



**UNIVERSIDAD
ANDRES BELLO**

Center for Bioinformatics and Integrative Biology
(CBIB)

Facultad de Ciencias Biológicas
Av. Republica 239, Santiago, Chile.

Introduction to molecular simulations

Dr. Fernando Danilo González Nilo
Dra. Valeria Marquez
Ing Bioinfo. Ingrid Araya

Innovation => Multidisciplinary discussion

Center for Bioinformatics and Integrative Biology (CBIB)
Escuela Ingeniería Bioinformática
Universidad Andres Bello

CBIB = Integration



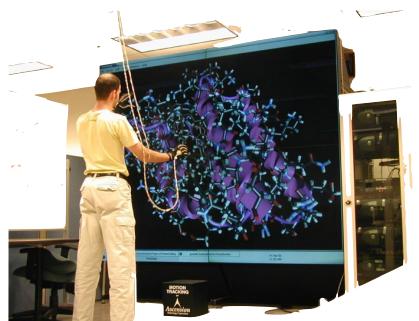
Cell Biology



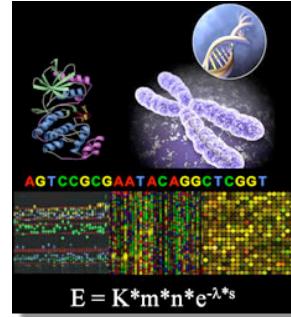
Electron Microscopy



Microbiology



Computational Biology

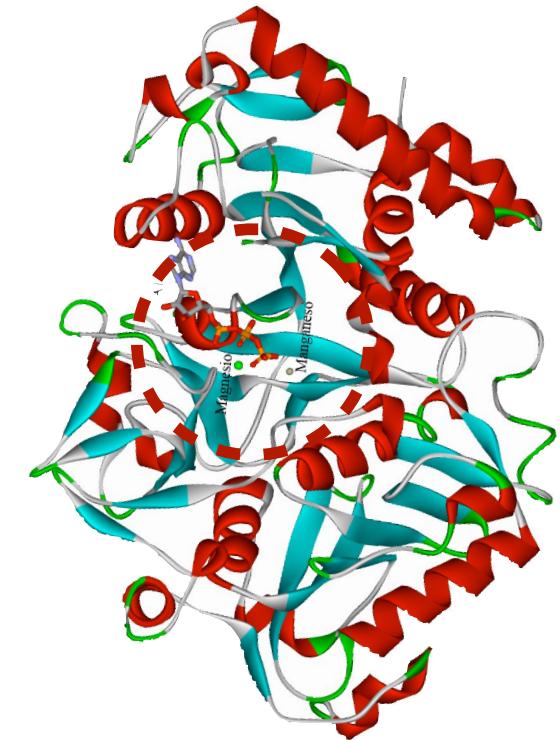
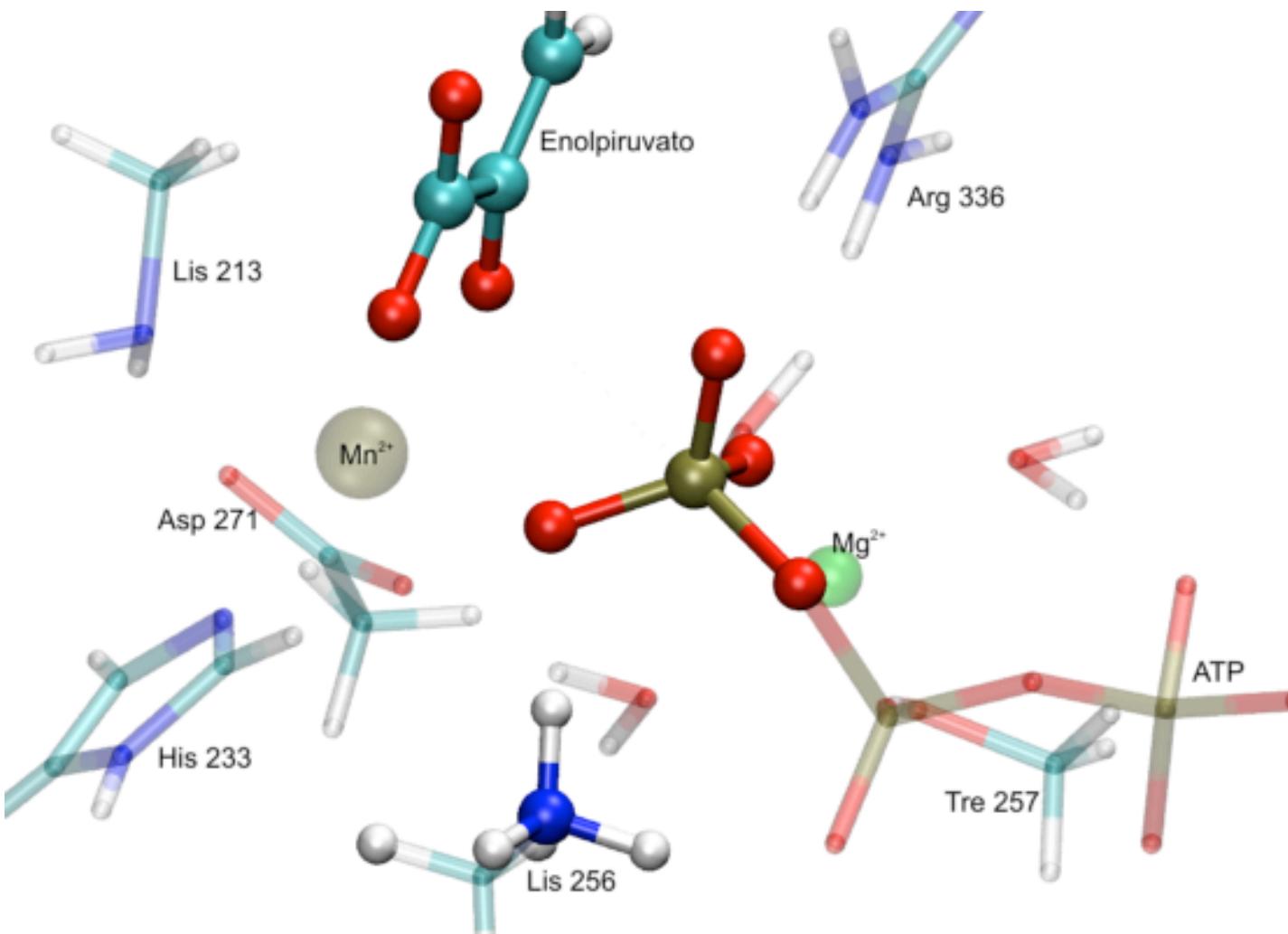


- ✓ 1.536 CORES
- ✓ 3 TB RAM Memory
- ✓ 12.9 Tera flops

Drug Discovery: Virtual screening

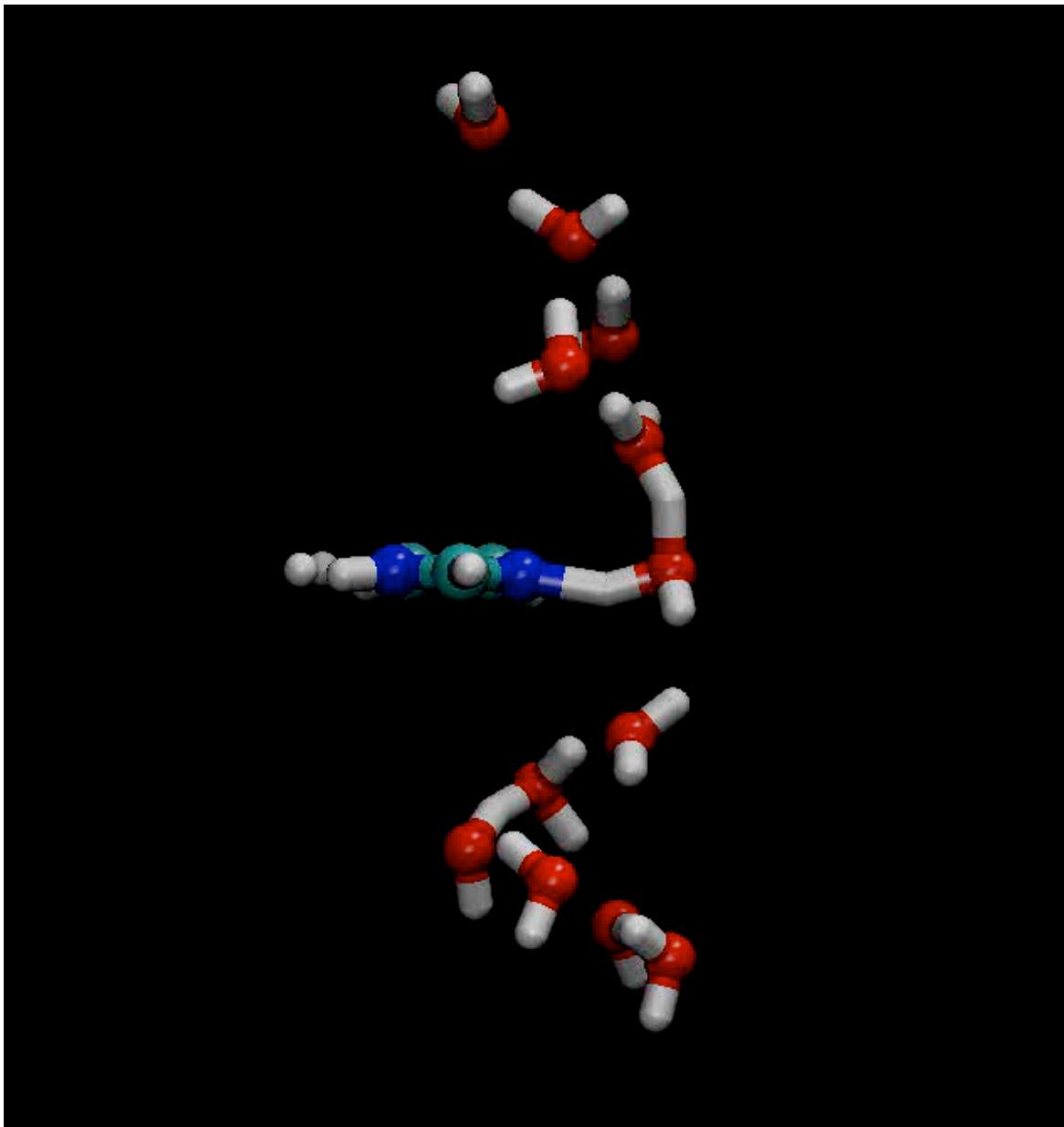
	Nº cores	
Nº small mol	32	1,536
2,054	0.1	0.0
4,000,000	289.4	3.0
14,000,000	1012.7	10.5

