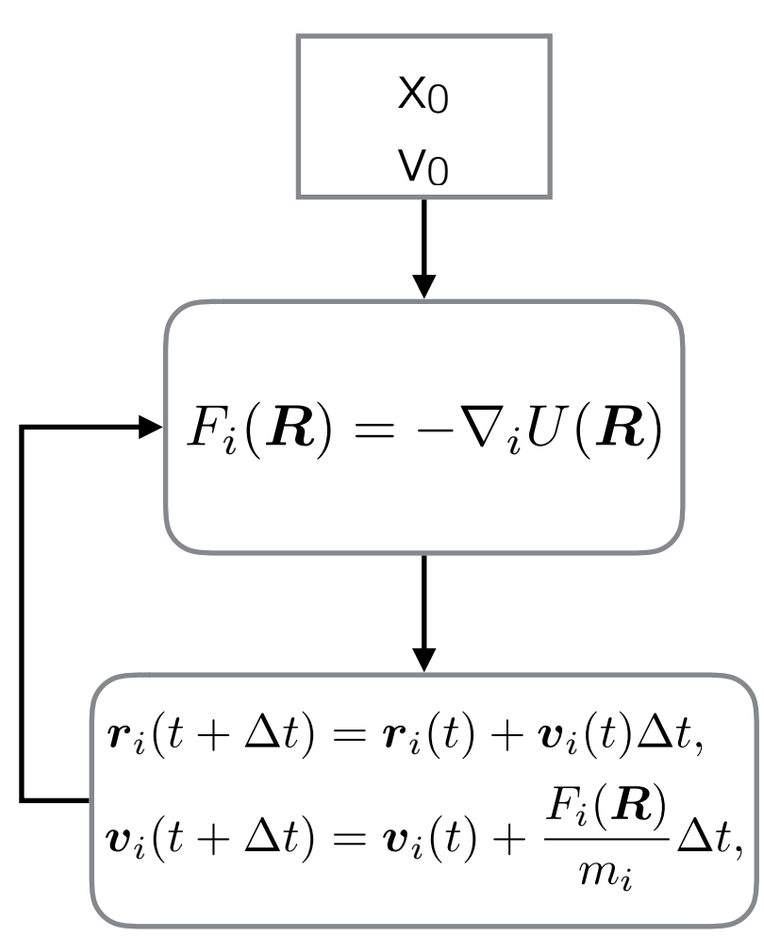
in practice velocities are updated using forces and are used to update positions

$$\boldsymbol{v}_i(t+\Delta t) = \boldsymbol{v}_i(t) + \frac{F_i(\boldsymbol{R})}{m_i}\Delta t,$$

$$\boldsymbol{r}_i(t+\Delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t)\Delta t$$

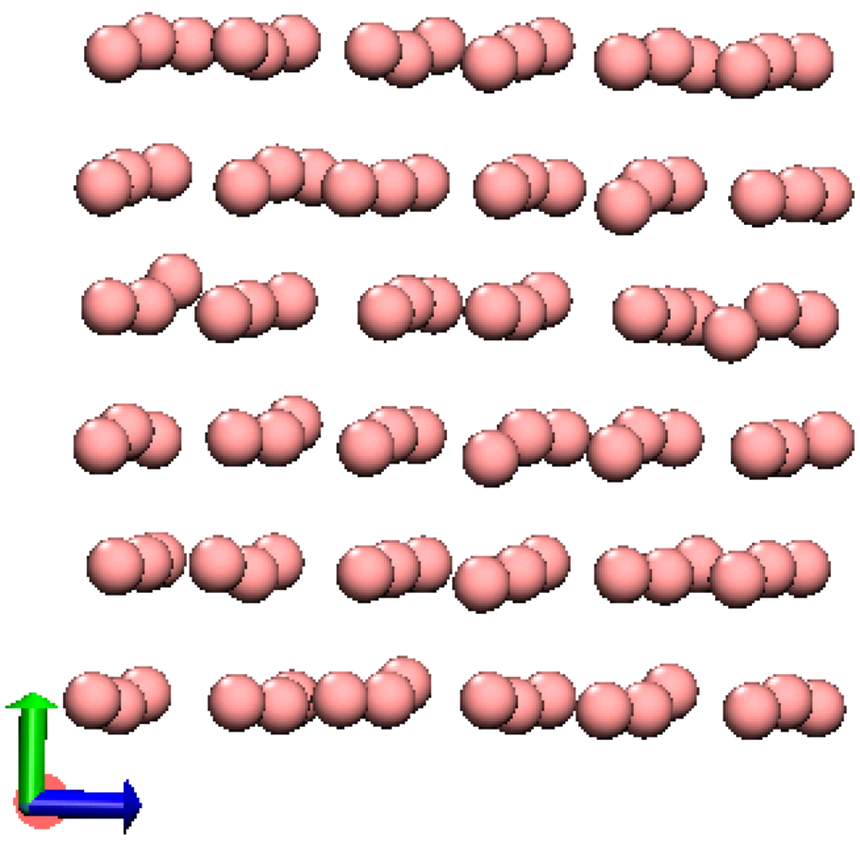
The microscopic world:
Distances in nm
Times in ps
Energies in kJ/mol

1. starting position and velocities
2. calculate forces
3. calculate new positions
4. calculate new velocities

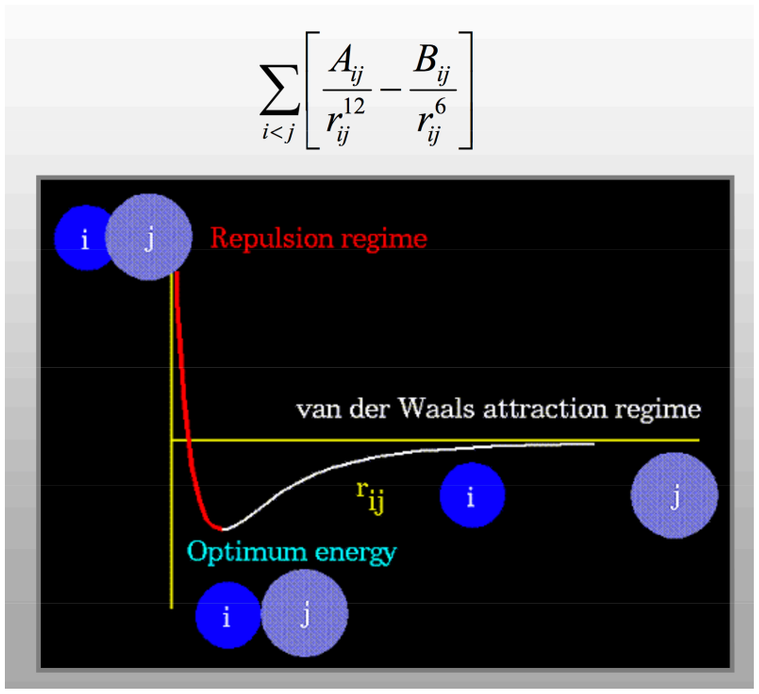


The Euler algorithm

let's take for example a set of 108 particles with LJ interactions (these can represent for example He atoms)

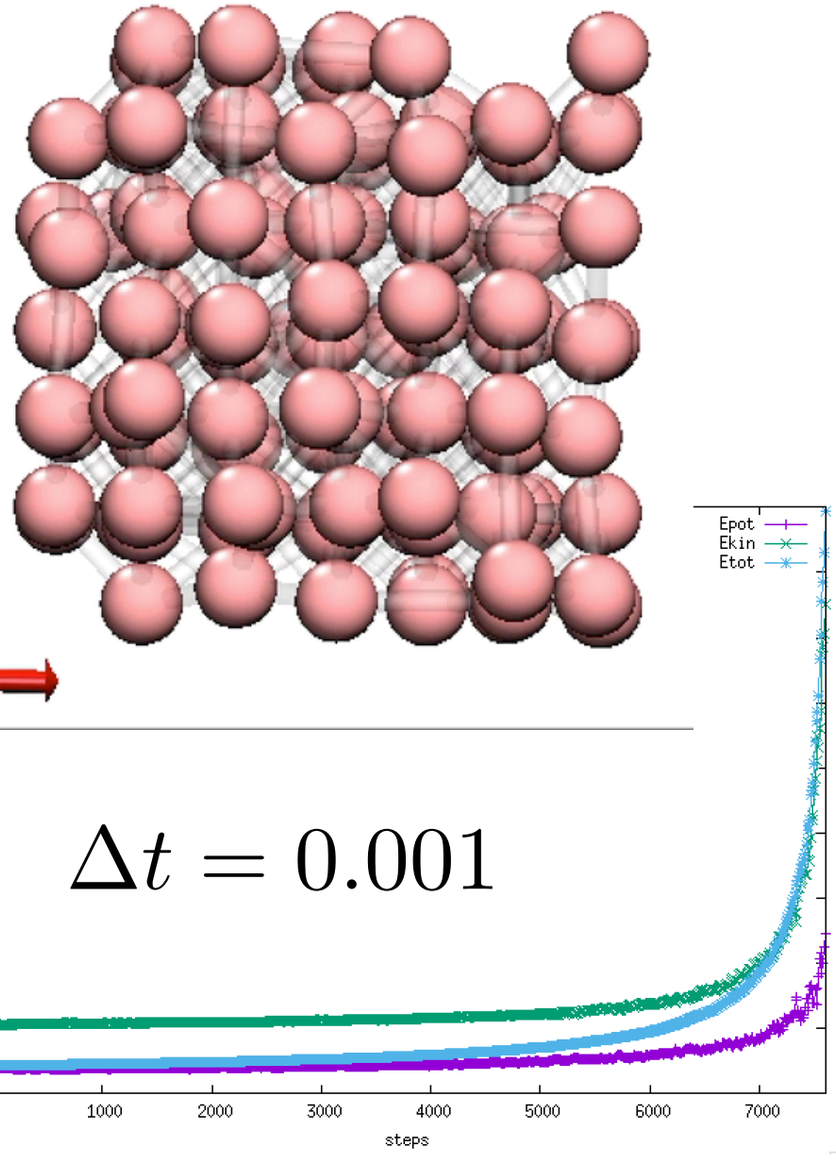


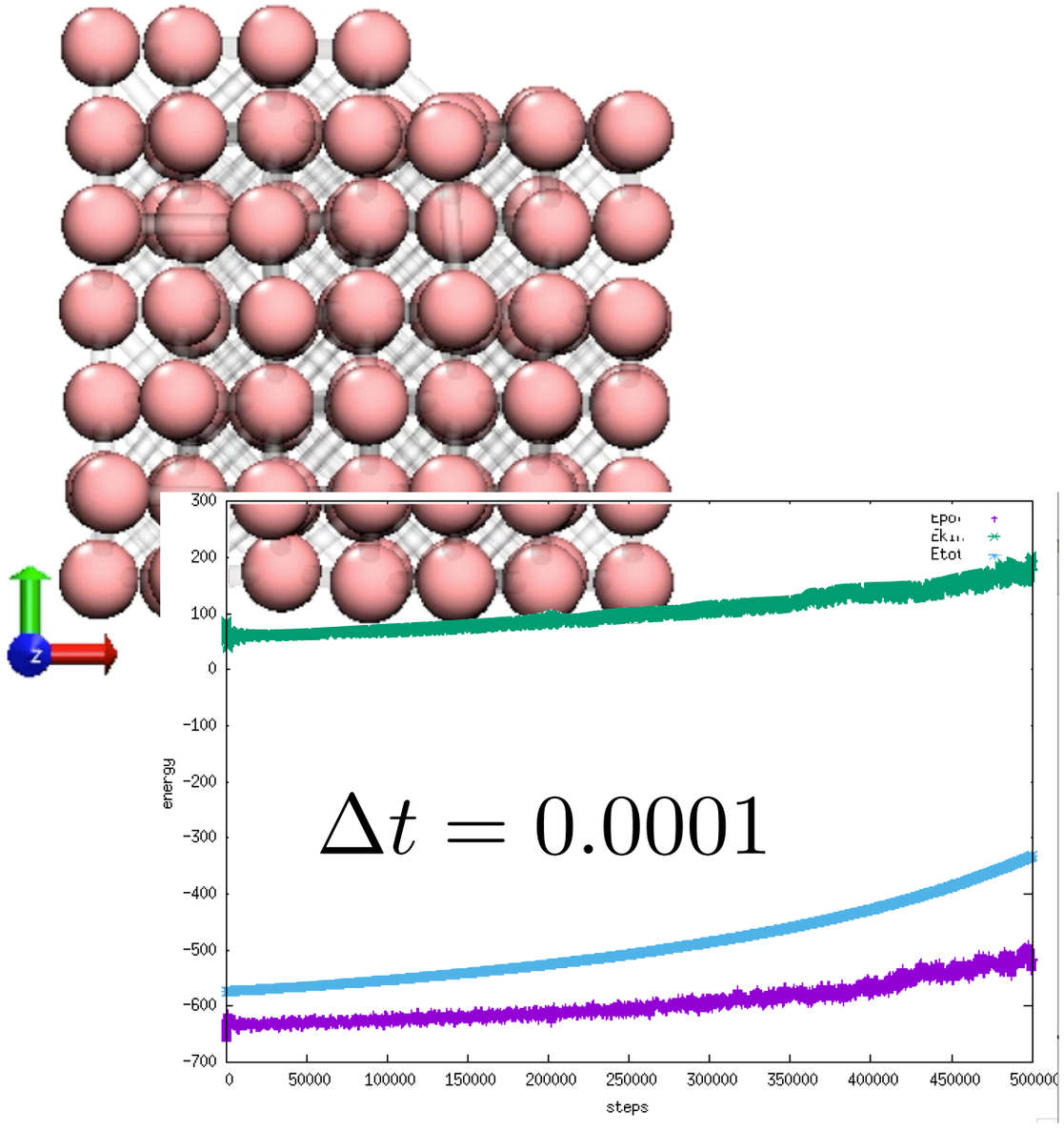
$$\sum_{i<j}\left[\frac{A_{ij}}{r_{ij}^{12}}-\frac{B_{ij}}{r_{ij}^{6}}\right]$$



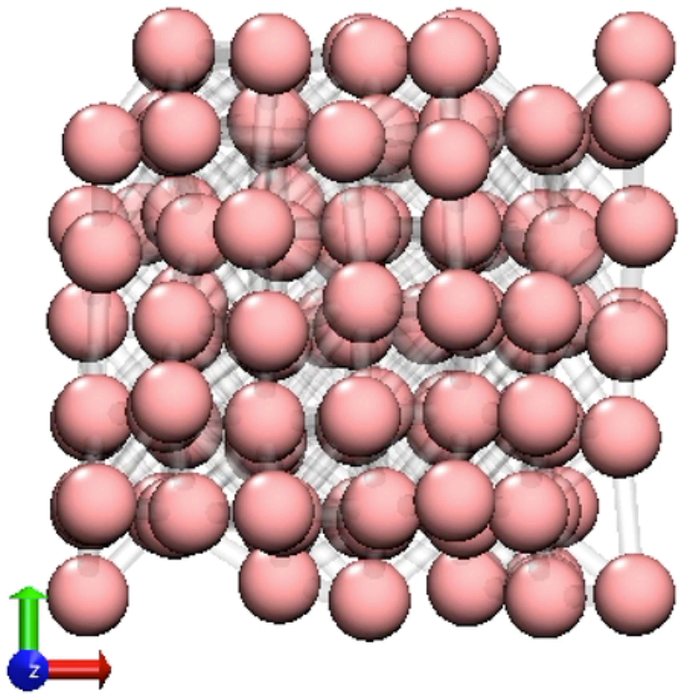
To **test if** the Euler algorithm is a good algorithm we could for example check whether the **total energy is conserved.**

Energy Conservation in Euler

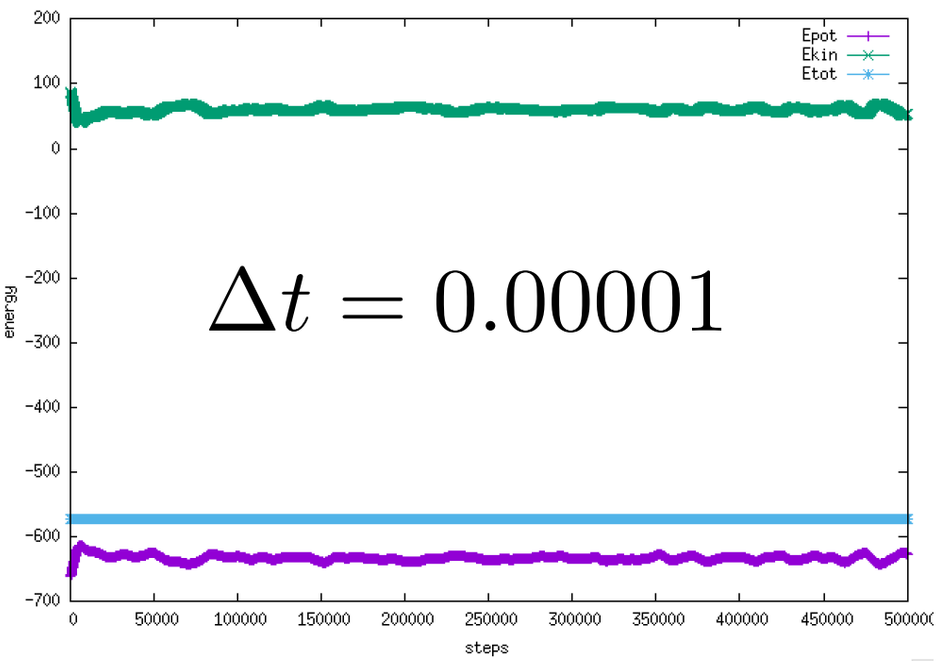




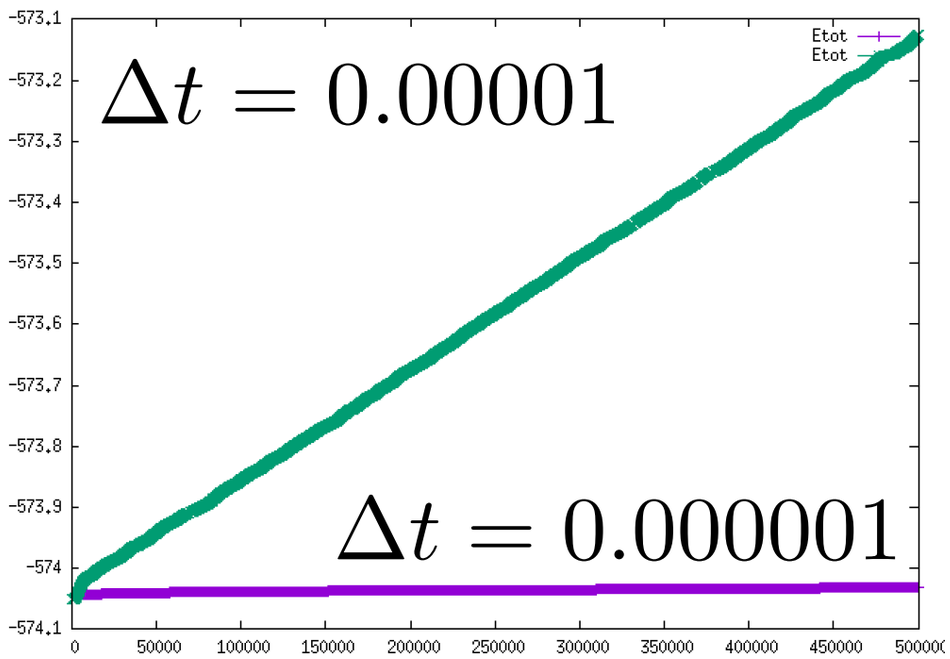
Energy Conservation in Euler



The Euler algorithm cannot conserve the energy, there is not a time step for which the algorithm can work correctly.
Can we explain why?



there is still a systematic drifts



Why Euler does not work?

The Euler algorithm cannot conserve the energy, there is not a time step for which the algorithm can work correctly. Can we explain why? There is a second important property in classical physics: **time reversibility**.

What happen if we start in r(t+dt) and we go back in time by dt?

$$
\begin{aligned}
r(t+\Delta t-\Delta t) &= r(t+\Delta t) - v(t+\Delta t)\Delta t = \\
&= r(t) + v(t)\Delta t - v(t+\Delta t)\Delta t = \\
&= r(t) + v(t)\Delta t - v(t)\Delta t - a(t)\Delta t^2 = \\
&= r(t) - a(t)\Delta t^2 \neq r(t)
\end{aligned}
$$

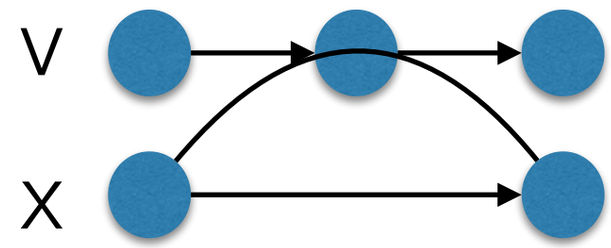
We are not back in r(t)! So this algorithm is not time-reversible!

