Molecular Dynamics Simulations of Proteins

Structure, Dynamics and Function

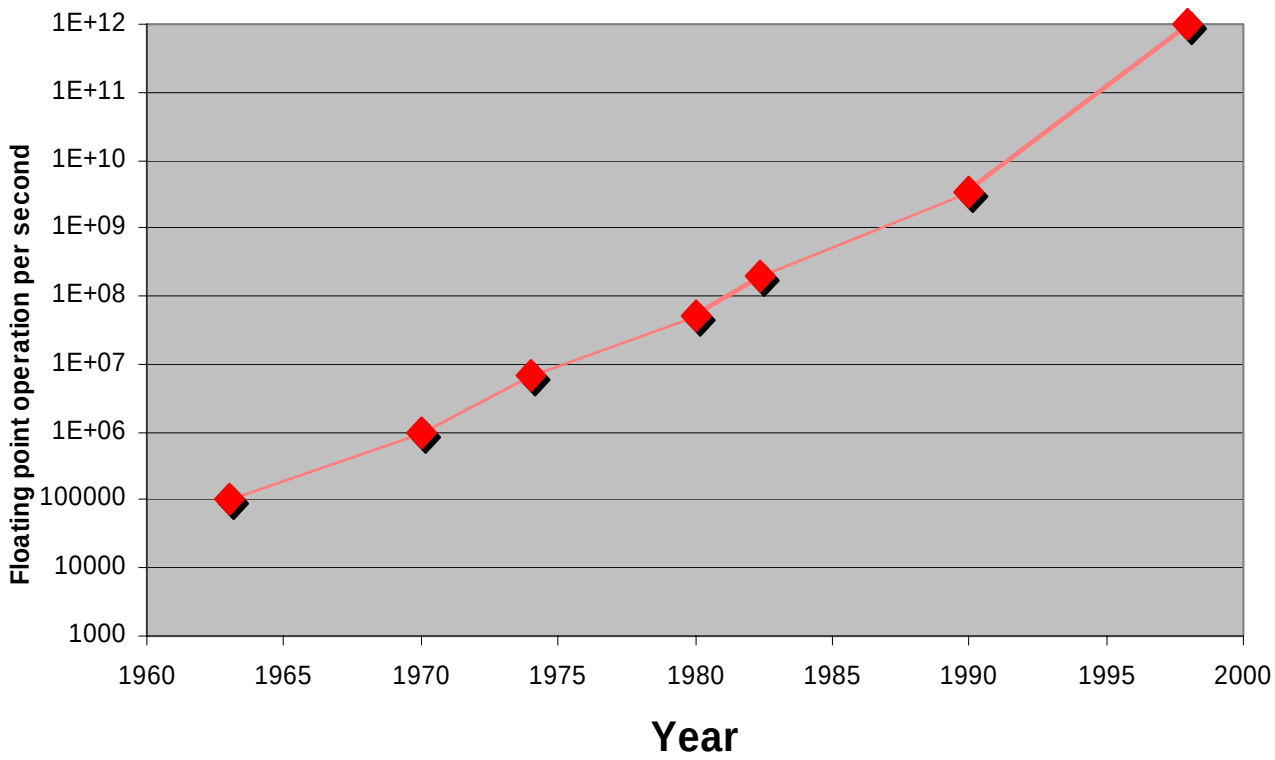
Summary

Method

What is Molecular Dynamics
Basis of the method
Present limits

Perspectives.

Increase of the computer power in the last 40 years



History of Molecular Dynamics simulations

Year	System	Simulated time
1957	Bidimensional rigid disks	10 ps
1964	Monoatomic Liquids	5 ps
1971	Molecular liquids	10 ps
1971	Melted salts	10 ps
1975	Simple polimers	20 ps
1977	Small protein in vacuo	20 ps

./.

Year	System	Simulated time
1982	Simple membrane model	20 ps
1983	Protein crystal	2 ps
1986	DNA in water	100 ps
1989	Complex DNA-Protein	100 ps
1993	Protein/DNA in solution	100 ps
1996	Protein/DNA in solution	10 ns
2000	Protein/DNA in solution	100 ns
2004	Protein/DNA in solution	1 μs
future	Reactions	100 μs
	Interactions between Macromolecules	1 ms
	Protein folding	100 ms

Development of different aspects of MD simulation in chemistry.

Aspect	Past (1980)	Present (2000)	Future
Accuracy of atomic positions	0.3 nm	0.1 nm	0.05 nm
Force Field	United Atoms	All Atoms	Polarizability
Environment	Vacuo	Solvent	Membrane
Time length	10 ps	1-100 ns	>100 ns
Dimensions	10^3 atoms	10^4-10^5 atoms	> 10^5 atoms

MD Applications

- Refininment of strucures obtained by experimental data (X-Ray or NMR)
- Prediction of equilibrium quantities and related thermodynamic quantities
- Time evolution of the system. Adequate and correct sampling is crucial
- Mixed QM/MM method to evaluate the electronic properties

MD Applications

Structural characterization

Dynamics of the systems

Ligand-receptor interaction

Effects of mutations

Models

Model	Included degrees of freedom	Removed degrees of freedom	Predictable quantities
Quantum	Nuclei, electrons	Nucleons	Reactions
Polarizable atoms	Atoms, dipoles	Electrons	Interactions between charged ligands
All atoms	Solvent and solute atoms	Dipoles	Idratation
Solute atoms	Solute atoms	Solvent atoms	Gas phase properties

Method

The applied law is:

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

Given $\mathbf{F}_i(t)$, $\mathbf{v}_i(t)$, $\mathbf{r}_i(t)$

We obtain $\mathbf{v}_i(t+\Delta t)$, $\mathbf{r}_i(t+\Delta t)$
(Δt is the *time step*)

We can thus calculate $F_i(t+\Delta t)$ and proceed iteratively

QUESTIONS

- How can we obtain the initial positions and velocities?
- How can we know the forces?
- How can we solve the equations of motion?
- How large can Δt be?
- How long can the simulated time be?
- Which quantities (P, T, V ..) can be controlled during the simulation?
- How can the correctness of the simulation be determined?
- Which comparisons with experimental data are possible?

Choice of the initial velocities

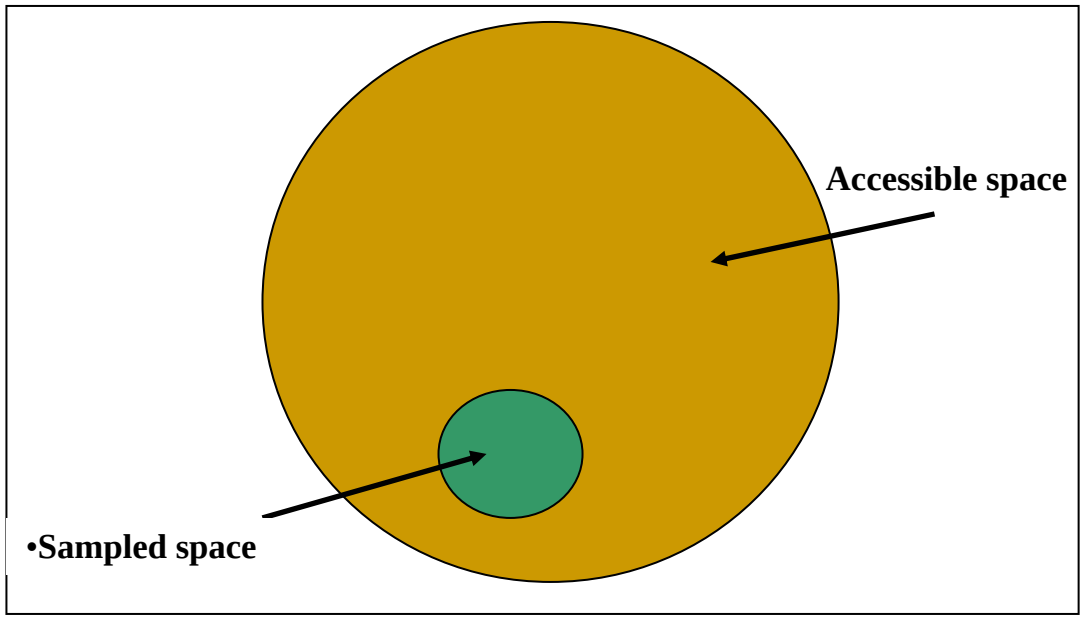
The initial velocities can be given by a Maxwell distribution:

$$\frac{1}{2}\sum m_i < v_i^2 >= \frac{3}{2} N k_B T$$

_____

Choice of the initial coordinates

The conformational space sampled in a simulation is much smaller than the total accessible space.



Choice of the initial coordinates

The initial coordinates can be obtained by:

Crystal structures

NMR Data

The Force Field

MD, MonteCarlo (MC) and Molecular Mechanics (MM) can use the same force field.

The force acting on atom ‘ i ’ is obtained by

$$\mathbf{F}_i = -\frac{\partial \mathbf{V}(\mathbf{r}_1,\ldots,\mathbf{r}_n)}{\partial \mathbf{r}_i}$$

A typical Force Field

$$V(r_1,.....,r_n) = \sum_{\text{bonds}} \frac{1}{2} \bullet K_b \bullet (b - b_0)^2 + \sum_{\text{angles}} \frac{1}{2} \bullet K_\vartheta \bullet (\vartheta - \vartheta_0)^2 +$$

$$\sum_{\substack{\text{improper} \\ \text{dihedrals}}} \frac{1}{2} \bullet K_\xi \bullet (\xi - \xi_0)^2 + \sum_{\text{dihedrals}} \frac{1}{2} K_\varphi [1 + \cos(n\varphi + \delta)] +$$

$$+ \sum_{\text{pairs}} \left\{ \varepsilon_{ij} \left[\frac{\sigma_{ij}}{r_{ij}} \right]^{12} + \left[\frac{\sigma_{ij}}{r_{ij}} \right]^{6} + \frac{q_i q_j}{4\pi\varepsilon \; r_{ij}} \right\}$$

- For the bond stretching the Morse potential is often used:

$$E(b) = D_e \{\exp[-A(b - b_o)] - 1\}^2$$

- Some force fields use the so called mixed terms:

$$\frac{1}{2} K_{\beta,\vartheta} \bullet (b - b_0)(\vartheta - \vartheta_0)$$

Hydrogen Bond

The hydrogen-bond is treated both

- as an electrostatic interaction

or

- as an explicit term.

Choice of the parameters

- The parameters are determined empirically to reproduce correct values of vibrational data, geometry of model compounds, free energy differences between rotamers, Debye-Weller factors, etc.…
- There are different force fields for different types of molecules: proteins, nucleic acid, inorganic species.

Charges

- A very common choice is the use of point charges. These are generally obtained by quantum calculations on model systems, or by empirical calculations based on the atom elecronegativity.

Polarizability

Polarizability is not usually used because of its computational requirements.