PEPCK: Phosphate transfer from ATP to Enolpyruvate



Influenza virus

M2 channel: proton transporter



Non-equilibrium Molecular Dynamics: eField

3

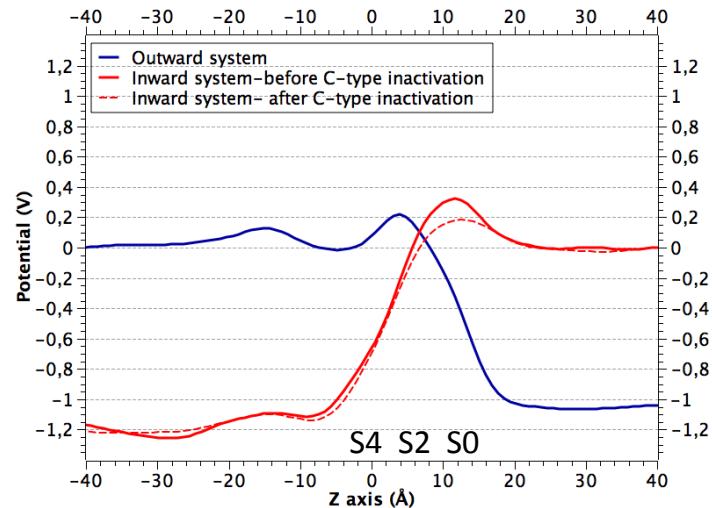


Fig 3 External electric fields effect on the electrostatic potential of Shaker P475D pore.

4

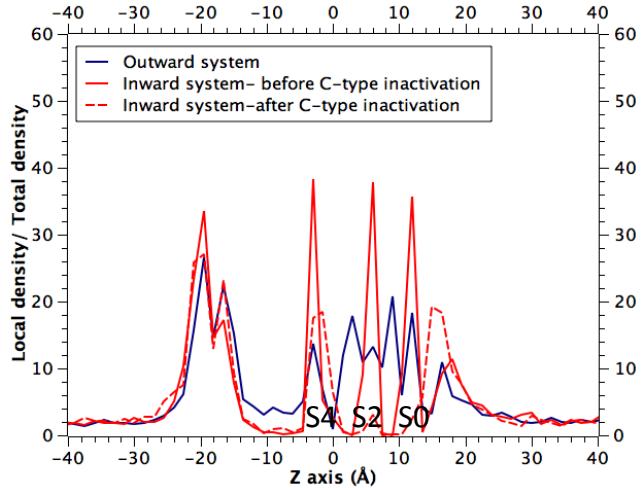
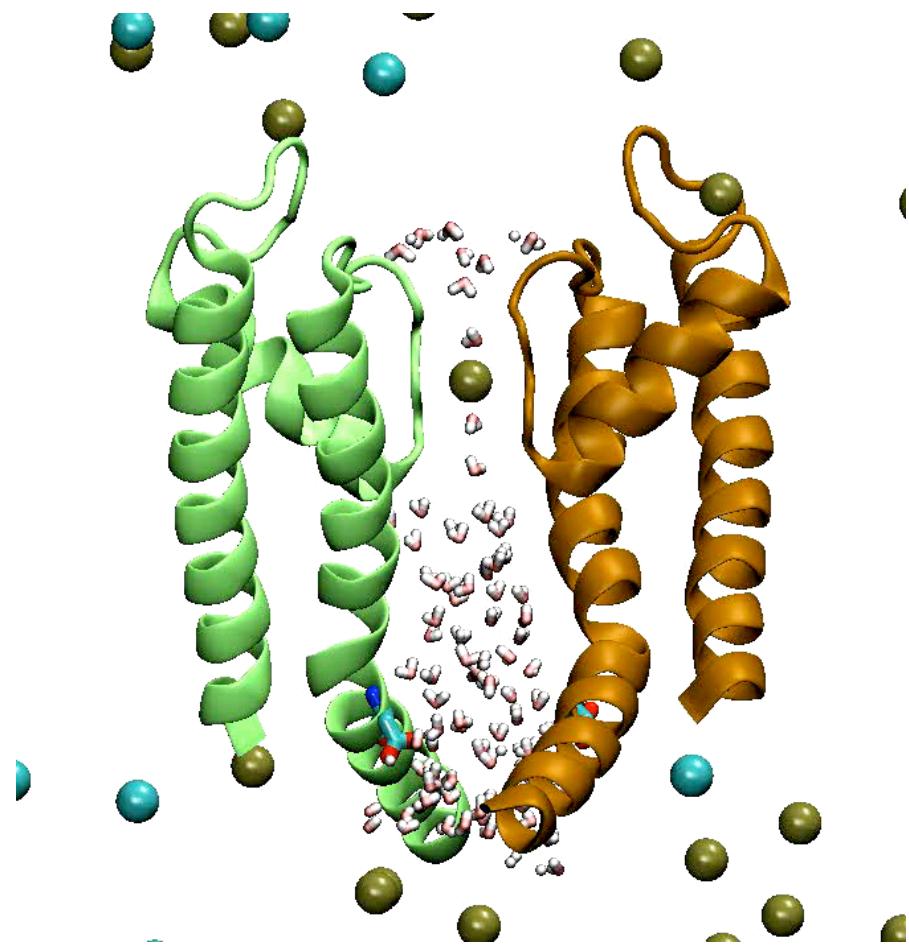


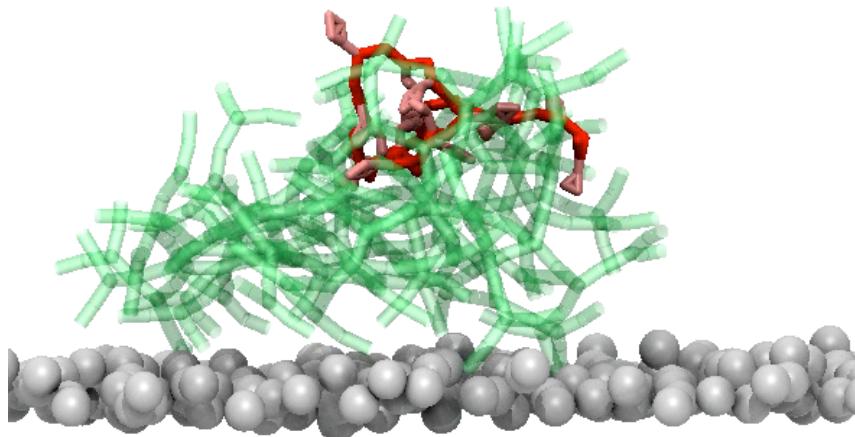
Fig 4. Modification of ion distribution inside the channel by eFields.



Experimental: 20 pA (100 mV)
Theoretical: 22.8 pA (100 mV)

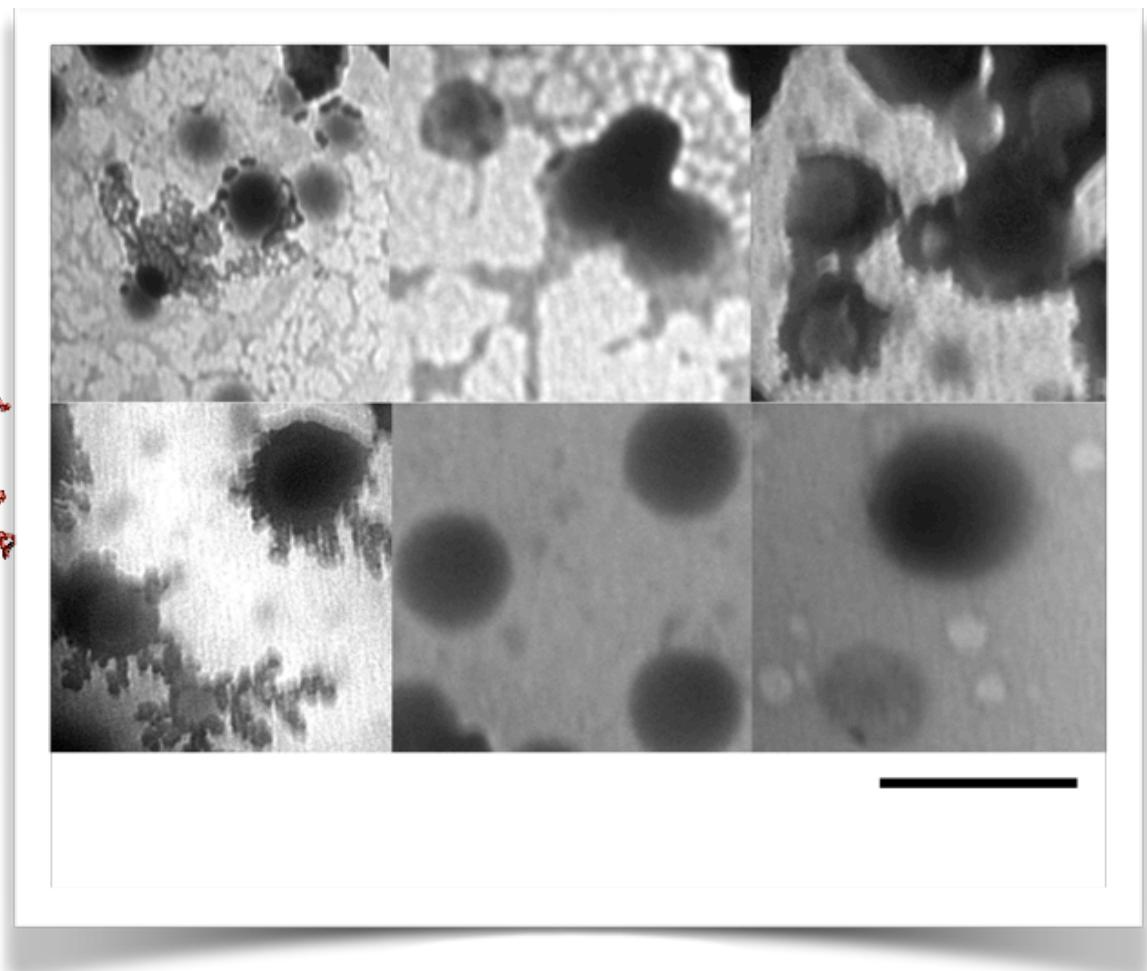
How is the process of transfection using XArg? Full atom molecular dynamics take too much time