A better algorithm: velocity Verlet

$$\vec{v}\left(t+\tfrac{1}{2}\,\Delta t\right)=\vec{v}(t)+\tfrac{1}{2}\,\vec{a}(t)\,\Delta t.$$

$$\vec{x}(t+\Delta t)=\vec{x}(t)+\vec{v}\left(t+\tfrac{1}{2}\,\Delta t\right)\,\Delta t.$$

$$\vec{v}(t+\Delta t)=\vec{v}\left(t+\tfrac{1}{2}\,\Delta t\right)+\tfrac{1}{2}\,\vec{a}(t+\Delta t)\Delta t.$$



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

$$\vec{v}(t+\Delta t)=\vec{v}(t)+\frac{\vec{a}(t)+\vec{a}(t+\Delta t)}{2}\Delta t.$$

Calculate more accurate velocities because then errors go on positions

The algorithm

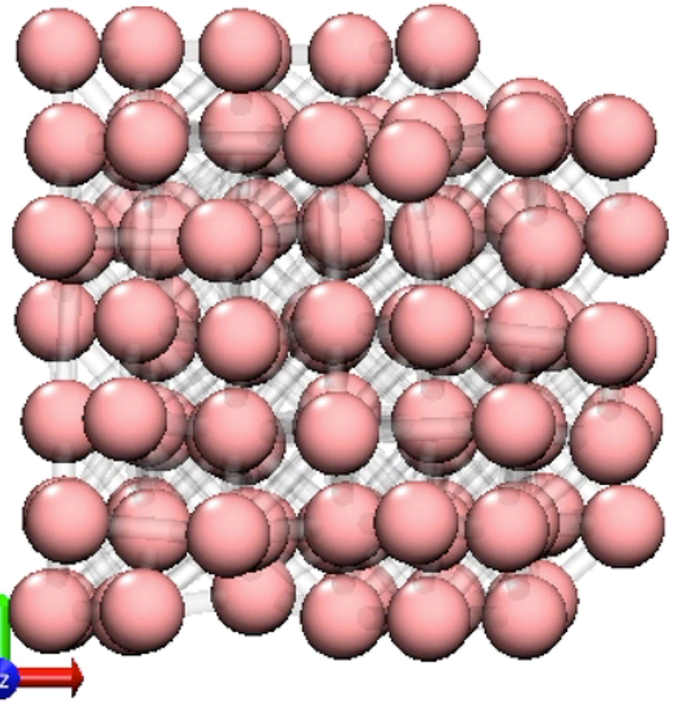
Start: x(0), v(0), f(0)

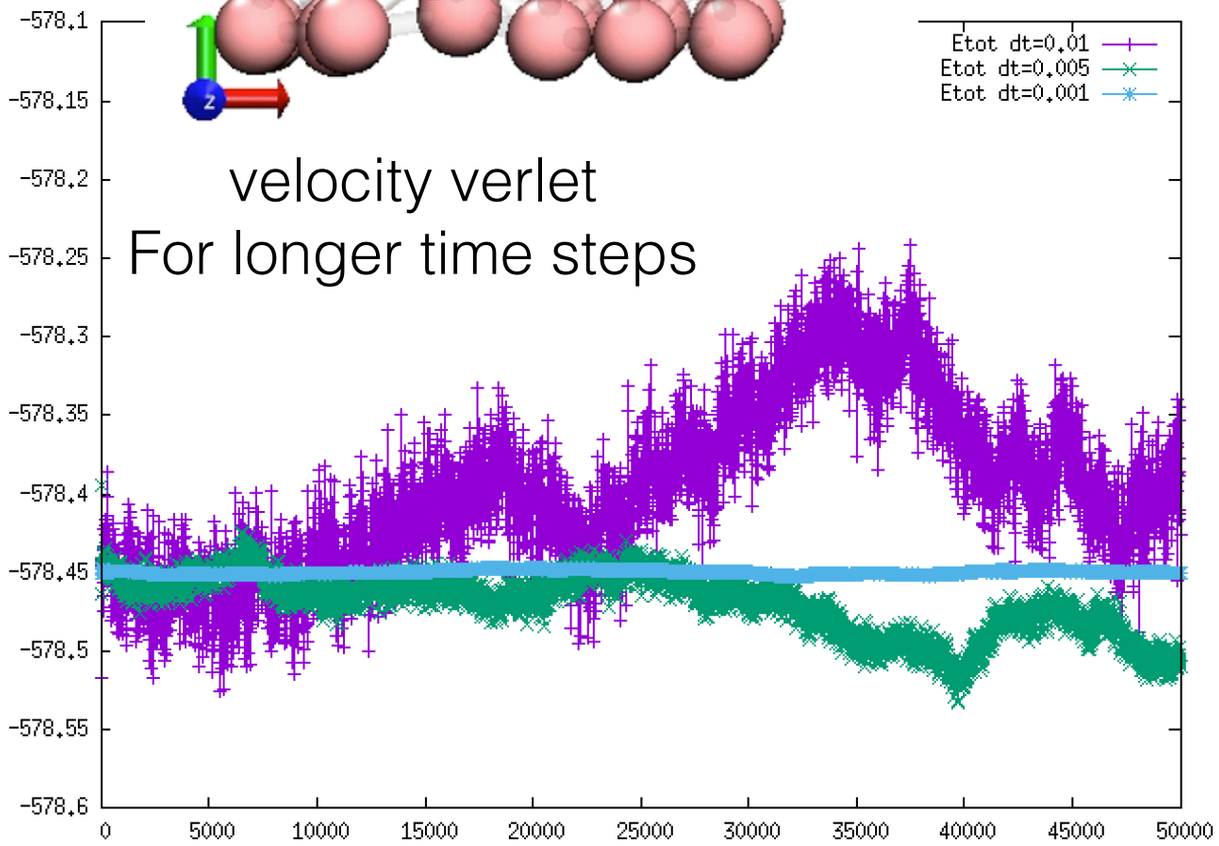
1. calculate velocities at time t+0.5dt: v(t+0.5dt)
2. calculate positions at time t+dt: x(t+dt)
3. calculate forces at time t+dt: f(t+dt)
4. calculate velocities at time t+dt: v(t+dt)

Proof of VV time reversibility

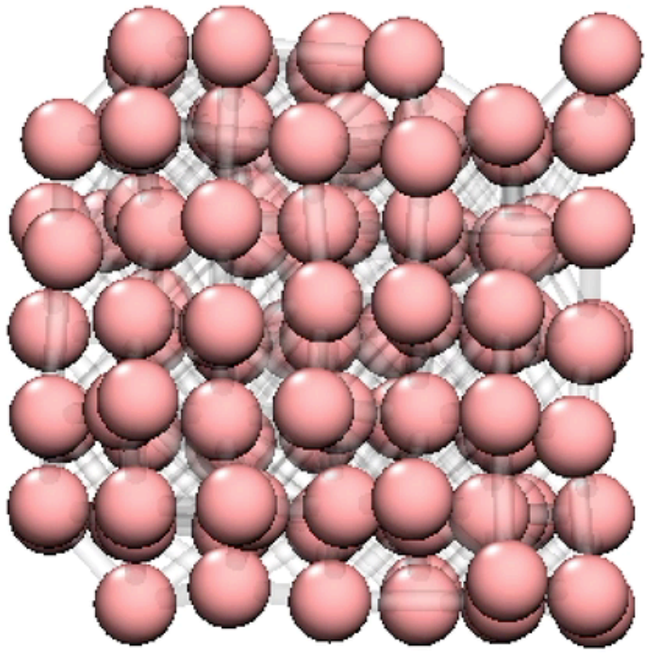
$$\begin{aligned}
r(t+\Delta t-\Delta t) &= r(t+\Delta t)-v(t+\Delta t)\Delta t+\frac{1}{2}a(t+\Delta t)\Delta t^2= \\
&= r(t)+v(t)\Delta t+\frac{1}{2}a(t)\Delta t^2-v(t)\Delta t+ \\
&-\frac{1}{2}a(t)\Delta t^2-\frac{1}{2}a(t+\Delta t)\Delta t^2+\frac{1}{2}a(t+\Delta t)\Delta t^2= \\
&= r(t)
\end{aligned}$$

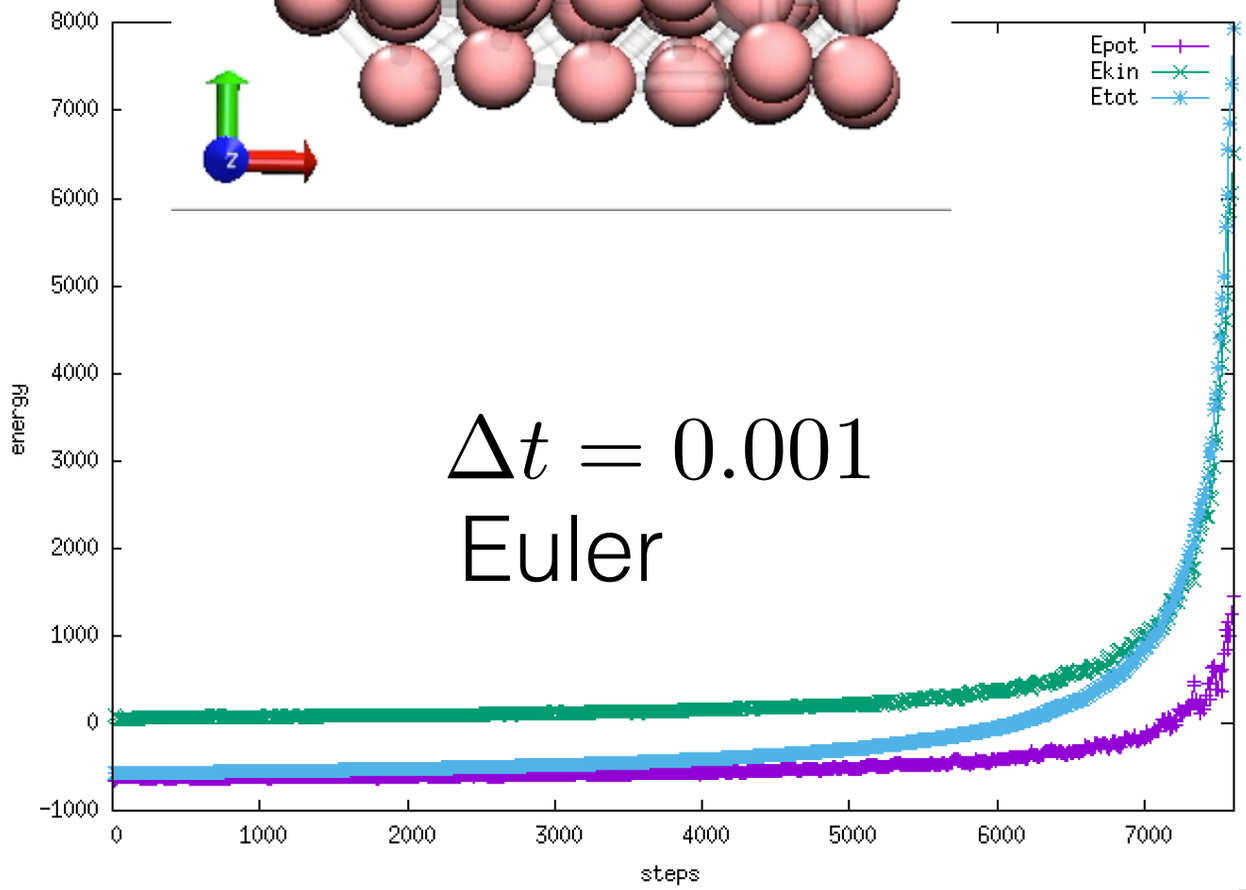
Velocity Verlet





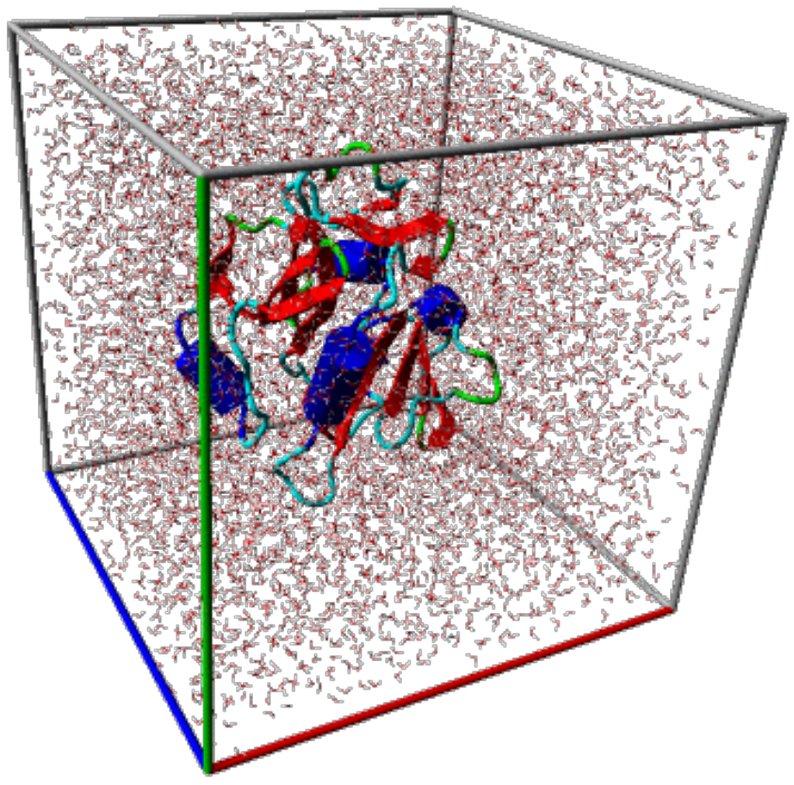
velocity verlet
For longer time steps





$\Delta t = 0.001$
Euler

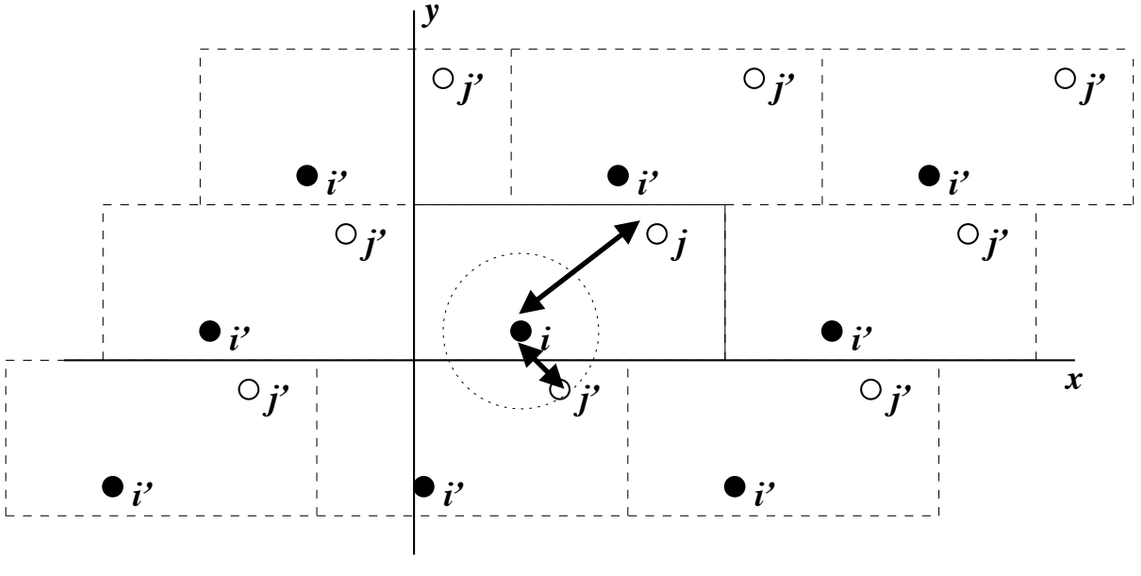
The simulation box



$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1]$$

$$+ \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right]$$

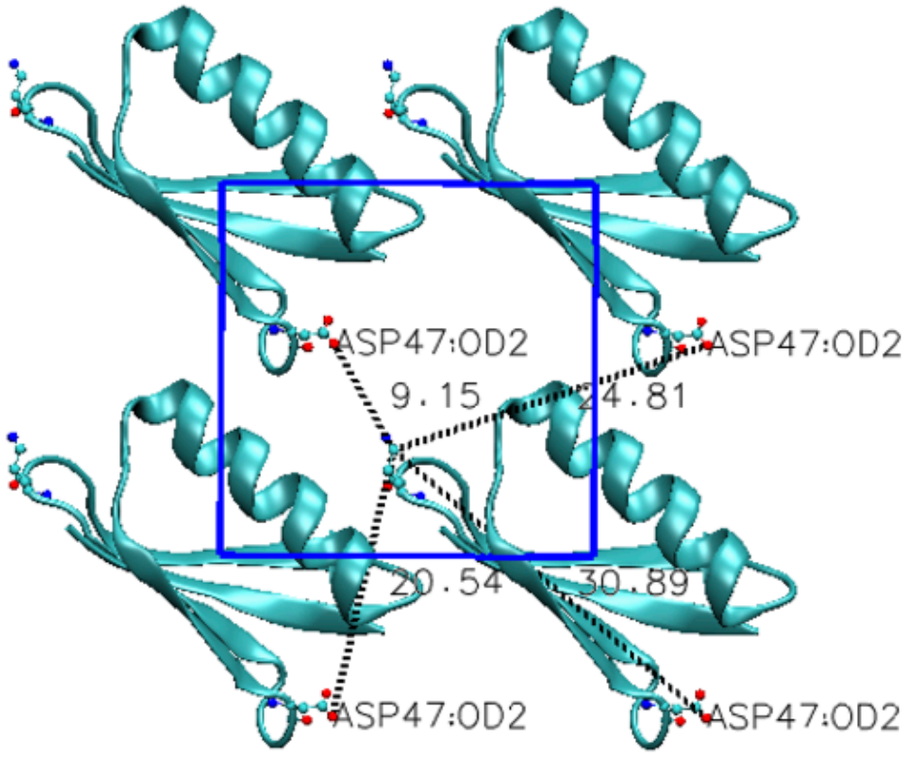
We need to calculate distances between atoms (N^2 calculations). So how big should the box be?



Minimum Image Convention

- the distance between to atoms is always the minimum with respect to all the neighbour images
- interactions can only be calculated up to a distance equal to half of the shortest side of the box

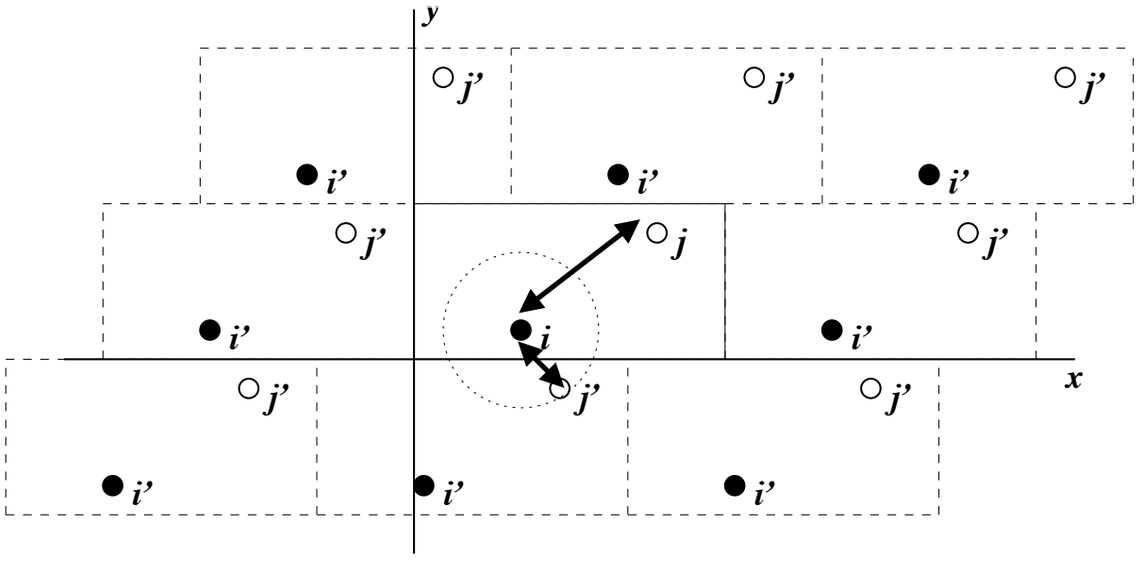
The simulation box



$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1]$$

$$+ \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right]$$

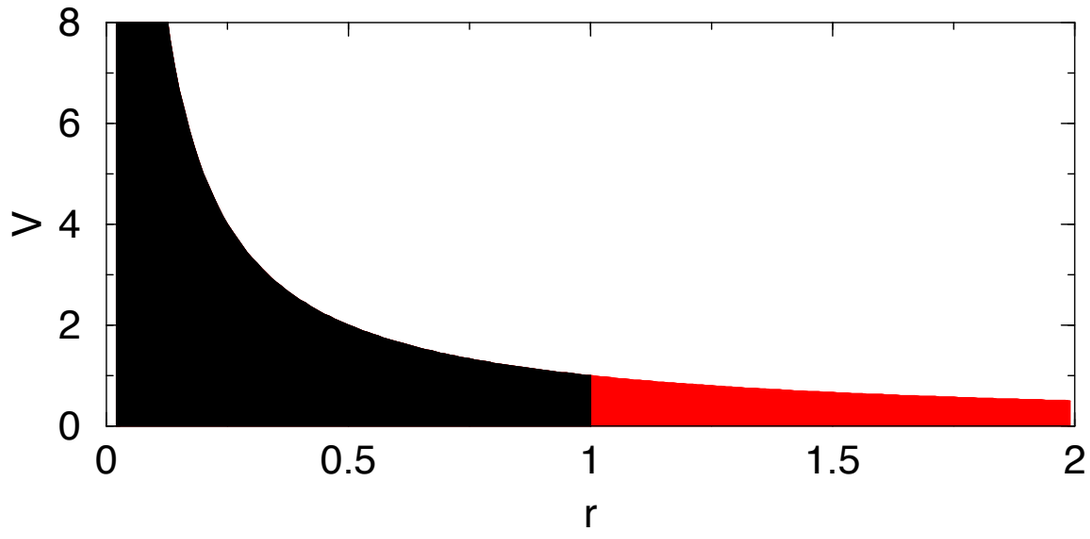
We need to calculate distances between atoms (N^2 calculations). So how big should the box be?



Minimum Image Convention

- the distance between to atoms is always the minimum with respect to all the neighbour images
- interactions can only be calculated up to a distance equal to half of the shortest side of the box

Coulomb Interactions



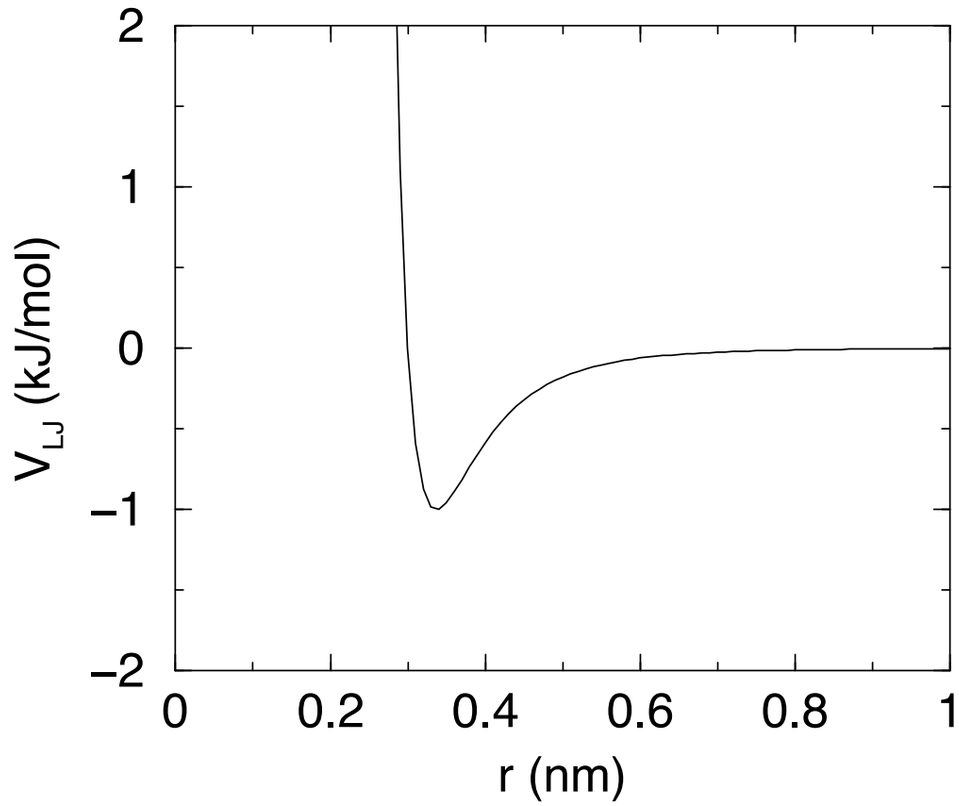
Cutting off interactions abruptly causes artefacts:

- In water: over-orientation of dipoles: anti-parallel before the cut-off, parallel beyond the cut-off
- Like charged ions will accumulate just beyond the cut-off

Some artefacts get worse with increasing cut-off distance!!!

This is because Coulomb goes as $1/r$, but the surface of the cut-off sphere goes as r^2

VdW interactions



LJ decays fast: $-r^{-6}$

but the dispersion contributes up to relatively long distances:

$$V_{disp} = \int_{r_c}^{\infty} \rho_n \langle C^{(6)} \rangle \, 4\pi \, r^2 \, r^{-6} \mathrm{d}r = -\rho_n \langle C^{(6)} \rangle \frac{4\pi}{3} \, r_c^{-3}$$

this is correct only for systems that can be considered homogenous farther than the cut-off (check the radial distribution function, is it 1 after the cut-off?)