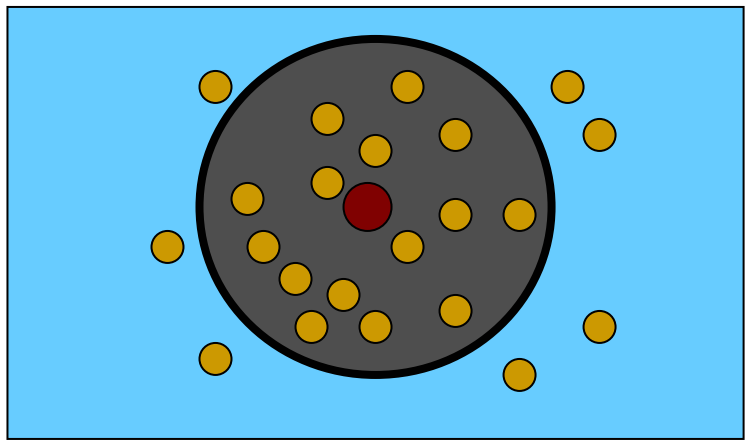
Computational time increases by a factor

2-3

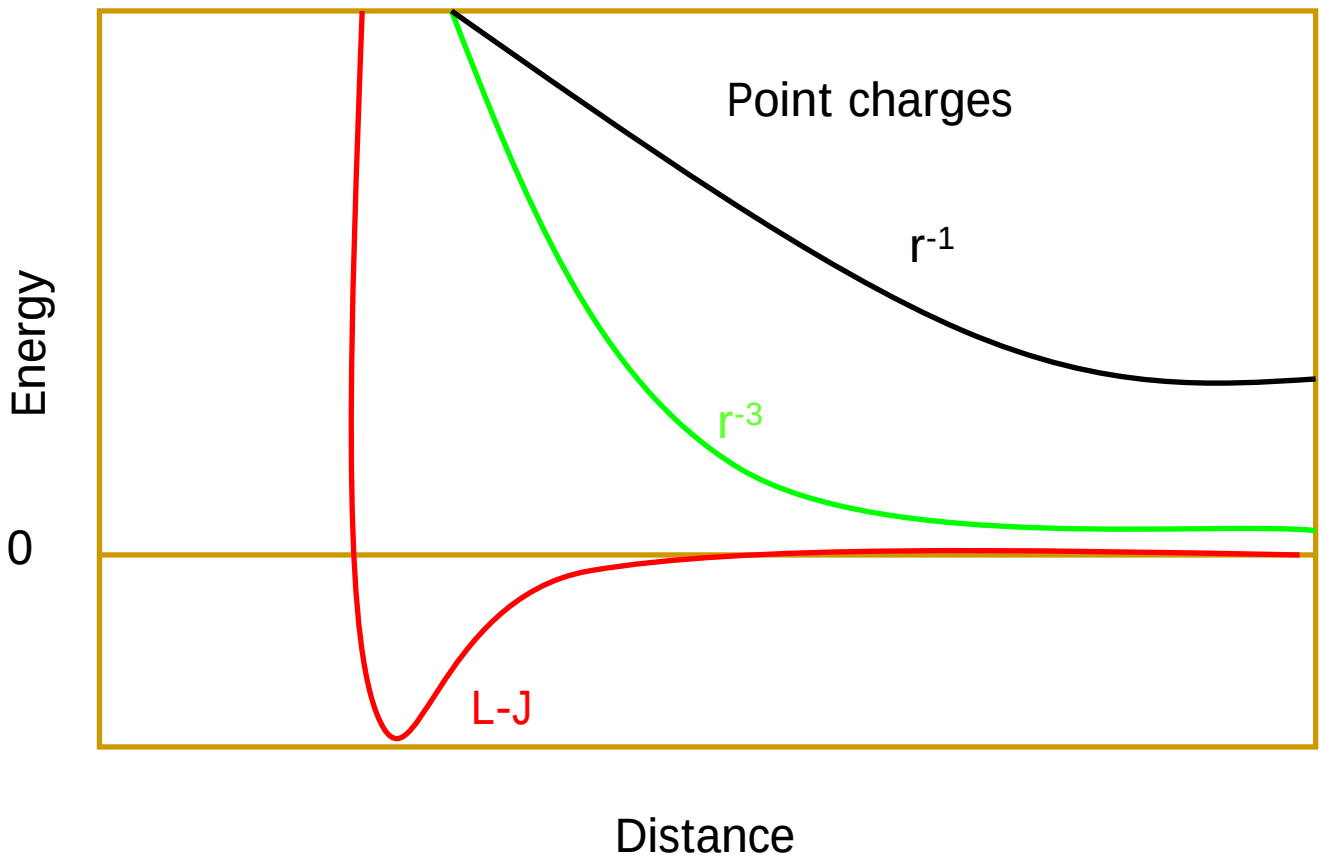
Interactions between non bonded atoms

The calculation of these interactions is the most time consuming task (~ the 95% of the total computational time).

The most simple tecnique to reduce the computational time is given by the so called cut-off method.



Interaction energies

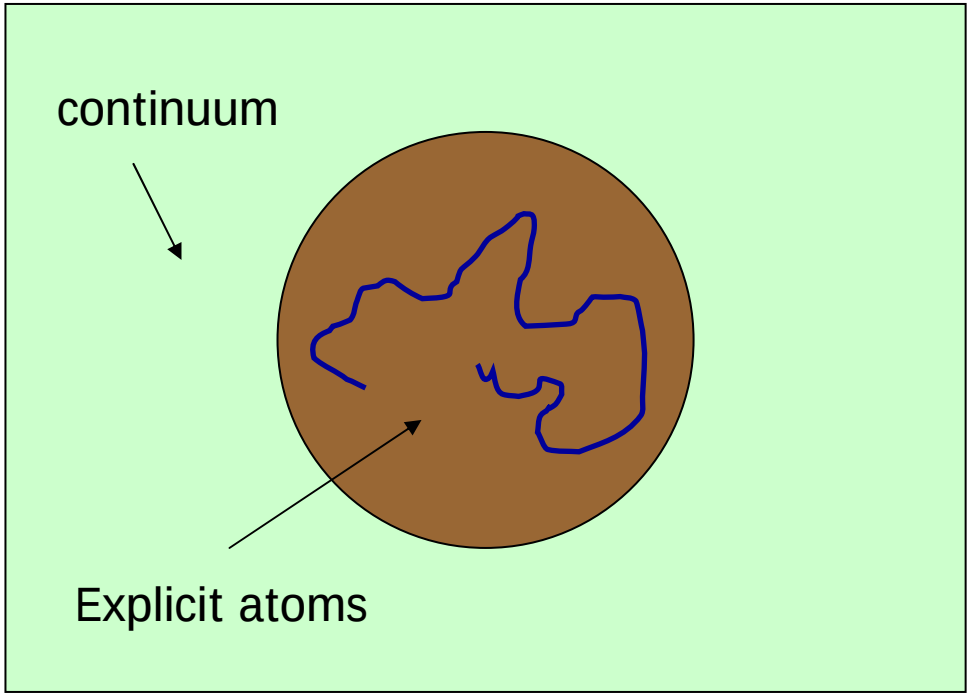


Electrostatic interactions

- Multipole expansion: the interaction between two charged groups can be written as the product of two multipole expansion: monopole-monopole (r^{-1}), monopole-dipole (r^{-2}), monopole-quadrupole and dipole-dipole (r^{-4}).
- It allows a large saving of computational time.

Mean field

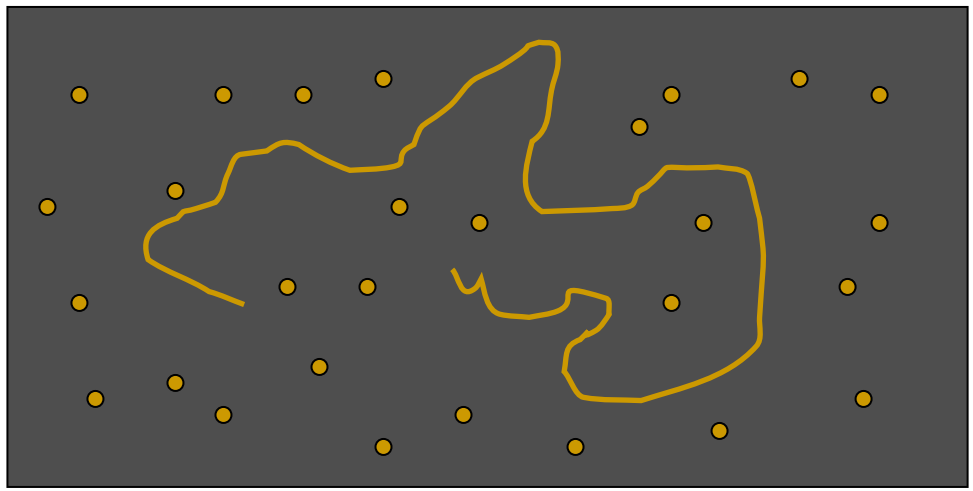
Treatment of a part of the system as a continuum.



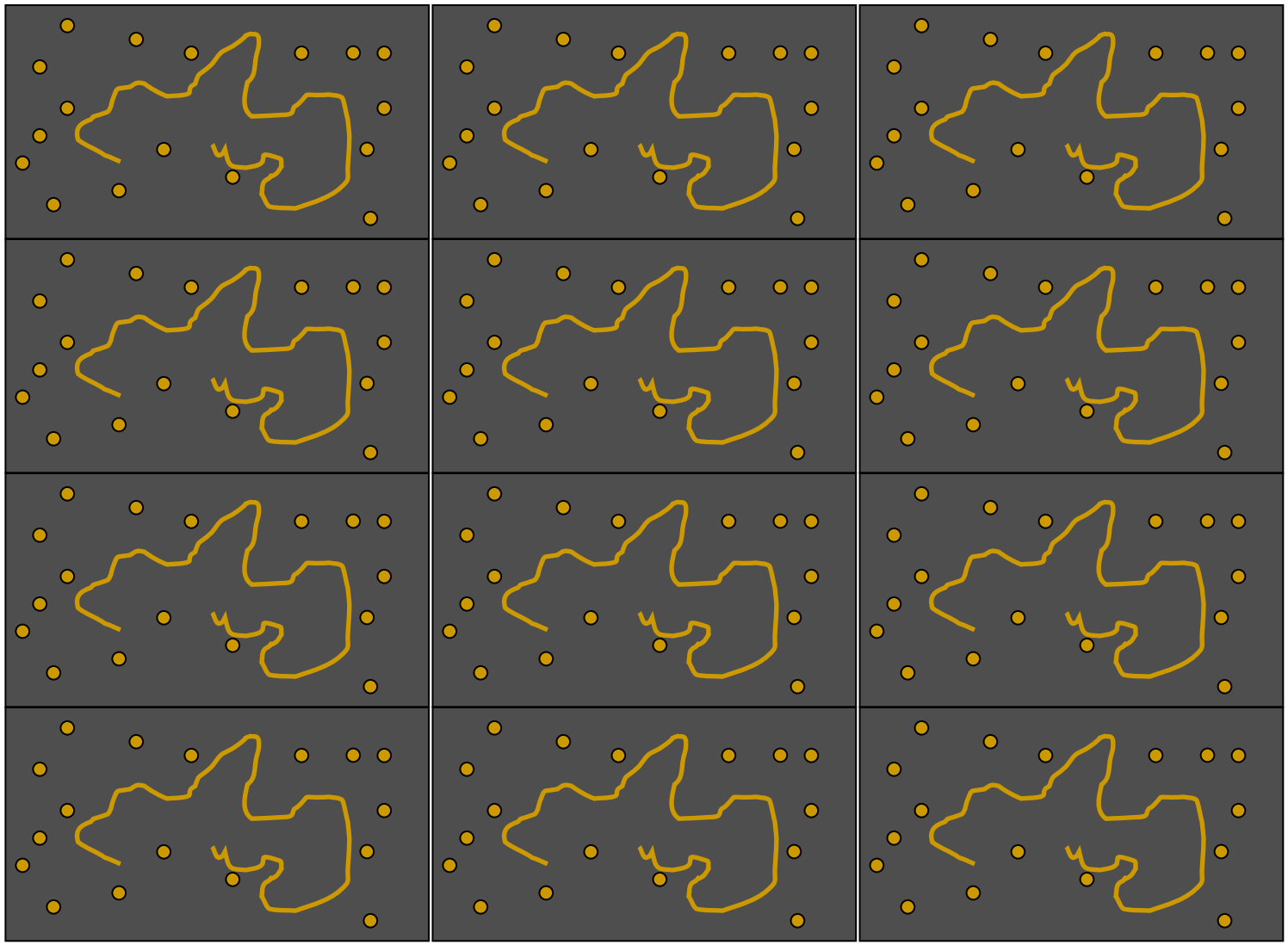
Boundary conditions

The simplest choice is the vacuum condition. This choice is correct for ideal gases only.