Transfection of DNA using CG molecular dynamics
2.2 microseconds



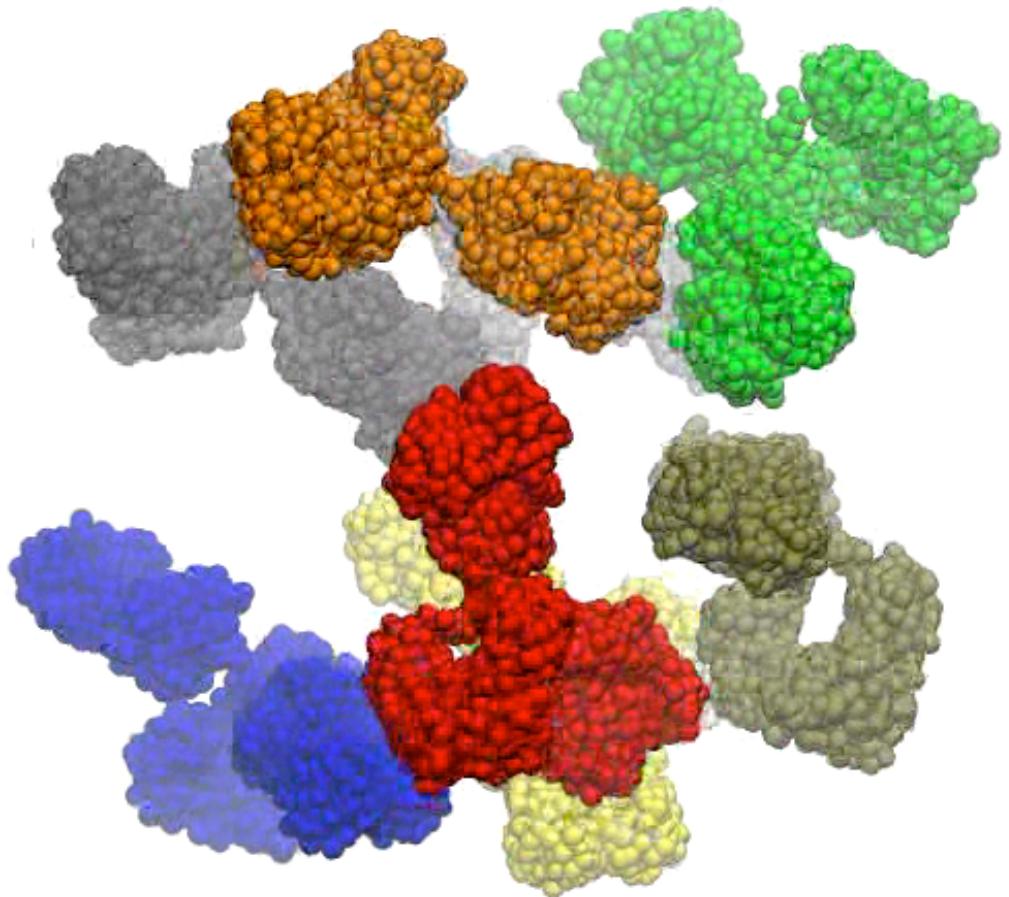
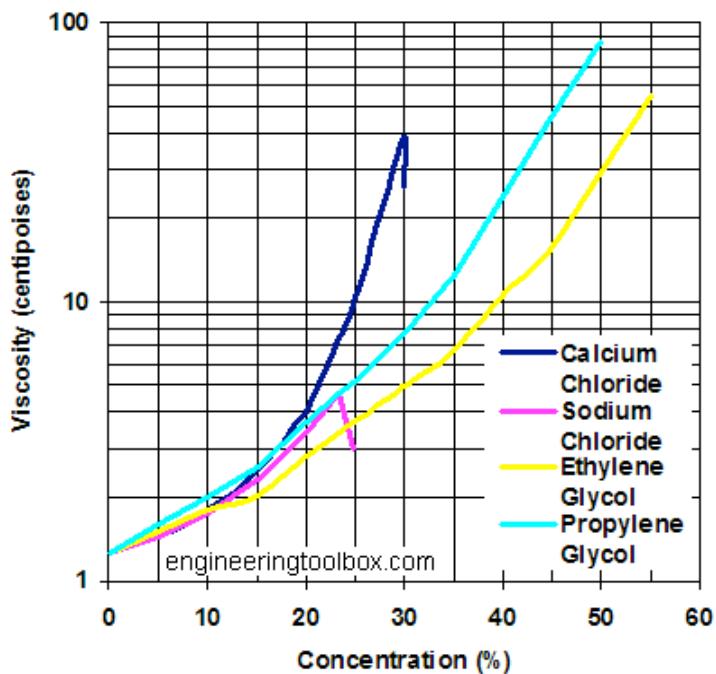
Drug Delivery Systems Based on Biopolymers

CIMIS Dr. Luis Velasquez



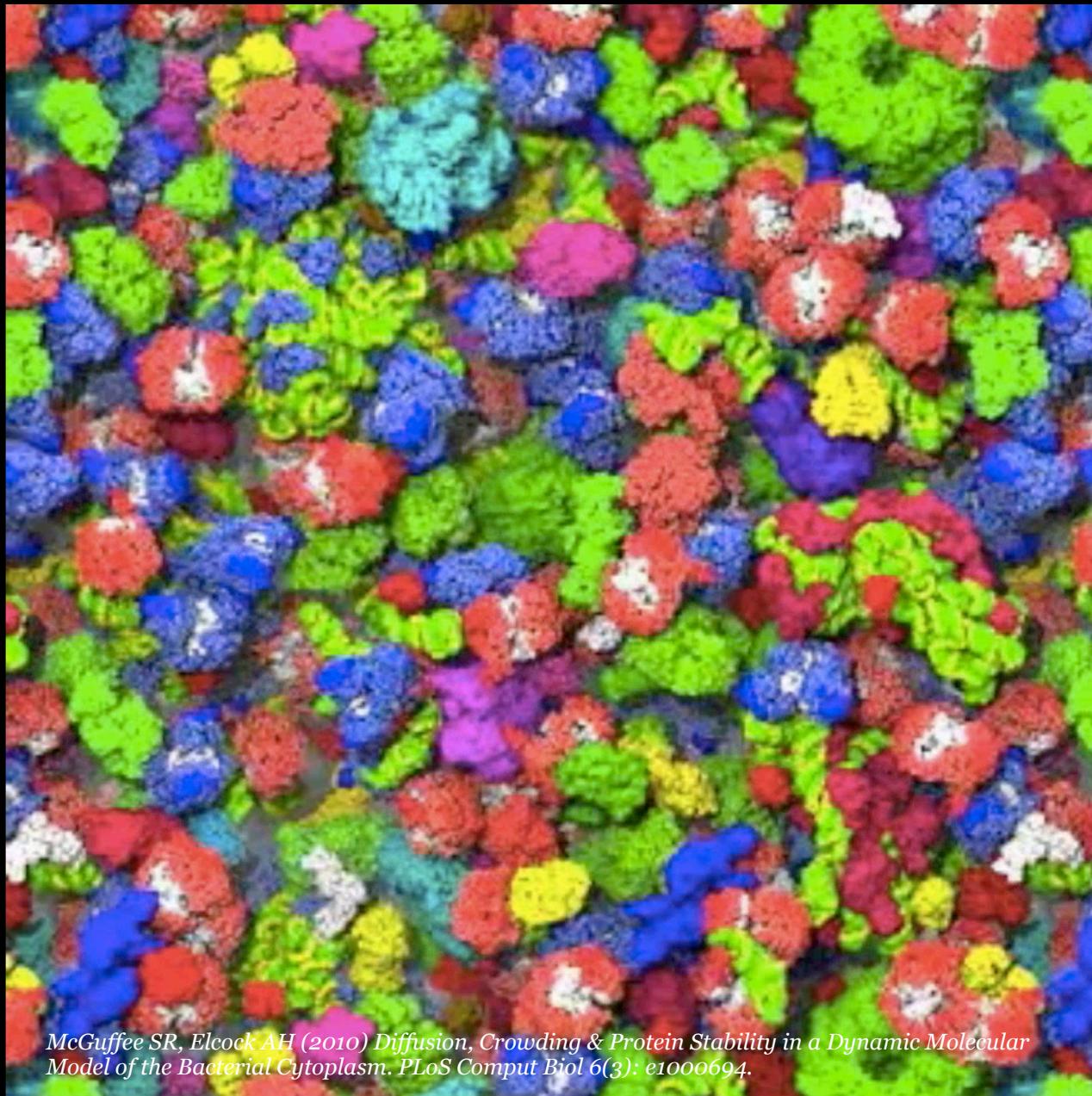
Case of study: Rheological properties are associated to the structure of the protein

Molecular simulations of crowded systems



Structural and dynamics properties are related with rheological properties.

Next Frontier



McGuffee SR, Elcock AH (2010) Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLoS Comput Biol* 6(3): e1000694.