Dealing with Temperature

If the goal of MD is to calculate an average quantity it does not matter if we calculate it at constant energy, or temperature, or other thermodynamic ensembles. In the thermodynamic limit average quantities do not depend on this choice. Fluctuation do, so if we are interested in probability distributions we need to chose the experimental conditions.

$$\langle K \rangle = \frac{3}{2} NkT$$

Equipartition Theorem gives us a definition for the temperature

We want an algorithm to obtain the correct average kinetic energy with an accuracy that is the standard error of the mean over our number of particles.

$$K_0 = \frac{3}{2} N_f kT,$$

$$K(t+\Delta t) = K(t) + (K_0 - K(t))\frac{\Delta t}{\tau_T} + 2\sqrt{\frac{K(t)K_0}{N_f}}\frac{dW}{\sqrt{\tau_T}},$$

Ornstein-Uhlenbeck process: a dynamic process whose equilibrium distribution is a gaussian.

$$v(t+\Delta t) = v(t+\Delta t)\sqrt{\frac{K(t+\Delta t)}{K(t)}}$$

At step 0 compute forces **{x(0), v(0), f(0)}**

1. update kinetic energy
2. update velocities
3. update positions
4. calculate forces
5. update velocities
6. update kinetic energy

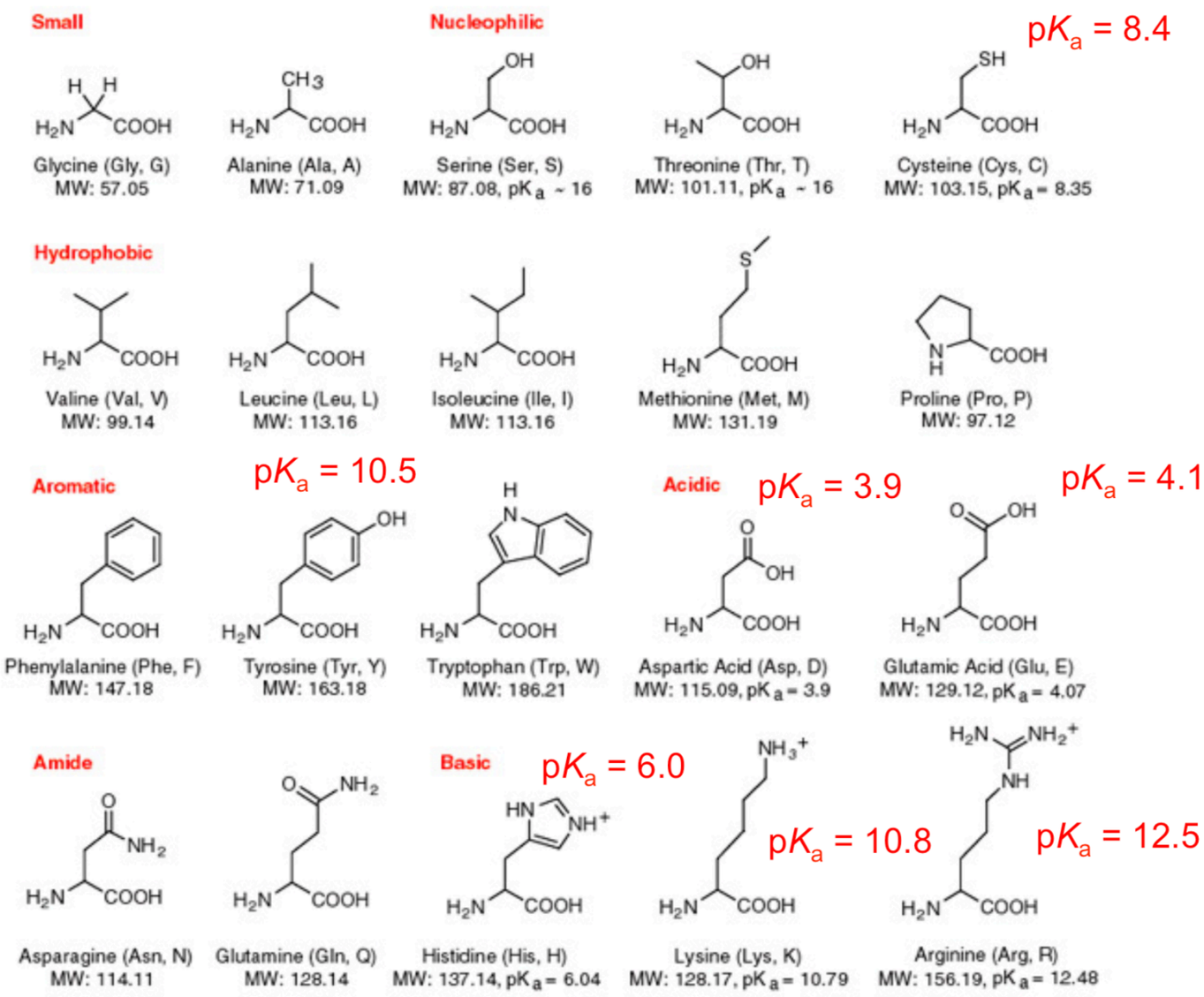
From the beginning…

Structure preparation: correct sequence, missing hydrogens, etc

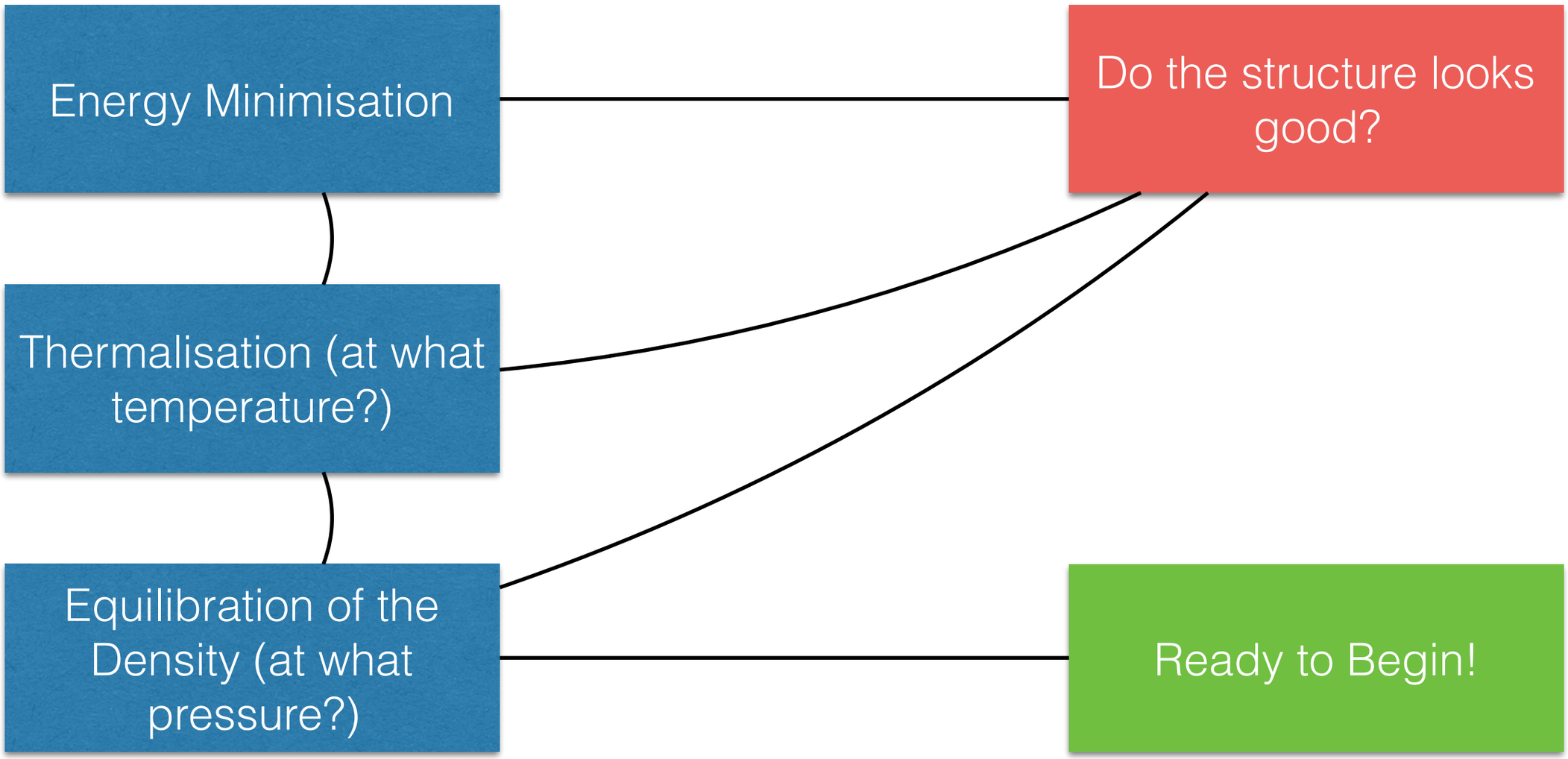
pH, set the protonation state of amino acids

Box size, is it too small? Is the box too large?

Solutions conditions, water, denaturants, salts, others



From the beginning…



From the beginning…

Ready to Begin!?

Can I observe the process I want multiple times on the time scale of my MD?

Too which extent the answer I am looking for could depend on my former choices?

Enhanced Sampling
Coarse Graining/
Simple models

Too bad…

Design of a control simulation/experiment

Analysis

In a MD simulation you **calculate position, velocities, forces for all the atoms** of a system for **every time-step**, once again with a simulation of 100AA (~1000 Atoms) in a box with 5000 water molecules (15000 atoms) this means 16000 atoms, that means 16000x(3+3+3)=144000 numbers per frame of simulation. This numbers uses 4 bytes each, this means 0.5Mb per frame.

If we run at 200ns/day, with a time step 0.002 ps, this means 100,000,000 frames per day, that means 50,000,000Mb/day = 50Tb/day ~ 5Tb/day if you zip the data.

We will see later why we don't need to save all these data, even if we should be careful about choosing the frequency at which we save the information

Distances: between any couple of atoms, or centre of masses of groups of atoms

RMSD: root mean square deviation of the positions with respect to a reference configuration

RMSF: root mean square fluctuations of the positions with respect to a reference configuration

Gr: Gyration radius, the distribution of the atoms around the centre of mass

DIH: dihedral angles

H-bonds: distance and angle

Coordination: how many atoms are closer than some distance to another atom

Analysis

Given that we have positions, velocities and forces for all the atoms this means that we can analyse any property that can be defined using these quantities:

If we run at 200ns/day, with a time step 0.002 ps, this means 100,000,000 frames per day, that means 50,000,000Mb/day = 50Tb/day ~ 5Tb/day if you zip the data.

We will see later why we don't need to save all these data, even if we should be careful about choosing the frequency at which we save the information

Surface: how large is the surface of the system

Density: the ratio between mass and volume

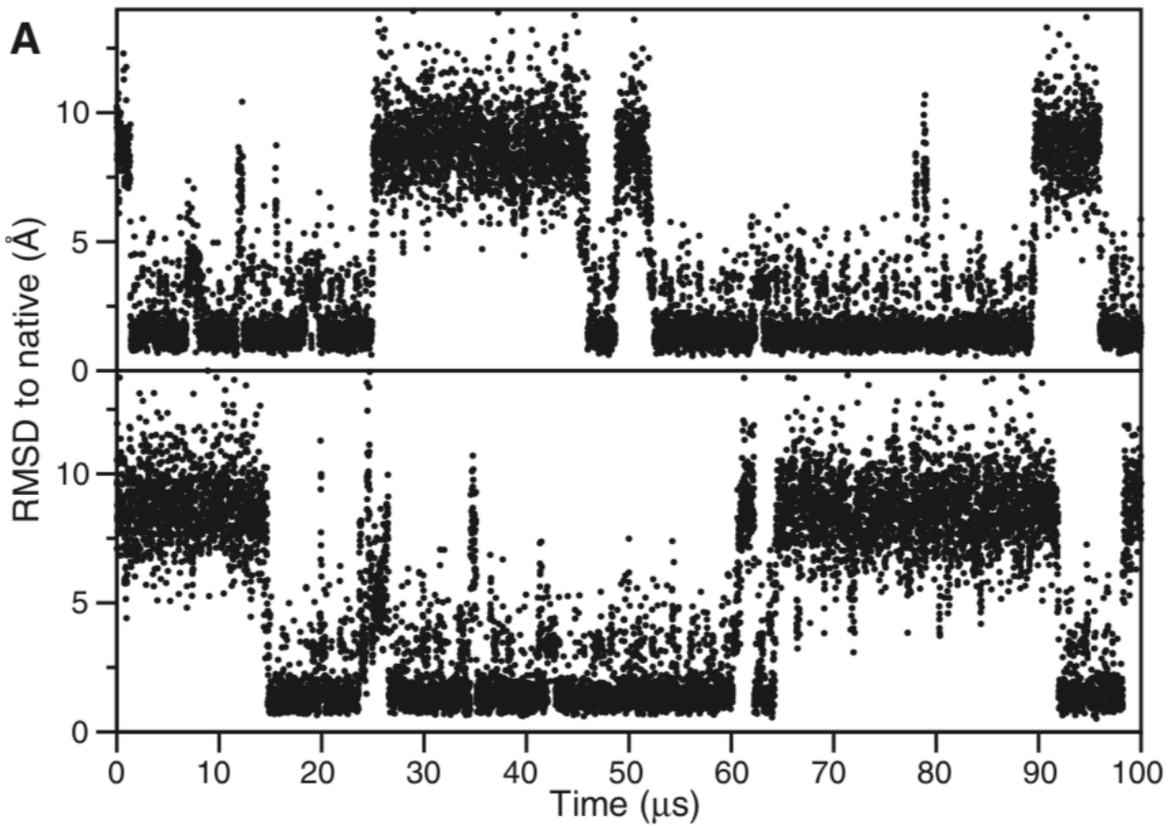
NMR: quantities like JCoupling, NOE and chemical shifts can be calculated using approximate functions

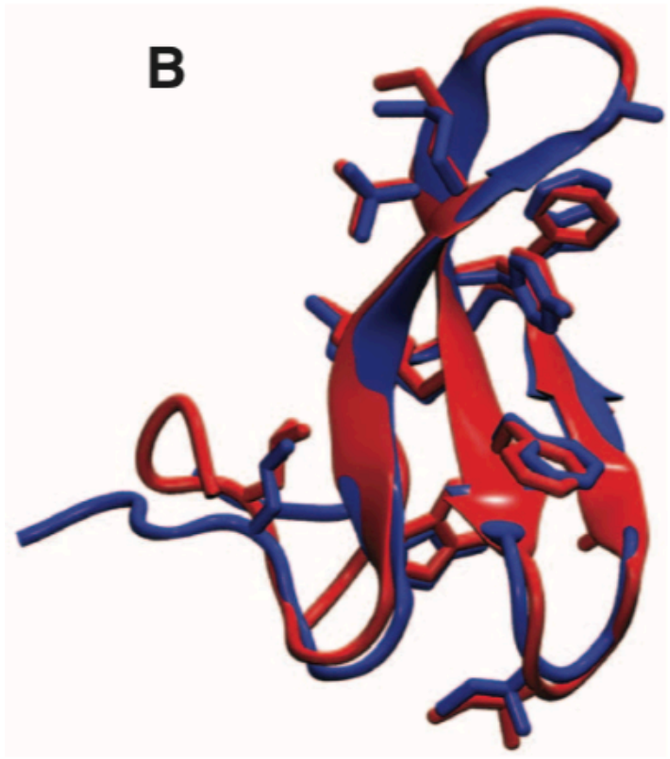
SAXS: root mean square deviation of the positions with respect to a reference configuration

But the key point is that we have the TIME EVOLUTION of these quantities

Analysis

But the key point is that we have the TIME EVOLUTION of these quantities

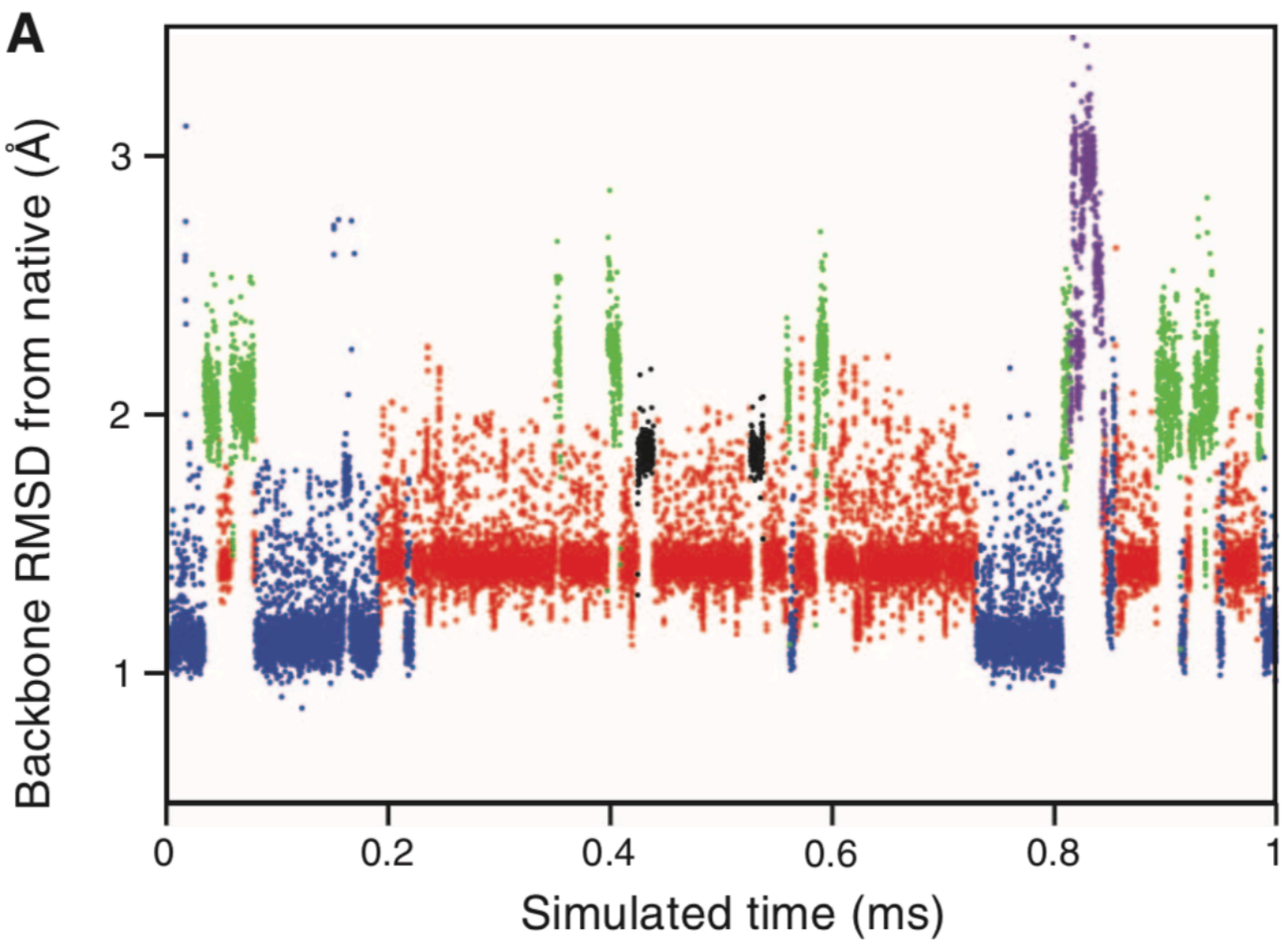


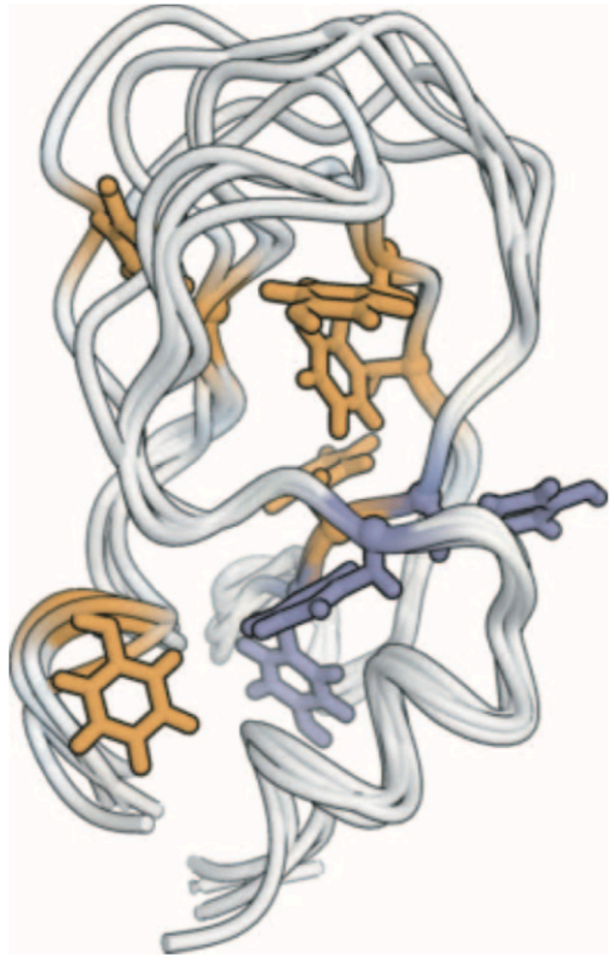


$$RMSD = \sqrt{\frac{1}{N}\sum_{i}^{N}(r_i - r_i^{ref})^2}$$ **After superimposition**

Analysis

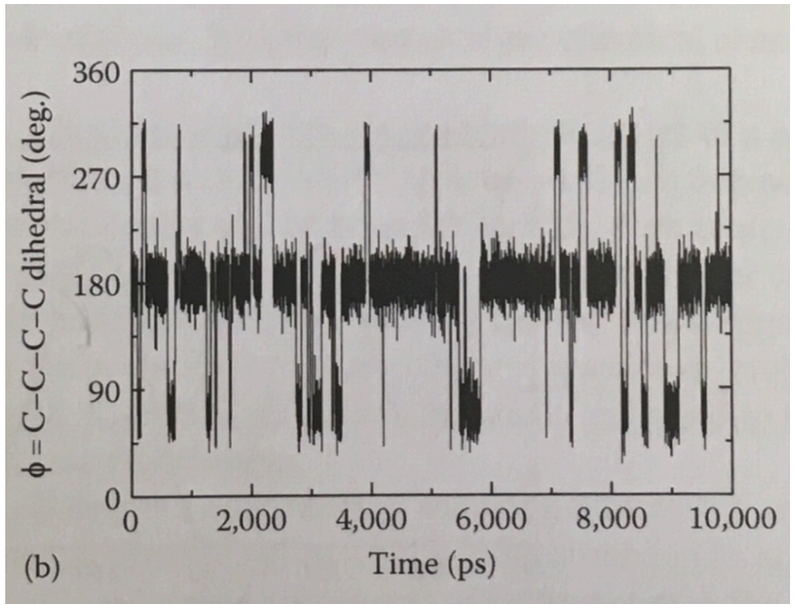
But the key point is that we have the TIME EVOLUTION of these quantities

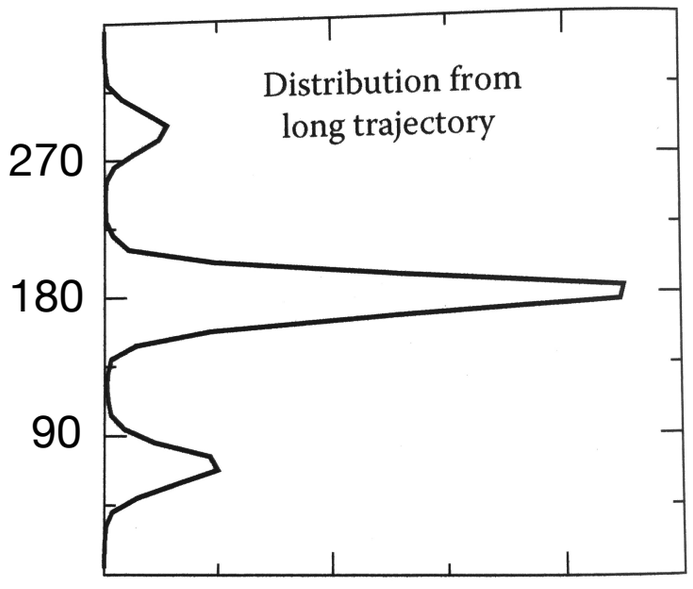


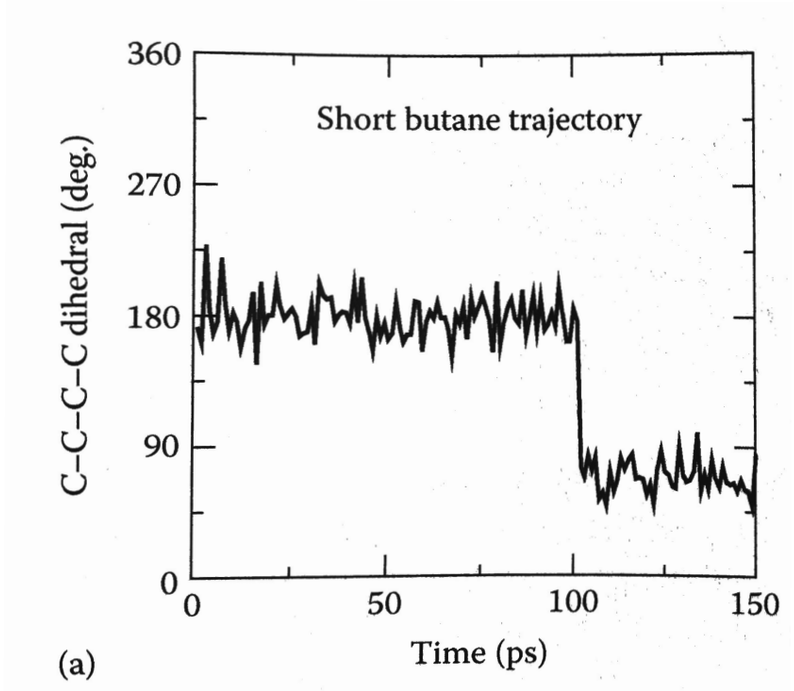


Time Averages and Correlated data

Time averages are the same of ensemble averages if the time is long enough.