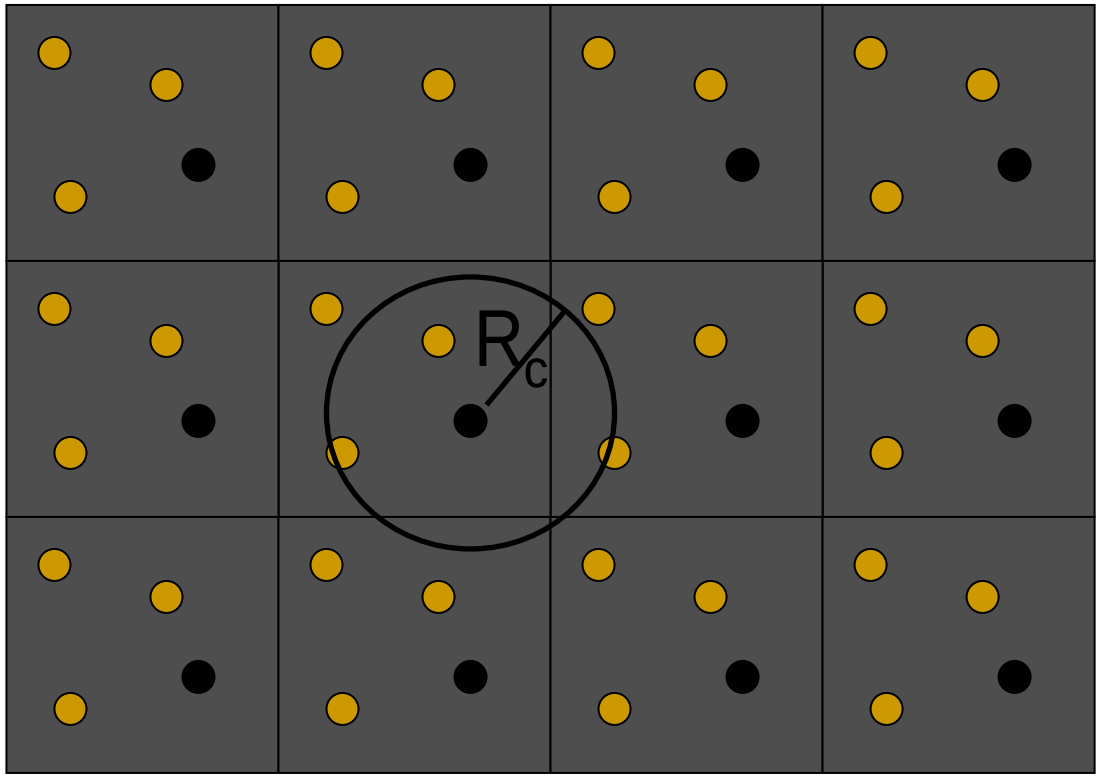
Vacuum boundary conditions are not correct for liquids, solutions or solids



Periodic boundary conditions

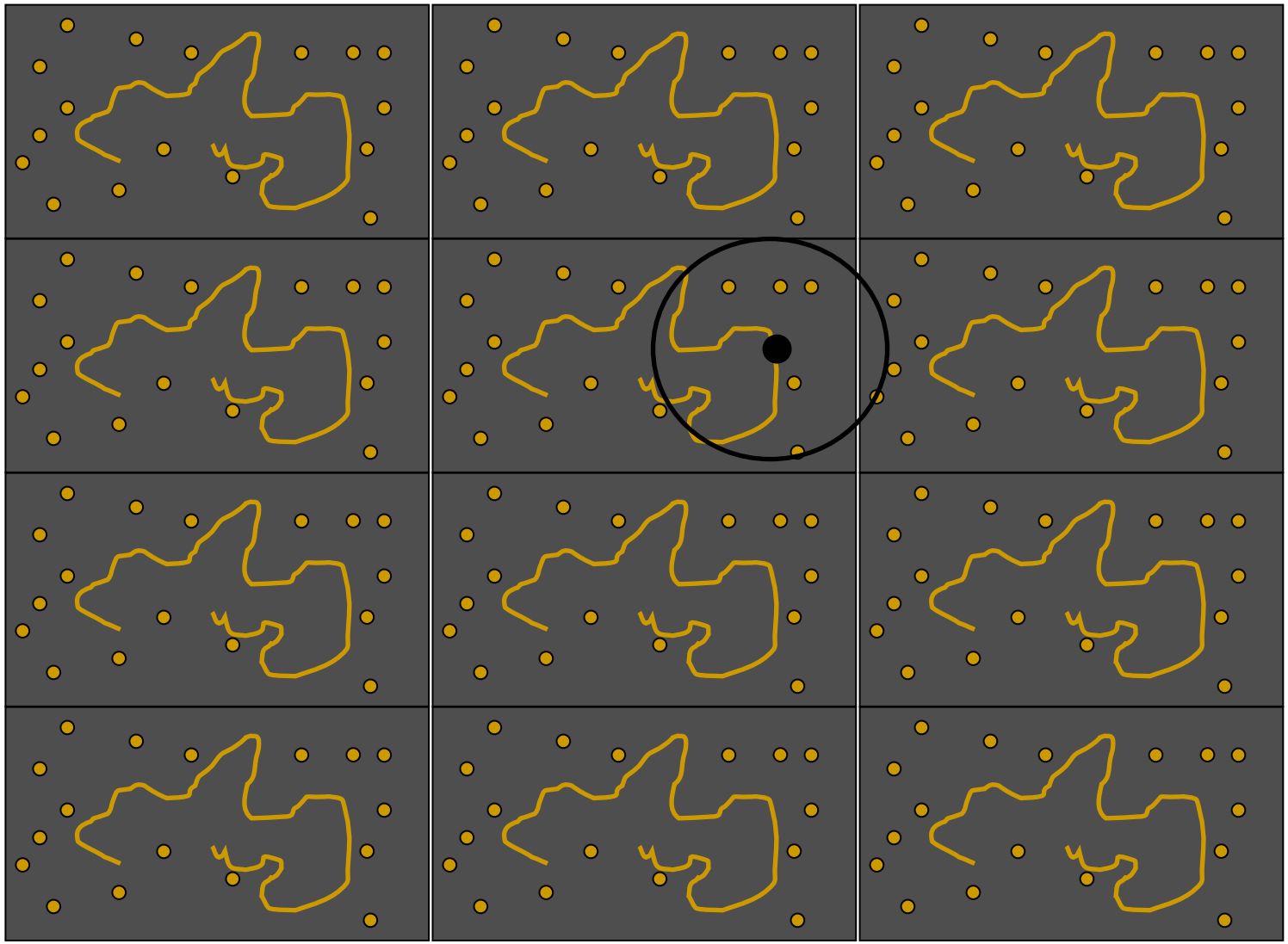


CUT-OFF with periodic boundary conditions



CUT-OFF

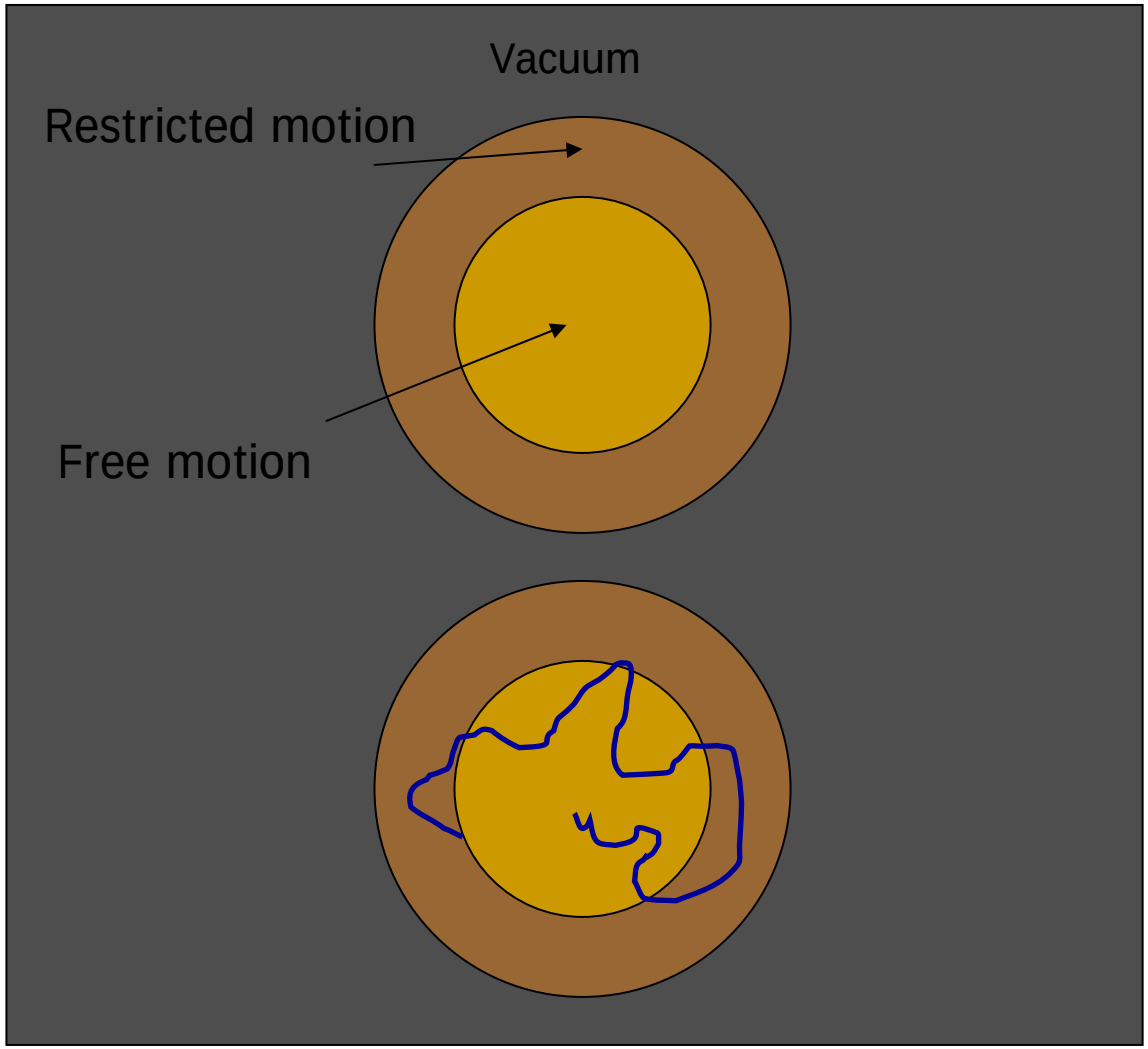
with periodic boundary conditions



ELECRTOSTATIC INTERACTIONS
Ewalds sum in periodic boxes

The sum of an infinite number of terms in a periodic system can be converted in a fast converging form that does not require a large computational effort.