Outline

- Classical mechanics, molecular mechanics
- Force field
- Energy minimization
- Molecular dynamics
- Solvent and boundary conditions
- Analysis tools

Molecular Simulations

(molecular microscope)

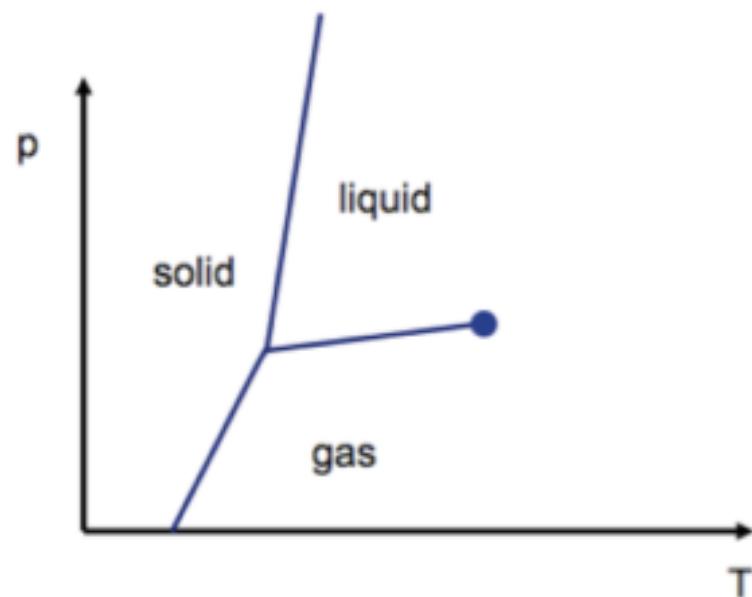
- Molecular simulations methods are now routinely used to investigate the structure, dynamics, surface properties and thermodynamics of :
 - biological systems (biochemistry, biofísica and nanobiotechnology)
 - bioinorganics complexes,
 - polymeric systems.
- The types of biological activity that have been investigated using molecular simulations include:
 - protein folding,
 - enzyme catalysis,
 - protein stability,
 - conformational changes associated with biomolecular function,
 - and molecular recognition of proteins, DNA, and membrane complexes.
- and provide the mean to carry out the following studies,
 - ✓ Drug & Molecular Design
 - ✓ Structural and dynamics characterization of protein at molecular level

Biological molecules exhibit a wide range of time scales over which specific processes occur; for example

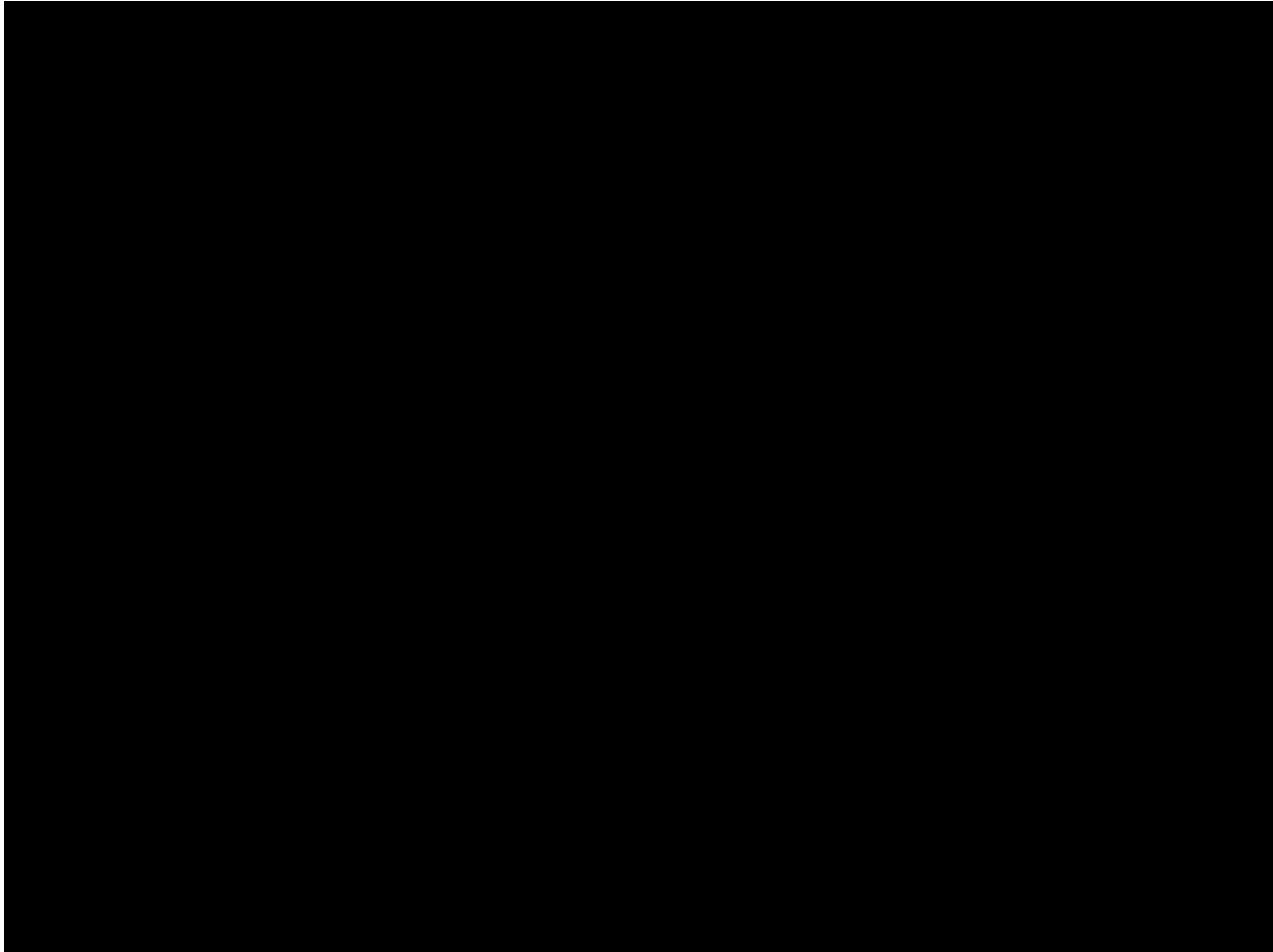
- Local Motions (0.01 to 5 \AA , 10^{-15} to 10^{-1} s)
 - Atomic fluctuations
 - Sidechain Motions
 - Loop Motions
- Rigid Body Motions (1 to 10 \AA , 10^{-9} to 1 s)
 - Helix Motions
 - Domain Motions (hinge bending)
 - Subunit motions
- Large-Scale Motions ($> 5 \text{ \AA}$, 10^{-7} to 10^4 s)
 - Helix coil transitions
 - Dissociation/Association
 - Folding and Unfolding

Simple Liquids

Phase diagram



The world we wish to explore.



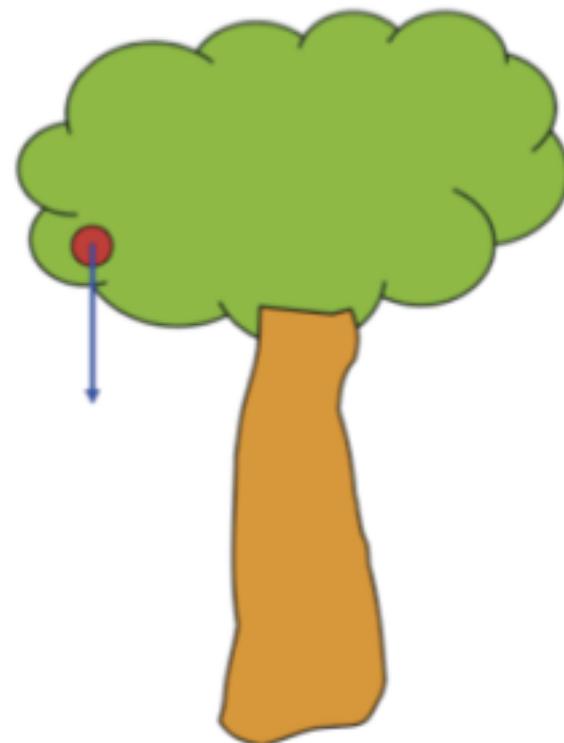
Historical notes

- 1687. Newton. Classical mechanics, equation of motion.
- 1782. Laplace. *Traité de mécanique analytique*.
- 1880. Boltzmann & Gibbs. Statistical mechanics.
- 1921. Schrödinger, Heisenberg & Dirac. Quantum mechanics.
- 1953. Metropolis. First Monte Carlo simulation of a system of rigid disks.
- 1953. Crick & Watson. Determination of the structure of DNA.
- 1957. Alder. First molecular dynamics simulation.
- 1960. Kendrew & Perutz. First determination of a protein structure by Xray crystallography.
- 1964. Rahman. First MD simulation with a realistic potential.
- 1977. Karplus. First simulation of a protein.
- 1985. Wütrich. First NMR protein structure.

Classical mechanics

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

$$\begin{cases} \mathbf{r}(t) = \mathbf{r}_0 + \mathbf{v}t + \frac{\mathbf{a}t^2}{2} \\ \mathbf{v}(t) = \mathbf{v}_0 + \mathbf{a}t \end{cases}$$



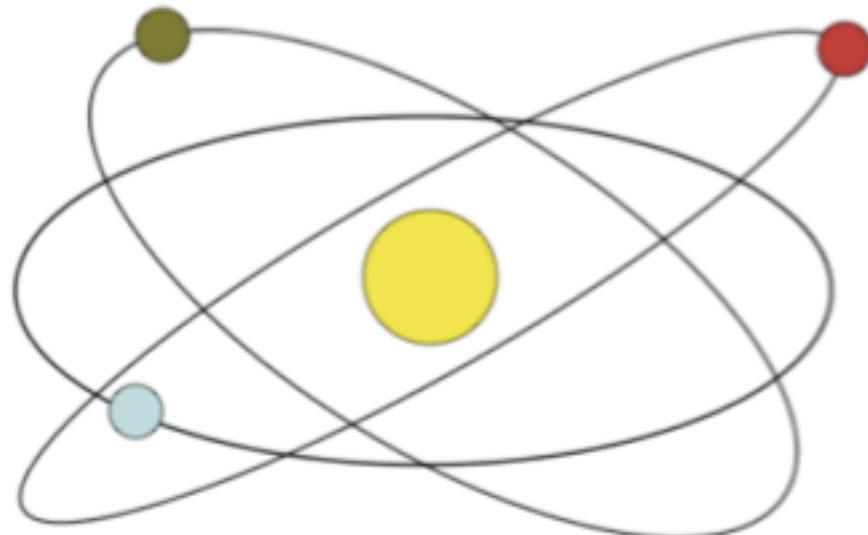
Newton (1687)

Classical mechanics

$$\mathbf{F}_i = -\nabla_i V$$

$$V \approx \frac{m_1 m_2}{|\mathbf{r}_1 - \mathbf{r}_2|}$$

$$\mathbf{r}_i(t)$$



Atoms and molecules

At atomic and subatomic level, classical laws are not applicable.