This because time evolution is correlated, that is two successive frames are not independent, but actually the latter is a consequence of the former

Time correlation

Autocorrelation function is a correlation:

$$< x(t_1)x(t_2) > - < x(t_1) >< x(t_2) >\neq 0$$

Let's now consider time differences, $\tau = t_2 - t_1$

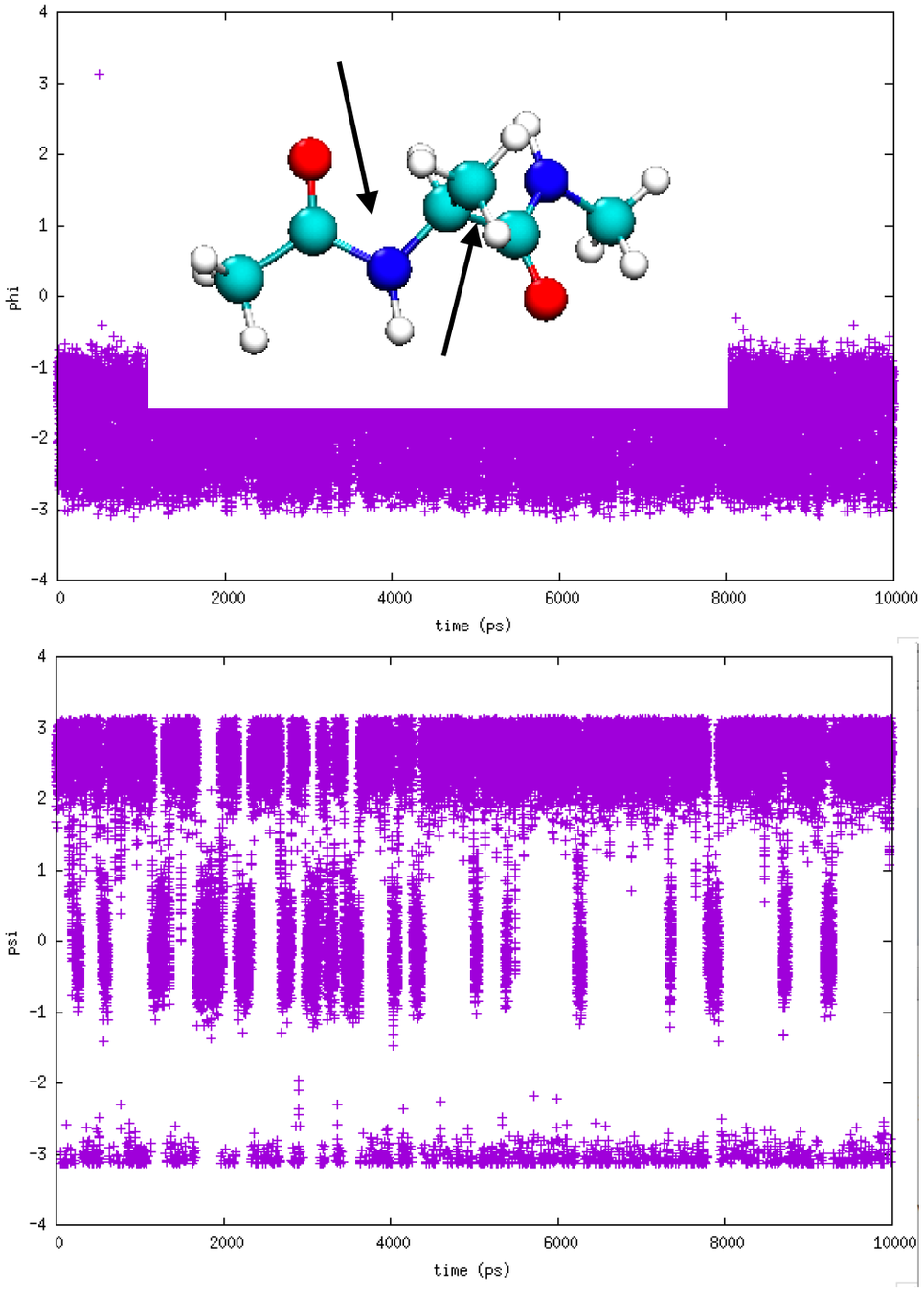
And consider the average over all times *t*

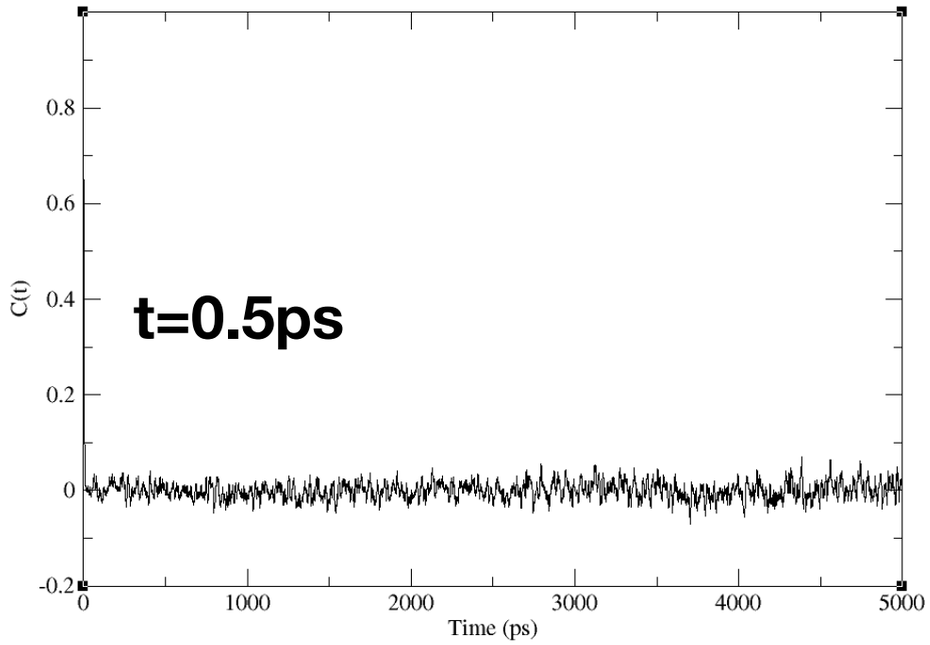
$$C_x(\tau) = \frac{< x(t)x(t+\tau) > - < x(t) >< x(t+\tau) >}{\sigma_x^2}$$

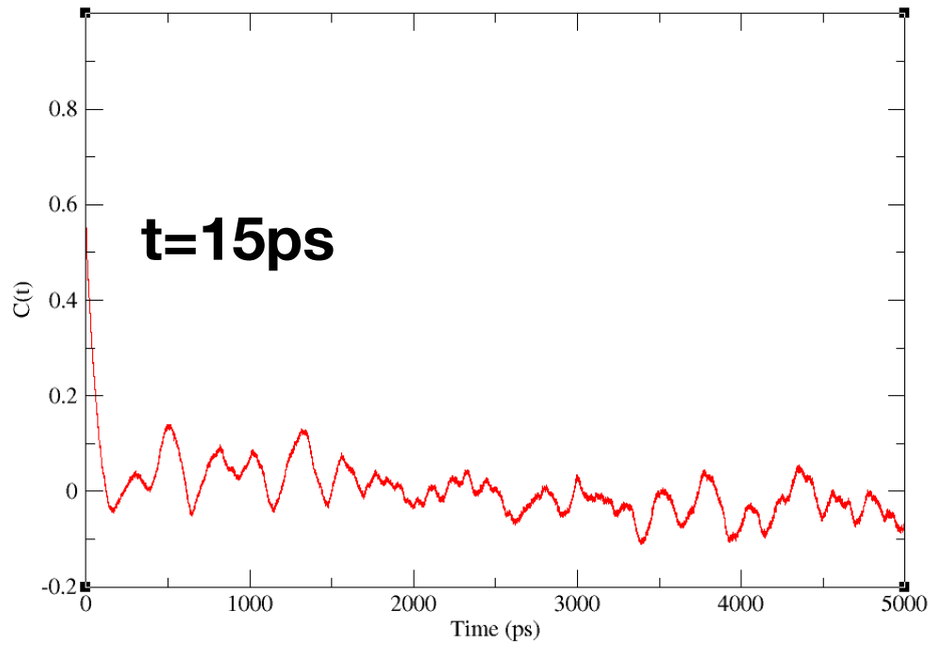
For tau=0 C(tau)=1

Correlation time is key to determine whether we are allowed or not to calculate averages over time

Time correlation



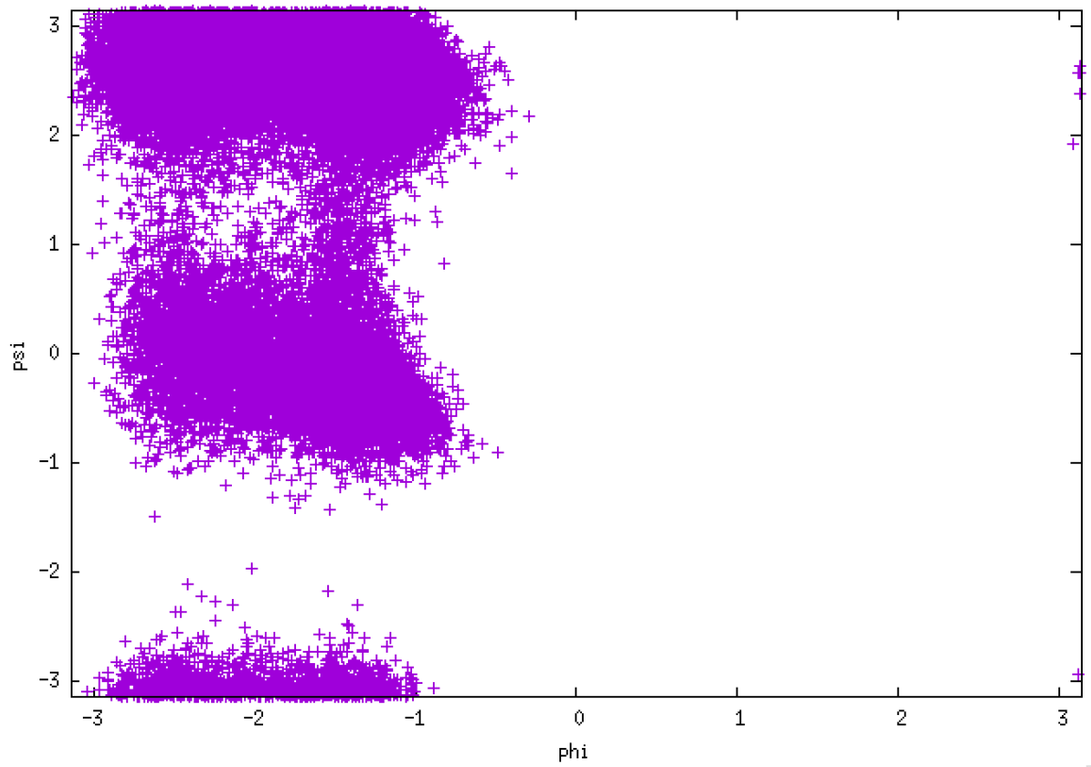


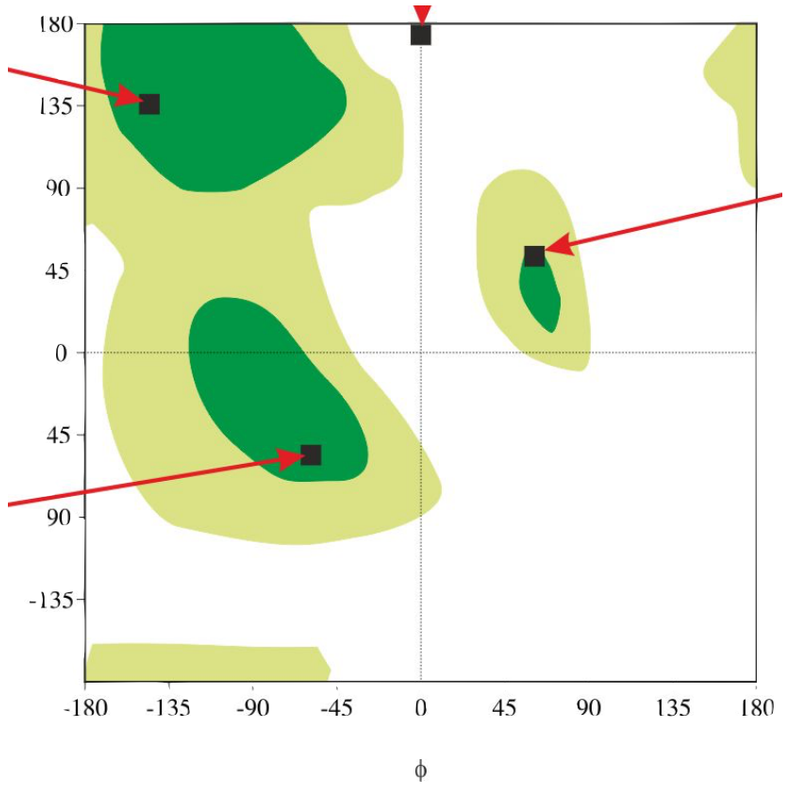


This is the time after which the data are not correlated anymore

Evaluating Convergence and Calculating Errors

Equilibrium: results do not depend any more by the length of the simulation. So averages over multiple blocks of the simulations at some point should give the same results (**limit**: this is always true for very short simulations and for trapped simulations)

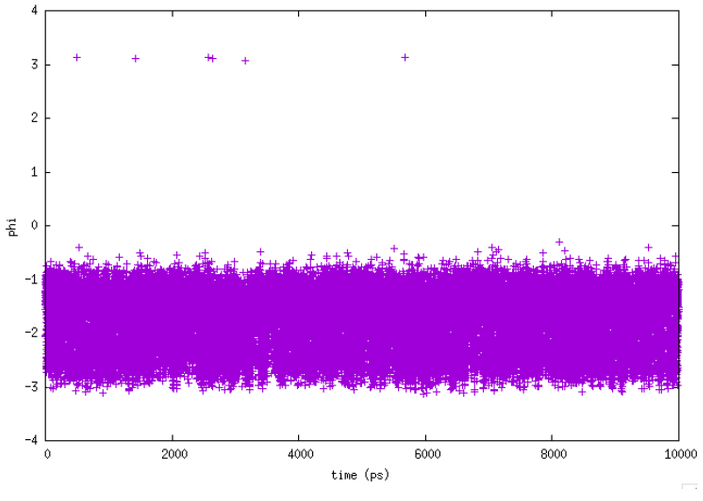




So probably we are still missing something!

Evaluating Convergence and Calculating Errors

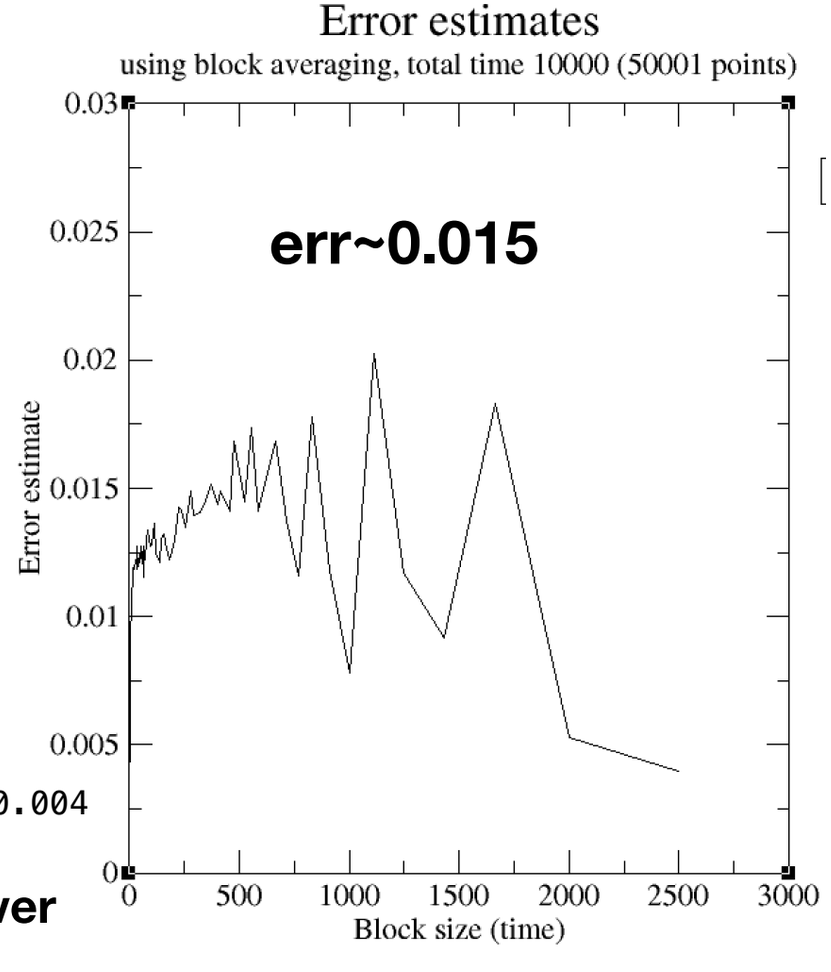
Equilibrium: results do not depend any more by the length of the simulation. So averages over multiple blocks of the simulations at some point should give the same results (**limit**: this is always true for very short simulations and for trapped simulations)



<phi>=-1.751

Calculate the average over four blocks:
0-2500, 2500-5000, …

<phi>=-1.756,-1.745,-1.759,-1,742

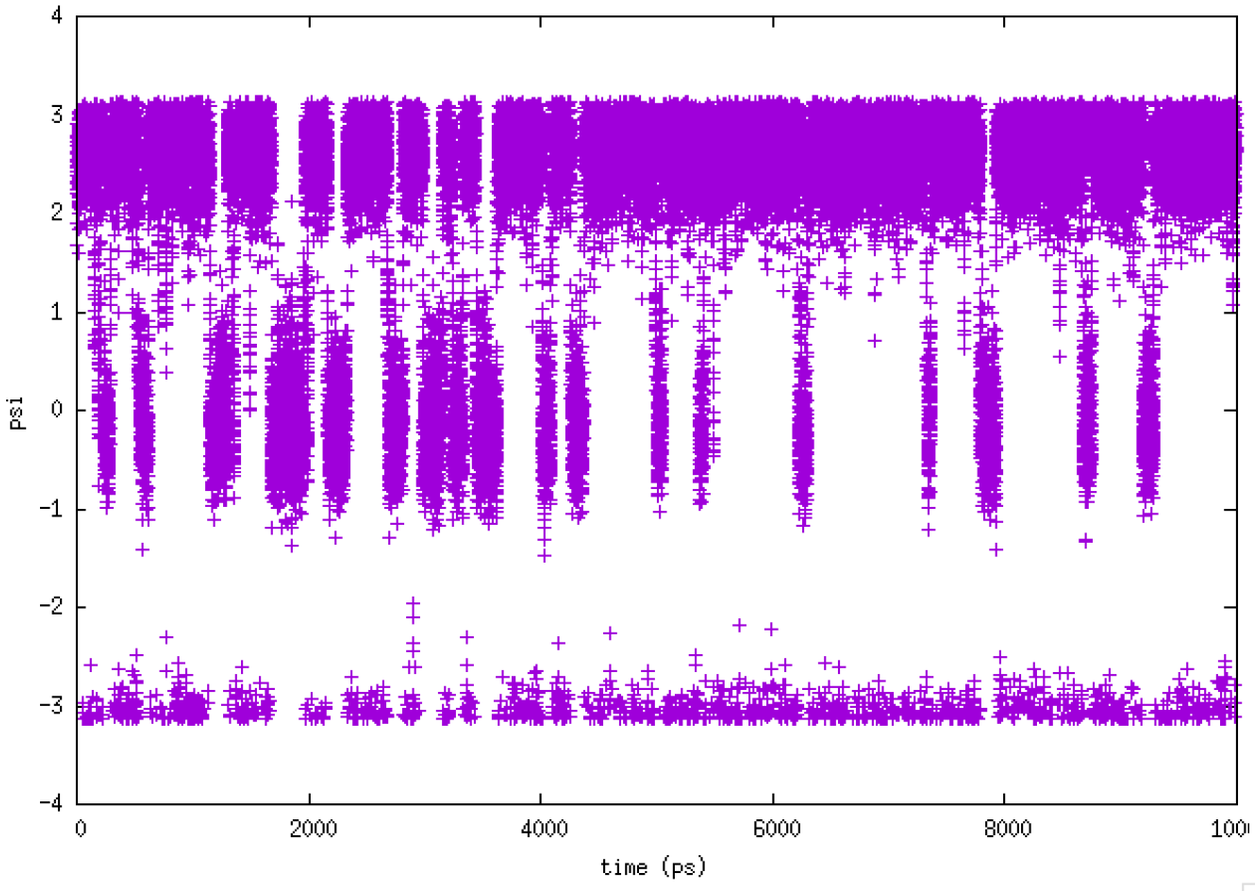


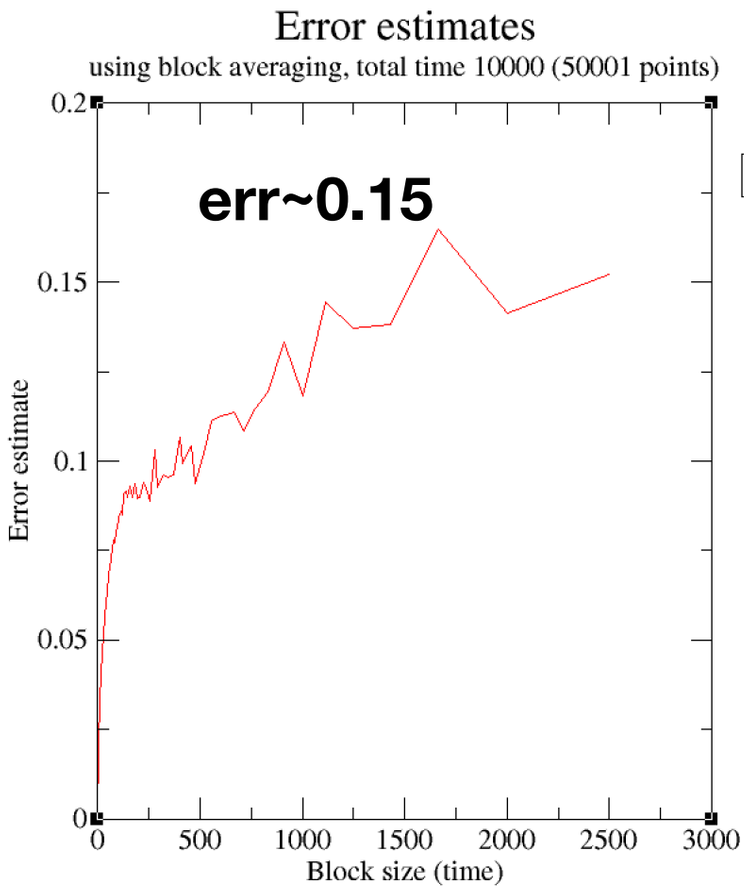
$$\text{Std-err} = \sqrt{\frac{var(f)}{N}} = \frac{\sigma}{\sqrt{N}} = \sqrt{\frac{1}{N^2 - N}\sum_{i=1}^{N}(f(x_i) - < f >)^2}$$

```
sqrt(((-1.756+1.751)^2+(-1.745+1.751)^2+(-1.759+1.751)^2+(-1.742+1.751)^2)/(4*(4-1)))=0.004
```