Spherical boundary conditions



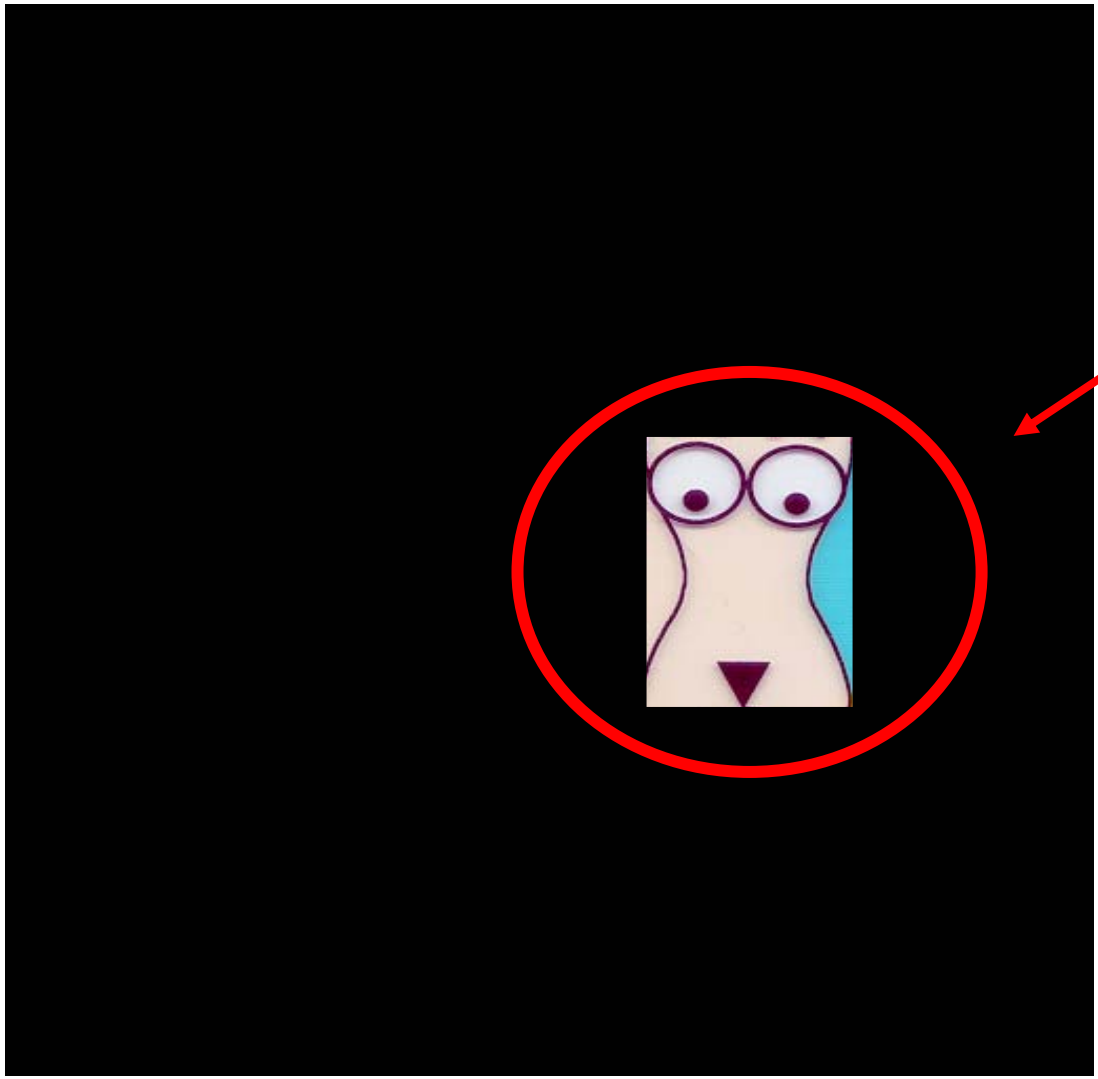
MD limits

- Time step
  - **2 fs with constant bond length simulations and light hydrogens**
  - **4 –6 fs with constant bond length simulations and heavy hydrogens**
  - **0.2 fs for variable bond length simulations**
- Size 10^4 -10^5 atoms

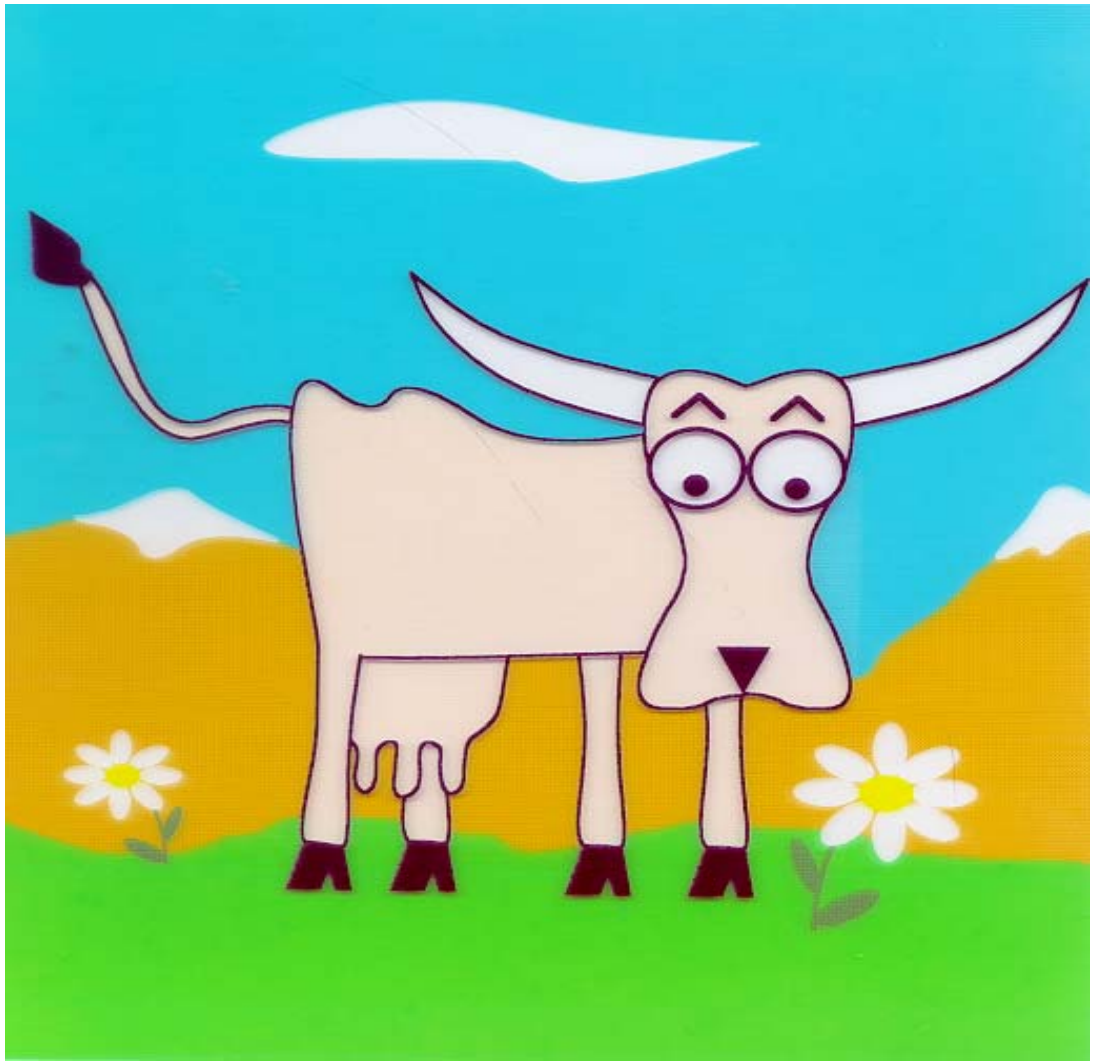
Sampling !!

- Simulated time 10^{2}-10^{3} ns
- Sampled space

In an attempt to avoid misleading conclusions…



It is important to perform adequate sampling



Increase of the sampled space: high temperature sampling

Di Nola, A., Berendsen, H. J. C., and Edholm, O. (1984) *Macromolecules*, **17**, 2044

- The advantage of MD with respect to other tecniques is the presence of the kinetic term that allows to overcome barriers of the order of magnitude of kT.
- High temperature simulation enhances the sampling of the conformational space.

Constrained dynamics (I)

SHAKE (Ryckaert , Ciccotti and Berendsen) is an iterative method that allows to perform a simulation with constant bond distances.

- Advantages
  - It is possible to increase the time step 4-5 times, up to Δt 2 fs.
- Disadvantages
  - Constraints reduce the flexibility of the system. Thus some properties can be affected.
  - SHAKE requires an iterative procedure which is difficult to be implemented on parallel codes.

Choice of the time step Δt

The time step must be:

- large to increase the total simulated time
- short to correctly integrate the equations of motion.

It must be must set according to the most rapidly varing forces (stretching)

$$\Delta t << \nu_{max}^{-1}$$

Some values ....

Time	Event
10 fs	Bond stretching and bending
40 fs	Other bnnd forces and short range non bonding interactions
1000 fs	Long range bonding interactions

For these reasons:

- Δt=0.2-0.5 fs with variable bond distances
- Δt= 1.0-2.0 fs with constant bond distances

IMPORTANT !!!