The **Schrödinger equation** replaces the Newton equation.

$$\mathbf{H}|\psi\rangle = E|\psi\rangle$$

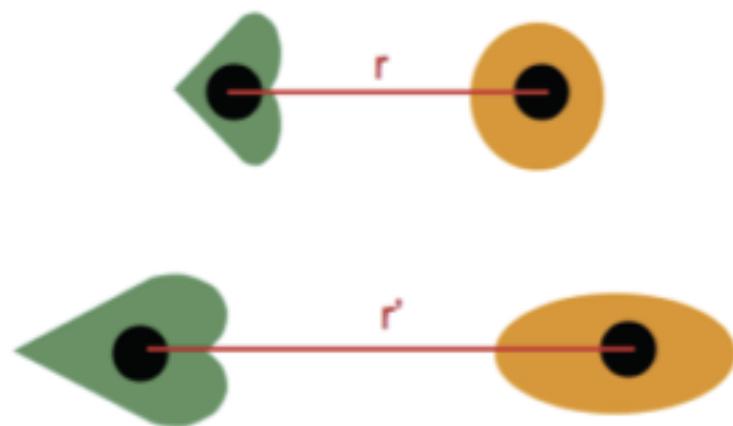
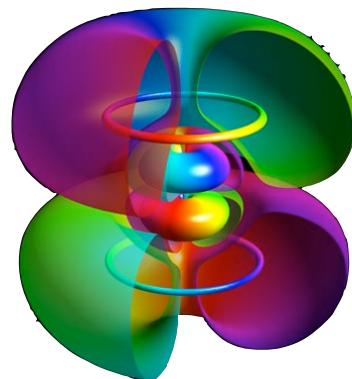
$$\mathbf{H} = -\frac{\hbar^2}{2m_1}\nabla_1^2 - \frac{\hbar^2}{2m_2}\nabla_2^2 - \dots + \mathbf{V}(r_1, r_2, \dots)$$

The Schrödinger equation can only be solved exactly for few very simple systems (harmonic oscillator, square well potential, the hydrogen atom).

$$V = \frac{Z}{4\pi\epsilon_0(r_1 - r_2)}$$

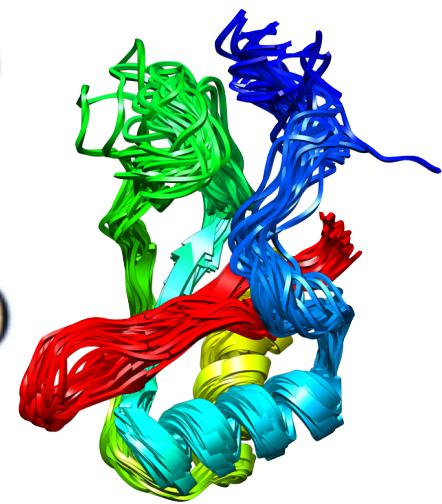
Atoms and molecules

Nuclei are much heavier than electrons: electrons move much faster than nuclei.
When studying the motion of electrons, nuclei can be considered fixed.
Vice-versa, when studying the movement of nuclei, the electron distribution associated with each atom can be considered fixed.
→ Born-Oppenheimer approximation.



$$V(r)$$

$$V(r')$$



Energy function (force field)

Goal: define a classical potential energy function (or force field) that allows accurate calculation of molecular properties.

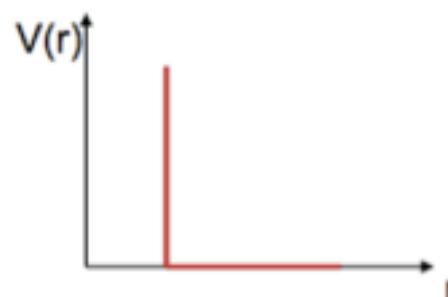
Energy functions are defined empirically based on experimental properties (structural, thermodynamic, spectroscopic).

Simplest case: pairwise additive potential energy

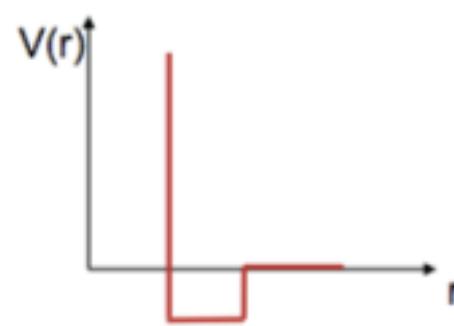
$$V(r_1, r_2, \dots) = \sum_{i=1}^N \sum_{j < i} v(r_i - r_j)$$

Energy function (force field)

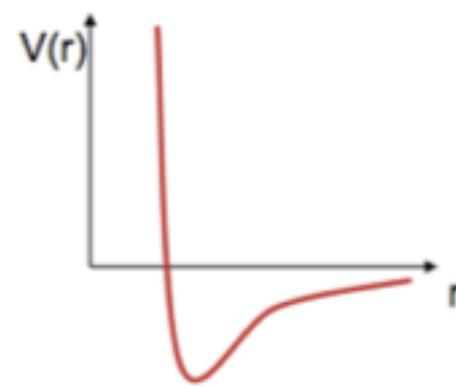
Simplest pairwise interactions:



Hard sphere



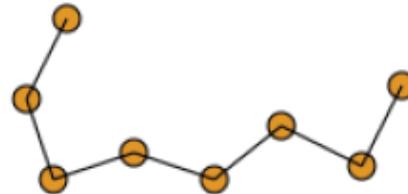
Square well



Lennard-Jones

Energy function (force field)

For **molecular systems** potential energy functions (force fields) defined as a sum of terms which include **bonded** and **non-bonded** interactions.

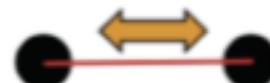


Energy function (force field)

Bonded interactions:

Intramolecular energy term associated with the deformation of the electronic structure of the molecule.

Bond stretch



Angle bend



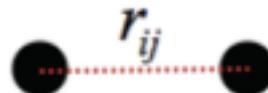
Dihedral torsion



Energy function (force field)

Non-bonded interactions:

Inter and intramolecular interactions arising from electrostatic interactions.



Coulomb interactions

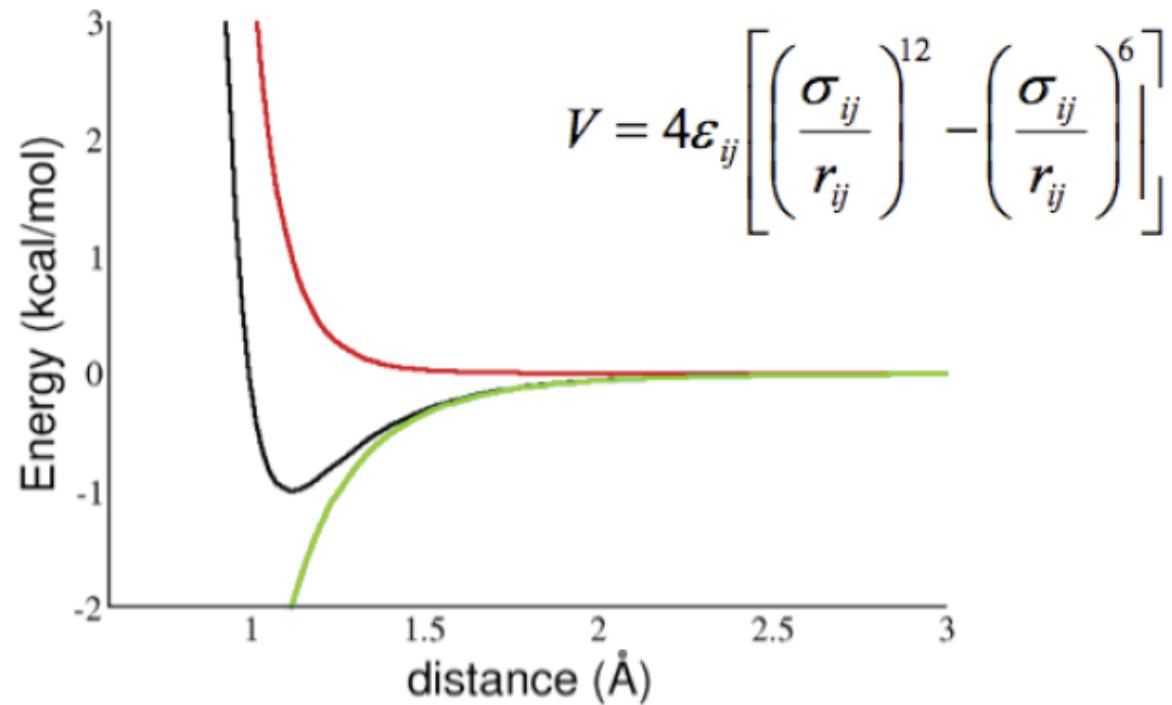
$$V = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

van der Waals interactions

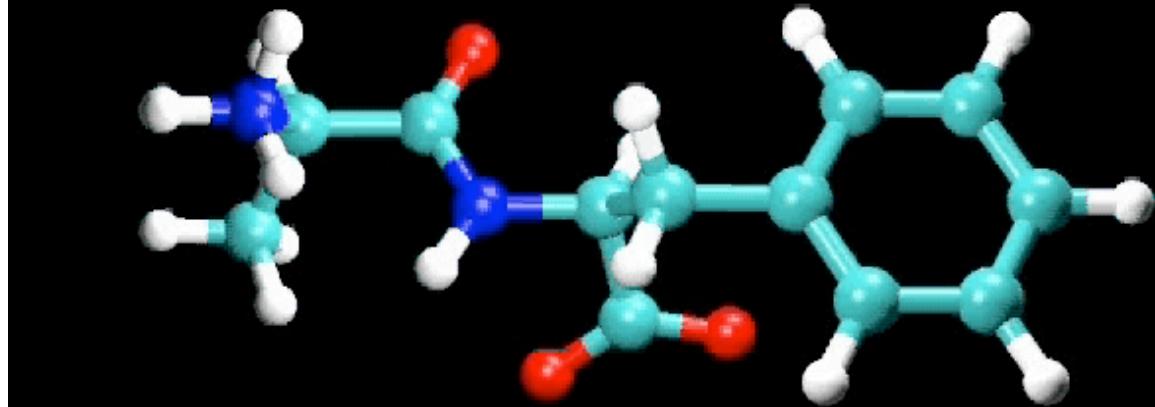
$$V = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

Energy function (force field)

van der Waals interactions



Summary:



CHARMM energy function

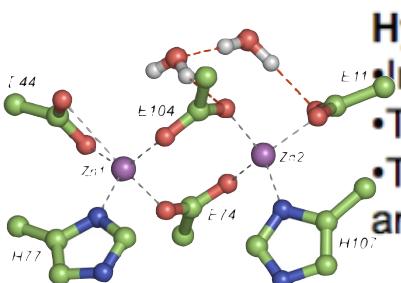
$$V = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_a (\theta - \theta_0)^2 + \sum_{dihedrals} k_d [1 + \cos(n\varphi - \delta)] + \sum_{i < j} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} + \sum_{i < j} 4\epsilon_y \left[\left(\frac{\sigma_y}{r_{ij}} \right)^{12} - \left(\frac{\sigma_y}{r_{ij}} \right)^6 \right]$$

(AMBER, GROMOS, OPLS,... all force fields are similar)

How can we treat phenomena which are inherently quantum mechanical such as enzymatic reactions, formations of covalent bonds, etc ?