Then you repeat the calculation for more and more blocks (i.e. averaging over shorter times)

Evaluating Convergence and Calculating Errors



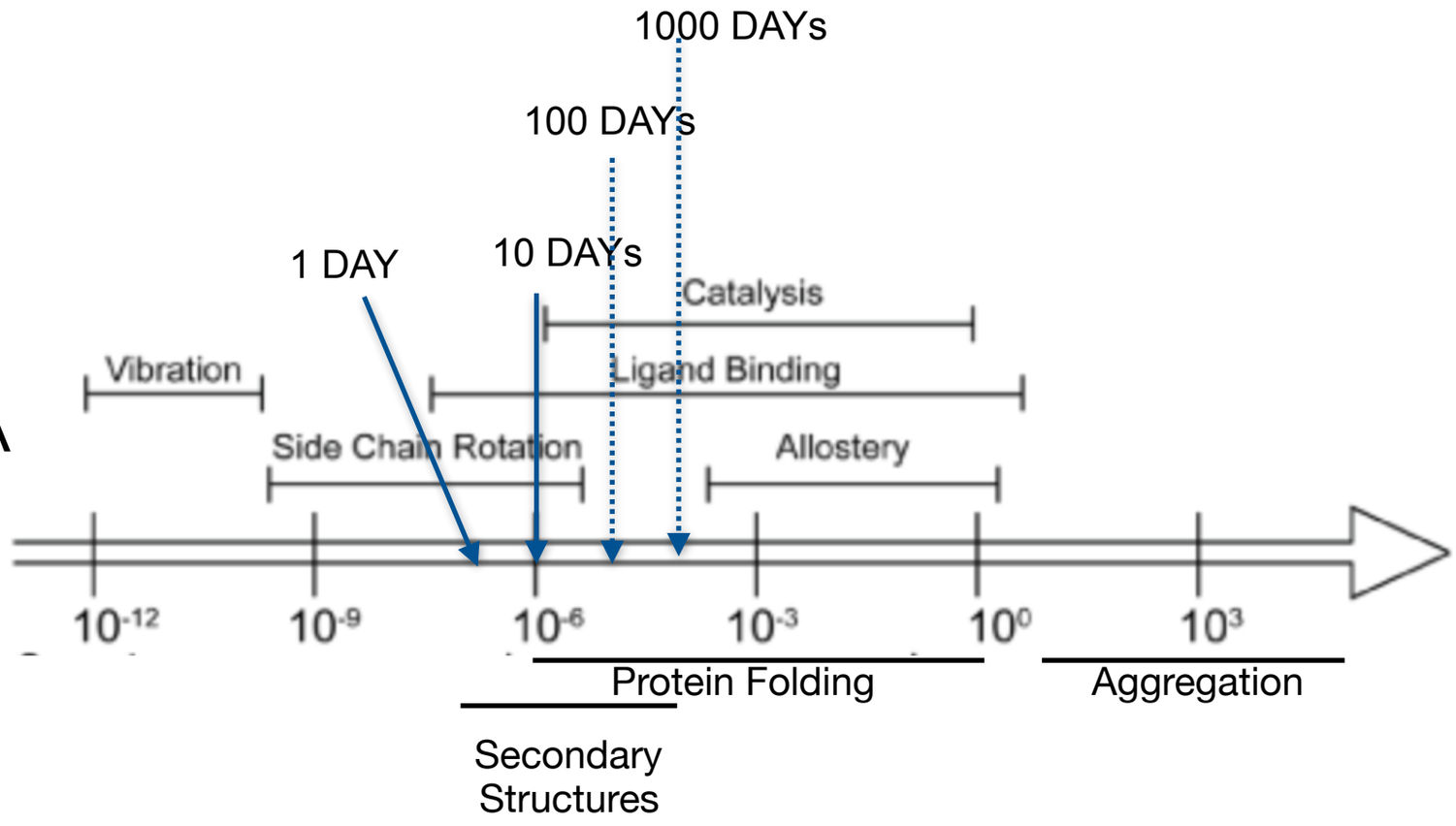


Even if we cannot say for sure that we have sampled everything that is relevant, we can say that at least locally we have converged estimates of the averages of these dihedral with some error.

Time step, time scales and probabilities

time-step ~ 10^{-15}s

an MD simulation for a 100AA protein can run at most at 200 ns/day



the probability of observing an event with a lifetime of 1ms with a simulation of 10us is ~ 1% in the case of a two state kinetics

More on statistical equilibrium

If a **close system** is in **equilibrium** here we mean that the **average** properties are the **same** if observed **over time** or **over space** (ensemble) or over both. Furthermore the behaviour of a single molecule in time (if the time is long enough) is the same of the behaviour of all the molecules at a given time. This is call **ERGODICITY** and is the basic principle that makes computer simulation useful!

This notion of equilibrium (if you want intra-molecular equilibrium) is also valid if we look at molecular interactions (inter-molecular equilibrium), that is for example the fraction of bound and unbound molecules in a solution should be the same of the fraction of time spent bound and free for a single couple of molecules.

Although dynamical measurements (movies) must agree with ensemble measurements (snapshots) the latter lack dynamical information: for example you can say in both case what is the fraction of a conformational state or of a bound state but in the first case you can also say how long molecules are in a given state, I.e. the life time or the rate of a reaction.

Non-equilibrium measurements refer to studies where the system is suddenly perturbed (by temperature, chemical agents, etc) and so **average quantities will change over time**.

Statistics and Mechanics

What is the connection between statistics and mechanics?

That is the Boltzmann equation: $pdf(x,v) \equiv \rho(x,v) \propto \exp\left[\frac{-E(x,v)}{k_B T}\right] = \exp\left[\frac{-U(x)}{k_B T}\right] \exp\left[\frac{-K(v)}{k_B T}\right]$

Let's look at this pdf in more detail:

$$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right] \qquad pdf(v) \equiv \rho(v) \propto \exp\left[\frac{-K(v)}{k_B T}\right] = \exp\left[\frac{-\frac{1}{2}mv^2}{k_B T}\right]$$

Conformations and velocities are independent. This means that we can study them separately. The distribution of the velocities does not affect the distribution of the configurations. Furthermore the distribution of the velocity is Gaussian, so it is easy to integrate it.

So for all practical purposes we can just ignore the velocity:

$$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

This gives a link between the energy of a conformation and the probability of observing it

Statistics and Mechanics

What is the connection between statistics and mechanics?

That is the Boltzmann equation: $$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

The probability of a conformation (microstate) is <u>proportional</u> (exponentially) to minus its energy and inversely proportional to k_BT (Boltzmann constant times Temperature).

T (K)	T (C)	k_BT (kJ/mol)	k_BT (kcal/mol)
273.15	0	2.27	0.54
293.15	20	2.44	0.58
300.00	26.85	2.49	0.60
310.15	37	2.58	0.62

1. Lower energy conformations are more likely than high energy ones.
2. Relative probabilities become more equal as temperature increases.

In principle we can normalise it and get the complete probability distribution, but let's think of a force field, to calculate the normalisation we need to calculate the energy for all possible conformations.

We have a problem of SAMPLING.

…Mechanics

The mechanical problem of the accessible time-scales in MD is translated in statistical terms in a SAMPLING problem. To estimate probabilities we need many many conformations. Convergence becomes the problem of understanding whether the conformations we got are enough to say something relevant.

$$pdf(x) \equiv \rho(x) = \frac{\exp\left[-U(x)/k_B T\right]}{\int_V \exp\left[-U(x)/k_B T\right] dx}$$

Given a sampling we can always estimate the pdf by normalising it over the sampled conformations

$$\hat{Z} = \int_V \exp\left[-U(x)/k_B T\right] dx$$

The normalisation constant is usually called (configurational) Partition Function

$$< g >= \frac{\int_V g(x) \exp\left[-U(x)/k_B T\right] dx}{\int_V \exp\left[-U(x)/k_B T\right] dx}$$

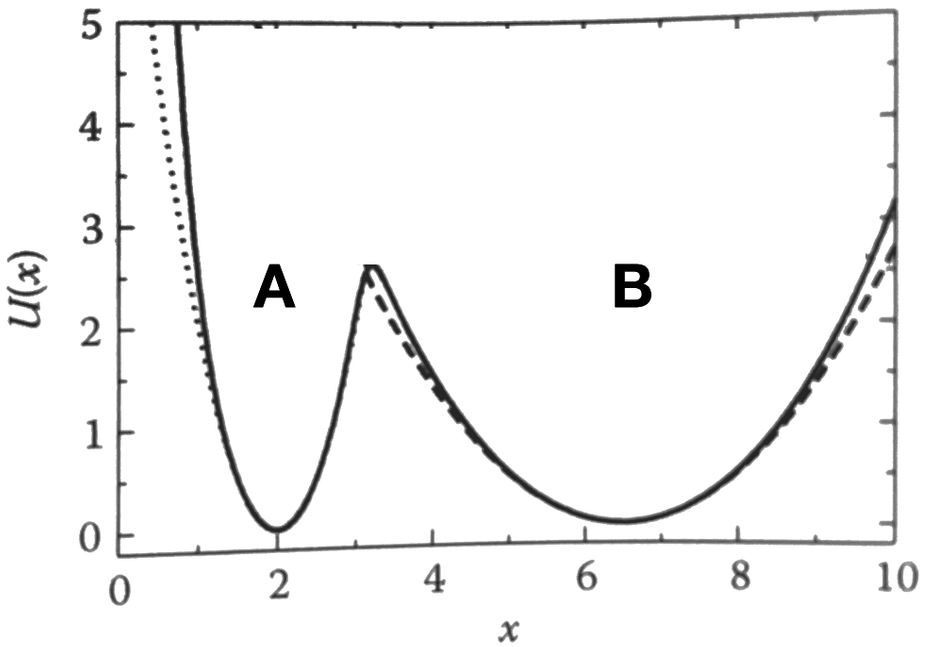
Given a sampling it is possible to calculate averages (e.g. the average RMSD, …)

In practice in MD/MC we don’t need to calculate any integral. What we should know is what is the pdf we are trying to estimate (the Boltzmann distribution).

States, Probabilities and Free-Energy

State: is a collection of configurations (microstates), ideally belonging to the same potential energy basin.

A simple 1D potential energy



Here we can visually define two states A and B for the two basins, e.g. all the conformations belonging to A and to B

$$p_A = \int_{V_A} \rho(x)dx \propto \int_{V_A} \exp[-U(x)/k_BT]dx$$

$$p_B = \int_{V_B} \rho(x)dx \propto \int_{V_B} \exp[-U(x)/k_BT]dx$$

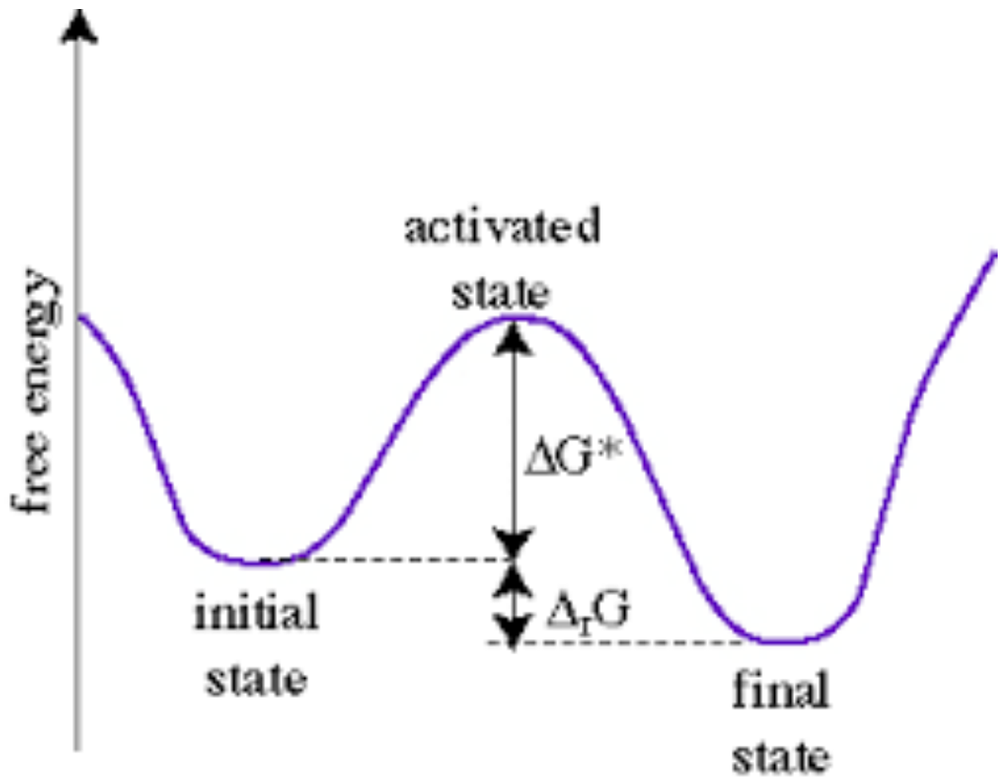
The Free-Energy is the "effective" energy of a state, i.e. the energy that will give the same probability

$$\frac{p_A}{p_B} = \frac{\int_{V_A} \exp[-U(x)/k_BT]dx}{\int_{V_B} \exp[-U(x)/k_BT]dx} \equiv \frac{\exp(-F_A/k_BT)}{\exp(-F_B/k_BT)} \qquad F_i \propto -k_BT \ln\left(\int_{V_i} \exp[-U(x)/k_BT]dx\right)$$

Time scales means Free Energy Barrier

$$k \propto \frac{k_B T}{h} e^{-\Delta G^{\ddagger}/k_B T}$$

Rate	Barrier(kcal/mol)
ps^{-1}	~1
ns^{-1}	~5
μs^{-1}	~10
ms^{-1}	~14
s^{-1}	~18



Two problems:

1. **Find the reaction coordinate, i.e. the optimal direction for the reaction**
2. **Cross the barrier**

Free Energy Methods

How can we tweak probabilities to make more likely the observation of important configurations?

Boltzmann distribution:

$$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

Force field:

$$V(r) = \sum_{bonds} k_b (b - b_0)^2 + \sum_{angles} k_\theta (\theta - \theta_0)^2 + \sum_{torsions} k_\phi [\cos(n\phi + \delta) + 1]$$
$$+ \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right]$$

Free Energy profile (potential of mean force):

$$PMF(f) \propto -k_B T \ln\left[\int \delta(f - f(x)) e^{-U(x)/k_B T} dx\right] = -k_B T \ln P(f)$$

Free Energy Methods

How can we tweak probabilities to make more likely the observation of important configurations?

Boltzmann distribution: $$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

1. The probability density function depends on the **Temperature**, so by changing the temperature we change the probability. In particular we increase the probability of observing high-energy conformations.

By increasing the temperature we are actually looking at different *pdf*. The key point is that we want to use a higher temperature to learn about the *pdf* at the temperature of interest. This concept fall under the name **reweighing**.

It is immediately evident that the strength of this approach is that it doesn't need any particular input from the user. This is essentially not knowledge based. On the negative side there is not a strong control, it does not allow to focus the sampling towards a particular goal.

Free Energy Methods

How can we tweak probabilities to make more likely the observation of important configurations?

Boltzmann distribution: $$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

2. The probability density function depends on the **Force Field**, so by changing the force field we change the probability. In particular we can decrease the interaction energy of some specific term.

$$V(r) = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{torsions} k_\phi[\cos(n\phi + \delta) + 1]$$
$$+ \sum_{\substack{nonbond \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right]$$

By decreasing the contribution of dihedral angles for example we speed up the motion of the backbone and side chains but we are actually looking at different *pdf*. The key point is that we want to use a different force field to learn about the *pdf* for the original force field of interest. This concept fall under the name **reweighing**.

Free Energy Methods

How can we tweak probabilities to make more likely the observation of important configurations?

Boltzmann distribution: $$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

3. The probability density function depends on the **Energy**, so by changing the energy we change the probability. In particular we can add a new constant energy term.

In particular we can add a potential to modify the *pdf* along a specific conformational parameter. For example we could try to flatten the *pdf* in some specific direction.

$$PMF(f) \propto -k_B T \ln\left[\int \delta(f - f(x)) e^{-U(x)/k_B T} dx\right] = -k_B T \ln P(f)$$

$$pdf'(x) \propto \exp\left[\frac{-(U(x) + V(f(x)))}{k_B T}\right]$$

This allows to increase the probability of observing configurations along specific reaction coordinates. Again since we will observe a new *pdf* we should make it in such a way to go back to the original one.

Free Energy Methods

How can we tweak probabilities to make more likely the observation of important configurations?

Boltzmann distribution: $$pdf(x) \equiv \rho(x) \propto \exp\left[\frac{-U(x)}{k_B T}\right]$$

4. The probability density function depends on the **Energy**, so by changing the energy we change the probability. In particular we can add a new time-dependent energy term.

In particular we can add a potential to modify the *pdf* along a specific conformational parameter. For example we could try to flatten the *pdf* in some specific direction.

One problem is that often we don't know the PMF along a coordinate *a priori* so we would like to add an adaptive potential (a time-dependent potential) that uses what we learn on-the-fly.

$$PMF(f) \propto -k_B T \ln\left[\int \delta(f - f(x)) e^{-U(x)/k_B T} dx\right] = -k_B T \ln P(f)$$

$$pdf'(x,t) \propto \exp\left[\frac{-(U(x) + V(f(x),t))}{k_B T}\right]$$

This allows to increase the probability of observing configurations along specific reaction coordinates. Again since we will observe a new *pdf* we should make it in such a way to go back to the original one.

Enhanced Sampling

Extended Ensemble Methods

1. Parallel Tempering

Collective variable Methods

3. Umbrella Sampling
4. Metadynamics

Parallel Tempering

At least in principle if we know all the configurations we can obtain the *pdf* at any temperature

$$pdf(x, T_1) = \frac{\exp(-U(x)/k_B T_1)}{\int \exp(-U(x)/k_B T_1) dx}$$

$$pdf(x, T_2) = \frac{\exp(-U(x)/k_B T_2)}{\int \exp(-U(x)/k_B T_2) dx}$$

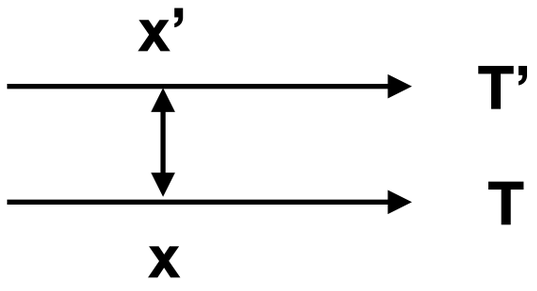
It is enough to recalculate the formula using a different temperature. But as we already discussed we are never in this condition. So this is useless.

What is instead the *pdf* for two copies of the same system at two different temperatures?

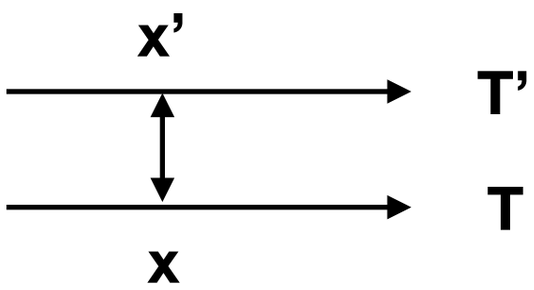
$$pdf(x, x', T, T') \propto \exp\left[\frac{-U(x)}{k_B T}\right] \exp\left[\frac{-U(x')}{k_B T'}\right]$$

This is the joint probability of observing two configurations from the same force field at two different temperatures.