The choice of Δt depends on the atomic mass. For hydrogens a shorter time step should be used

Multiple time step (MTS)

Use of different time steps for

Different force types

Different atoms

Different distance between interacting atoms

Advantages

- All the degrees of freedom are taken into account
- Gain in computational time
- Easy implementation in a parallel code

Position Verlet integration scheme

Taylor expansion of $\mathbf{r}(t+\Delta t)$ and $r(t-\Delta t)$

up to 3rd order gives:

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + \Delta t^2 a(t)$$

Velocities are given by

$$v(t) = \frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t}$$

Leap Frog integration scheme

The integration formula is:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t+\frac{\Delta t}{2})\Delta t + o(\Delta t^3)$$

Stochastic Dynamics (SD)

SD is an extension of MD. An SD trajectory is obtained integrating the Langevin equation for each atom:

$$\frac{d^2 r(t)}{dt^2} = \frac{F}{m} + \frac{R}{m} - \gamma \frac{dr(t)}{dt}$$

With respect to the Newton's equation there are two additional terms:

The stochastic force **R**

The friction term γ

Related by the following expression:

$$\langle R \rangle = 6mTk_B\gamma$$

Pressure and temperature control

- Constraint methods: T(t) is exactly rescaled to the reference temperature T_0, operating on velocities. The simulation creates a canonical distribution in the coordinates but not in velocities
- Stochastic methods: the velocity of each atom are randomly modified simulating collisions whith a particle of the temperaturebath in which the particles are embedded

Pressure and temperature control II

- Extended system methods: in these methods a new degree of freedom is added to the system, simulating an external temperature bath. The metod yields to a canonical distribution in the coordinates and velocies, but is very sensitive to the choiche of the parameter that governs the heat exchange between system and bath.

Pressure and temperature control III

- Weak coupling method: in this method (Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W., Di Nola, A. and Haak, J. R. (1984), *J. Chem. Phys.*, **81**, 3684) the motion equation are modified to allow a first order relaxation of T(t) to T_o This is a stochastic method with null stochastic force and variable friction coefficient.

The Berendsen bath

- The coupling with the external bath is obtained rescaling the velocity of each single particle according to a parameter λ. The strength of the system-bath interaction is determined by a time constant τ and follows an exponential law:

$$\frac{dT}{dt} = \frac{1}{\tau}\left[T_0 - T(t)\right]$$

Isothermal Gaussian coupling

With this method the temperature is kept constante using the following motion equations:

$$\frac{\partial U(p,\ q)}{\partial p_i} = \dot{q}_i$$

$$\frac{\partial U(p,\ q)}{\partial q_i} = -\dot{p}_i + \gamma p_i$$

where γ is a friction coefficient given by

$$\gamma(r,\ p) = \frac{\sum_{i=1}^{N} p_i f_i / m_i}{\sum_{i=1}^{N} p^2{}_i / m}$$

Weak coupling methods applied to pressure control

- The first order relaxation of P(t) to P_0 follows the equation

$$\frac{dP(t)}{dt} = \frac{P_0 - P(T)}{\tau}$$

- For an isotropic system the pressure is defined

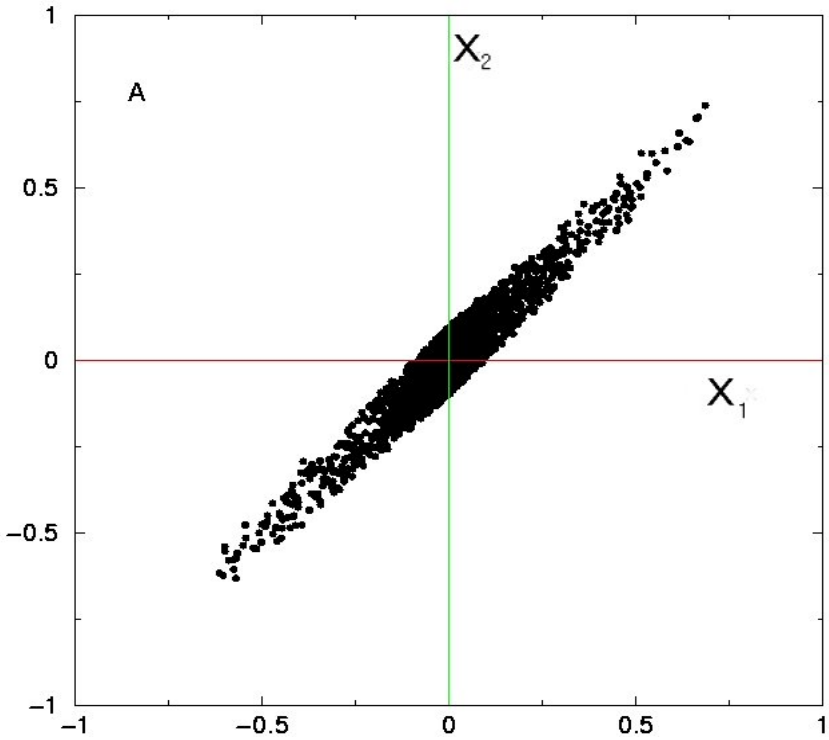
$$P(t) = \frac{2}{3}\frac{E_{kin}(t) - \Theta(t)}{V_{box}(t)}$$

where $V_{box}(t)$ is the volume of the computational box and $\Theta(t)$ is the Virial of the forces:

$$\Theta(t) = -\frac{1}{2}\sum_{i<j} r_{ij}(t) \bullet F_{ij}(t)$$

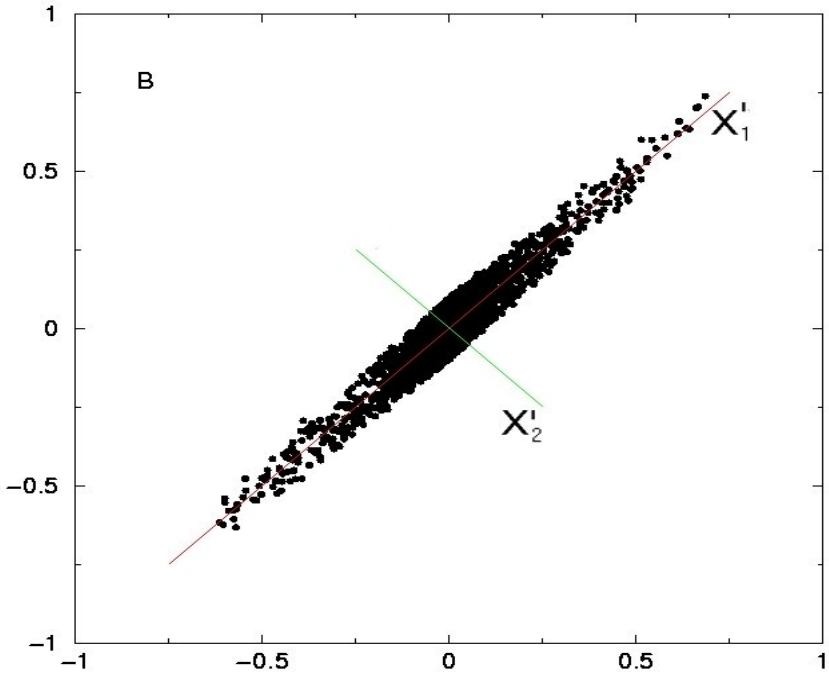
Essential Dynamics of proteins

- This method has been independently proposed by
  - A. Amadei, A. B. Linssen, H. J. C. Berendsen
    - Proteins: Struct. Funct. Gen. 17:412-425 (1993)
  - A. E. Garcia
    - Phys. Rev. Lett. 68:2696-2699 (1992)



$$\left\langle \left(X_1 - \langle X_1 \rangle\right)^2 \right\rangle \qquad \left\langle \left(X_1 - \langle X_1 \rangle\right) \bullet \left(X_2 - \langle X_2 \rangle\right) \right\rangle$$

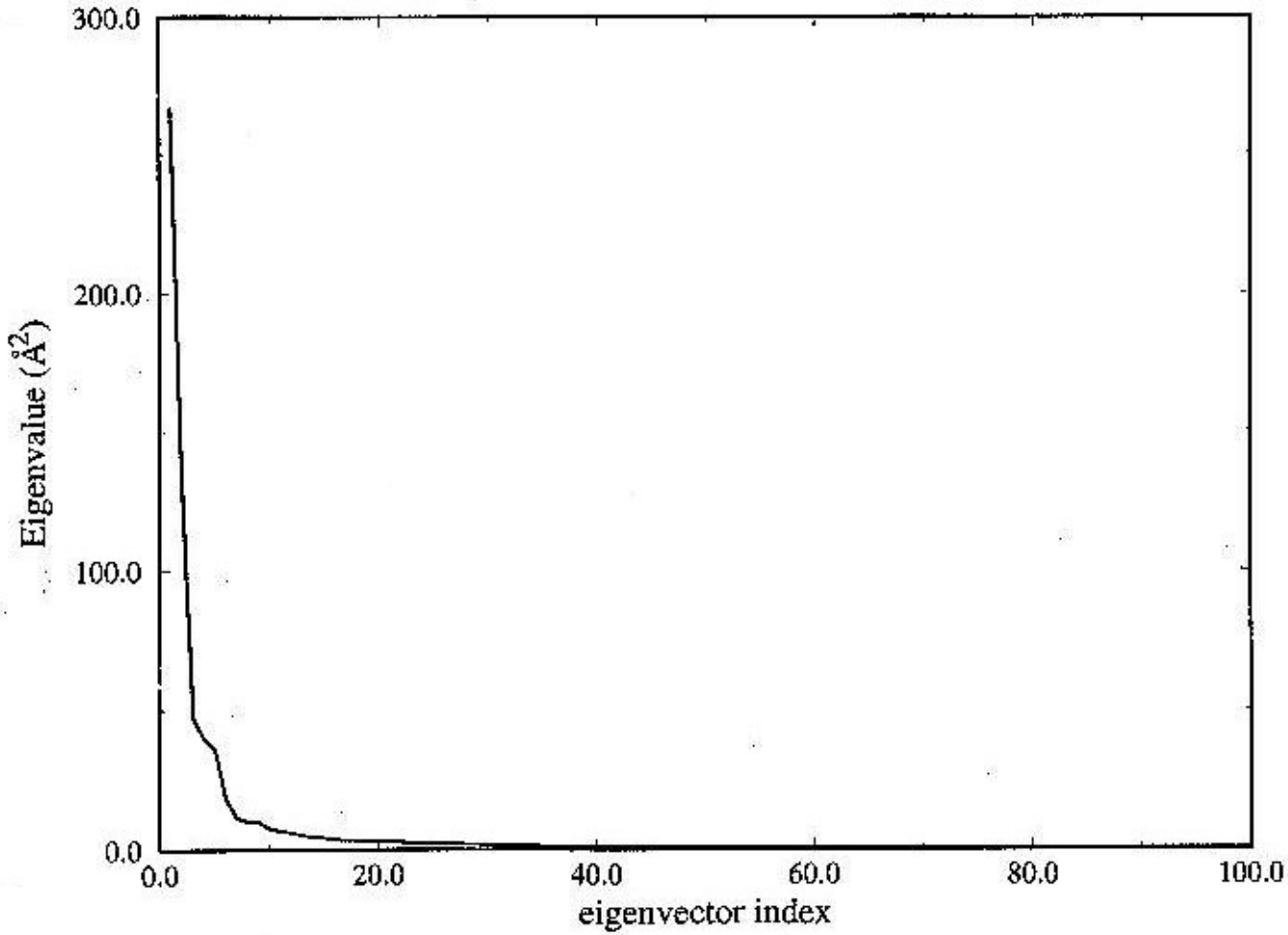
$$\left\langle \left(X_1 - \langle X_1 \rangle\right) \bullet \left(X_2 - \langle X_2 \rangle\right) \right\rangle \qquad \left\langle \left(X_2 - \langle X_2 \rangle\right)^2 \right\rangle$$