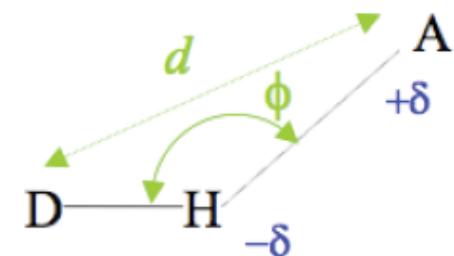
Derived interactions

Some interactions are often referred to as particular interactions, but they result from the two interactions previously described, i.e. the electrostatic and the van der Waals interactions.



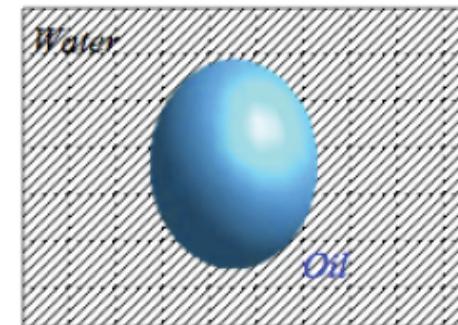
Hydrogen bonds (HB)

- Interaction of the type D-H \cdots A
- The origin of this interaction is a dipole-dipole attraction
- Typical ranges for distance and angle: $2.4 < d < 4.5\text{\AA}$ and $180^\circ < \phi < 90^\circ$



Hydrophobic forces

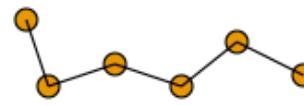
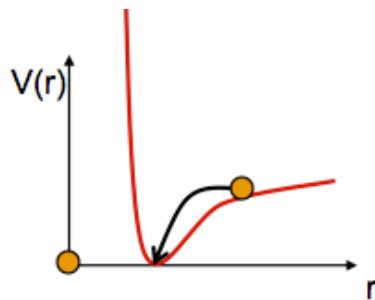
- Collective effect resulting from the energetically unfavourable surface of contact between the water and an apolar medium (loss of water-water HB)
- The apolar medium reorganizes to minimize the water exposed surface



Parametrization of force fields

- Using quantum mechanical calculations (small compounds)
- Reproduction of experimental vibrational spectra
- Reproduction of thermodynamic quantities
- Optimization of geometry
- ...

Energy minimization



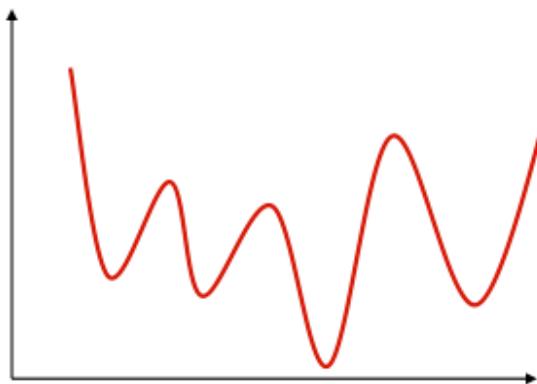
For all except the simplest systems the potential energy is a complicated multidimensional function of the coordinates.

$$f(x_1, x_2, x_3 \dots)$$

$$\frac{\partial f}{\partial x_i} = 0$$

$$\frac{\partial^2 f}{\partial x_i^2} > 0$$

Energy minimization



Main algorithms:

- Steepest Descent: local, finds the closest minimum.
- Conjugate Gradient: searches in a region of the conformational space.

Molecular dynamics

Newton, Lagrange, Hamilton equation of motion



$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i = -\nabla_i V \quad (\dot{\mathbf{r}}_i \equiv \mathbf{v}_i ; \ddot{\mathbf{r}}_i \equiv \dot{\mathbf{v}}_i \equiv \mathbf{a}_i)$$

The numerical solution of the equation of motion requires a discretization of time

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \delta t \dot{\mathbf{r}}_i(t) + \frac{1}{2} \delta t^2 \ddot{\mathbf{r}}_i(t) + \frac{1}{6} \delta t^3 \dddot{\mathbf{r}}_i(t) + \dots$$

timestep δt should be small (1 or 2 fs)

Molecular dynamics

Verlet algorithm

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \delta t \dot{\mathbf{r}}_i(t) + \frac{1}{2} \delta t^2 \ddot{\mathbf{r}}_i(t) + \frac{1}{6} \delta t^3 \dddot{\mathbf{r}}_i(t) + \dots$$

$$\mathbf{r}_i(t - \delta t) = \mathbf{r}_i(t) - \delta t \dot{\mathbf{r}}_i(t) + \frac{1}{2} \delta t^2 \ddot{\mathbf{r}}_i(t) - \frac{1}{6} \delta t^3 \dddot{\mathbf{r}}_i(t) + \dots$$

$$\mathbf{r}_i(t + \delta t) + \mathbf{r}_i(t - \delta t) = 2\mathbf{r}_i(t) + \delta t^2 \ddot{\mathbf{r}}_i(t) + \delta t^4 \dots$$

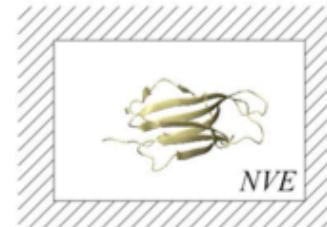
$$\mathbf{r}_i(t + \delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \delta t) + \delta t^2 \ddot{\mathbf{r}}_i(t) + \delta t^4 \dots$$

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \delta t) - \mathbf{r}_i(t - \delta t)}{2\delta t}$$

Various thermodynamic ensemble

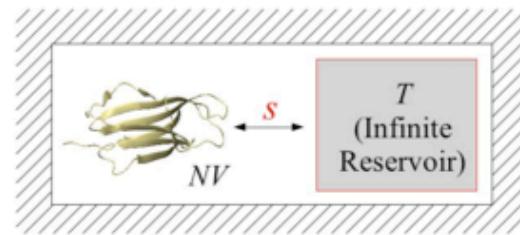
1) Microcanonical ensemble (constant N, V, E)

$$H(\mathbf{p}, \mathbf{q}) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i s^2} + \phi(\mathbf{q})$$



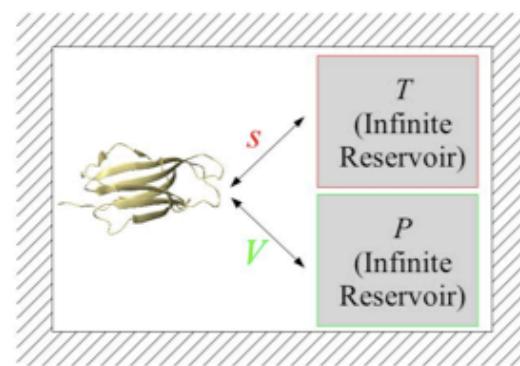
2) Canonical ensemble (constant N, V, T)

$$H(\mathbf{p}, \mathbf{q}, p_s, s) = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i s^2} + \phi(\mathbf{q}) + \frac{\mathbf{p}_s^2}{2Q} + (3N+1)kT \ln s$$



3) Isothermic-isobaric ensemble (constant N, P, T)

$$H = \sum_i^N \frac{\mathbf{p}_i^2}{2m_i s^2 V^{2/3}} + \phi(V^{1/3} \mathbf{q}) + \frac{\mathbf{p}_s^2}{2Q} + (3N+1)kT \ln s + \left[\frac{\mathbf{p}_V^2}{2W} + P_{ex} V \right]$$



Statistical Mechanics

Thermodynamic potentials

$$A(N,V,T) = E - TS \quad (\text{Helmholtz free energy})$$

$$G(N,p,T) = E + pV - TS \quad (\text{Gibbs free energy})$$

Statistical Mechanics

Canonical distribution:

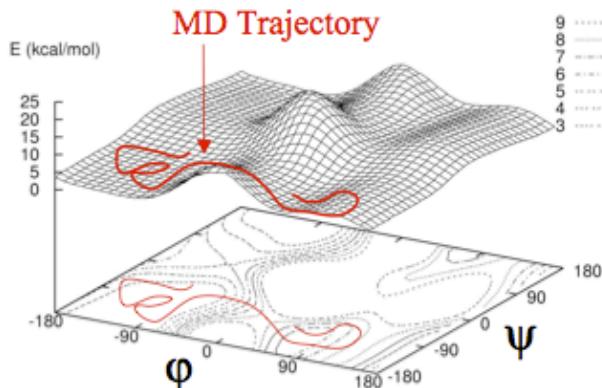
$$\rho(\mathbf{r}, \mathbf{p}) = \frac{1}{Q} e^{-\frac{H(\mathbf{r}, \mathbf{p})}{k_B T}}$$

probability density

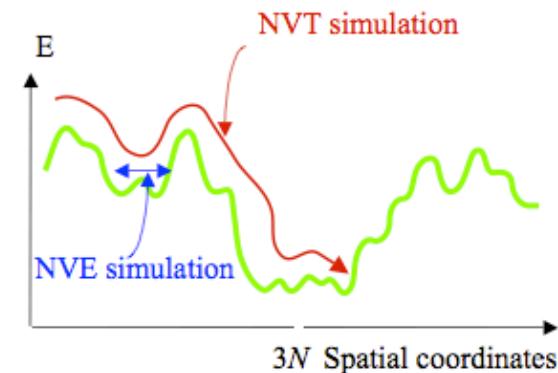
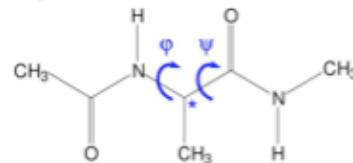
$$Q = \int d\mathbf{r} d\mathbf{p} e^{-\frac{H(\mathbf{r}, \mathbf{p})}{k_B T}}$$

partition function

Ergodicity principle



Dialanine



Protein



$$\langle O \rangle_{\text{Ensemble}} = \frac{1}{Z} \int O(\varphi, \psi) e^{-\beta E(\varphi, \psi)} d\varphi d\psi = \frac{?}{\tau} \int_0^\tau O(t) dt = \langle O \rangle_{\text{Time}}$$