In principle we can take conformations obtained from the simulation at high temperature and bring them at low temperature and vice versa. We could do it only if we are not going to change the equilibrium:



Parallel Tempering



The probability of exchange is the same as the joint probability for x'->x and x->x' at their own temperatures

$$w(x \to x', T)P(x, T) = w(x' \to x, T)P(x', T)$$
$$w(x \to x', T')P(x, T') = w(x' \to x, T')P(x', T')$$

At equilibrium the flux in one direction is the same of the flux in the opposite

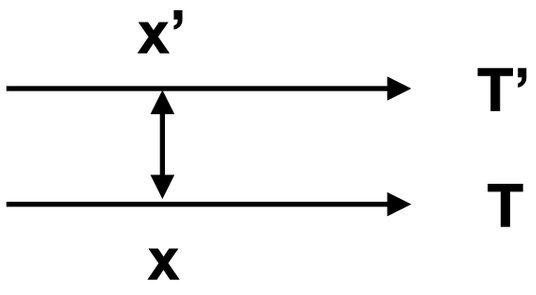
Taking the joint probability:

$$w(x \to x', T)P(x, T)w(x' \to x, T')P(x', T') = w(x' \to x, T)P(x', T)w(x \to x', T')P(x, T')$$

$$\frac{w(x \to x', T)w(x' \to x, T')}{w(x' \to x, T)w(x \to x', T')} = \frac{P(x', T)P(x, T')}{P(x, T)P(x', T')} = \frac{\exp\left[-U(x')/\,k_B T - U(x)/k_B T'\right]}{\exp\left[-U(x)/\,k_B T - U(x')/k_B T'\right]} =$$

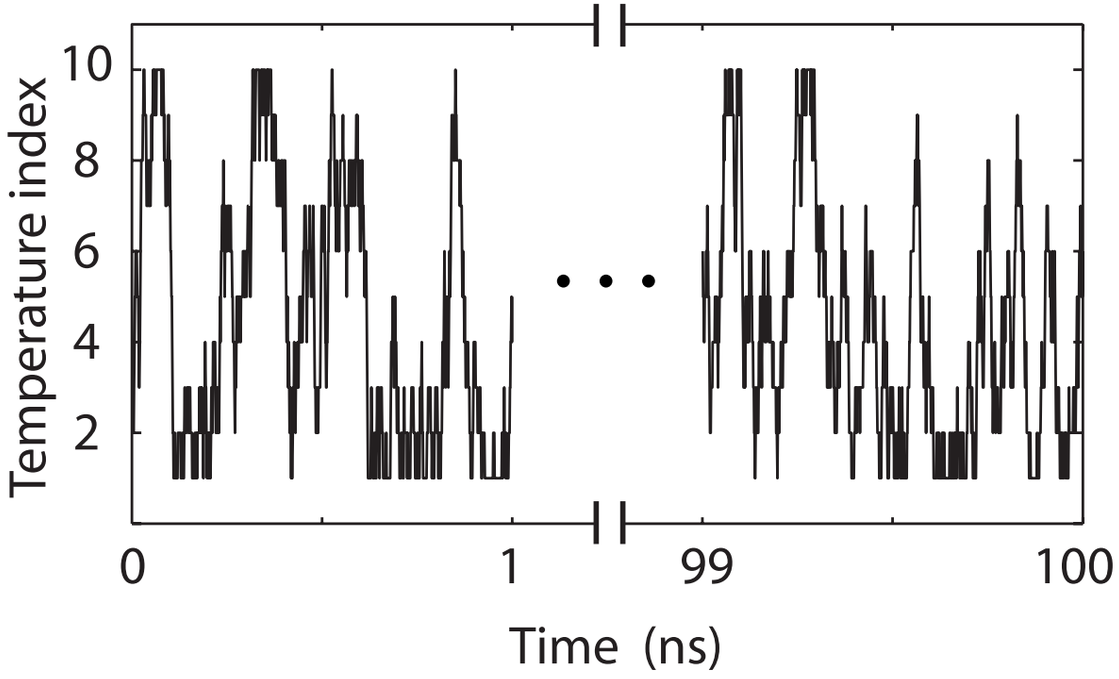
$$\exp\left[-\frac{(U(x) - U(x'))}{k_B}\left(\frac{1}{T'} - \frac{1}{T}\right)\right]$$

Parallel Tempering



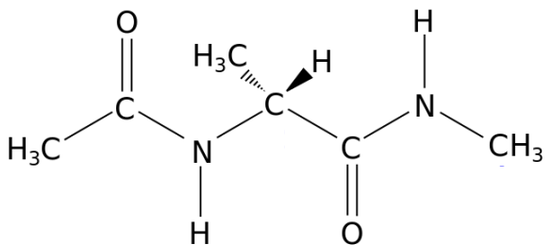
Apart from the math what happen in practice is that we have two (many) MD at many temperatures, and with some frequency an exchange is attempted between configurations at different temperature. This is done using a **Metropolis-Monte-Carlo simulation**, with a probability of acceptance given by:

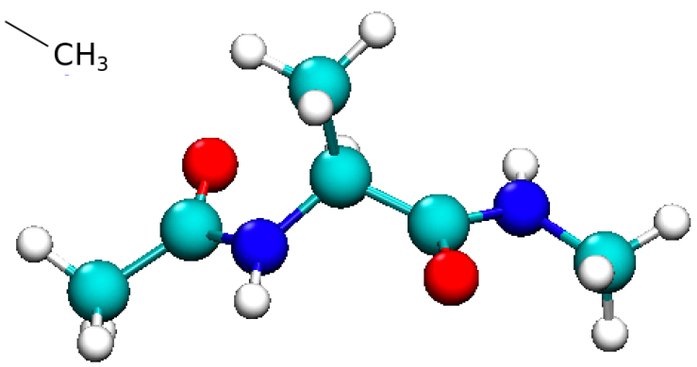
$$P_{acc}(x \leftrightarrow x') = \min\left(1, \exp\left[-\frac{(U(x) - U(x'))}{k_B}\left(\frac{1}{T'} - \frac{1}{T}\right)\right]\right)$$

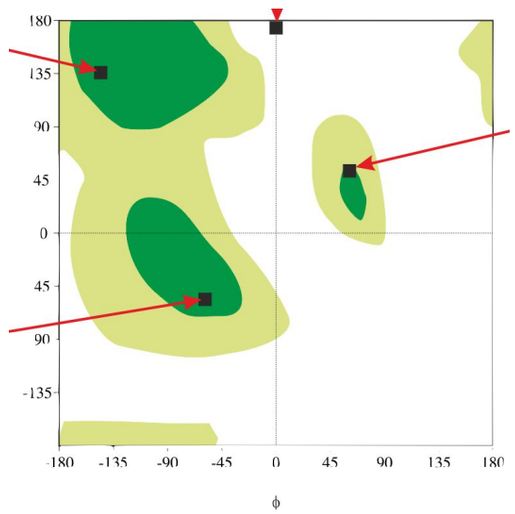


So a simulation will move in time and temperature.
The result is that one obtains at once the temperature dependence of the system. The cons is that the probability of exchange decreases a lot with the system size, so for large system one could easily need hundreds of replicas. The other cons is that lack of control of the sampling. For example if we are interested in conformational changes and not in protein unfolding, the temperature can be dangerous because the protein can unfold and it can take a long time before refolding.

Parallel Tempering

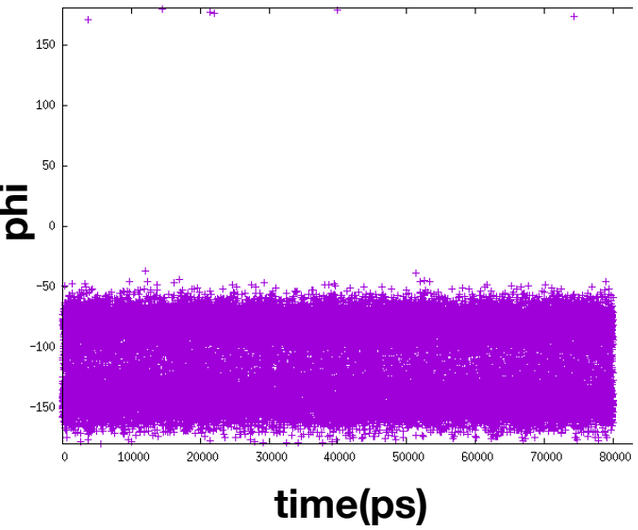


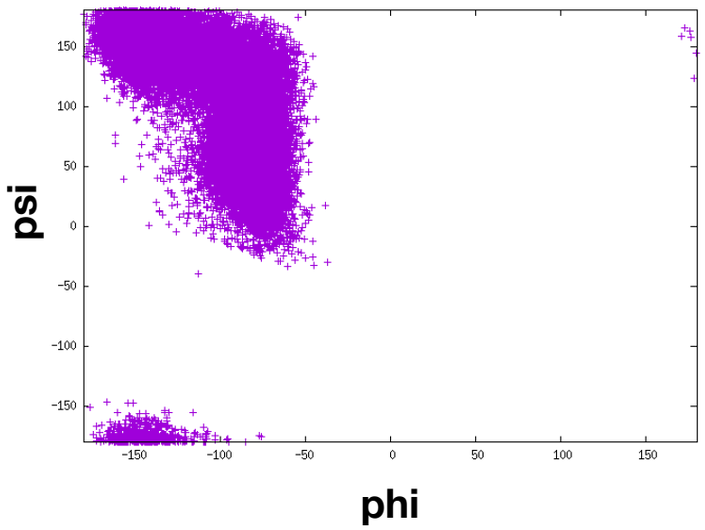




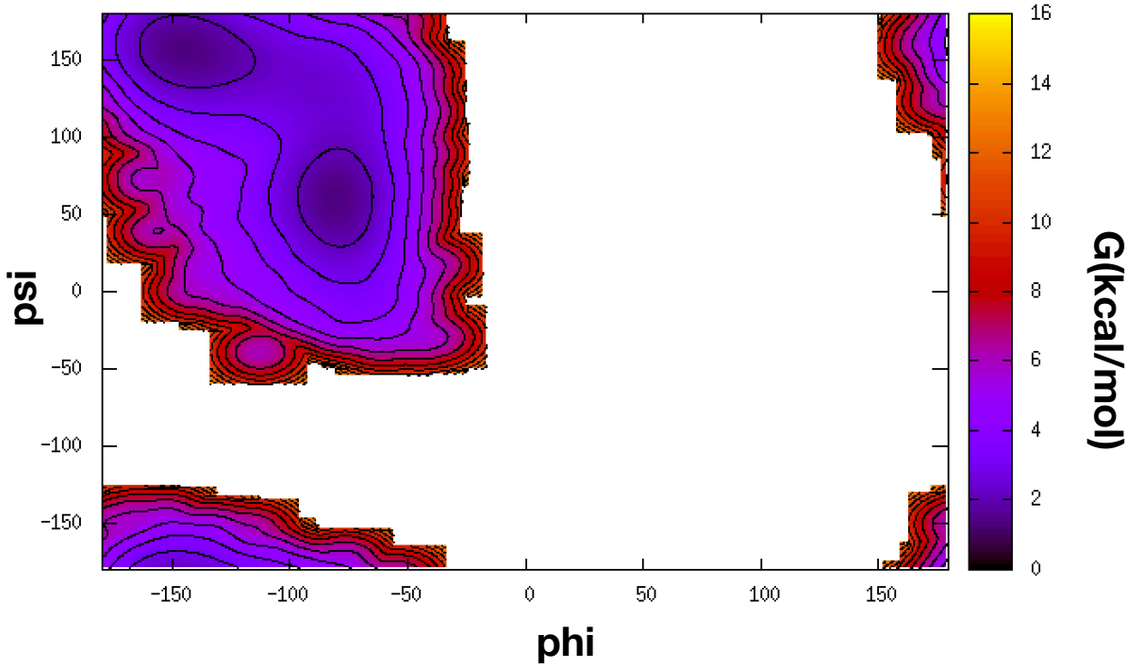
Alanine dipeptide:
ACE-ALA-NME
The expectation for this system is to sample three relevant conformations.

A relatively long MD (80ns) at 300K samples only the left region:

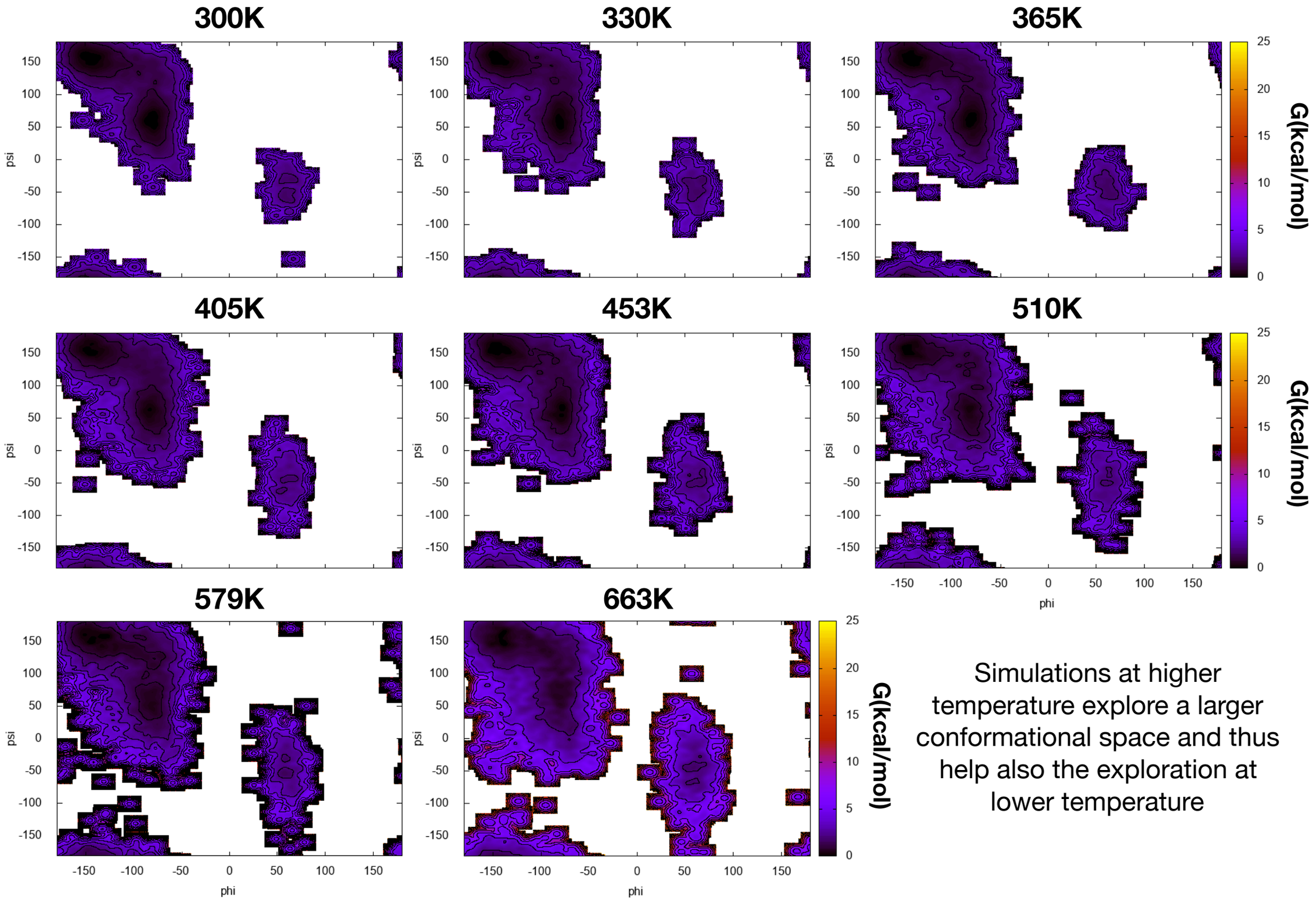




A longer simulation could sample the right region as well, but how longer? Alternatively here we run 8 simulations at 8 different temperature for 10ns each (the same total simulation time)

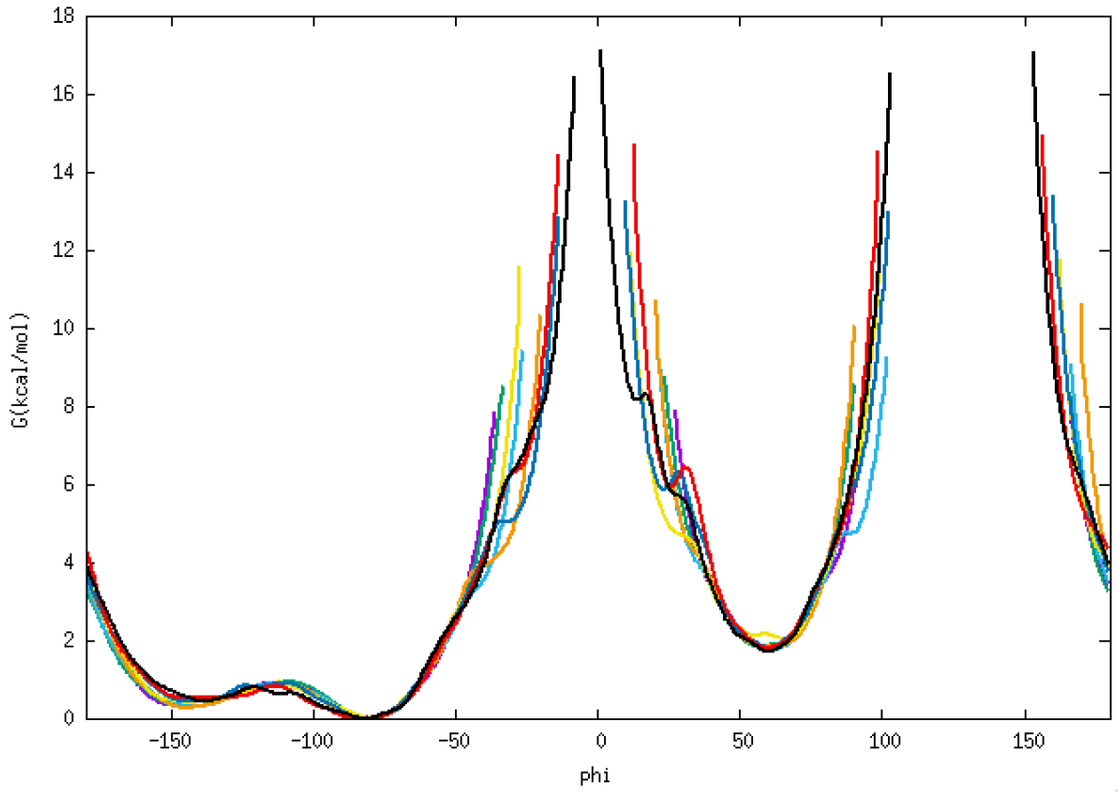


Parallel Tempering



Simulations at higher temperature explore a larger conformational space and thus help also the exploration at lower temperature

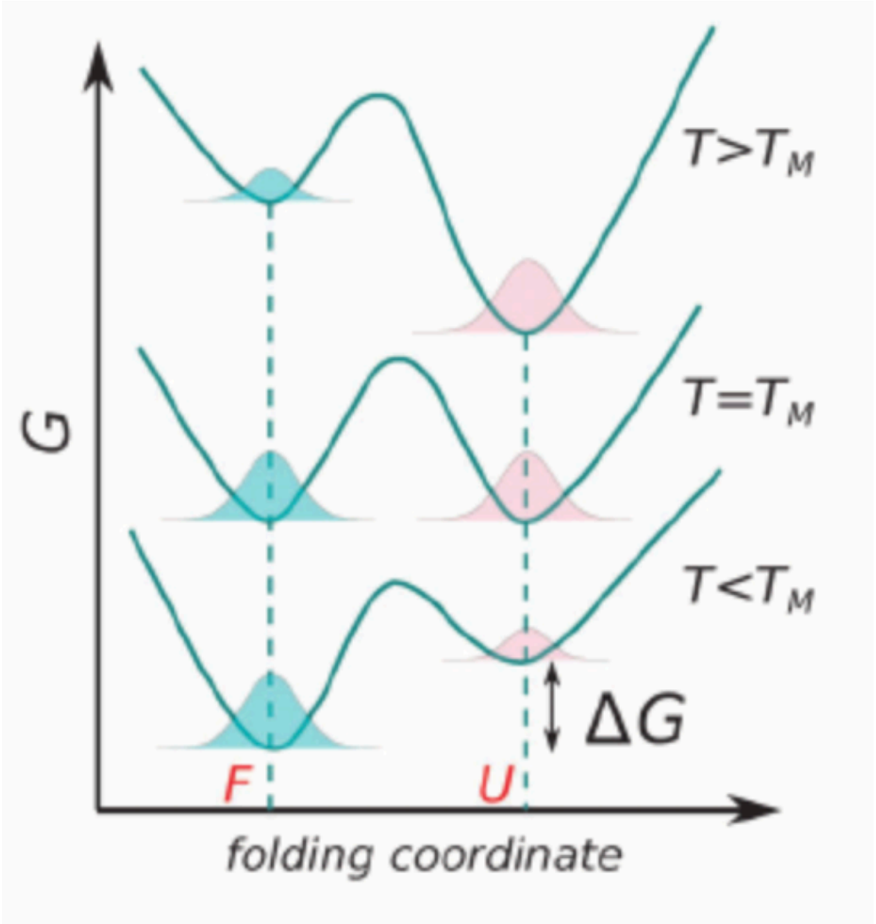
Parallel Tempering



By looking a the free energy projected only on phi it is clear that the free energy in this case does not change with the temperature, so using higher temperature here is particularly efficient. Furthermore form the barrier estimate of ~10kcal/mol we could say that we would have needed a 1us simulation to sample this conformational change! We got the result with 80ns so we have speed up the simulation by a factor >10

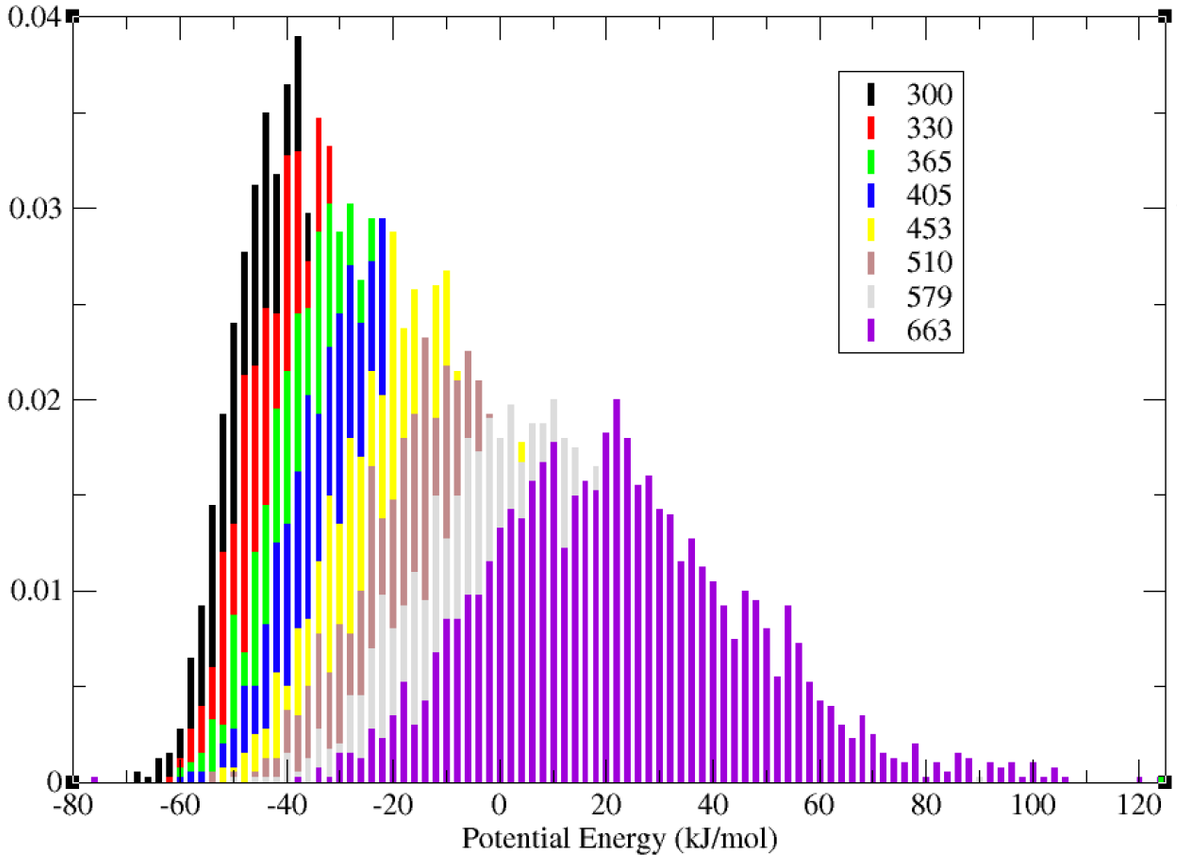
This is an extreme case, usually the free energy will change with the temperature, think at proteins, the higher the temperature the more is gonna be populated the unfolded state. In this case the high temperature will not help too much.

This is the first limitation of PT.



Parallel Tempering: Pros and Cons

1. **Free Energy changes with temperature**
2. **The number of replicas needed increases with system size:**



In order to have a good probability of exchange

$$P_{acc}(x \leftrightarrow x') = \min\left(1, \exp\left[-\frac{(U(x) - U(x'))}{k_B}\left(\frac{1}{T'} - \frac{1}{T}\right)\right]\right)$$

The histograms of the potential energy should overlap significantly (~10-30%), here it means that we could have used even less replicas, the overlap is very high, but for real system in water (here we are in vacuum) the temperature difference needed can become of the order ~ 1-10K.