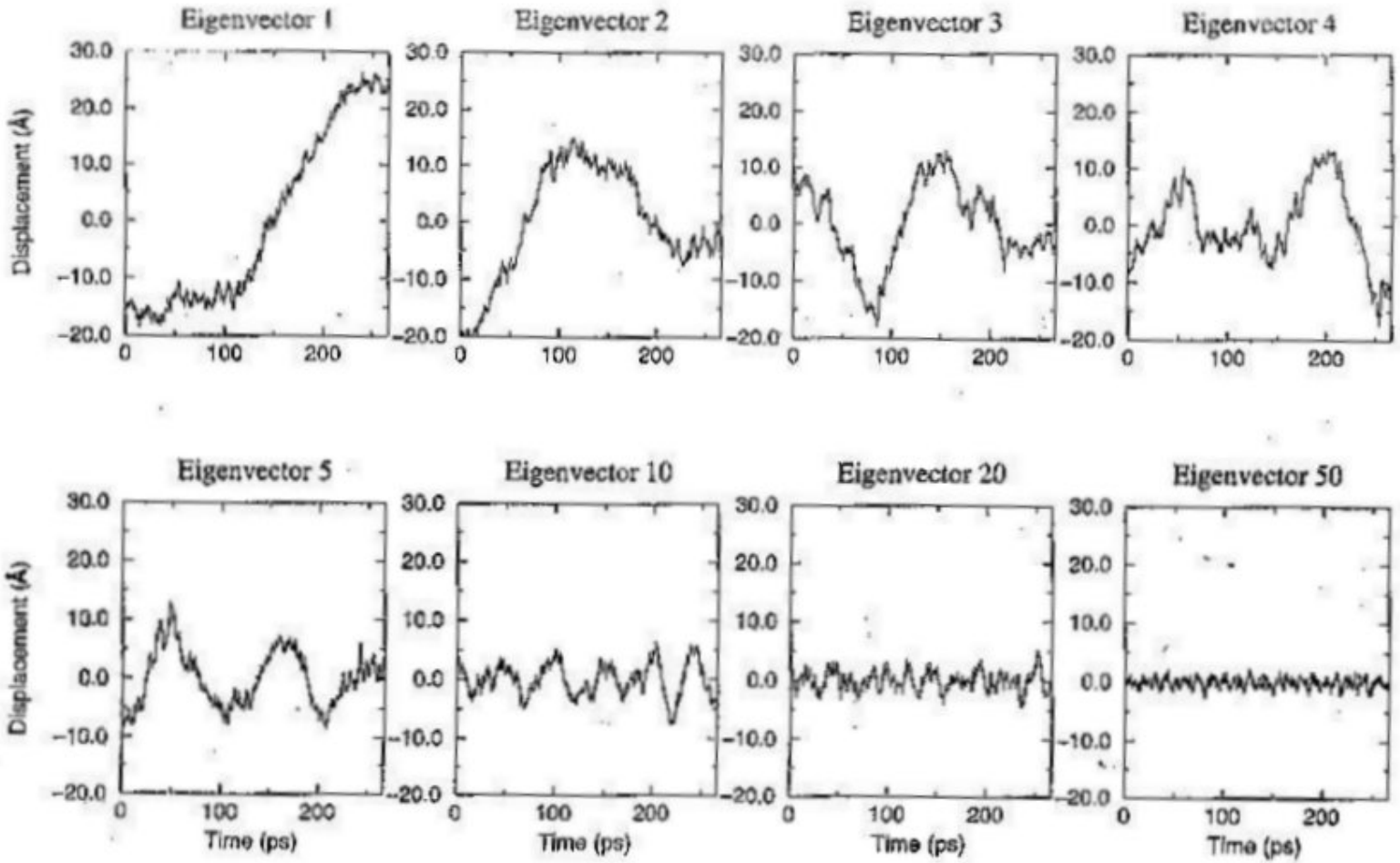


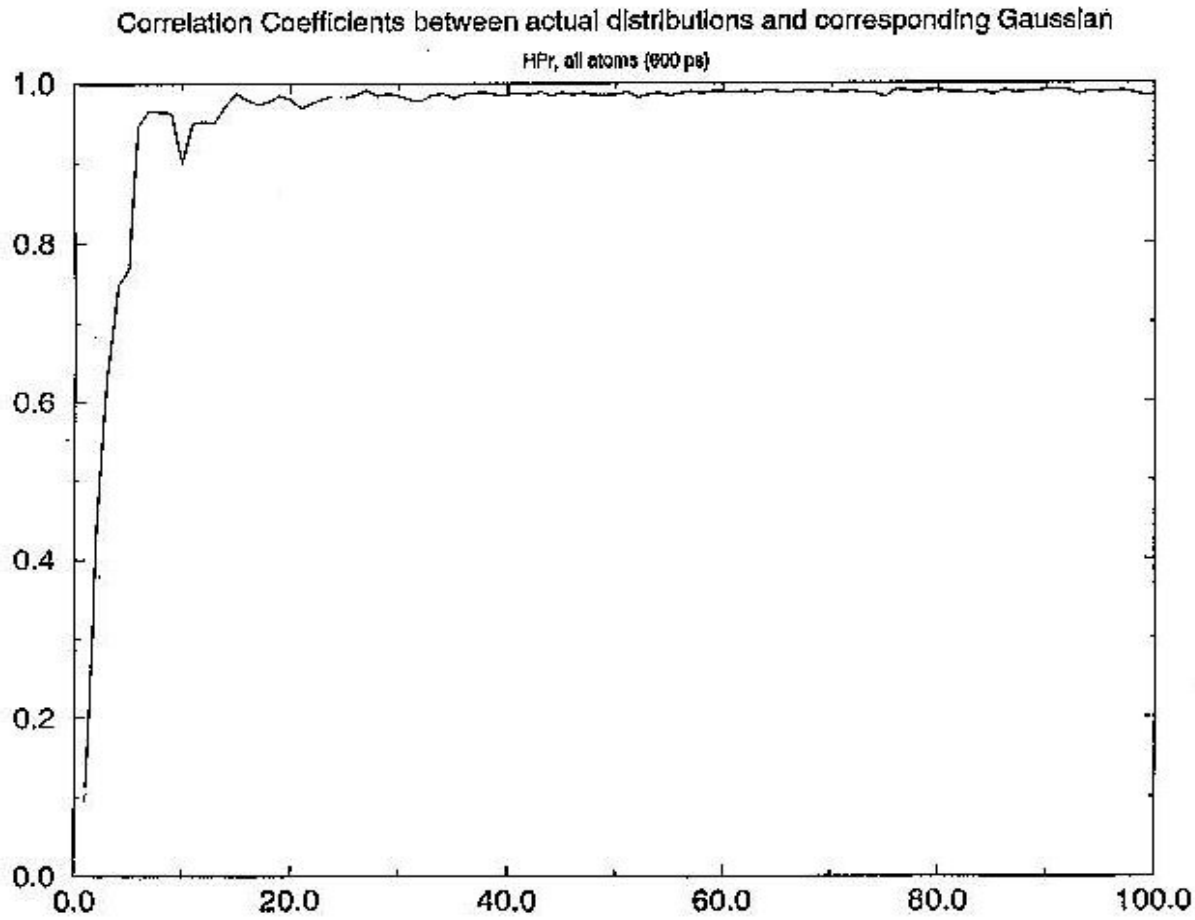
$$X'_1 = a_{11}X_1 + a_{12}X_2$$
$$X'_2 = a_{21}X_1 + a_{22}X$$

$$\left\langle (X'_1 - \langle X'_1 \rangle)^2 \right\rangle \approx 0$$
$$\approx 0 \qquad \approx 0$$

- To define the essential subspace one needs a long MD simulation of the protein including solvent
- X(t) is the 3N-dimensional vector describing the trajectory with traslations of, and rotations around the center of mass removed
- X(t)-<X> is the 3N-dimensional vector of the atomic displacements on respect to the average
- $(X_i(t)-\langle X_i\rangle)(X_j(t)-\langle X_j\rangle)$ is the (i,j) element of the covariance matrix C
- $\Lambda=T^{-1}CT$ or $C=T\,\Lambda\,T^{-1}$ indicates a diagonal that yields to the eigenvectors and eigenvalues


300.0
200.0
100.0
0.0
Eigenvalue (Å²)
0.0
20.0
40.0
60.0
80.0
100.0
eigenvector index


Eigenvector 1
Eigenvector 2
Eigenvector 3
Eigenvector 4
Eigenvector 5
Eigenvector 10
Eigenvector 20
Eigenvector 50
Displacement (Å)
Time (ps)
30.0
20.0
10.0
0.0
-10.0
-20.0
0
100
200


Correlation Coefficients between actual distributions and corresponding Gaussian
HPr, all atoms (600 ps)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
20.0
40.0
60.0
80.0
100.0

Docking

- Docking is a metodology with which we try to reproduce the interactions between molecules in a complex

The objective is to reconstruct a molecular complex starting from dissociates molecules

Docking methods

- Both protein and ligand rigid
- Rigid protein and flexible ligand
- Both protein and ligand flexible
- Explicit solvent molecules

Molecular Dynamics Docking (MDD)

A. Di Nola, D. Roccatano, H. J. C. Berendsen, *Proteins: Structure, Function, and Genetics*, **19**, 174-182 (1994).

M. Mangoni, D. Roccatano and A. Di Nola, *Proteins: Structure, Function, and Genetics*, **35(2)**, 153-16 (1999).

$$E_{kin} = \frac{1}{2}\sum_i m_1 v_i^2$$

The velocity can be decomposed as:

$$\mathbf{v}_i = \mathbf{v}_{c,i} + \mathbf{v}_{CM} \quad \text{where} \quad \mathbf{v}_{CM} = \frac{\sum_i m_i \mathbf{v}_i}{M}$$

$\mathbf{v}_{c,i}$ is the velocity of the atom with respect to the center of mass

$\mathbf{v}_{CM}$ is velocity of the center of mass

Molecular Dynamics Docking (MDD)

The kinetic energy can be written as:

$$E_{kin} = \frac{1}{2}\sum_i m_i v_i^2 = \frac{1}{2}\sum_i m_i v_{c,i}^2 + \frac{1}{2} M v_{CM}^2$$

The last two terms corresond to the internal and c.o.m. kinetic energies, resp.

They can be coupled to two different thermal baths

Molecular Dynamics Docking (MDD)

So we have two different temperatures for the rigid body motion and for the internal motions of the ligand

We can also have a different temperature for the protein and water molecules.