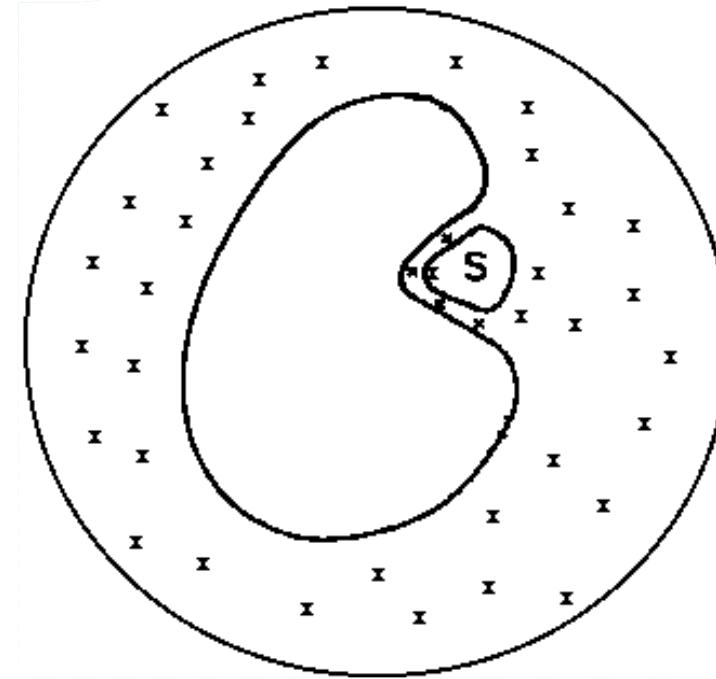
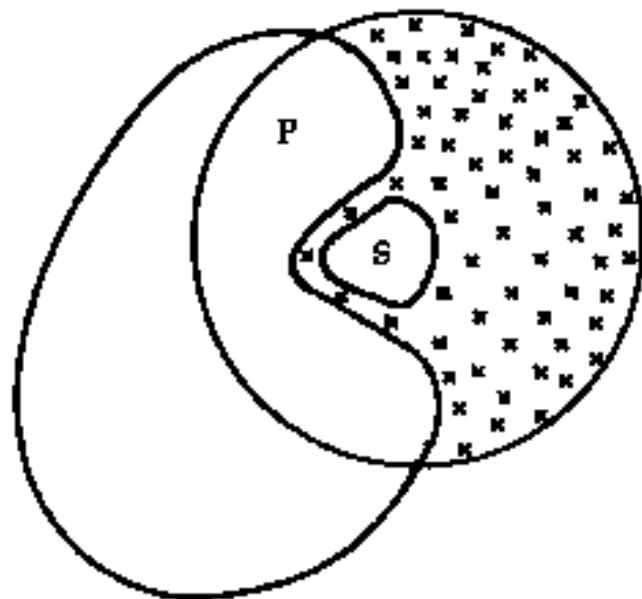
Initial conditions

The result of a simulation *should not* depend on the initial conditions.
In practice we chose the initial condition so that we are as close as possible to equilibrium.

Initial positions: crystal lattice if we want to simulate a crystal or a liquid, experimental structure for a complex molecule.

Initial velocities: from the Boltzmann (Gaussian) distribution.

Solvent and boundary conditions

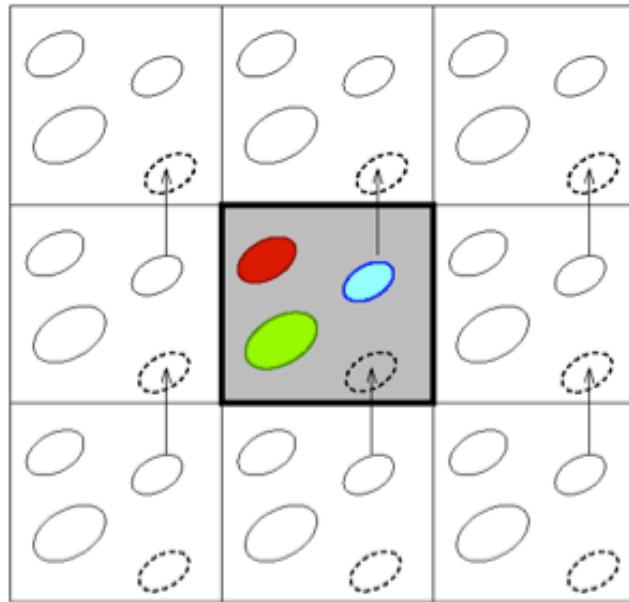


There exist numerous cases where one may not wish to use periodic boundary conditions. In some cases, the use of periodic boundary conditions requires the use of a prohibitively large number of water molecules.

With the increase in computer power, it has become much more feasible to incorporate water molecules in the simulation. The simplest way is to surround the protein or just a part of the protein with a sphere of water. Boundary potentials have been developed which restrain the water molecules to a sphere while maintaining a strong semblance to bulk water. Structural and thermodynamics properties when calculated under these conditions indicate that the water still behaves as bulk water. This usually involves much fewer water molecules than in a periodic boundary simulation and is often sufficient.

Boundary conditions

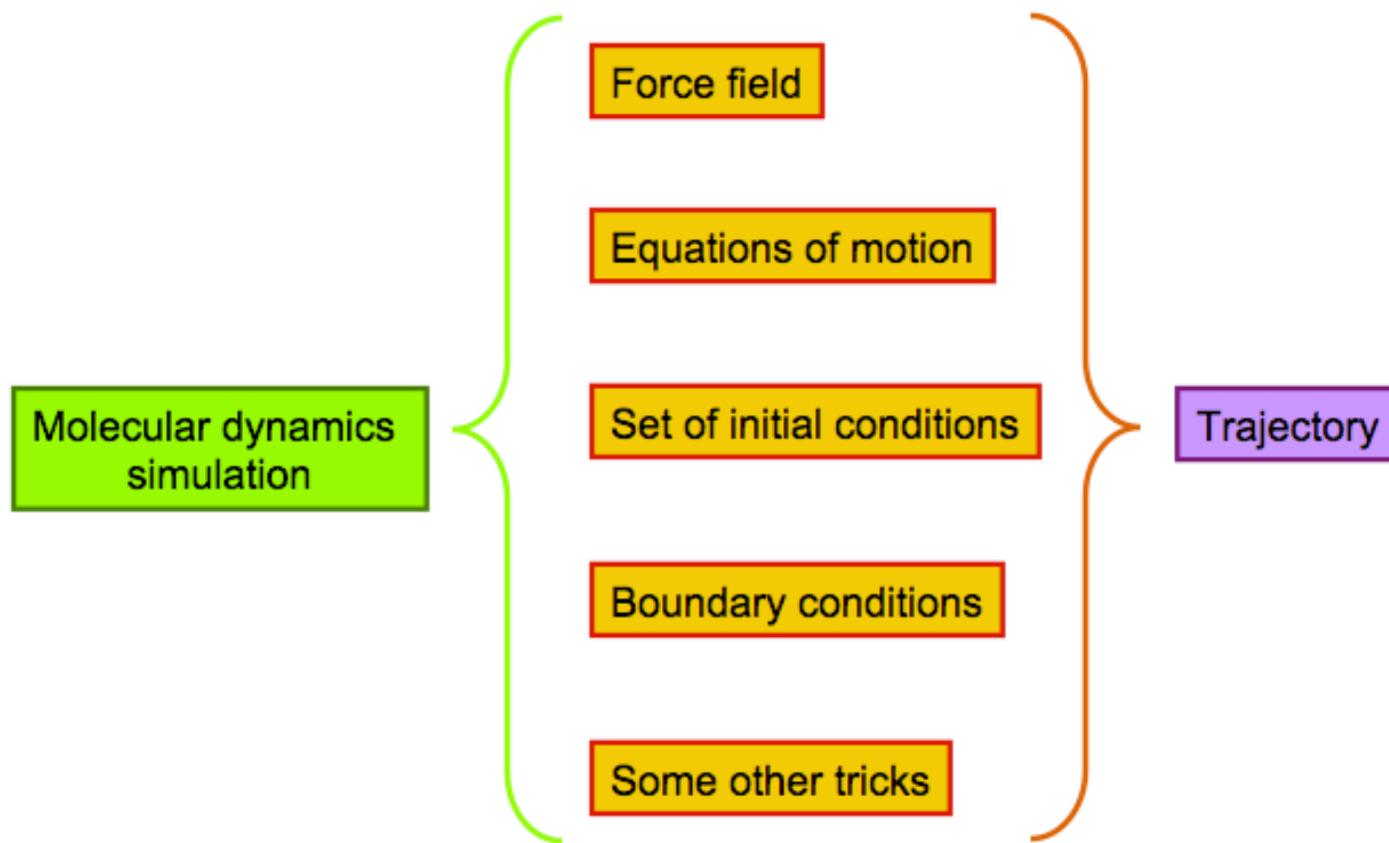
Fixed boundaries introduce strong boundary effects. *Periodic* boundary conditions are a commonly used alternative.



Alternatives:
Spherical (SSBP) and Generalized (GSBP) Solvent Boundary Potentials.