3. **The speed up is general, it is not focus on a specific process of interest, so in principle we could be faster**

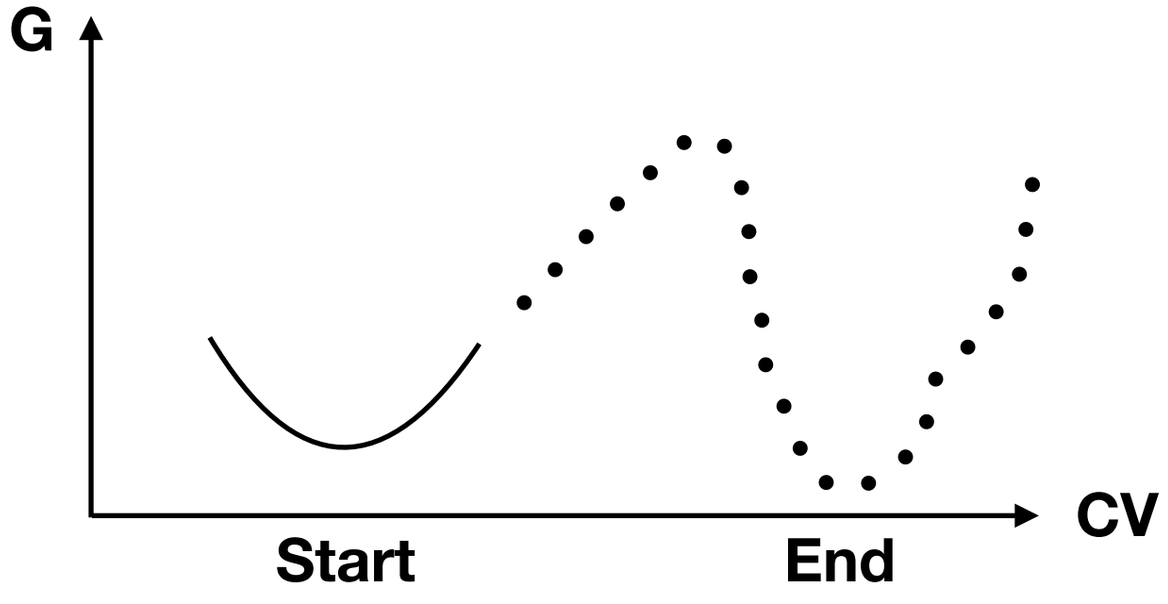
Umbrella Sampling

The former methods allow speeding up the sampling without addition of knowledge.
Oftentimes we have ideas on how a process should work

For example:

1. **alanine dipeptide: the conformational change is related to phi-psi;**
2. **ligand-binding: the binding requires a decrease in the distance ligand-binding-site**
3. **maybe one knows the initial and final structure of a conformational change**
4. **…**

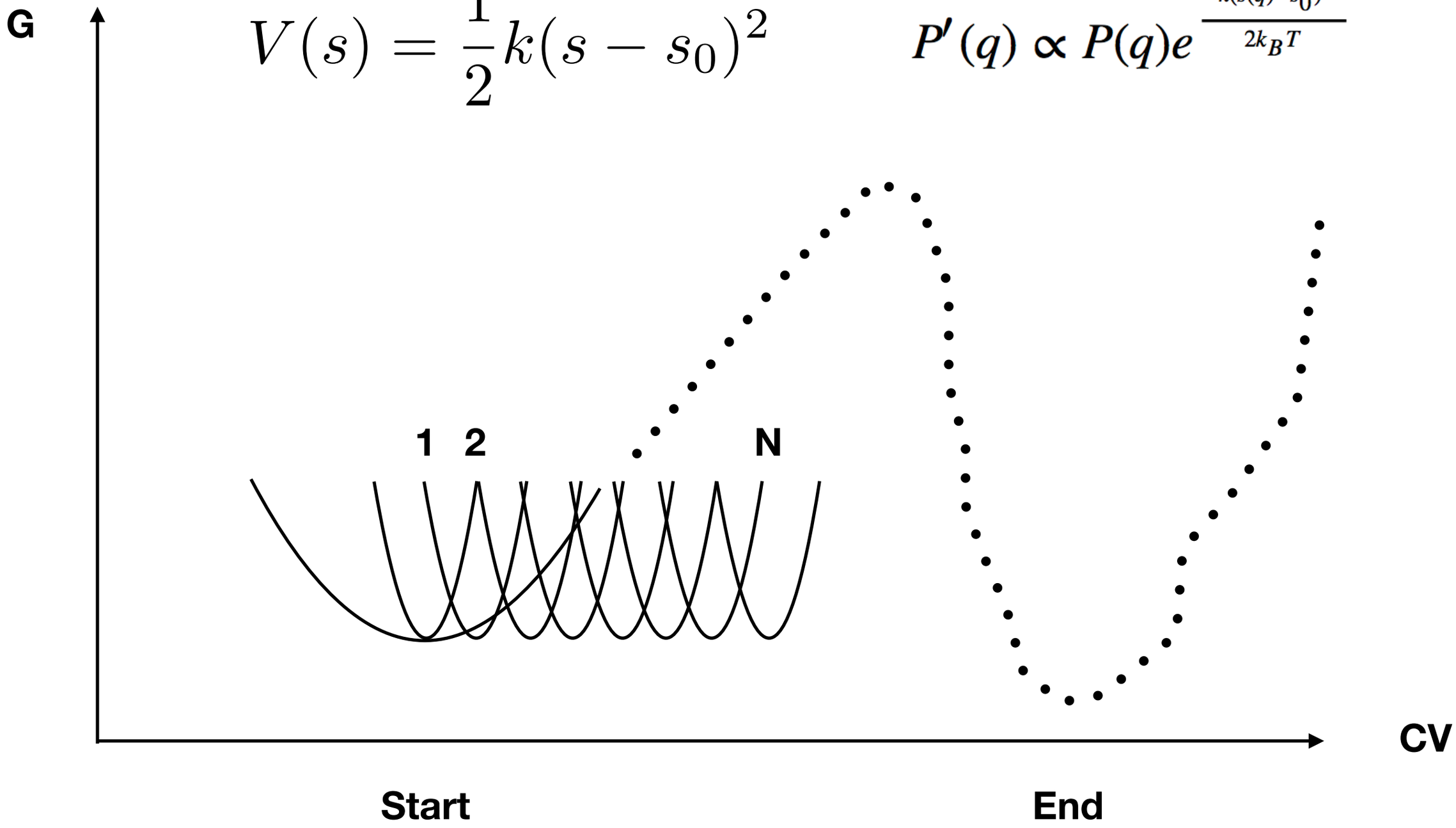
How can we get the free energy profile over one (few) conformational parameters?



Umbrella Sampling

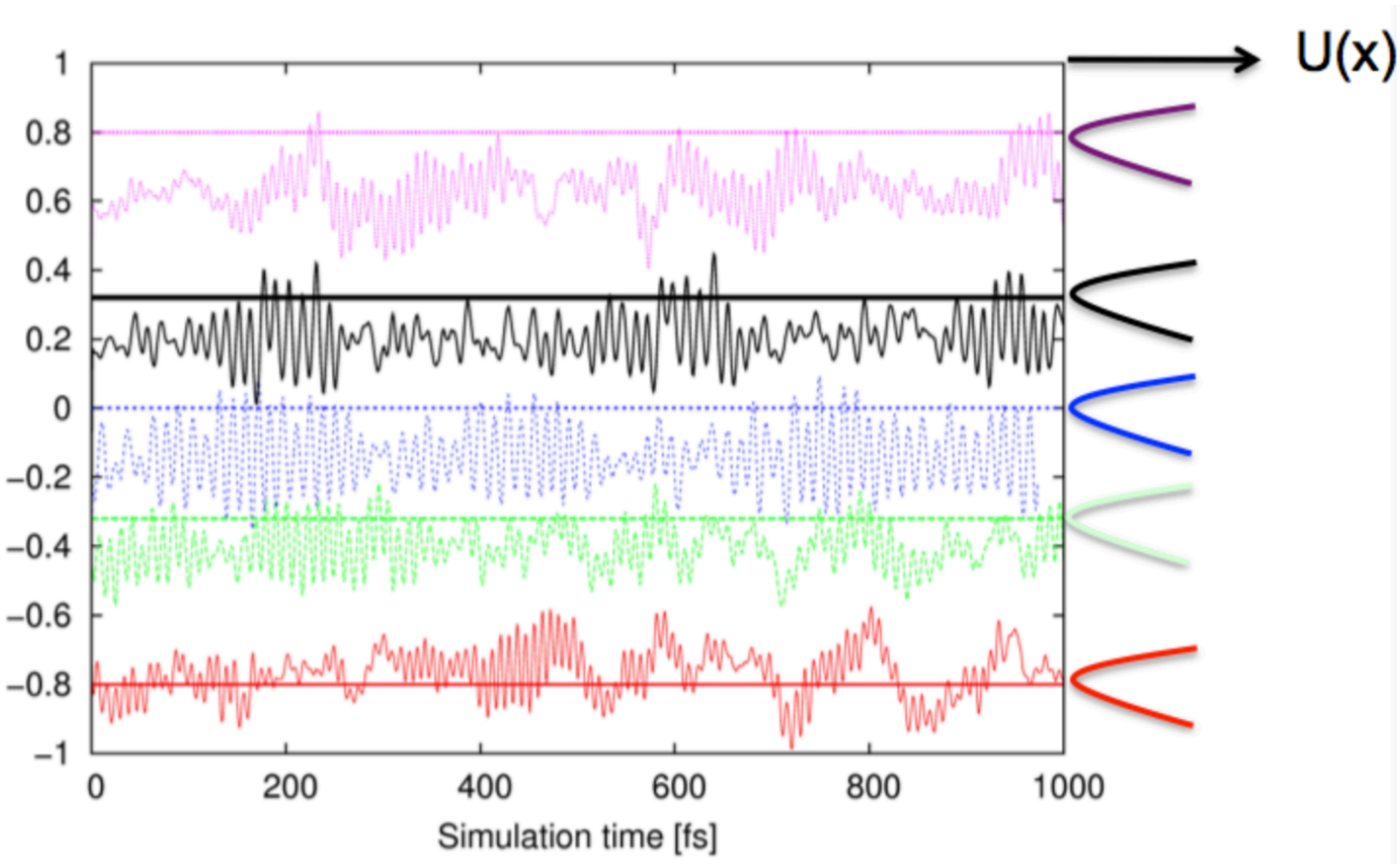
$$V(s) = \frac{1}{2}k(s - s_0)^2$$

$$P'(q) \propto P(q)e^{\frac{-k(s(q)-s_0)^2}{2k_BT}}$$



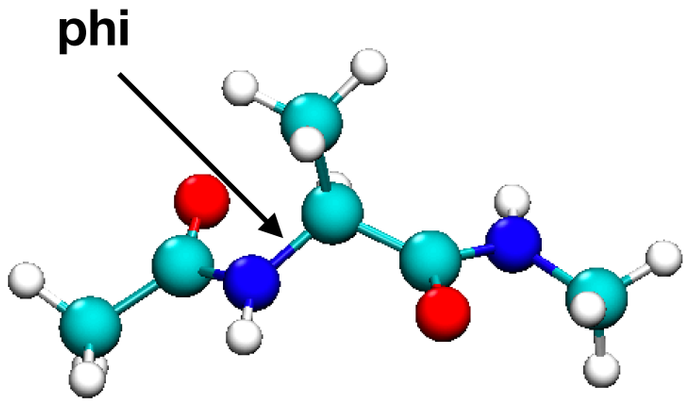
Umbrella Sampling

Many simulations are performed each one centred around a specific value of a conformational parameter

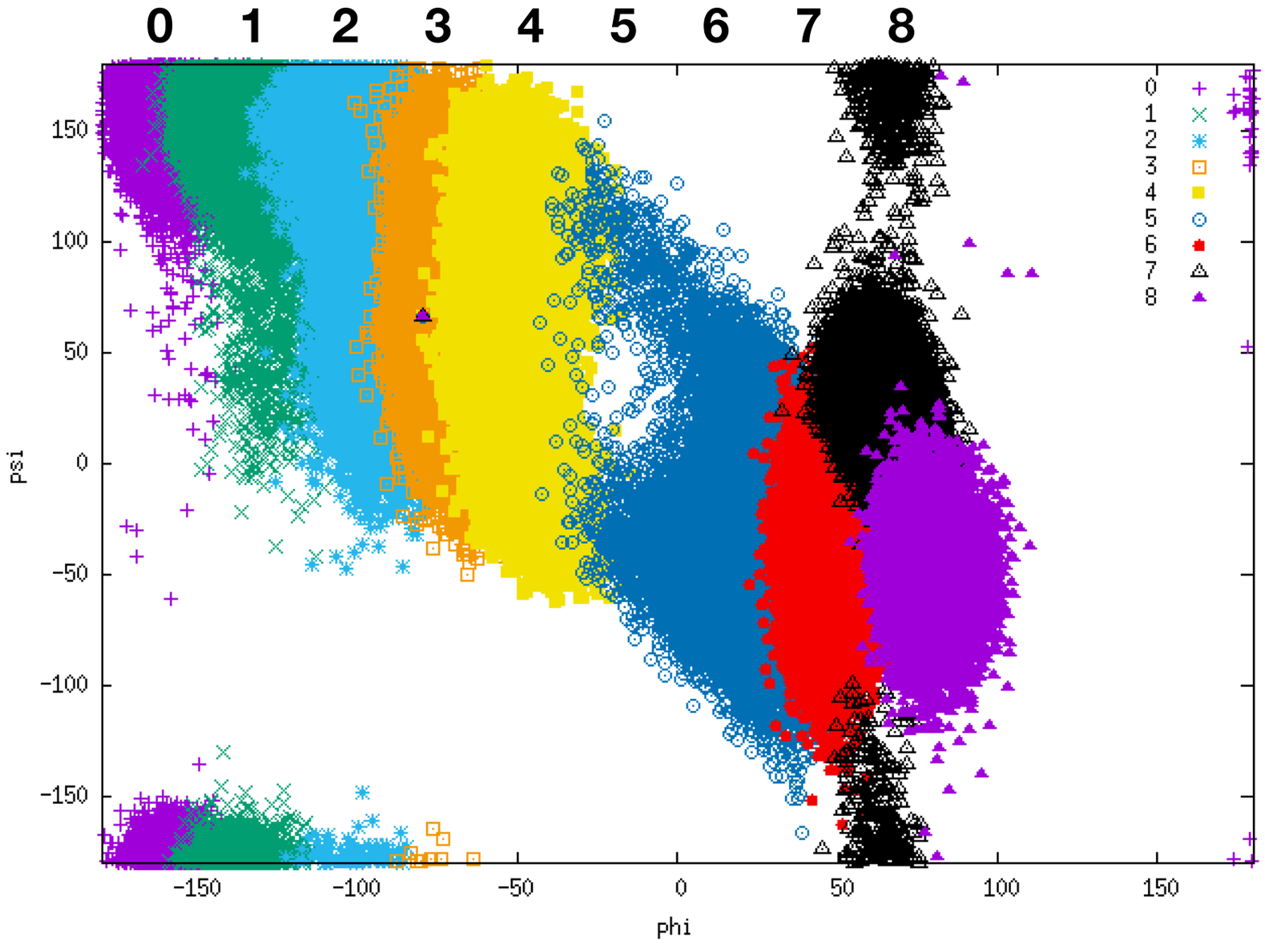


Umbrella Sampling

Many simulations are performed each one centred around a specific value of a conformational parameter



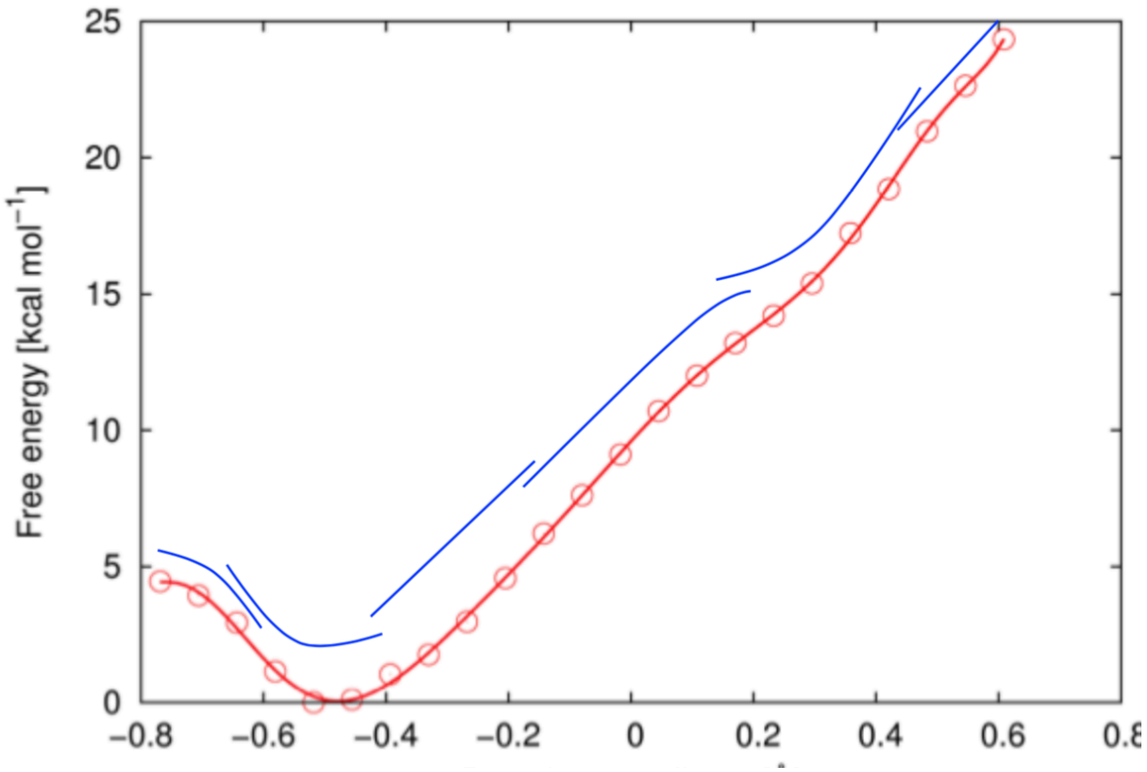
In this way we can force the simulation to sample regions that would not normally sample

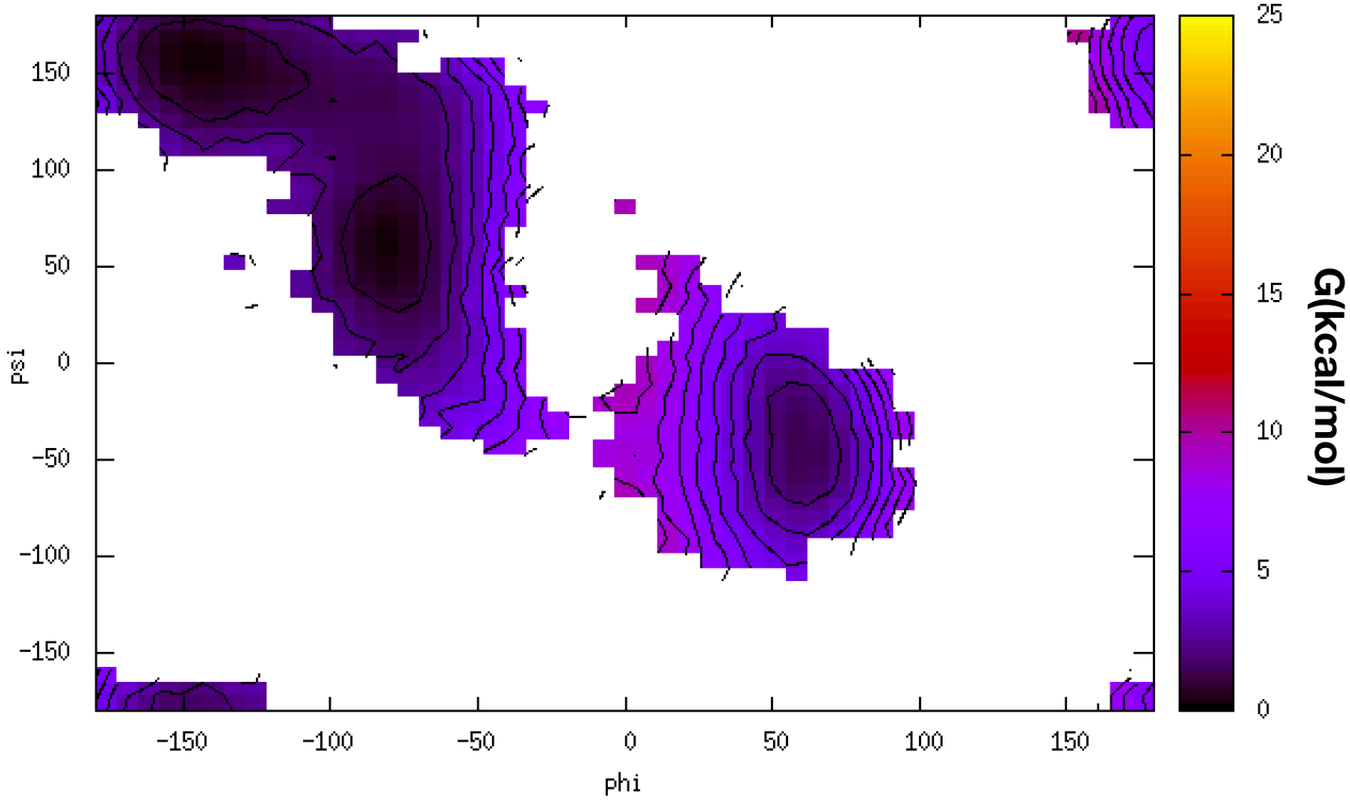


Umbrella Sampling

The problem is how to use this to go back to the original force-field behaviour?

Qualitatively the idea is that each simulation will give a local estimate and that we need to merge them together:

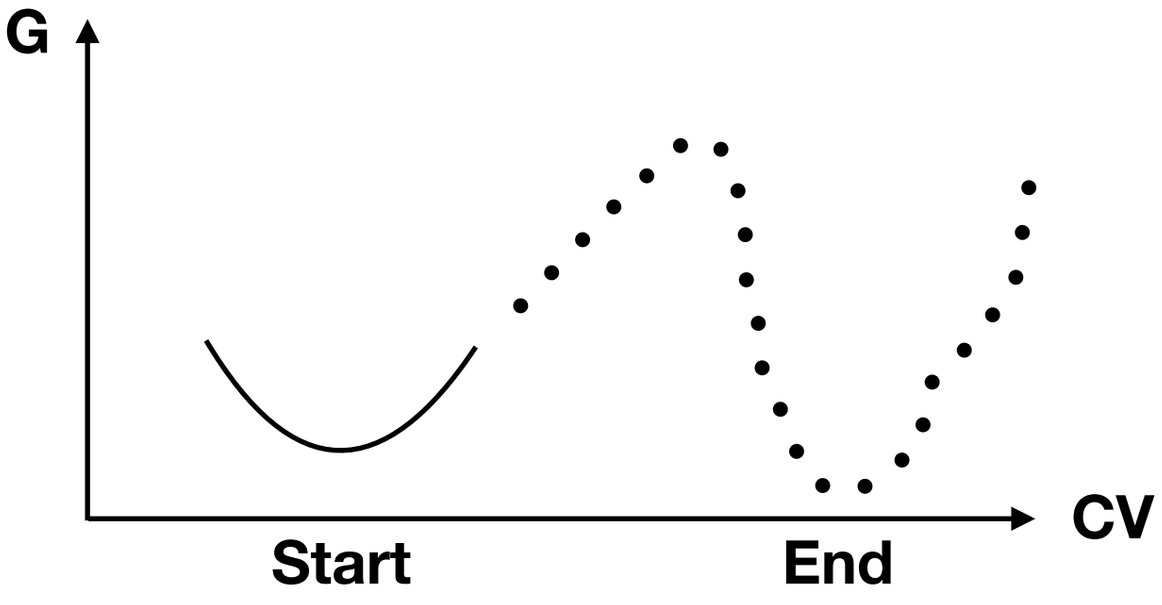




Metadynamics

$$pdf'(x,t) \propto \exp\left[\frac{-(U(x)+V(f(x),t))}{k_B T}\right]$$

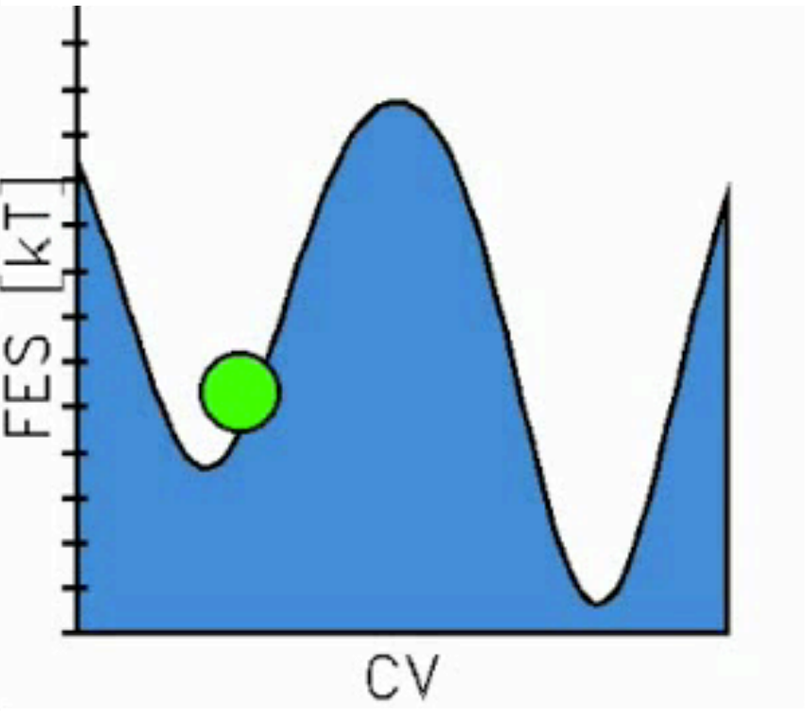
This means that we learn by the simulation that we have already run and modify the potential along a specific direction



As a last case we will see how we can build a bias to speed up the simulation that learns by going.

What is the information we are learning during the course of a simulation?

Metadynamics

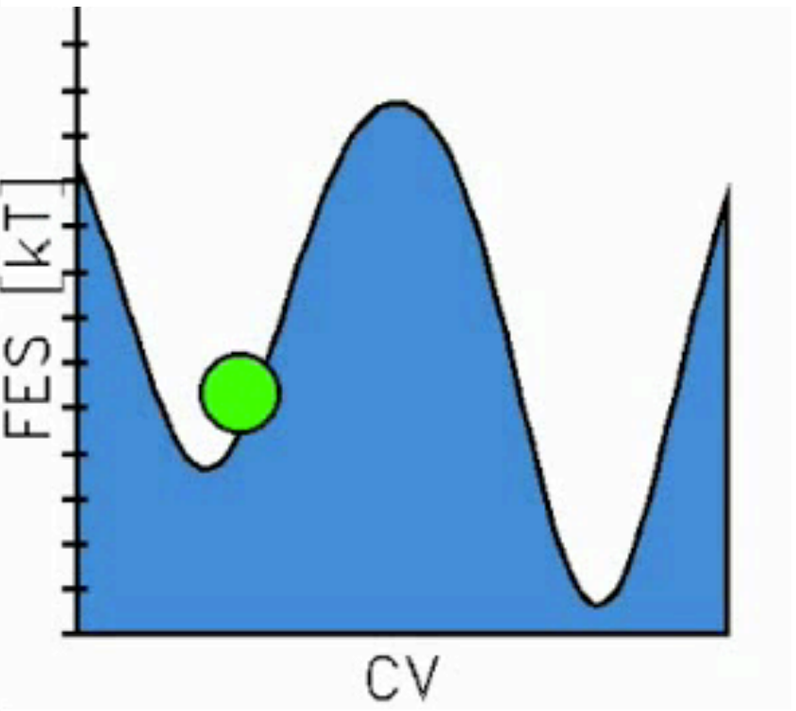


In a standard MD the probability of visiting a conformation is constant

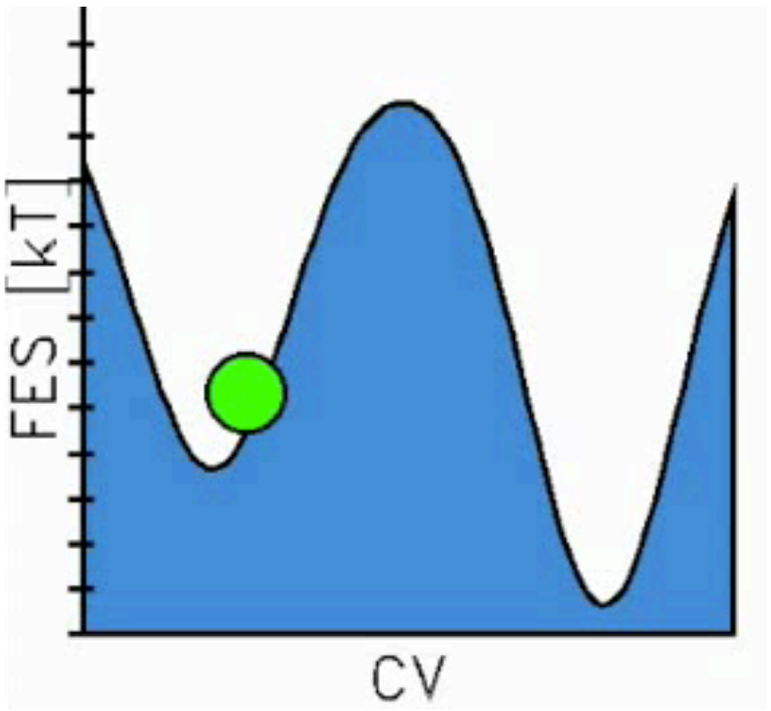
This means that if two states are separated by a barrier it will be unlikely to cross the barrier (low probability) and visit a different state

Metadynamics

$\dot{V}(s, t) = 0$



$\dot{V}(s, t) = \omega e^{-(s-s(t))^2/2\sigma_s^2}$



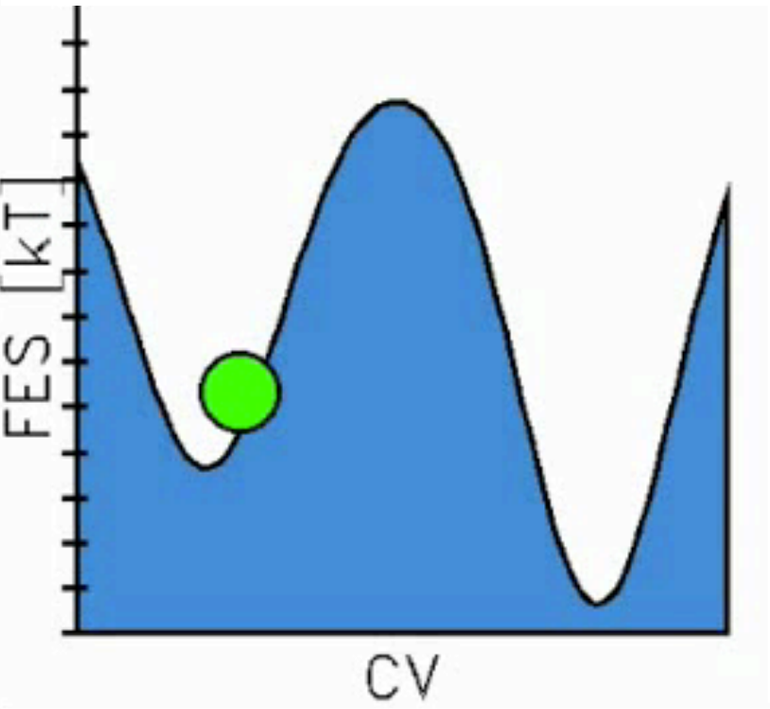
The original idea of metadynamics was to try to make the probability of visiting any conformation equal. But this result in making likely also very uninteresting configurations.

We can add a bias proportional to the time spent in a particular region. The problem is this never ends.

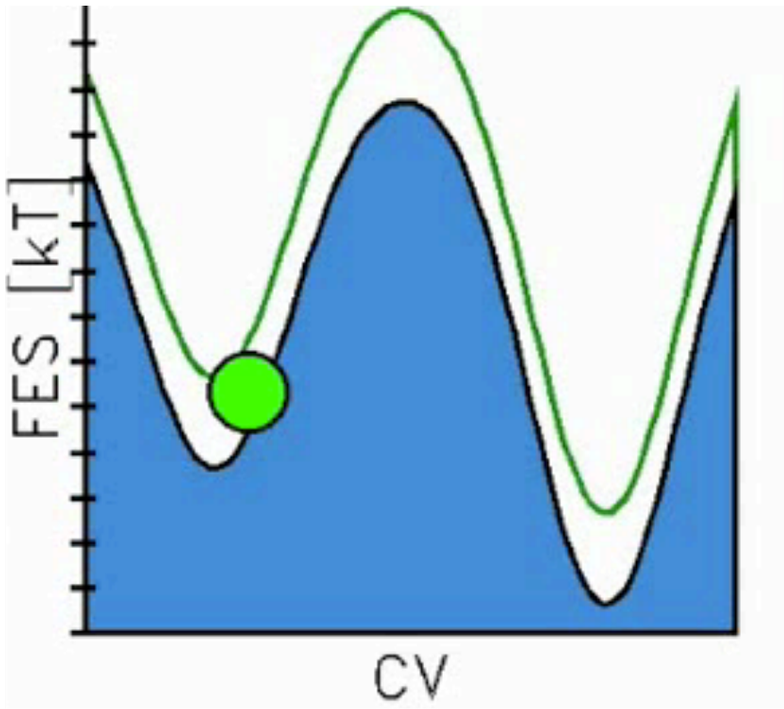
Metadynamics

$\dot{V}(s, t) = 0$

$\dot{V}(s, t) = \omega e^{-[V(s,t)/\Delta T]} e^{-(s-s(t))^2/2\sigma_s^2}$

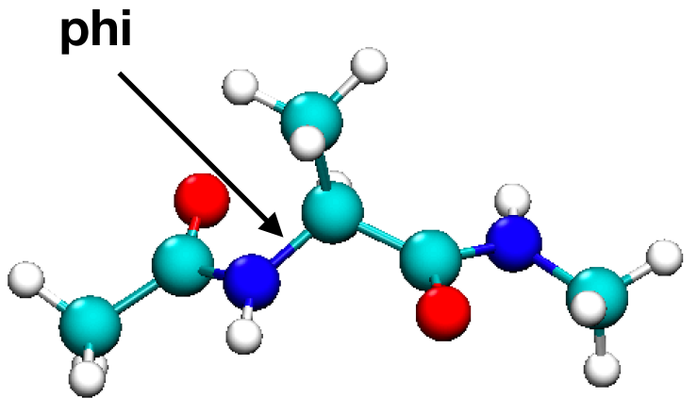


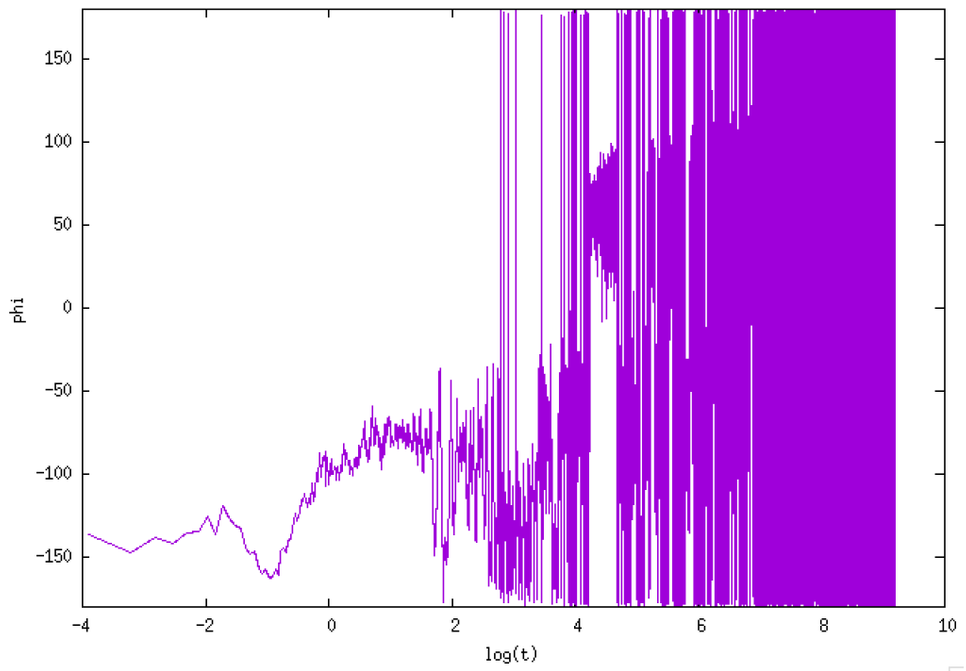
In Well-Tempered Metadynamics the idea is to increase to rescale the probability only of conformations up to some energy defined from a parameter ΔT

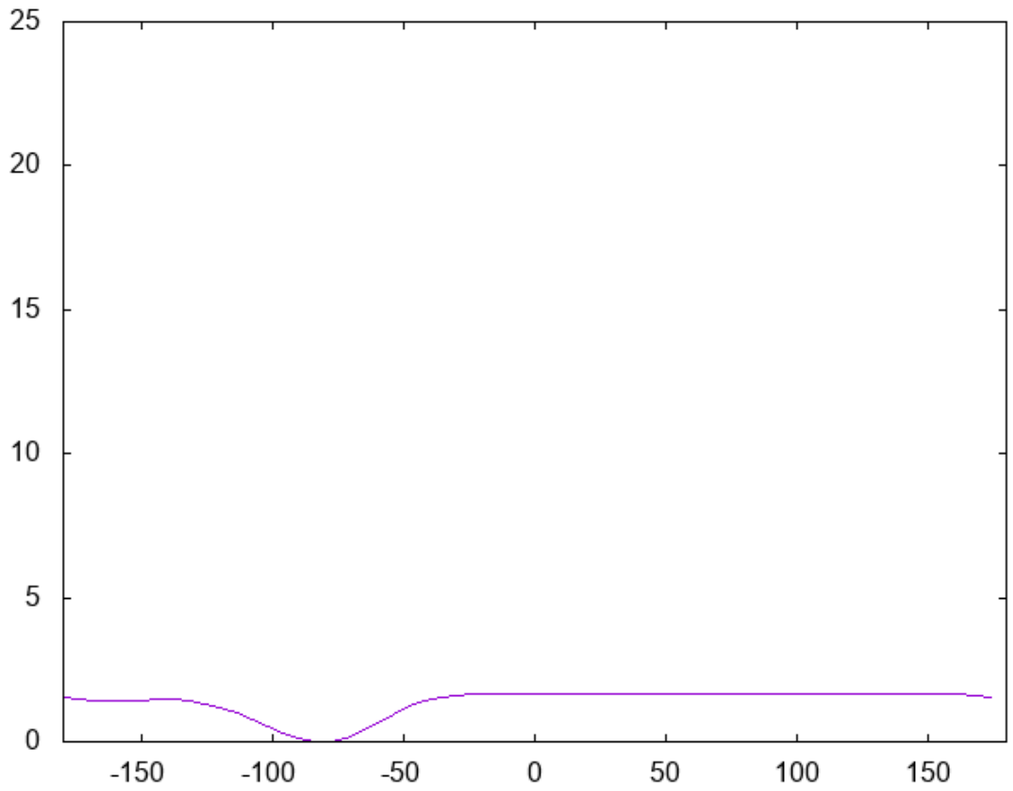


We can add a bias proportional to the time spent in a particular region. But counting every addition as 1/t

Metadynamics

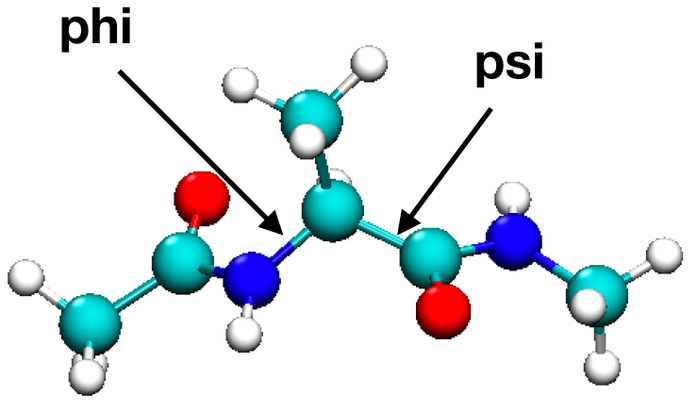




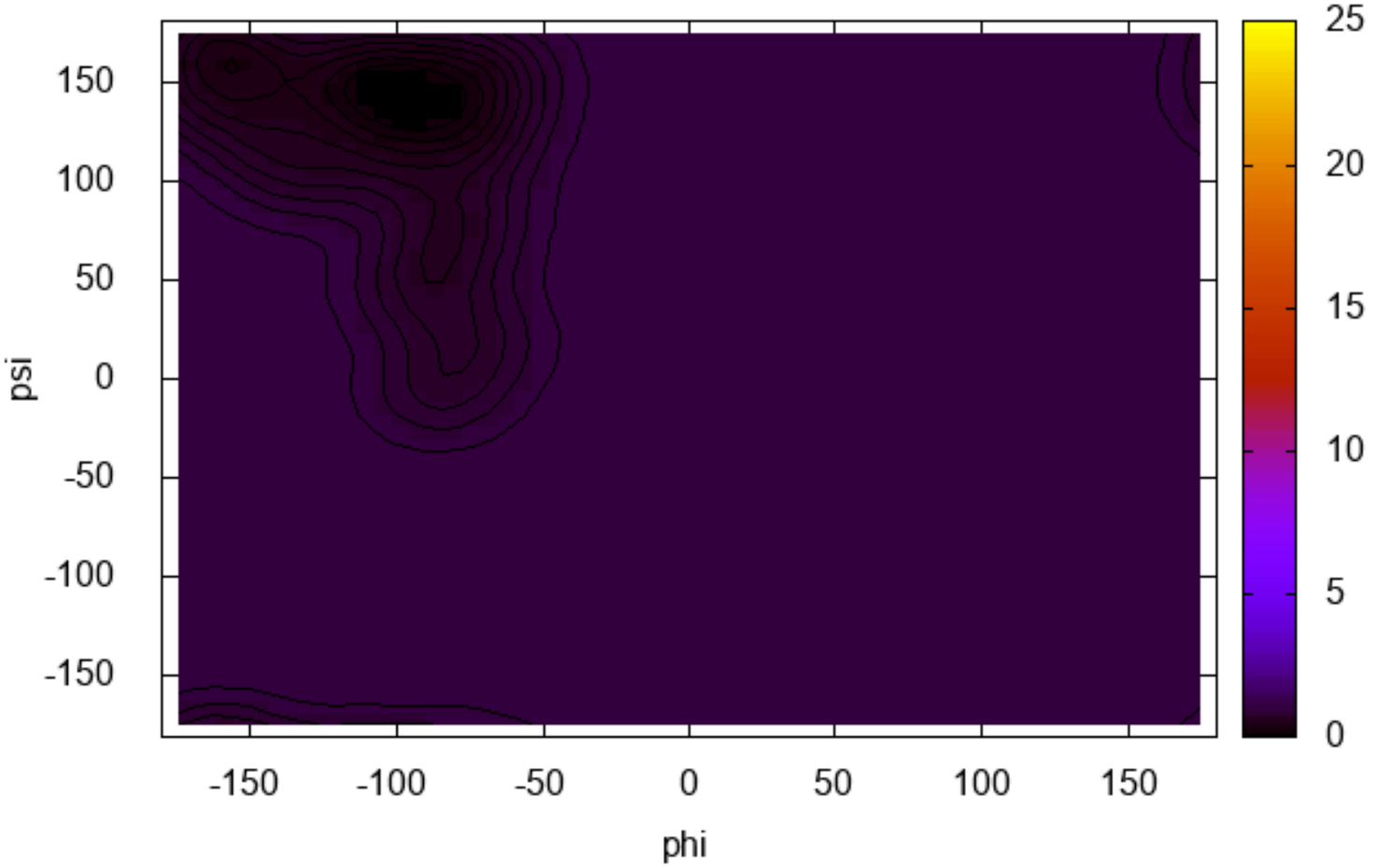


Now this is done with a single 10ns simulation (so 8x faster than PT or US)

Metadynamics

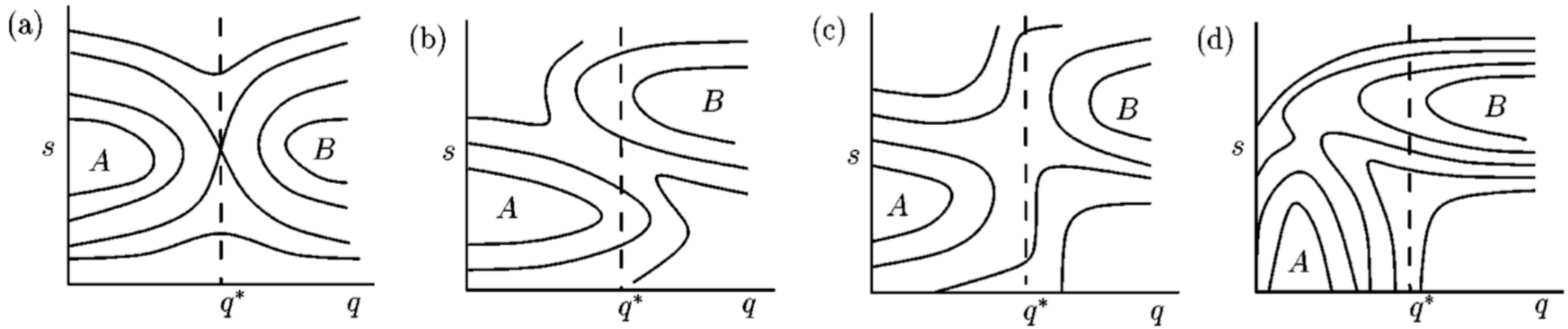


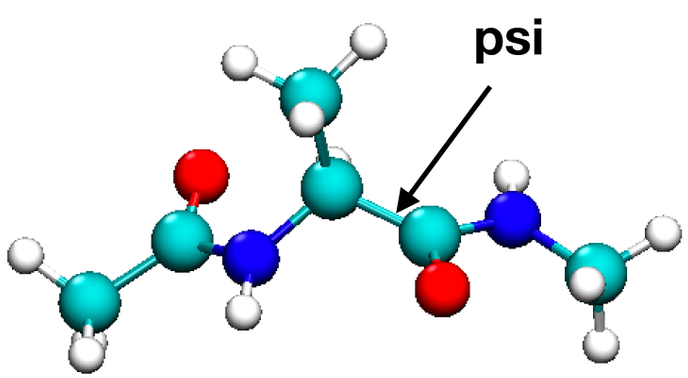
We can now easily run it in more than 1D (2 or 3, not much more)



Choosing CV can be tricky

Projections are tricky:





With US and MTD the big issue is not anymore how to speed up the sampling but how to choose a good reaction coordinate. We cannot choose many because the method is exponentially slower with the number of CVs.

Conclusions

Molecular dynamics simulation can provide a time-resolved picture of biological phenomena at atomistic resolution.

The resulting model are affected by at least three sources of error

- **Quality of the physical model employed**
- **Extent of the sampling**
- **Correctness in the extrapolation of the information**

When they work correctly they allow

- **identifying relevant forces at play in a process**
- **predict the effect of external agents a process**
- **visualise processes in action**
- **provide structural representation of statistical processes**