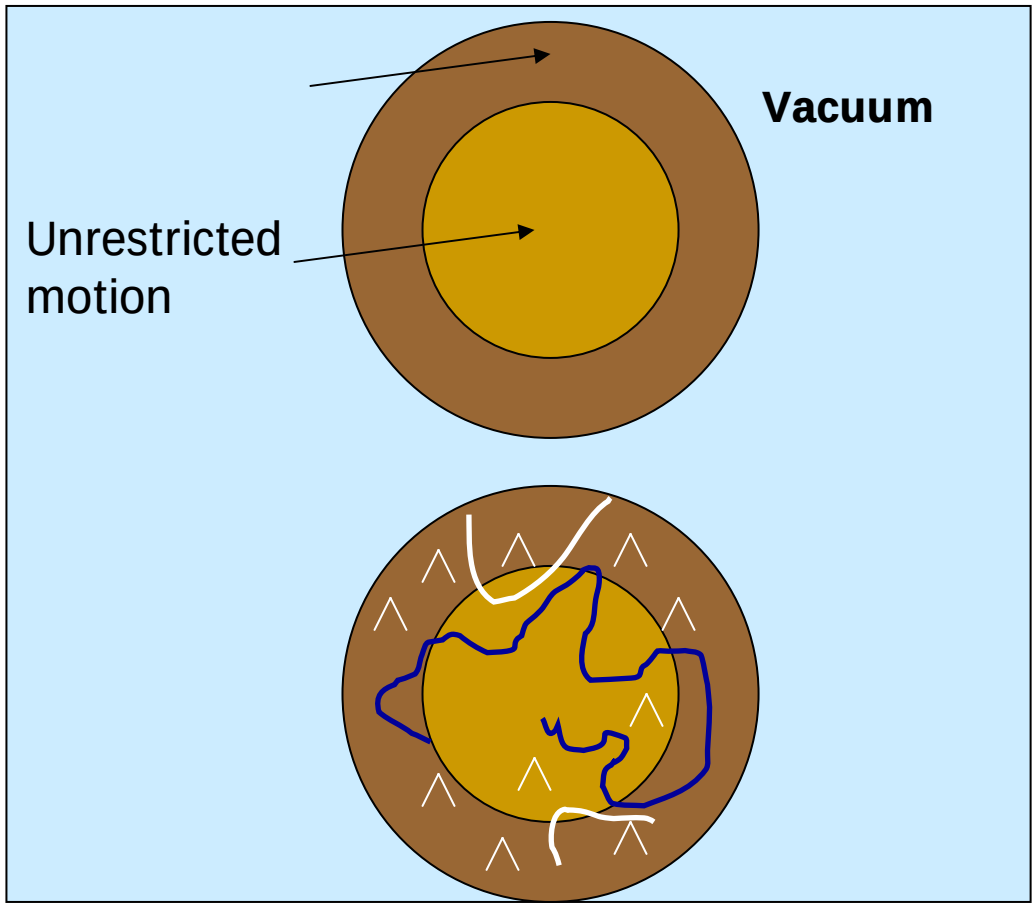
In solution it is also necessary to have a different weights for the ligand-protein interactions

Docking of the phosphocholine to the immunoglobulin McPC603 in solution

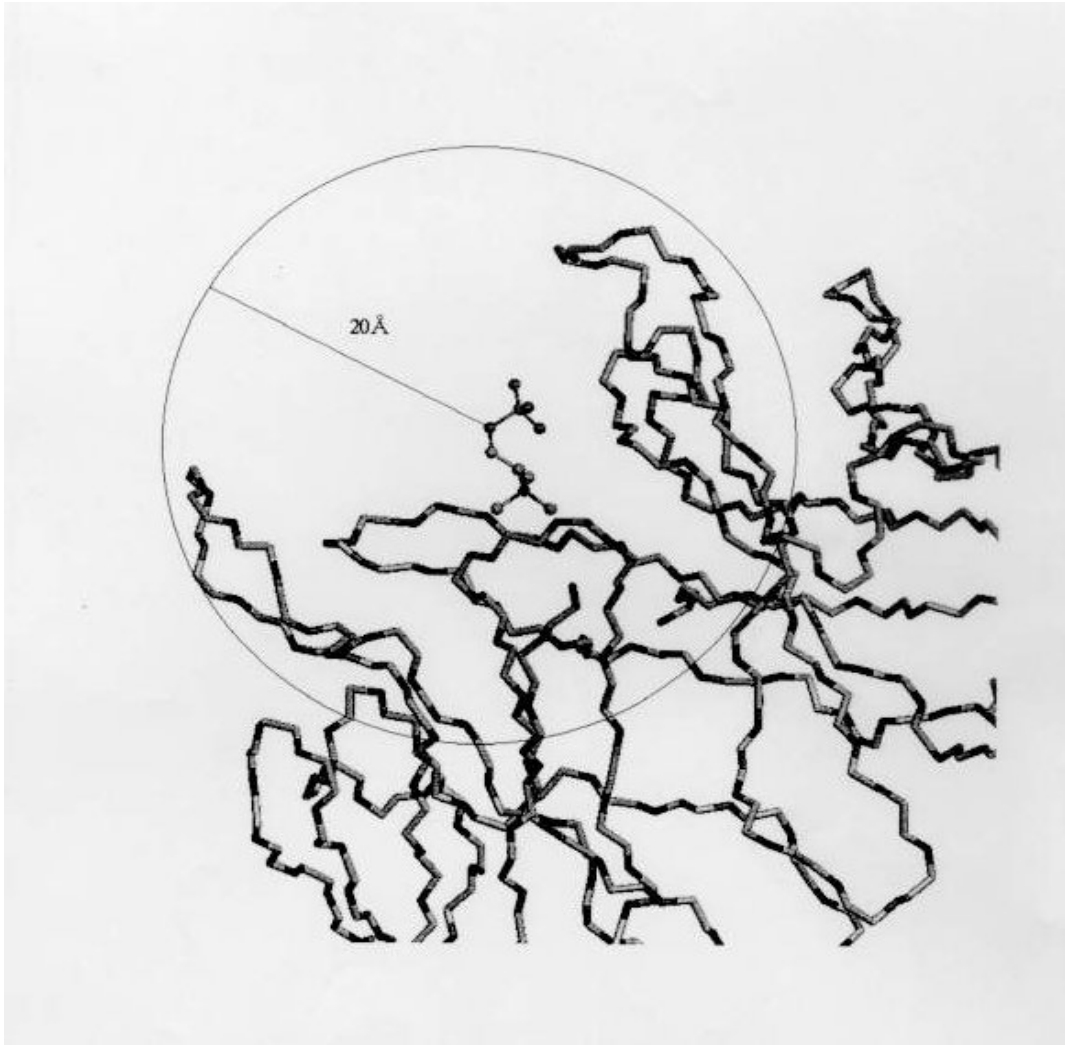
M. Mangoni, D. Roccatano and A. Di Nola, *Proteins: Structure, Function, and Genetics*, **35(2)**, 153-162 (1999).

Spherical boundary conditions

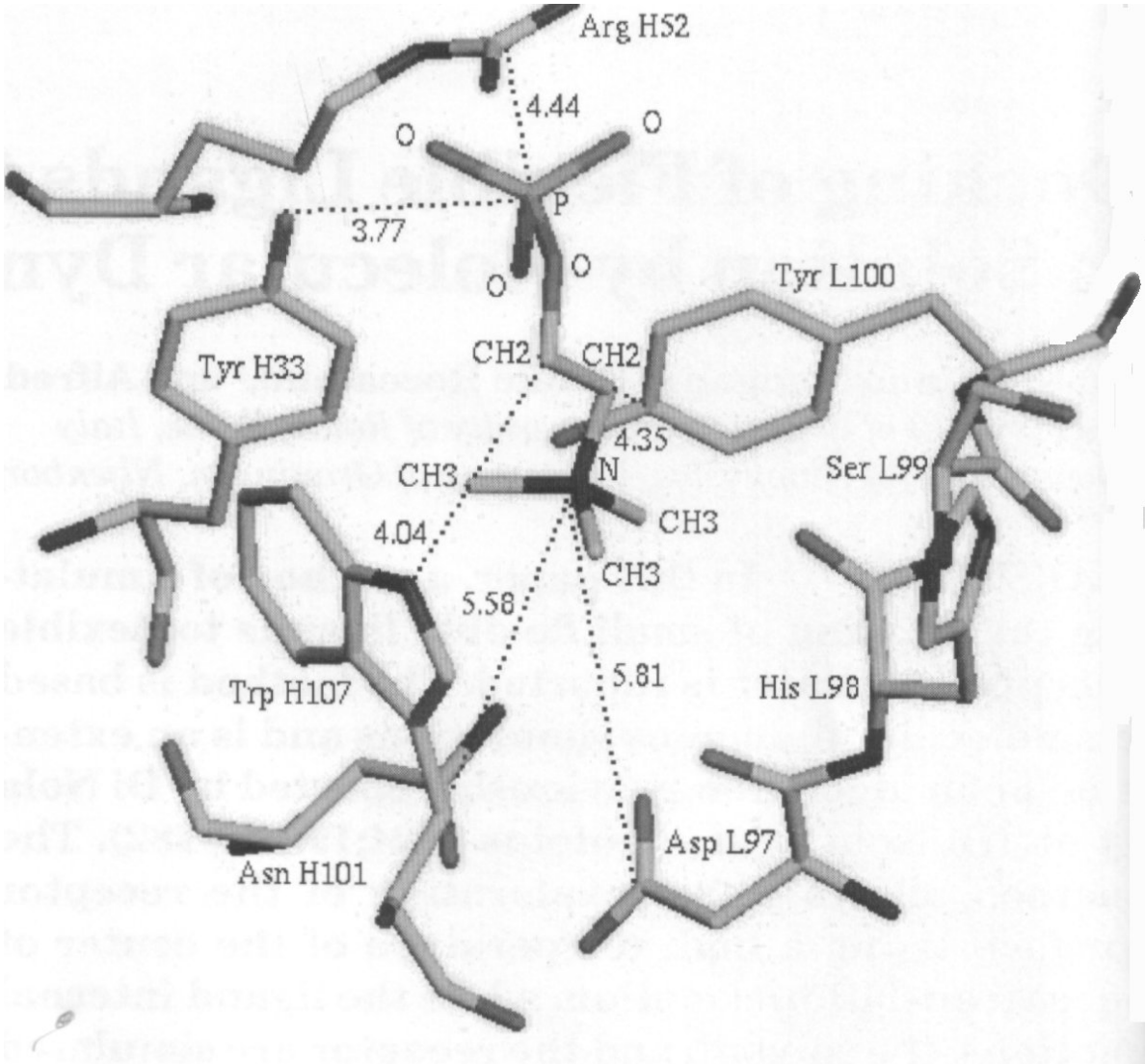
Brunger, A. T., Brooks, C. L., Karplus, M. K. (1985), *Proc. Natl. Acad. Sci., USA 82, 8458*