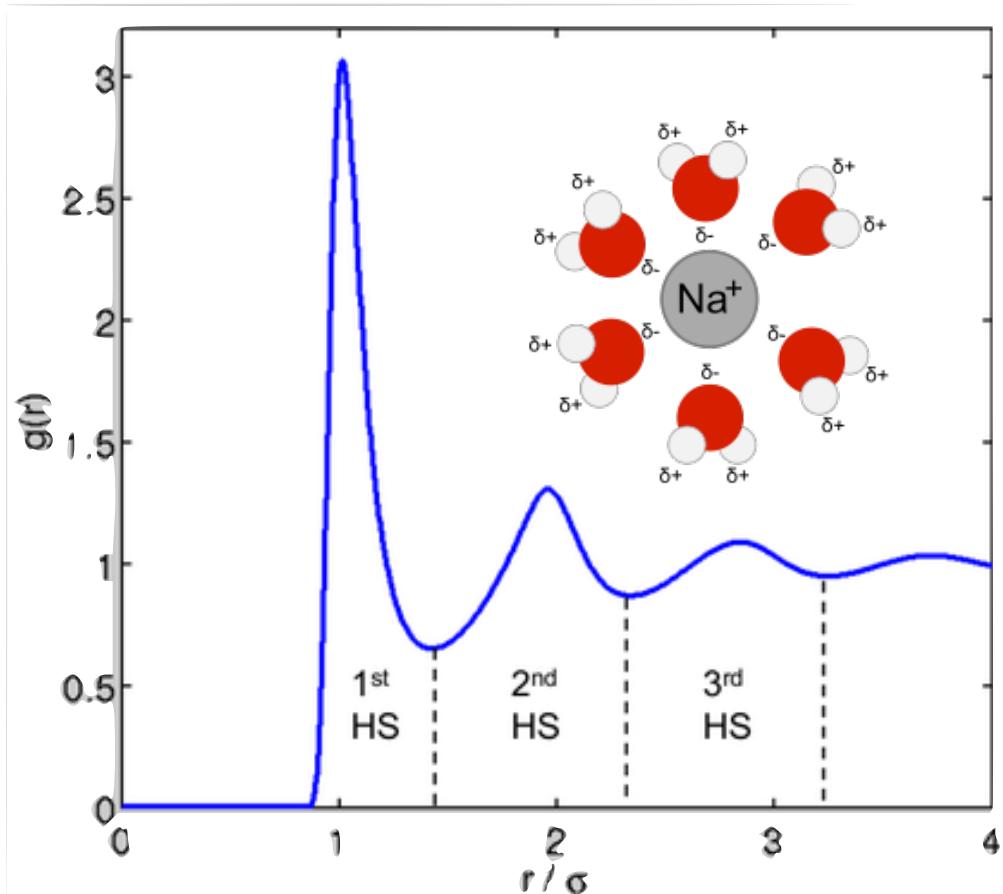
Periodic boundary conditions enable a simulation to be performed using a relatively small number of particles in such a way that the particles experience forces as though they were in a bulk solution. See, for example, the two dimensional box. The central box is surrounded by eight neighbors. The coordinates of the image particles, those found in the surrounding box are related to those in the primary box by simple translations. The simplest box is the cubic box. Forces on the primary particles are calculated from particles within the same box as well as in the image box. The cutoff is chosen such that a particle in the primary box does not see its image in the surrounding boxes.

Summary

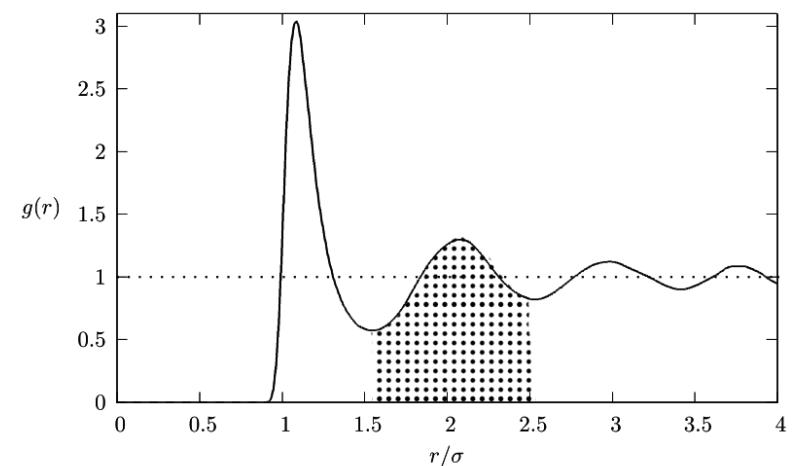
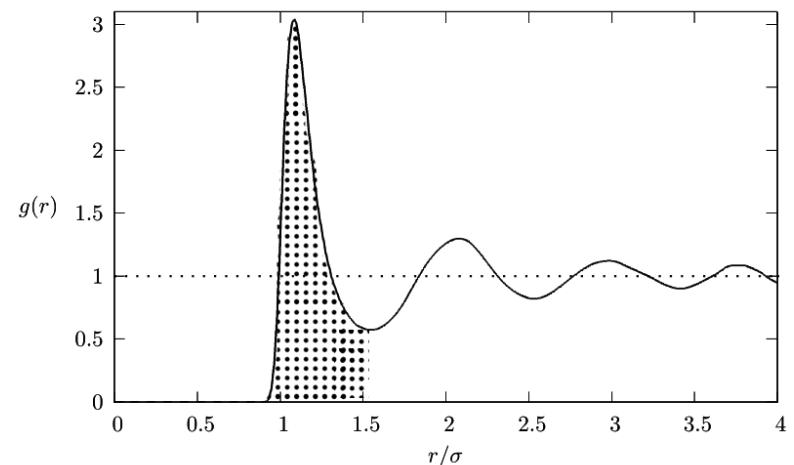


Analysis tools

Radial Distribution function (RDF)



$$g(r) = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle$$



Analysis tools

I) Mean Energy

$$\langle E \rangle = \frac{1}{N} \sum_{i=1}^N E_i$$

II) RMS difference between two structures

$$RMS = \sqrt{\left\langle (r_i^\alpha - r_i^\beta)^2 \right\rangle} = \sqrt{\frac{1}{N_i} \sum_i (r_i^\alpha - r_i^\beta)^2}$$

$$RMS_i^{fluct} = \sqrt{\frac{1}{N_f} \sum_f (r_i^f - r_i^{ave})^2}$$

III) RMS fluctuations

$$B_i = \frac{8}{3} \pi^2 (RMS_i^{fluct})^2$$

note the relation between the RMS fluctuations and the B factors;

$$RadiusGyration = \sqrt{\frac{1}{N_i} \sum_i (r_i - r_{cm})^2}$$

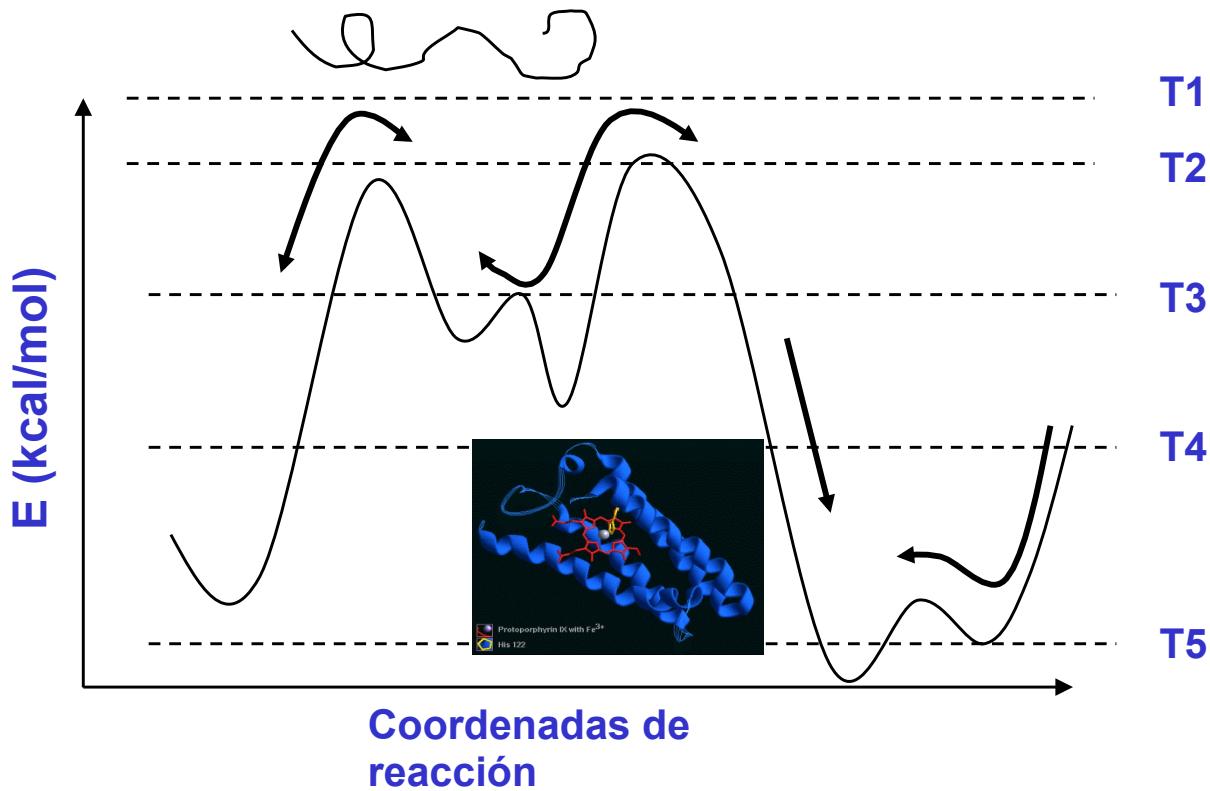
IV) radius of gyration

where $r_i - r_{cm}$ is the distance between atom i and the center of mass of the molecule

Center for Bioinformatics and Integrative Biology

From a molecular dynamics simulation, time dependent properties such as correlation functions can also be

Simulated annealing



T1
T2
T3
T4
T5

$$D = \frac{1}{6t} \langle |r_i(t) - r_i(0)|^2 \rangle$$

Diffusion coefficient

Donde $T_1 > T_4$

$$C_v = \frac{\langle E - E_{avg} \rangle}{k_B T^2}$$

Heat capacity

Reference

- Molecular Modelling Principles and Applications,
Andrew R. Leach