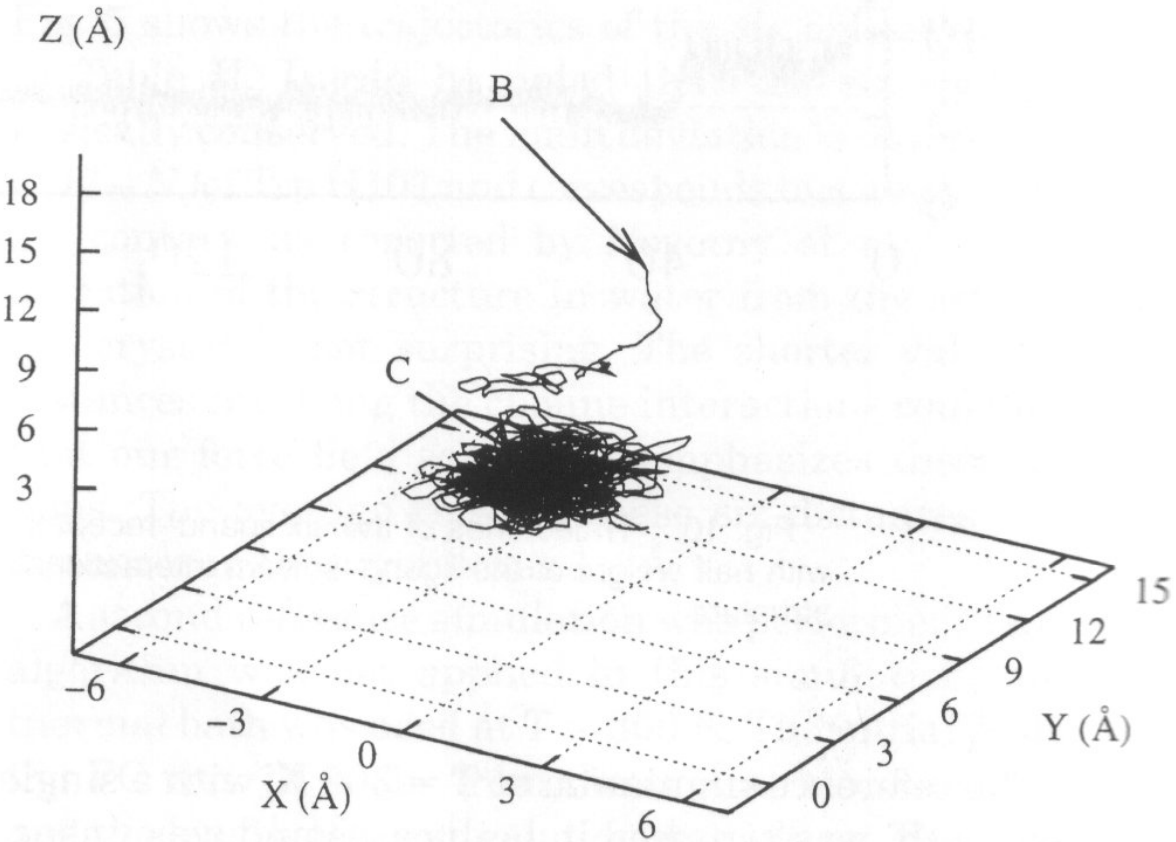


Position of the PC in the crystal complex



Main interactions of the PC in the crystal



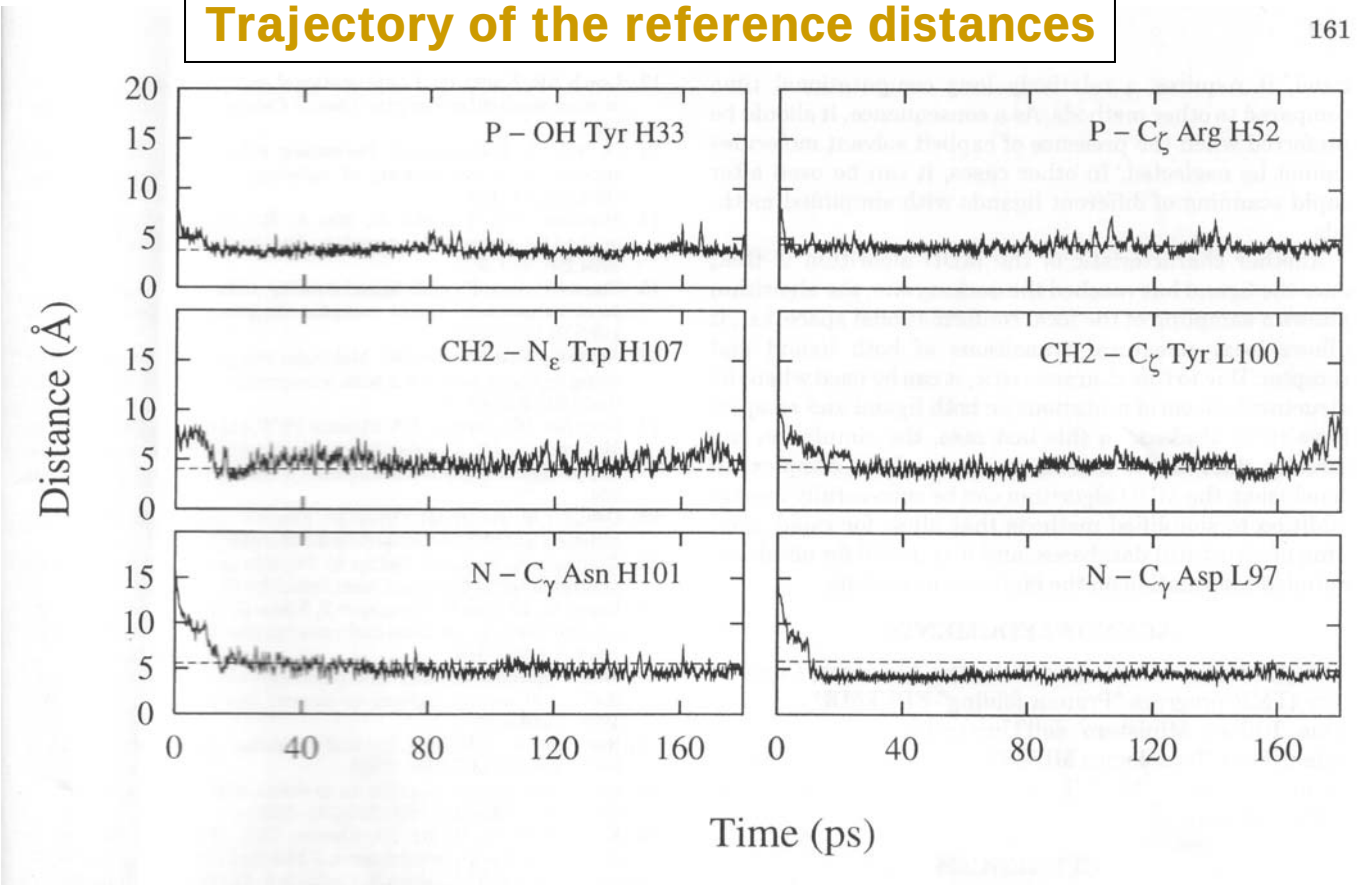


Trajectory of the center of mass of the PC during the 180 ps simulation.

B is the staring position.

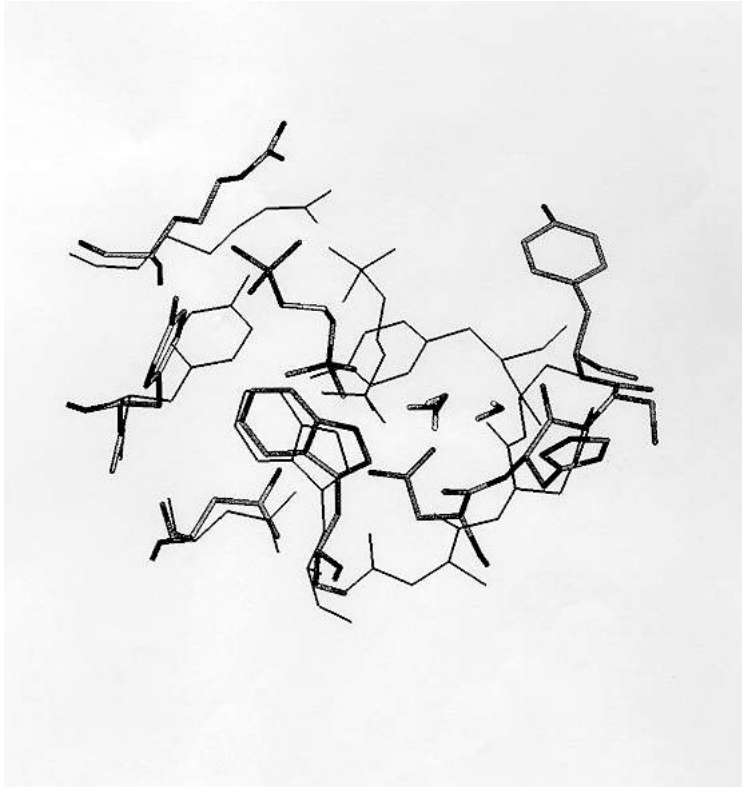
C is the PC position in the crystal

Trajectory of the reference distances



Dashed lines represent the distances in the crystal

Superimposition of the x-ray (thin) and MD (thick) structures



The free energy problem

We define

$$A = kT \ln\left\langle \exp\left[+\beta E(\mathbf{X})\right]\right\rangle$$

where

$$\left\langle \exp\left[+\beta E(\mathbf{X})\right]\right\rangle = \sum \exp\left[+\beta E(\mathbf{X})\right] P(\mathbf{X})$$

and $\mathrm{P}(\mathbf{X})$ is the probability of the state $\mathbf{X}$

We can notice that the high energy states have low probability P but high exponential. So

- The free energy converges slowly
- A very long sampling is required
- The frequence is substituted to probability and thus all the quantities are calculated as an average.

The coupling method approach

If we suppose tha the free energy has a dependence on a parameter λ, we can write

$$A(\lambda)=-kT\ln Z(\lambda).$$

Three different methods use this expression as the starting point of free energy calculations:

- Thermodynamic integration (TI)
- Perturbation Method (PM)
- Potential of Mean Force

Thermodynamic Integration

Starting from the simple relation

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} d\lambda$$

we can easily arrive to the following expression

$$\frac{\partial A(\lambda)}{\partial \lambda} = \left\langle \frac{\partial E(X^N, \lambda)}{\partial \lambda} \right\rangle_\lambda$$

where the subscript λ indicates an ensemble average over the following function

$$P(X^N, \lambda) = \frac{\exp\left[-\beta E(X^N, \lambda)\right]}{\int \ldots \int \exp\left[-\beta E(X^N, \lambda)\right] dX^N}$$

In this way we obtain the final relation of thermodynamic integration method

$$\Delta A = \int_0^1 \left\langle \frac{\partial E\left(X^N, \lambda\right)}{\partial \lambda} \right\rangle_\lambda d\lambda$$

- λ can be identified with volume, temperature, pressure changes , even if it is not restrict, ed to thermodynamic variables. In fact, a lot of analytical continuations between initial and final states can be used

- For example, ideal topological transition can be associated to λ: