www.bsc.es

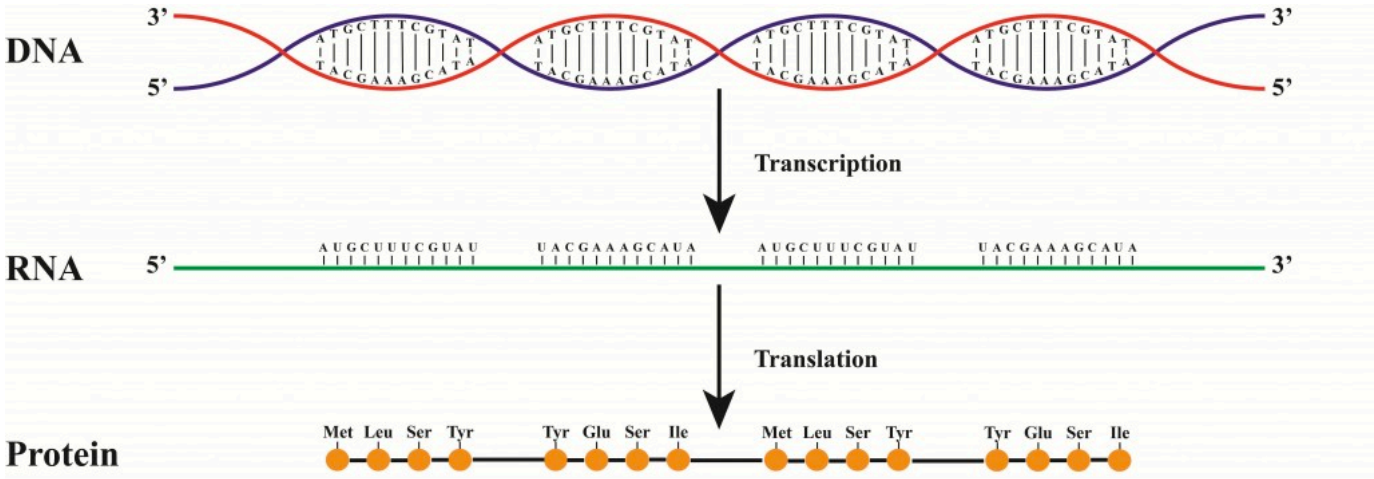
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

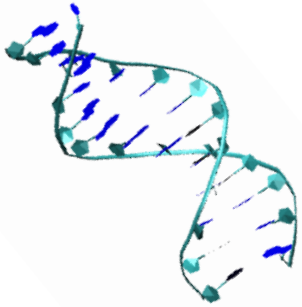
MD Simulation Setup

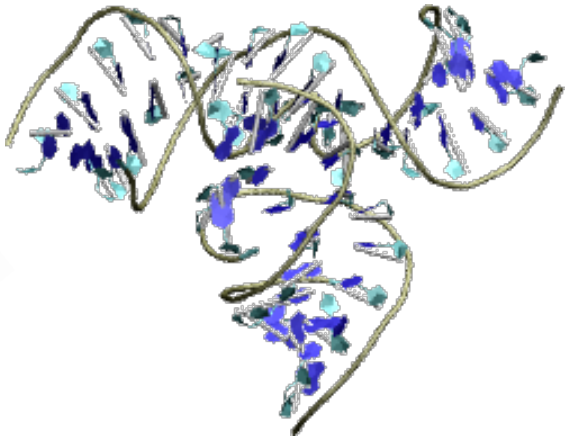
PATC Simulation Environment for Life Sciences, BSC 2020

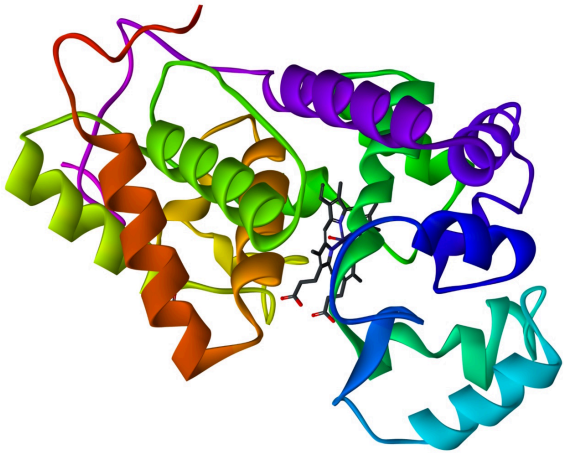
Federica Battistini

Introduction: DNA – RNA – Proteins









Molecular dynamics simulation is a computational method to study and describe equilibrium and dynamics properties of a biological system.

Introduction: Molecular Dynamics

MD simulations:

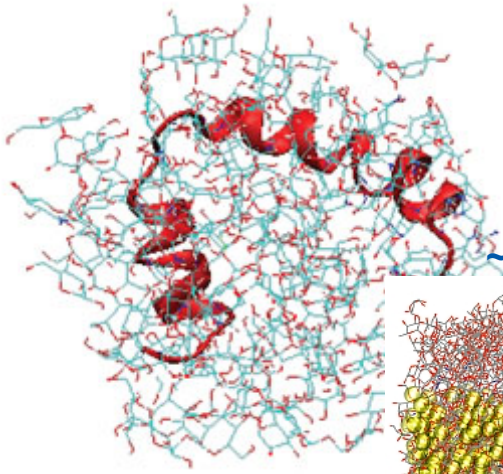
- To study the motion for atoms and molecules and to obtain the time evolution of the system.

- To gain deeper insights in terms of atomic interactions in biological processes involving proteins, DNA, etc, and simulate their behavior with techniques that obey the rules of physics.

- To generate configurations of the system (by integration of Newton's laws of motion) and information at the microscopic level – atomic positions and velocities that can be translated intro macromolecular, thermodynamic properties.

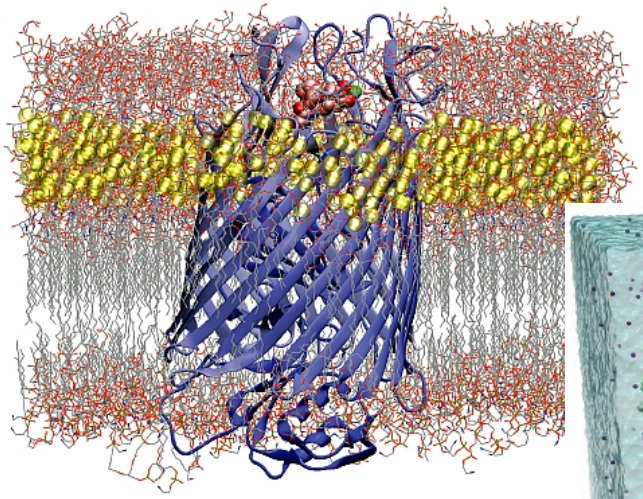
- To connect structure and function by providing additional information to X-ray crystallography and NMR

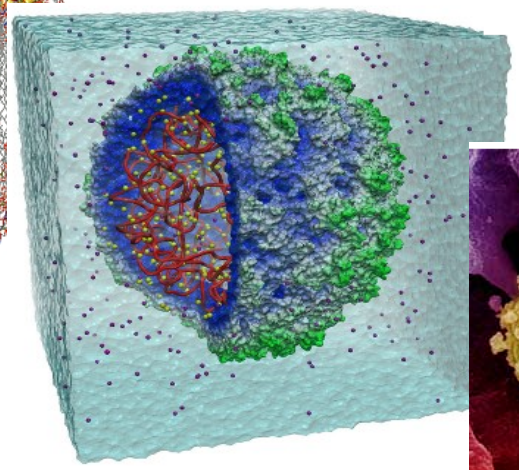
Introduction: Molecular Dynamics



Small peptides
~ 1.000 atoms

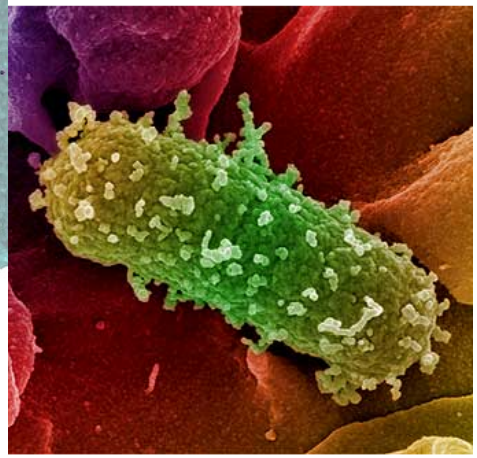
Macromolecular complexes
~ 50.000 to ~ 1.000.000 atoms



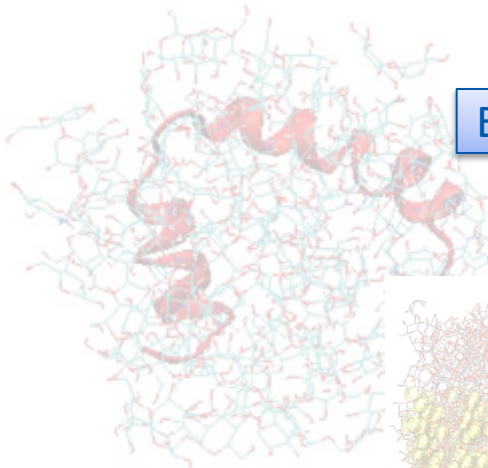


Virus
several millions of atoms

Bacteria/cells?
7e+9 carbon atoms



Introduction: Molecular Dynamics

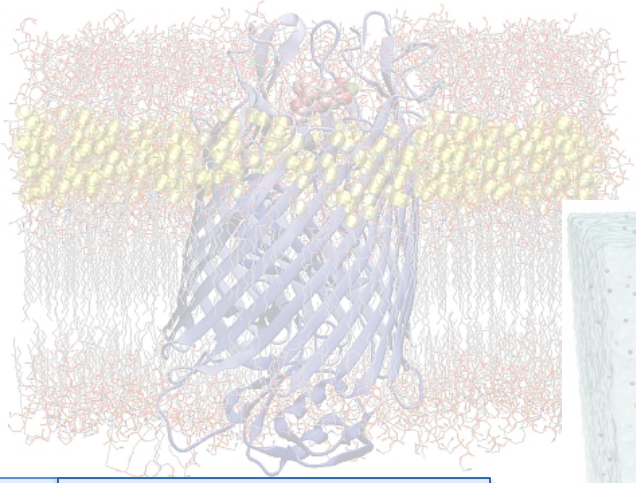


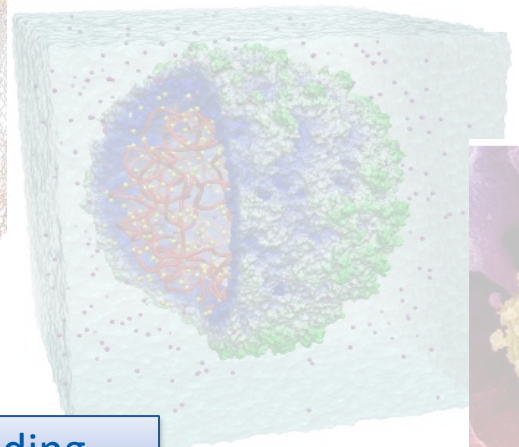
Small peptides
~ 1.000 atoms

fs | 100 fs - ps | µs | ms

Electronic exitations | Vibrations | Rotations | Conformational Transitions

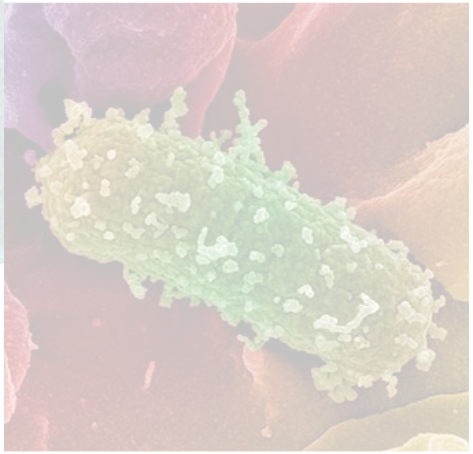
Macromolecular complexes
~ 50.000 to ~ 1.000.000 atoms





Virus
several millions of atoms

Bacteria/cells?
7e+9 carbon atoms



ps

e-tranfer reactions | Enzymatic reactions

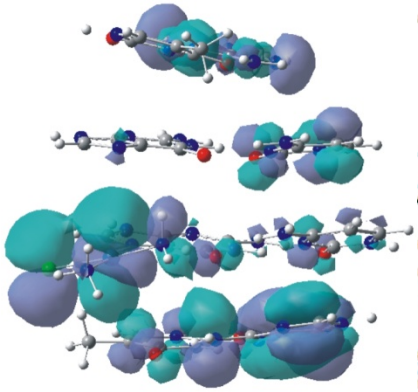
ns | µs - s

Peptide folding | Protein folding

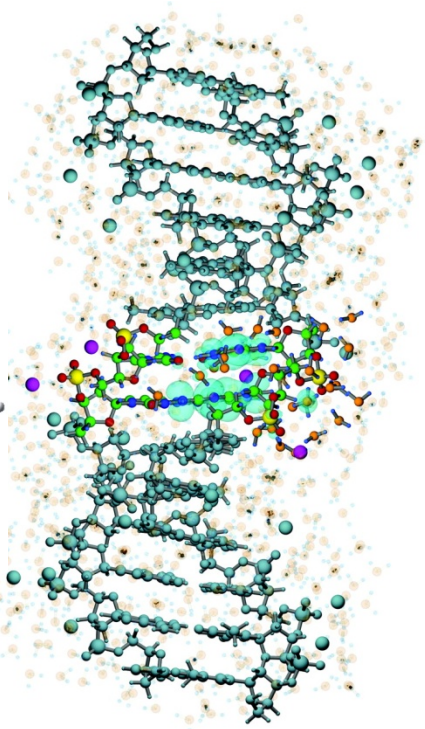
µs

DNA melting

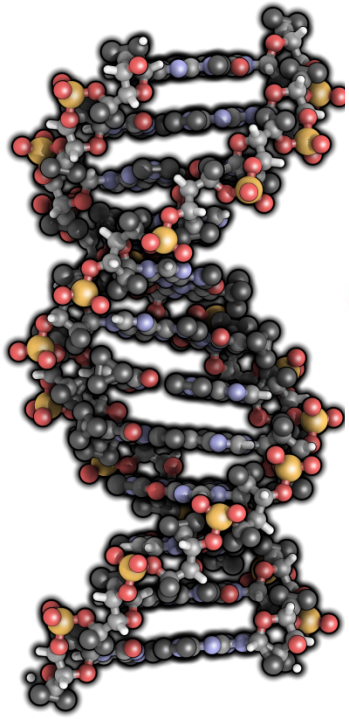
Introduction: Molecular Dynamics



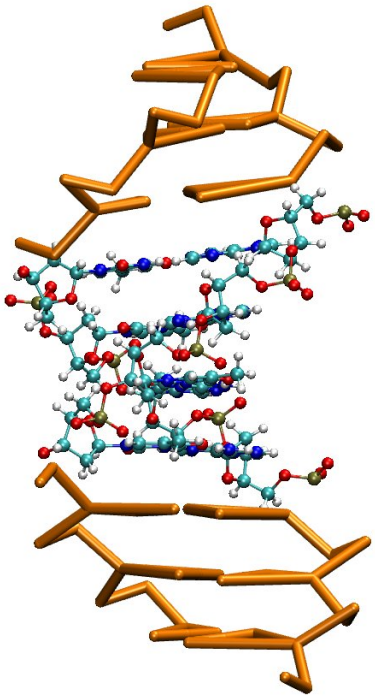
Quantum Mechanics
Dozens to hundreds of atoms
Few ps



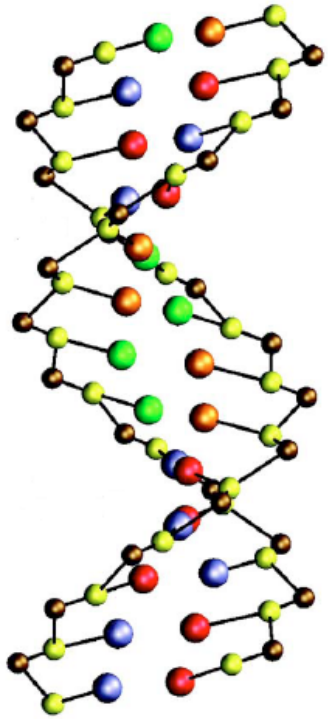
QM/MM
Dozens to thousands of atoms
Dozens of ps



Molecular Mechanics
Thousands to (if) few millions of atoms
Hundreds of ns to few μs



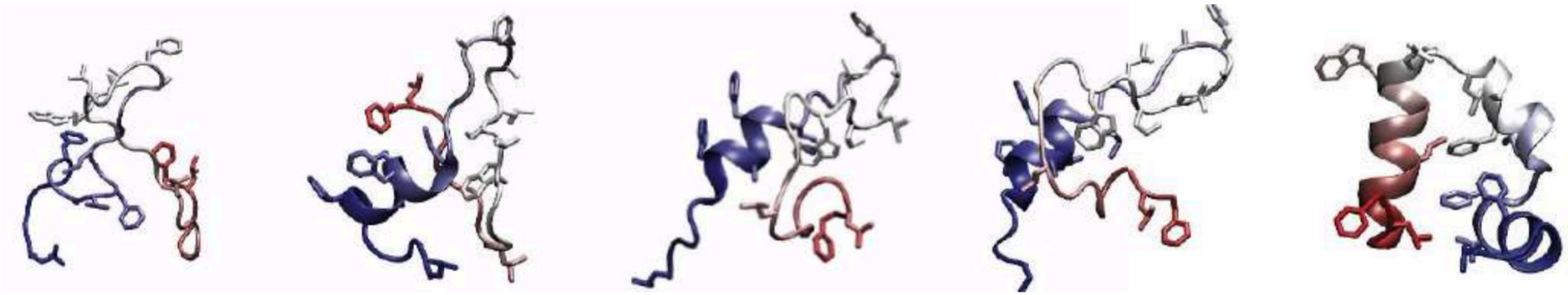
MM/CG
few millions of "atoms"
Hundreds of μs to (if) few ms



Coarse Grained
Several millions of "atoms"
Dozens of μs to ms

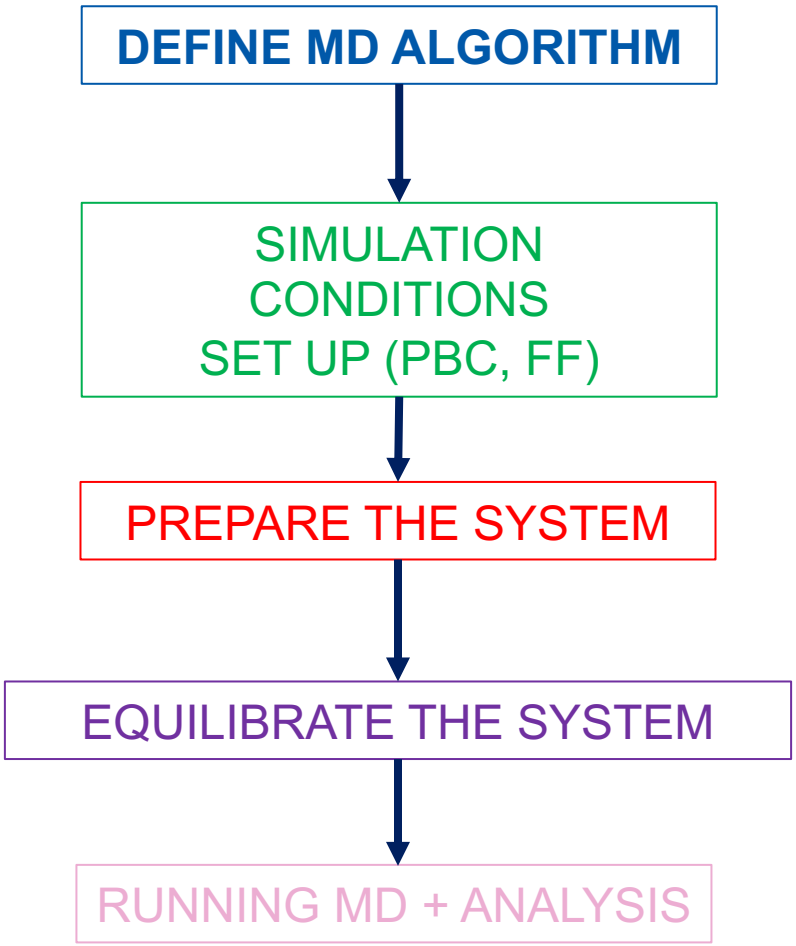
Introduction: Molecular Dynamics

The protein-folding problem is one of the major challenges of molecular biology.
MD simulation to solve protein folding in a solution for at least microseconds



MD simulation of the folding of the villin headpiece, one of the fastest-folding proteins
Nature 451, 240, 2008

MD Simulation Step by Step



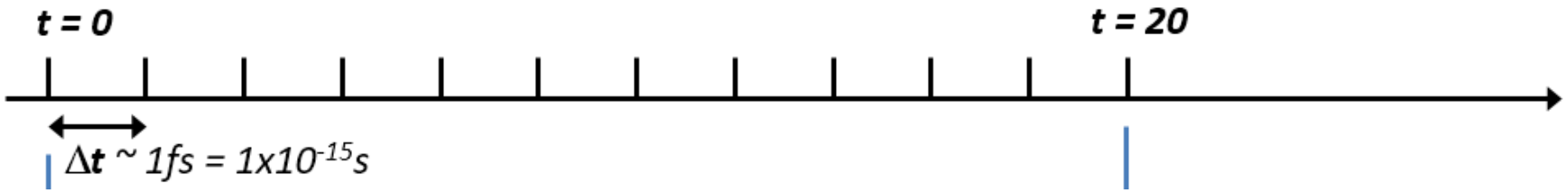
MD Simulation Step by Step

DEFINE MD ALGORITHM

Define a set of initial conditions algorithm to describe:

- velocities of all particles in the system
- interaction potential for deriving the forces among all the particles.

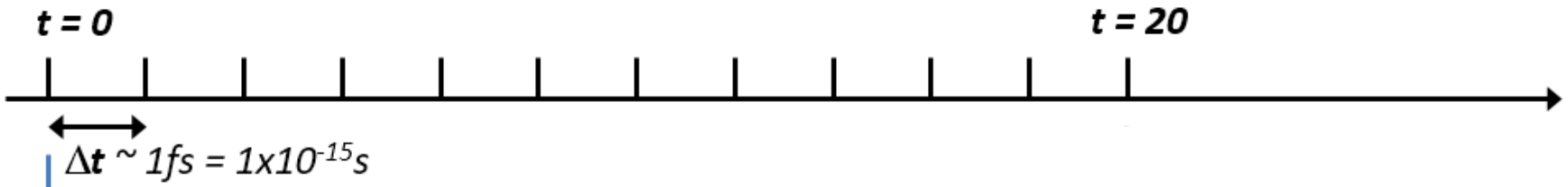
Define an Algorithm

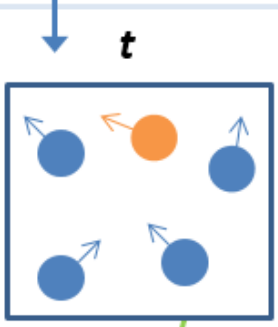


Divide time into discrete time steps

- Dynamics simulations are limited by the highest frequency vibration
- Ideally the timestep should be 1/10 highest frequency
- In most cases C-H bond stretching (10^{-14} s) is the fastest mode
- SHAKE algorithm – impose constraints on the bond lengths. It averages out the highest frequency vibrations.

Define an Algorithm



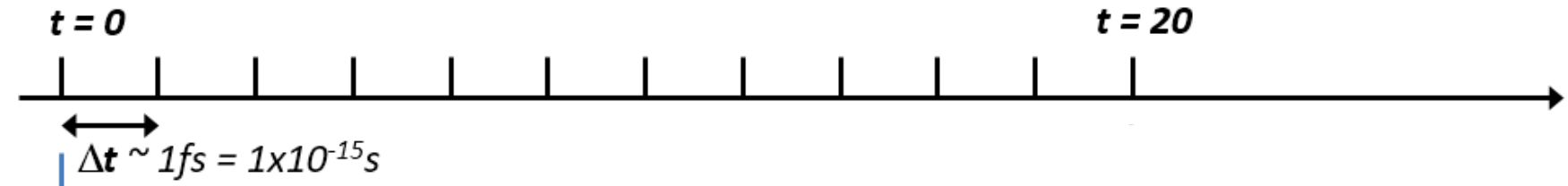


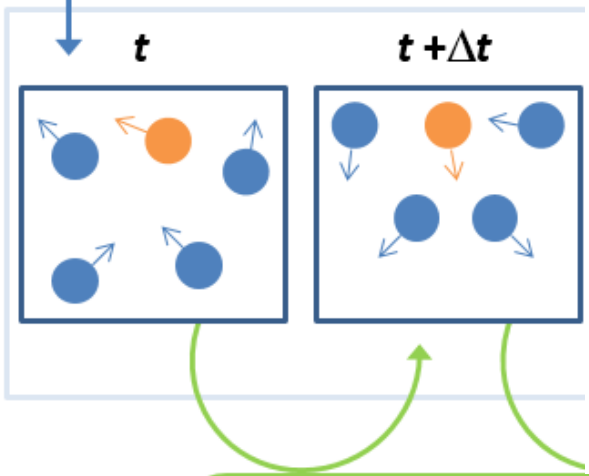
Calculate the forces

$$f_{x,j}(R(t)) = -\frac{\partial U(R(t))}{\partial x_j}$$

$$x_j(t+\Delta t) \approx 2x(t) - x(t-\Delta t) + \frac{f_{x,j}(t)}{m}\Delta t^2$$

Define an Algorithm



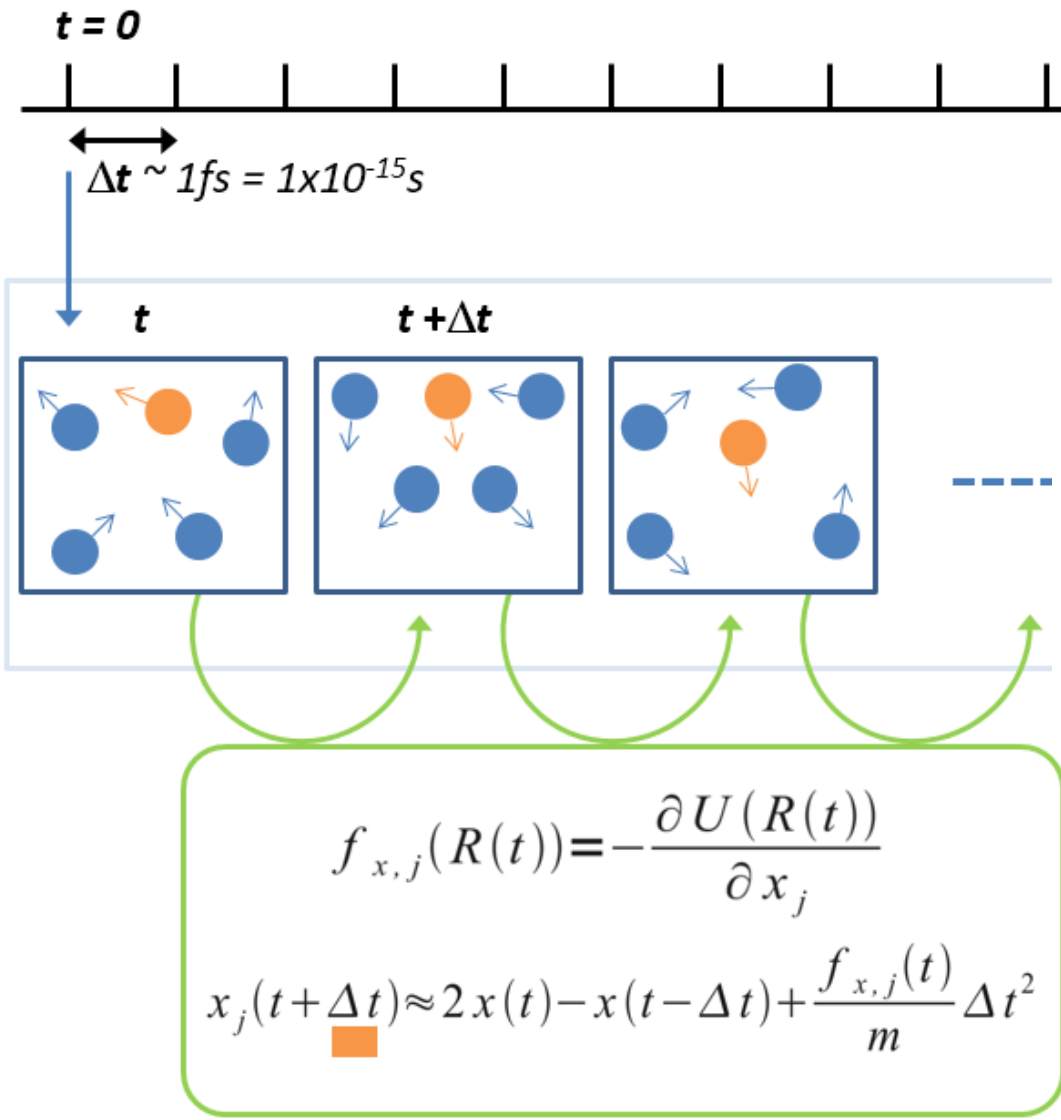


Move Atoms

$$f_{x,j}(R(t)) = -\frac{\partial U(R(t))}{\partial x_j}$$

$$x_j(t+\Delta t) \approx 2x(t) - x(t-\Delta t) + \frac{f_{x,j}(t)}{m}\Delta t^2$$

Define an Algorithm

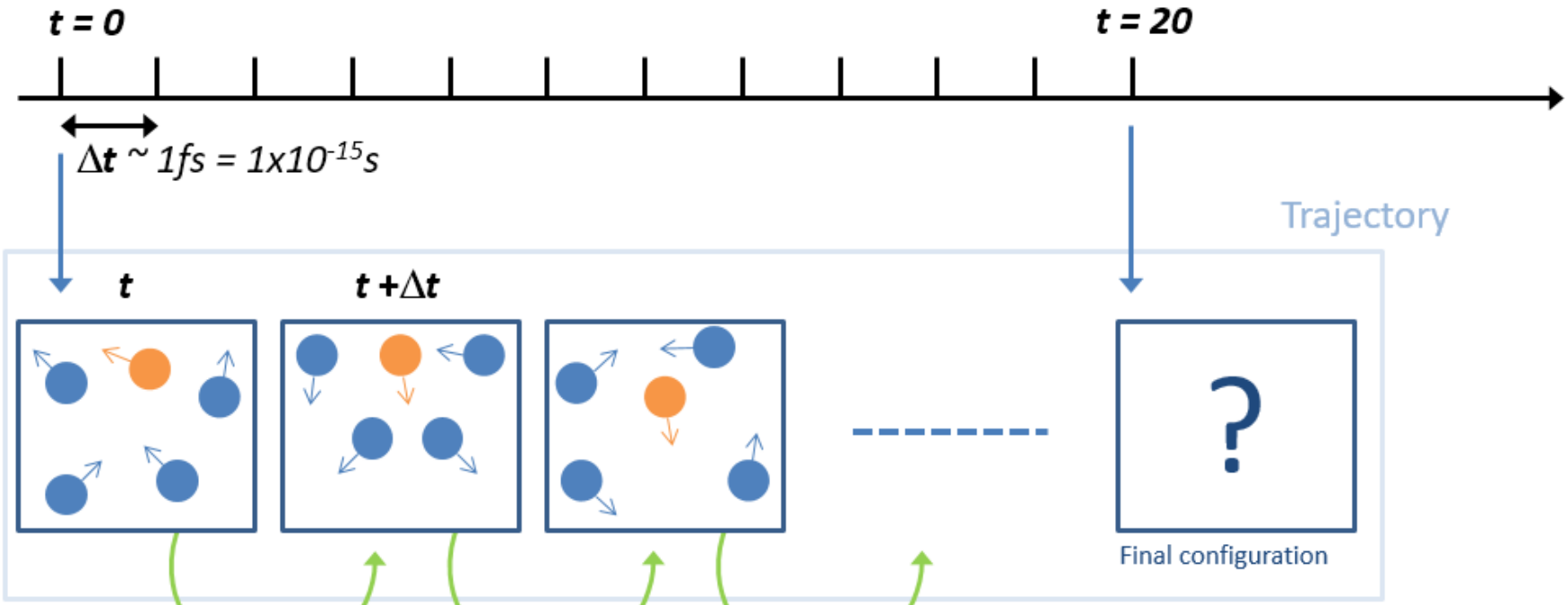


$$f_{x,j}(R(t)) = -\frac{\partial U(R(t))}{\partial x_j}$$

$$x_j(t+\Delta t) \approx 2x(t) - x(t-\Delta t) + \frac{f_{x,j}(t)}{m}\Delta t^2$$

Integrate….
…Integrate…
…Integrate …
Newton's laws of motion

Define an Algorithm



$$f_{x,j}(R(t)) = -\frac{\partial U(R(t))}{\partial x_j}$$

$$x_j(t+\Delta t) \approx 2x(t) - x(t-\Delta t) + \frac{f_{x,j}(t)}{m}\Delta t^2$$

$$U = \sum_{bonds} k_b \left(r_{ij} - r_{eq}\right)^2 + \sum_{angles} k_\theta \left(\theta - \theta_{eq}\right)^2 +$$

$$\sum_{dihedrals} \frac{V_k}{2}\left[1 + \cos\left(n_k\varphi - \gamma_k^{eq}\right)\right] +$$

$$\sum_{l}^{N}\sum_{l>m}^{N}\left\{4\varepsilon\left[\left(\frac{\sigma}{r_{lm}}\right)^{12} - \left(\frac{\sigma}{r_{lm}}\right)^{6}\right] + \frac{q_l q_m}{\in r_{lm}}\right\}$$

Define Force Field

A **Force Field** refers to the form and parameters of mathematical functions used to describe the potential energy (*U*) of a system of particles

Components:

- **Mathematical Functions**
- **Atom Types**
- **Parameters**

$$\vec{F}_i = -\nabla U_{\vec{r}_i}(\vec{r}_1, ..., \vec{r}_n)$$

$$U = E_{bond} + E_{angle} + E_{dihedral} + E_{elec} + E_{vdw}$$

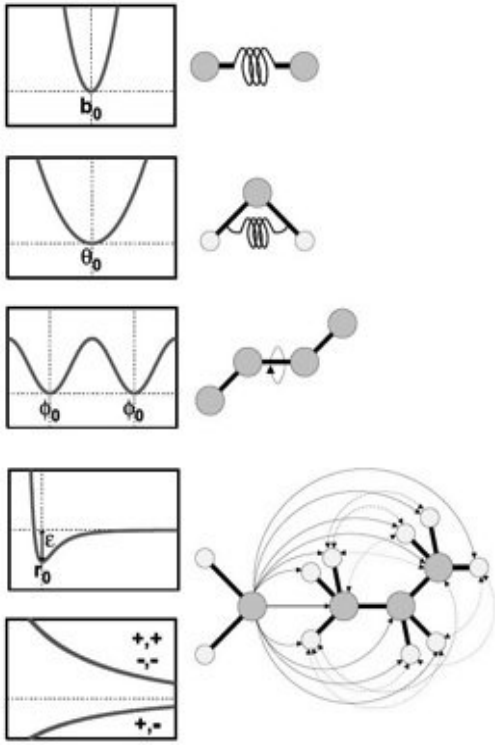
$$E_{bond} = \sum_{bonds} K_r (r - r_0)^2$$

$$E_{angle} = \sum_{angles} K_\theta (\theta - \theta_0)^2$$

$$E_{dihedral} = \sum_{diheds} \sum_{n} \frac{V_n}{2}(1 + \cos(n\Phi - \gamma))$$

$$E_{vdw} = \sum_{i} \sum_{j>i} \left(A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-6} \right)$$

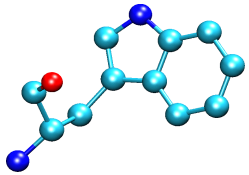
$$E_{elec} = \sum_{i} \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$



Reiβer. *J. Chem. Theory Comput.* **2017**,13:2367-2372.

Define Force Field

Force Field

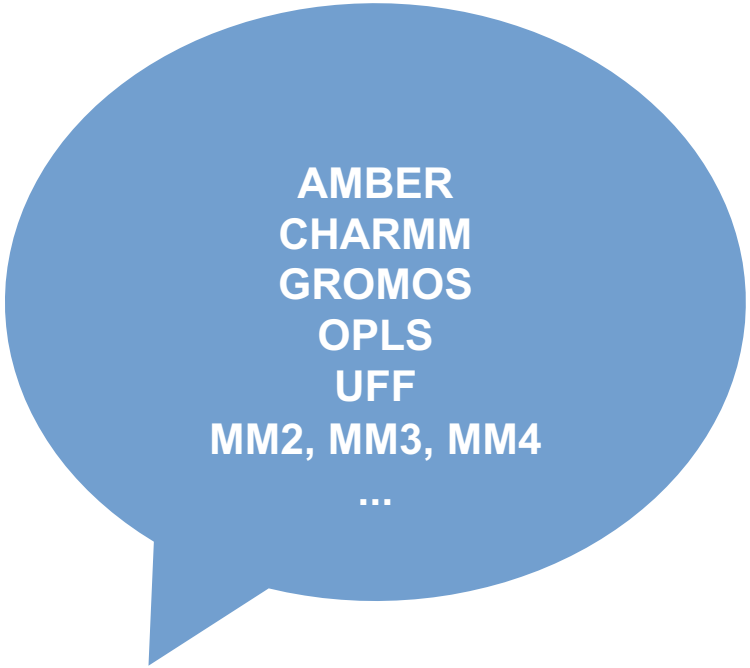


In molecular dynamics each molecule is described as a series of charged points (atoms) linked by springs (bonds).

To describe the time evolution of bond lengths, bond angles and torsions, also the non-bonding van der Waals and elecrostatic interactions between atoms.

The forcefield is a collection of equations and associated constants designed to reproduce molecular geometry and selected properties of tested structures.

Define Force Field

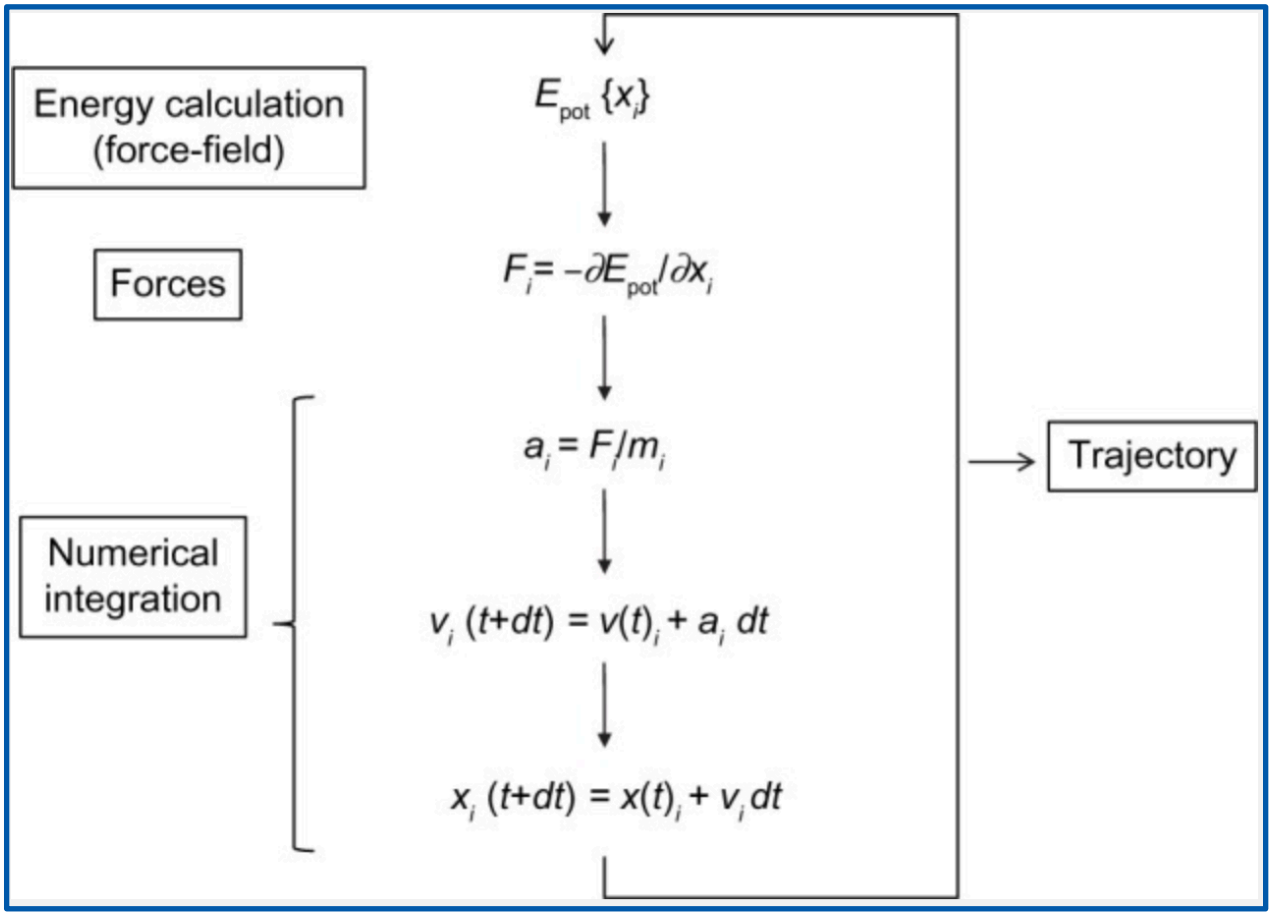


Classical

The force field jungle

MD Simulation Time

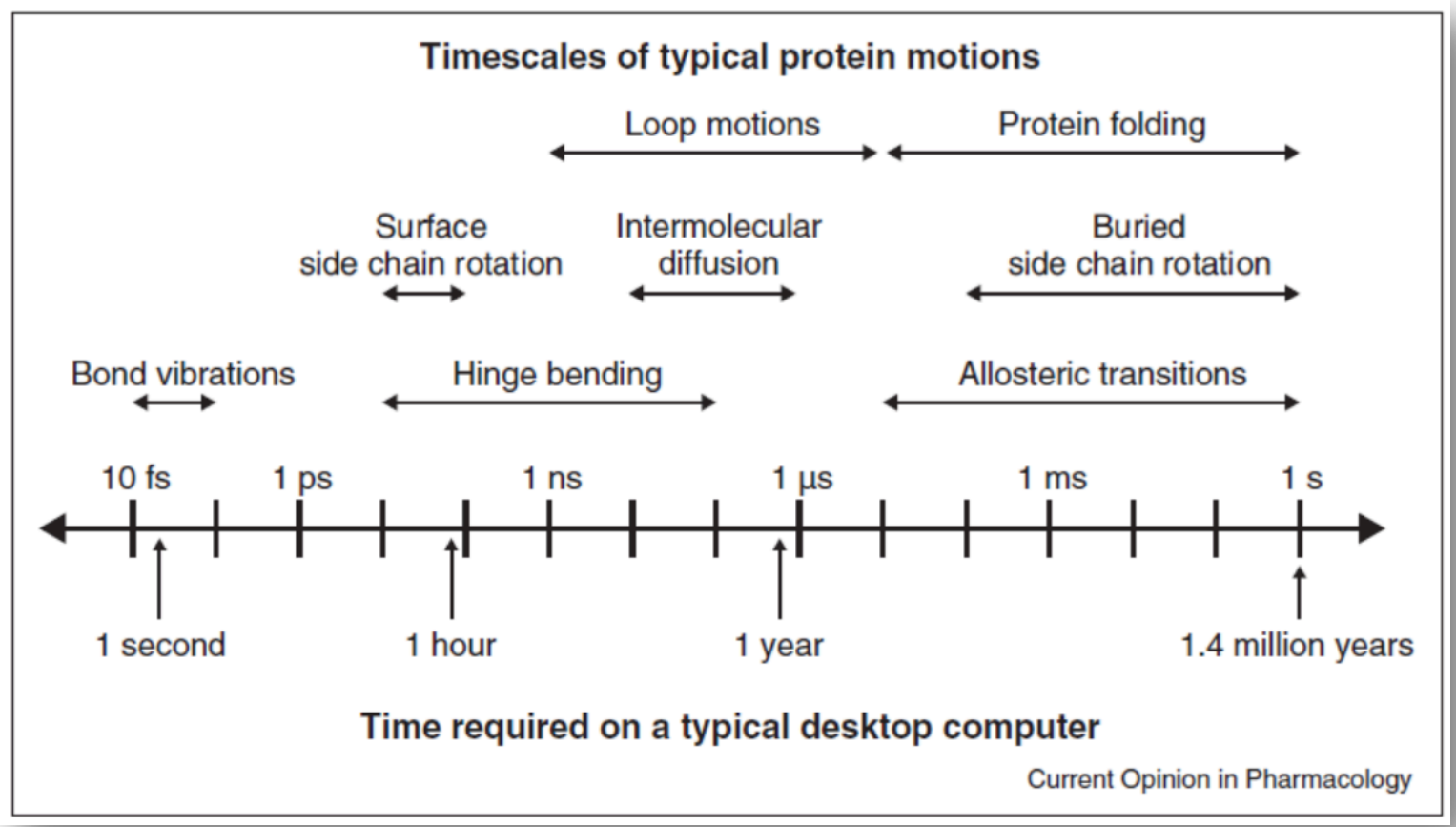
MD Simulation time?



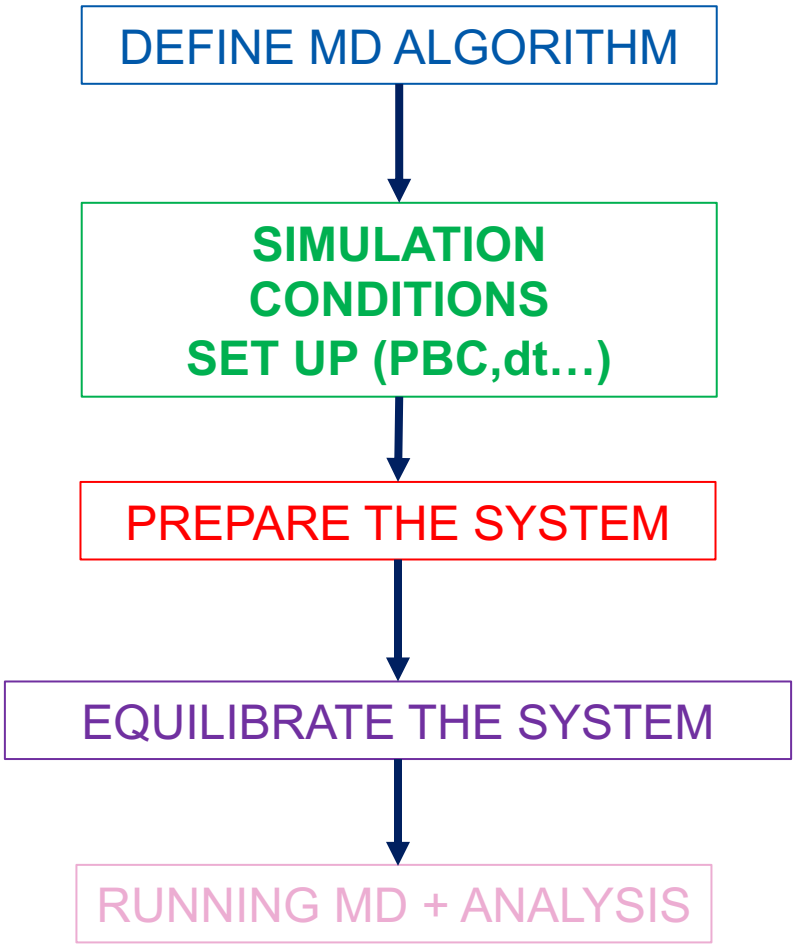
Hospital et al, Adv Appl Bioinform Chem, 2015, 10:37-47

MD Simulation Time

MD Simulation time?



MD Simulation Step by Step



Simulation Conditions Set Up

Simulation Cell

Boundary Condition

Constructing neighboring cells

Simulation Conditions Set Up

Simulation Cell

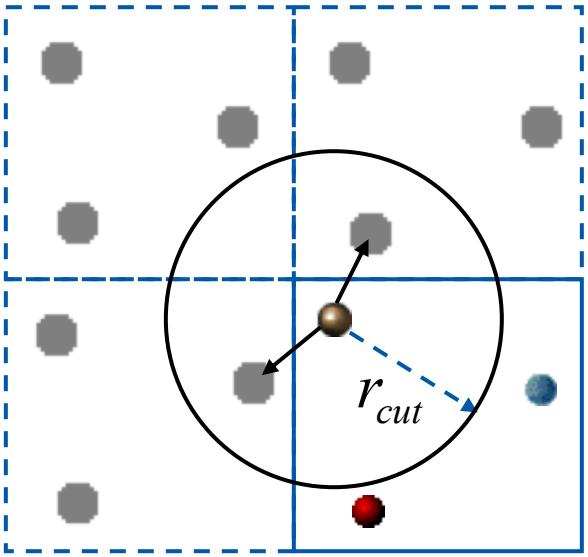
Boundary Condition

Periodic boundary conditions enable a simulation to be performed using a relatively small number of particles. The box containing the system is replicated infinitely in all directions to give a periodic array

https://www.youtube.com/watch?v=eRLIJq0CrXg

https://www.youtube.com/watch?v=sHpLY-cow-I

Constructing neighboring cells



$$r_{cut} < \frac{L}{2}$$

Simulation Conditions Set Up

Simulation Cell

Boundary Condition

Constructing neighboring cells

Initial atom velocities

MD Time step

Temperature Control

Simulation Conditions Set Up

Simulation Cell

Boundary Condition

Constructing neighboring cells

Initial atom velocities

MD Time step

$$f_{x,j}(R(t)) = -\frac{\partial U(R(t))}{\partial x_j}$$

$$x_j(t+\Delta t) \approx 2x(t) - x(t-\Delta t) + \frac{f_{x,j}(t)}{m}\Delta t^2$$

Simulation Conditions Set Up

Simulation Cell

Boundary Condition

Constructing neighboring cells

Initial atom velocities

MD Time step

Temperature Control
P, V conditions

Specific MD conditions (temperature, pressure, volume)

Constant energy, constant volume (NVE)

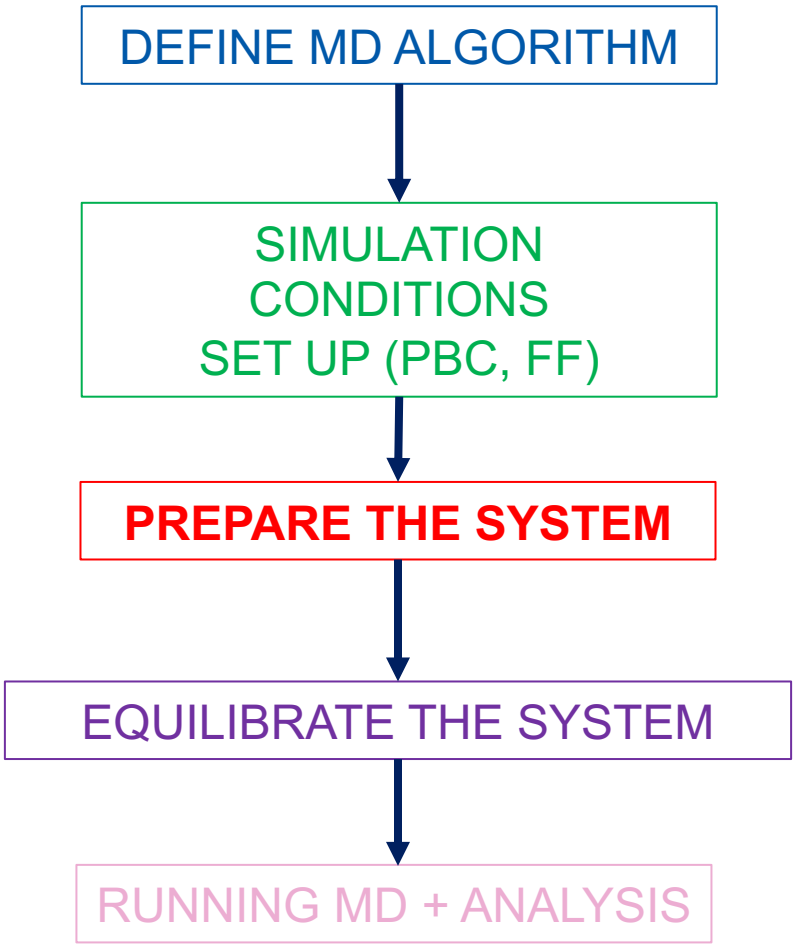
Constant temperature, constant volume (NVT)

Constant temperature, constant pressure (NPT)

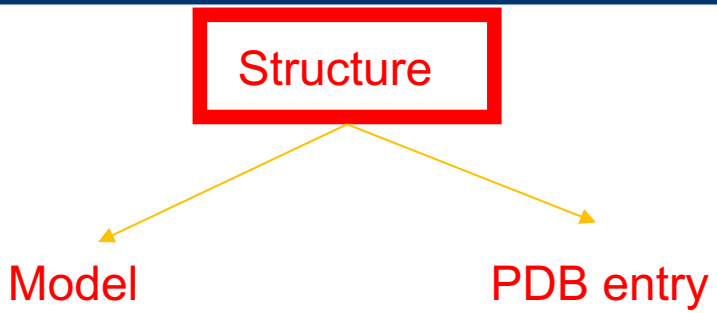
Choose the ensemble that best fits your system and start the simulations

Canonical ensemble, where to keep T constant: the most common thermostats are those of Berendsen and Andersen. System in contact with a large (infinite) heat bath with constant T_{bath}, to interchange energy and keep T constant, rescaling it at each step.

MD Simulation Step by Step



Prepare the System



Prepare the System

PDB entry

Prepare the System: Structure

PDB entry

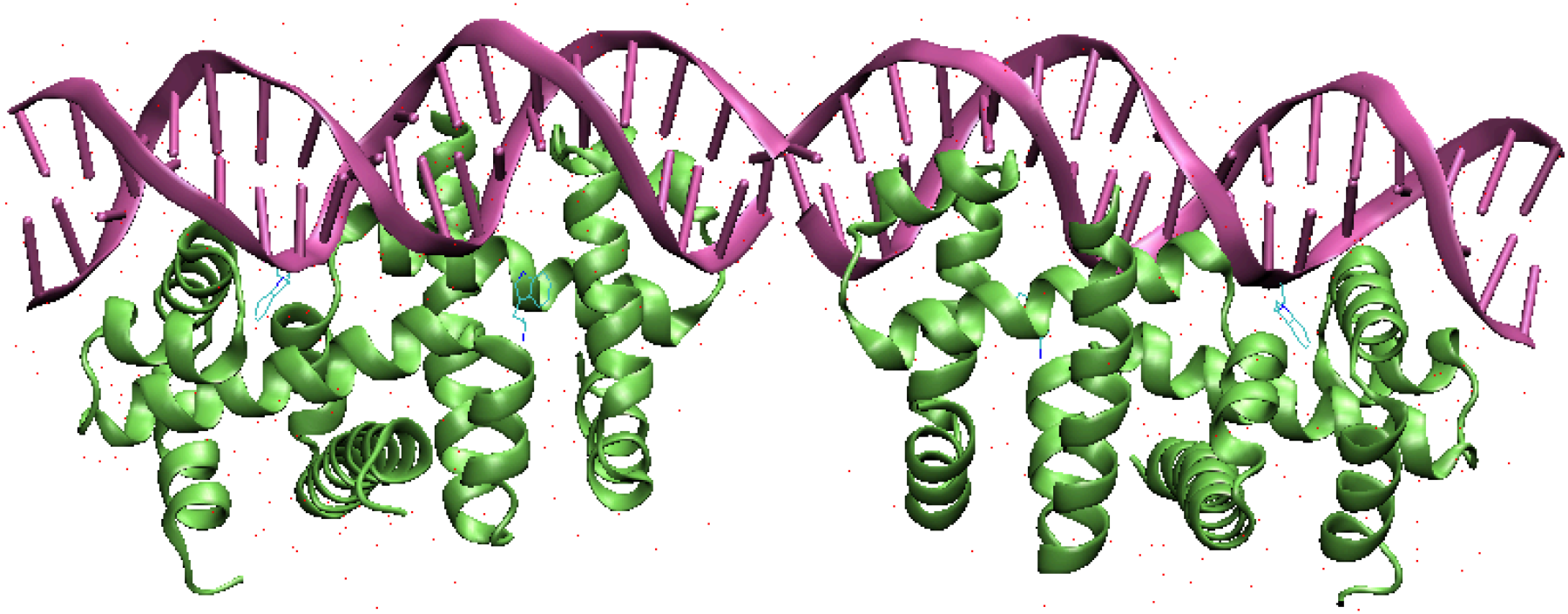
```
HEADER    TRANSCRIPTION/DNA                       30-AUG-92   1TRO
TITLE     CRYSTAL STRUCTURE OF TRP REPRESSOR OPERATOR COMPLEX AT ATOMIC
TITLE    2 RESOLUTION
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: DNA (5'-
COMPND   3 D(*TP*GP*TP*AP*CP*TP*AP*GP*TP*TP*AP*AP*CP*TP*AP*GP*T P*AP*C)-3');
COMPND   4 CHAIN: I, J, K, L;
COMPND   5 ENGINEERED: YES;
COMPND   6 MOL_ID: 2;
REMARK 470
REMARK 470 MISSING ATOM
REMARK 470 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 470 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 470 I=INSERTION CODE):
REMARK 470   M RES CSSEQI  ATOMS
REMARK 470     LYS A 106    CG   CD   CE   NZ
REMARK 470     GLU E  13    CG   CD   OE1  OE2
ATOM      1  O5'  DT I   1      39.210  -1.192  38.093  1.00 43.50           O
ATOM      2  C5'  DT I   1      38.300  -2.028  38.913  1.00 39.82           C
ATOM      3  C4'  DT I   1      37.682  -0.940  39.802  1.00 37.42           C
ATOM      4  O4'  DT I   1      38.758  -0.136  40.197  1.00 35.06           O
ATOM      5  C3'  DT I   1      36.792   0.012  39.000  1.00 35.46           C
ATOM      6  O3'  DT I   1      35.486  -0.472  38.842  1.00 36.33           O
ATOM      7  C2'  DT I   1      36.998   1.343  39.667  1.00 36.52           C
ATOM      8  C1'  DT I   1      38.277   1.241  40.462  1.00 34.06           C
```

Prepare the System

PDB entry

↓

Clean PDB



Prepare the System

PDB entry

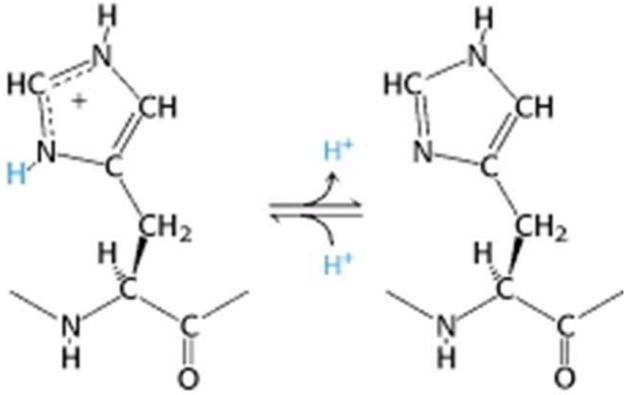
Clean PDB

Add missing atoms

Add H

Add heavy atoms

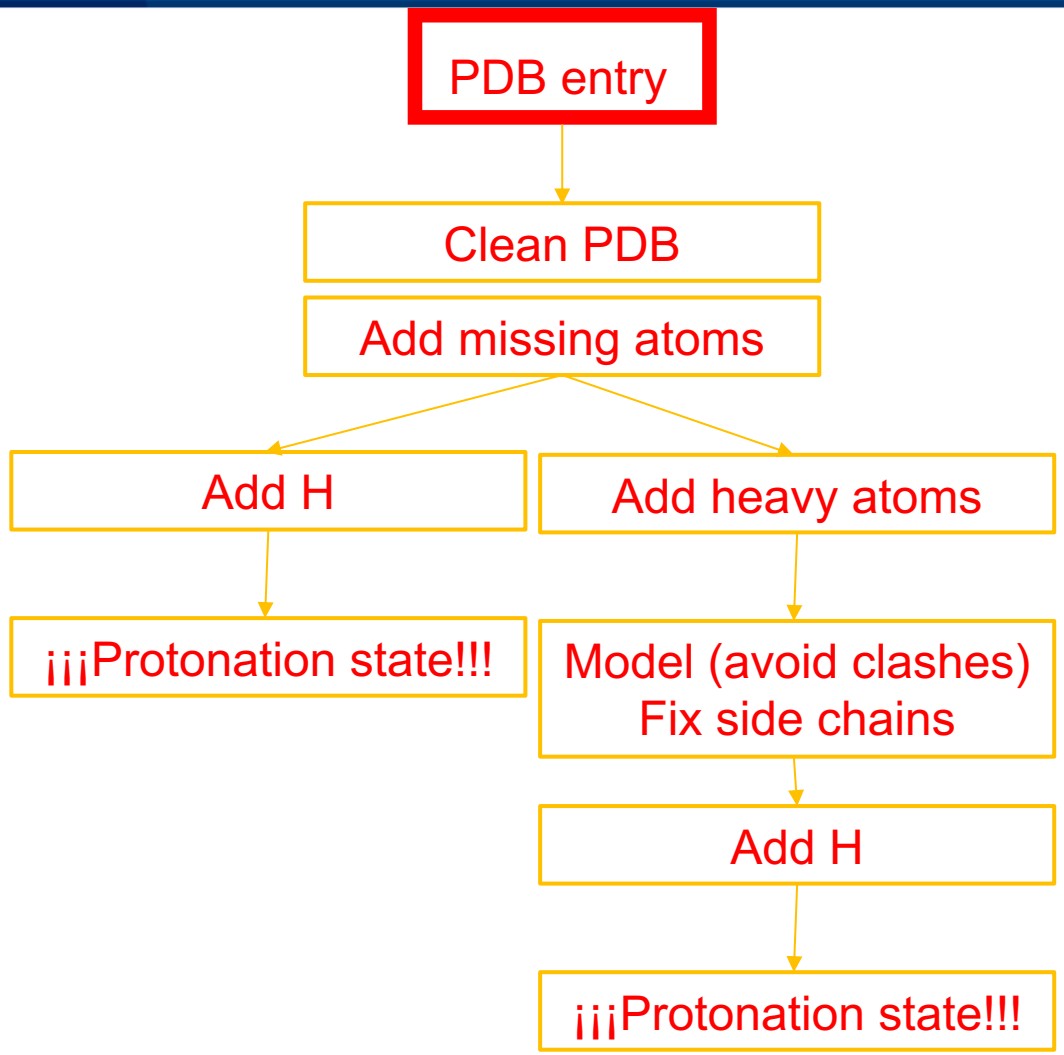
¡¡¡Protonation state!!!



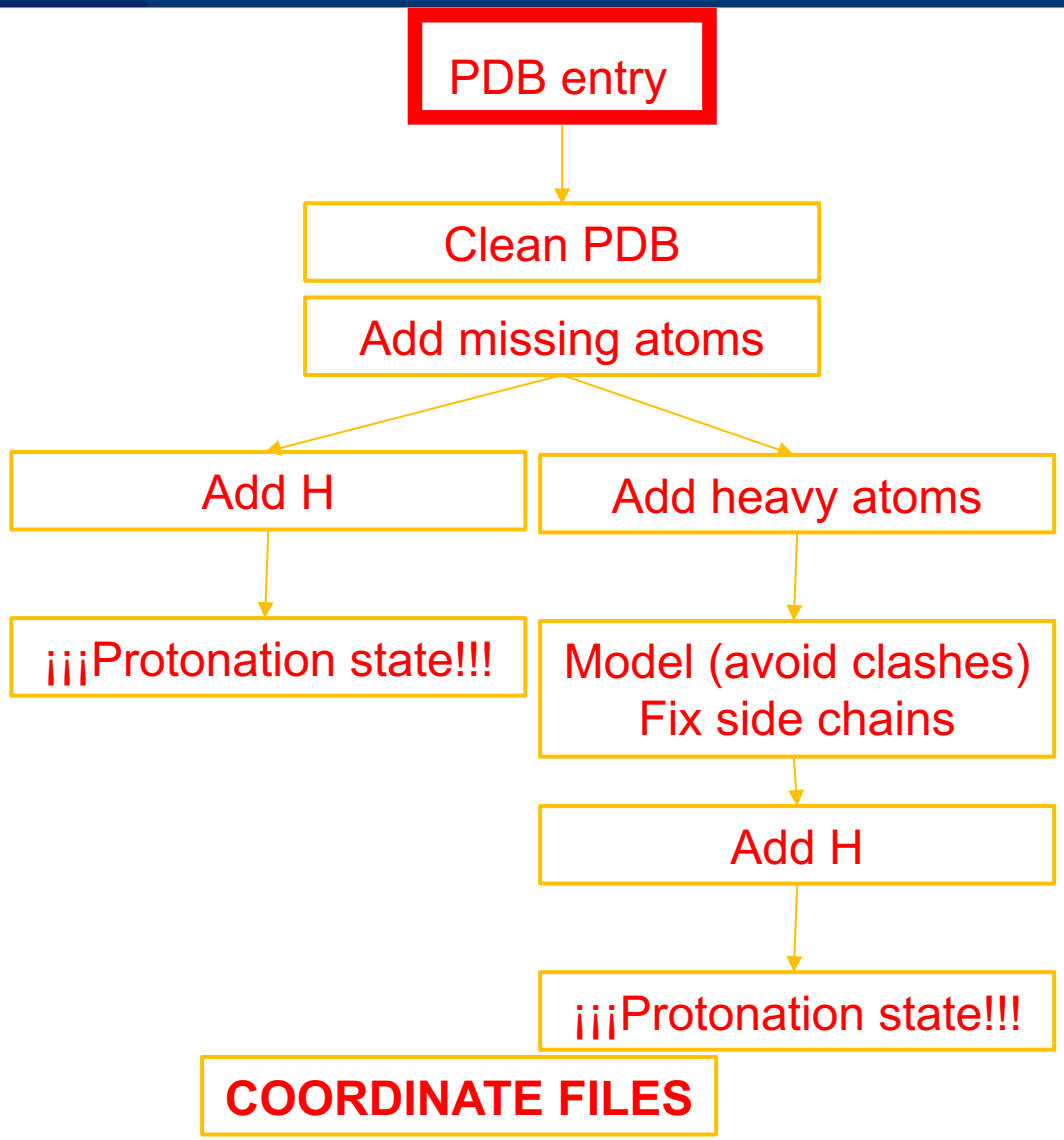
$pH < pK_a$ $pH > pK_a$

Histidine

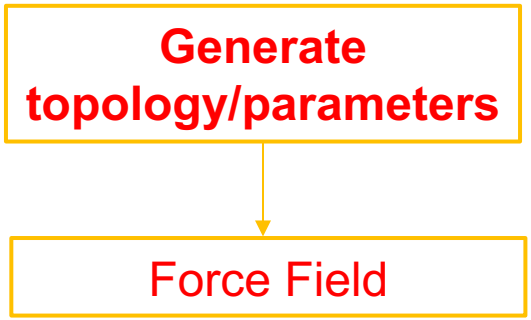
Prepare the System



Prepare the System



Prepare the System

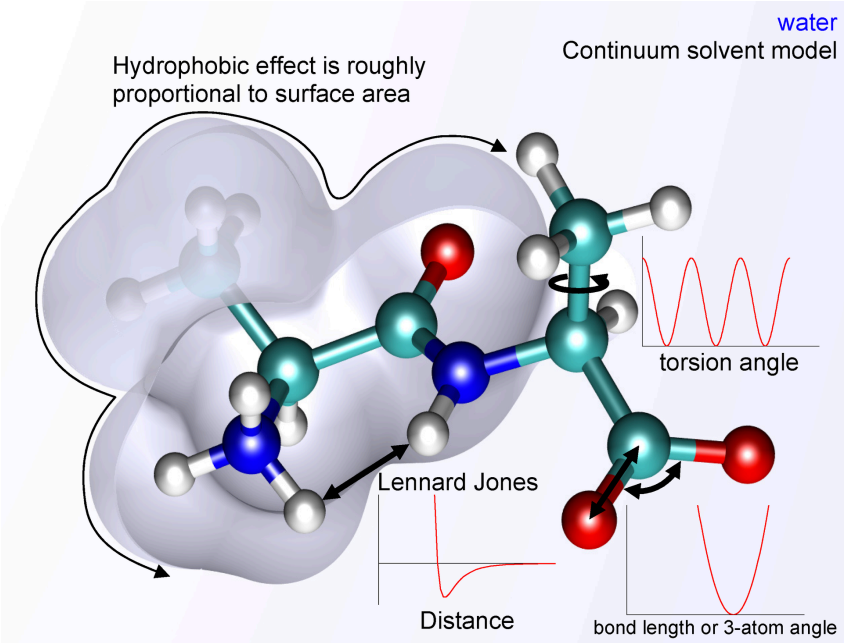


Prepare the System

Generate topology/parameters

↓

Force Field



For each atom type:

- charges for each atom
- connectivities between atoms, bonds, angles, dihedrals (GEOMETRY)
- force constants necessary to describe the bond energy, angle energy, torsion energy, nonbonded interactions (van der Waals and electrostatics)

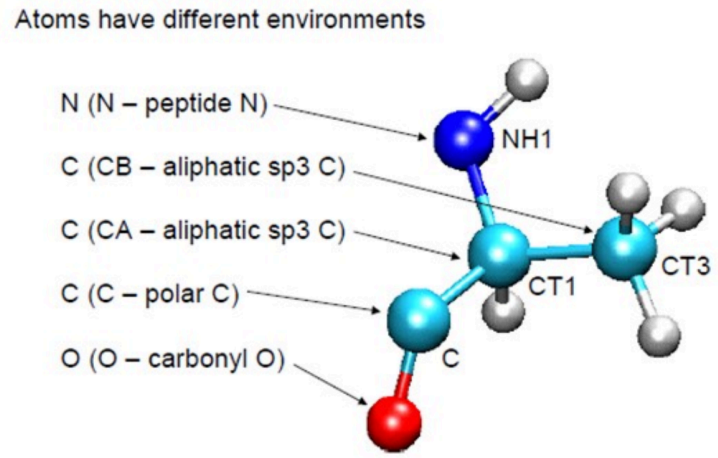
By Edboas - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=4194424

Prepare the System

Generate topology/parameters

↓

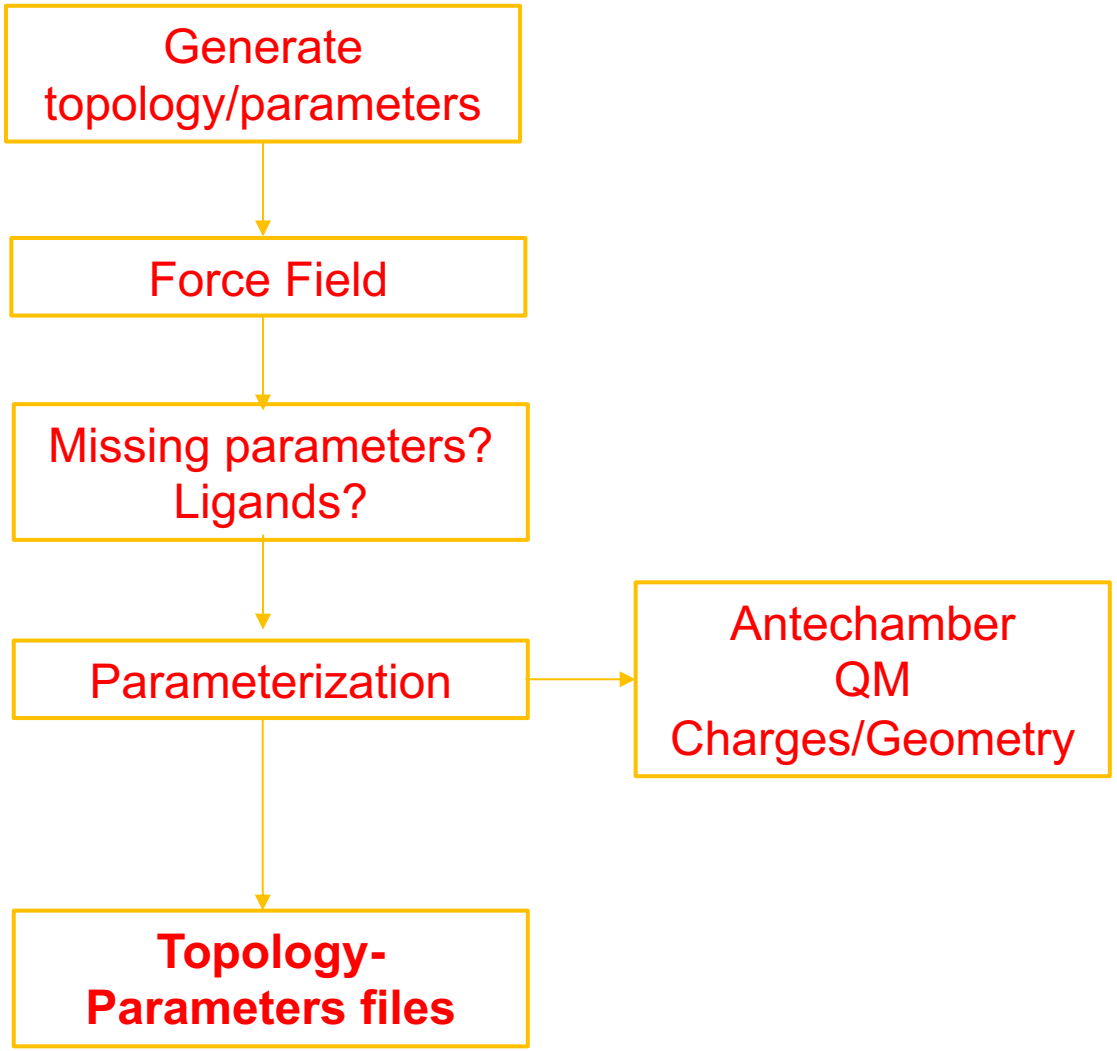
Force Field



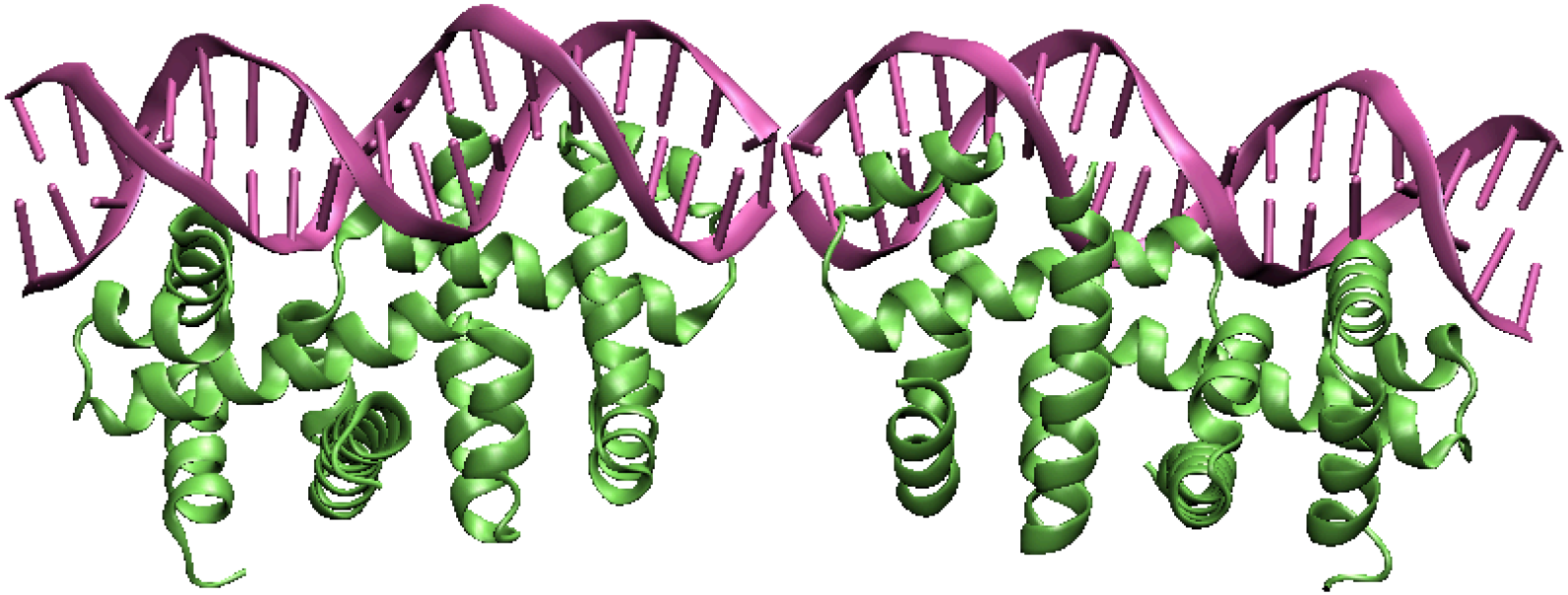
For each atom type:

- charges for each atom
- connectivities between atoms, bonds, angles, dihedrals (GEOMETRY)
- force constants necessary to describe the bond energy, angle energy, torsion energy, nonbonded interactions (van der Waals and electrostatics)

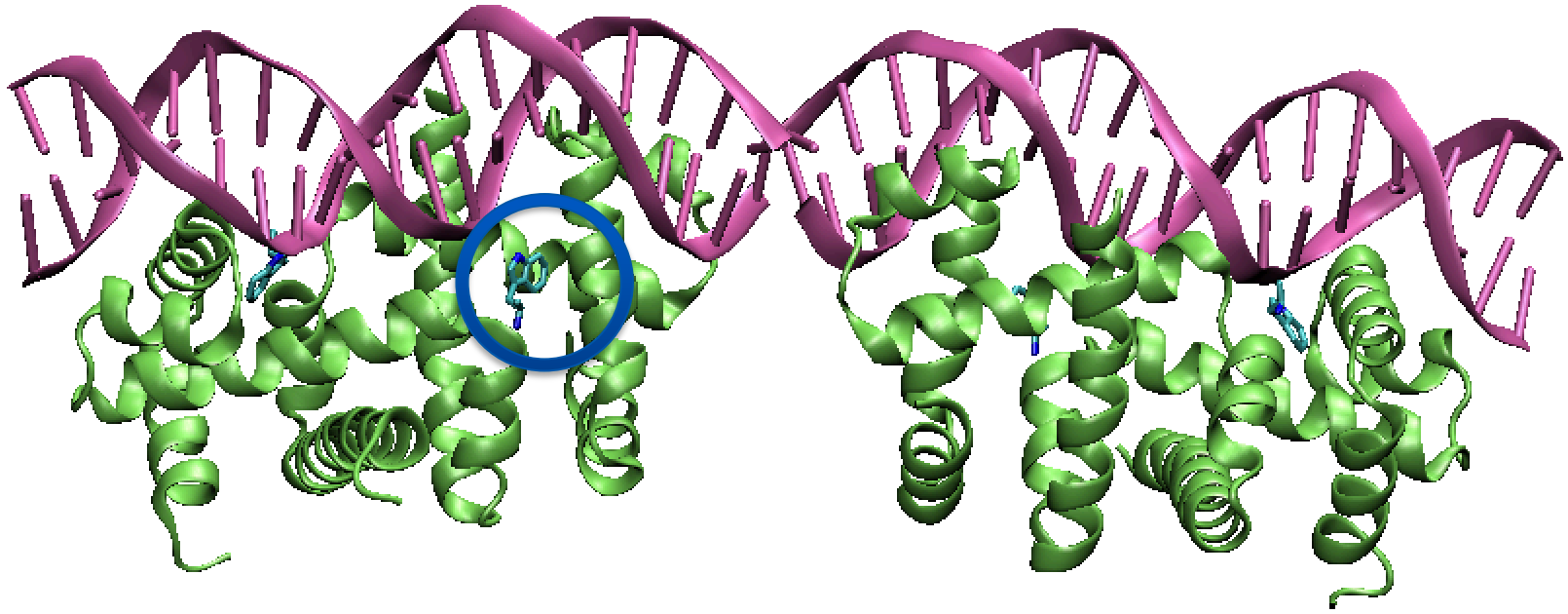
Prepare the System



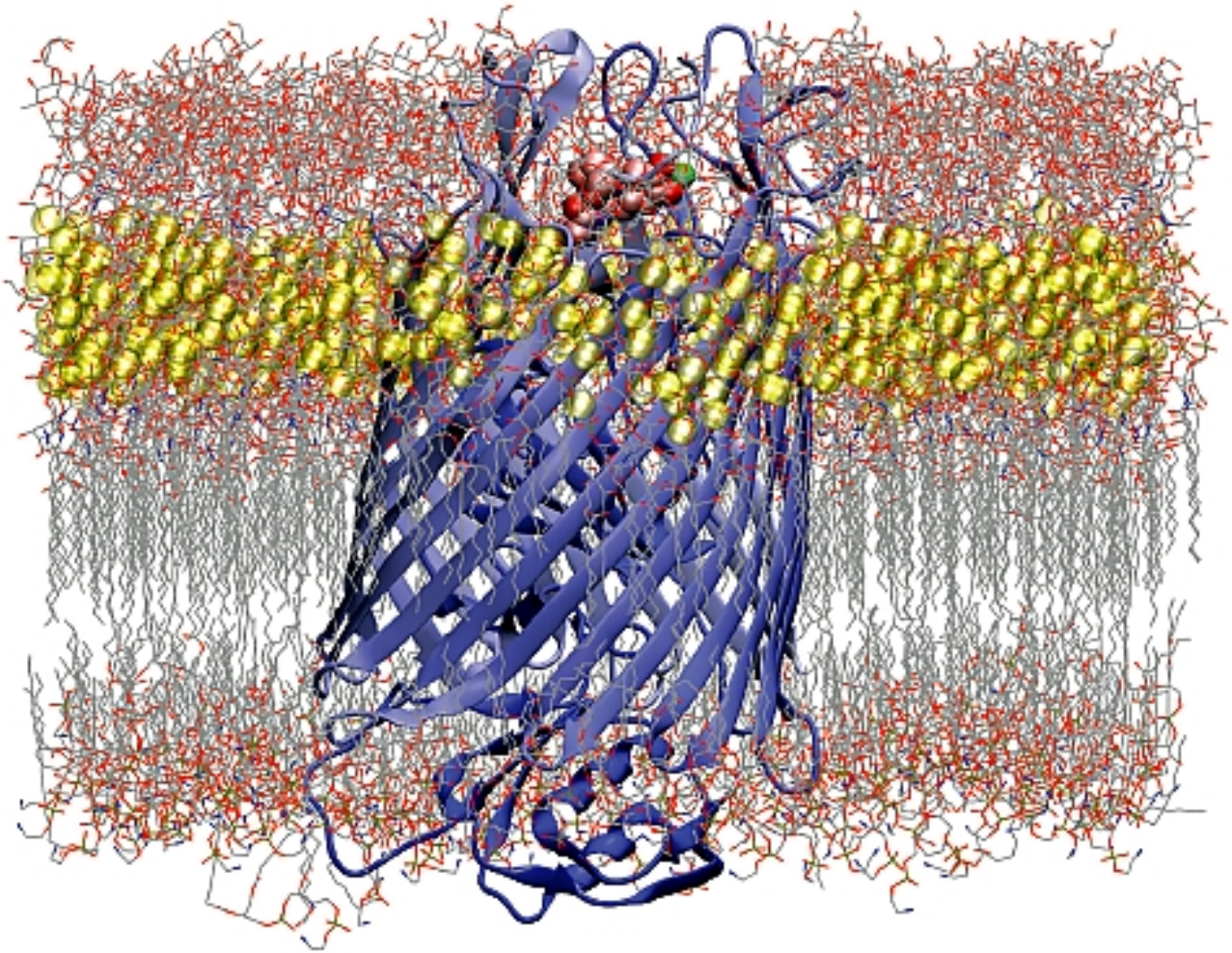
Initial Complex



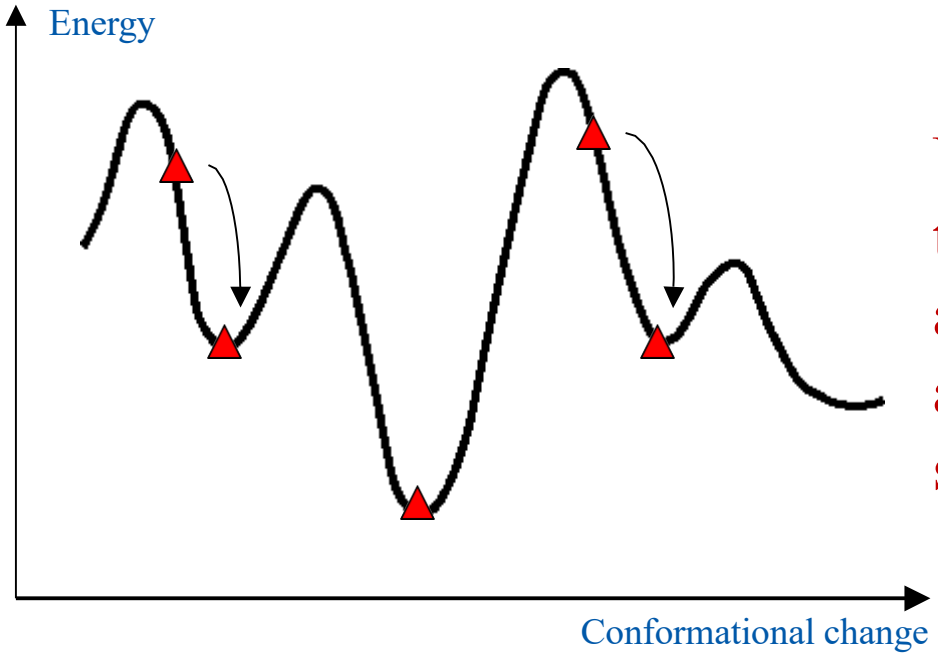
Add Ligands or Membrane



Add Ligands or Membrane

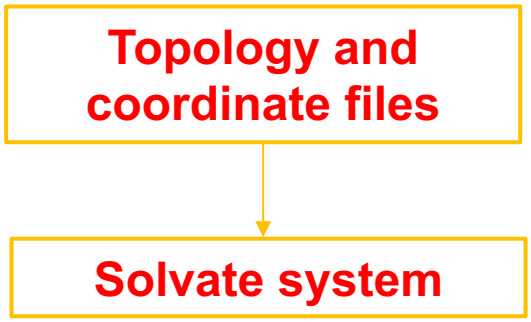


Prepare the System: Minimization



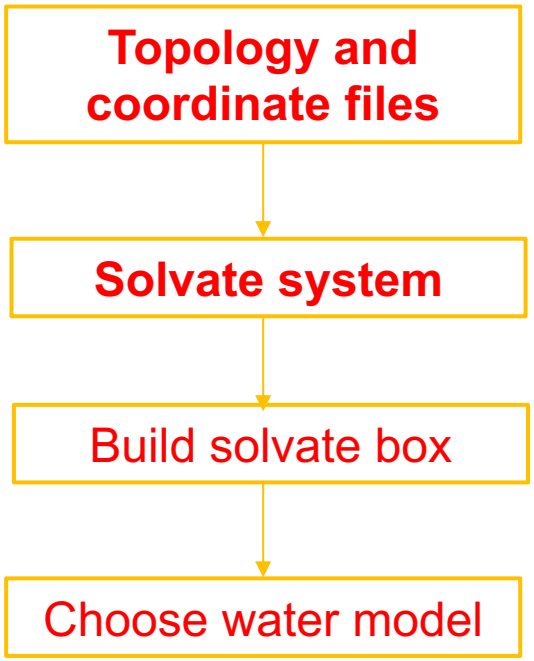
Using the forcefield, calculate the energy of the system and run a process of **minimization** to arrive to a **conformation** of the system with lower energy.

Prepare the System



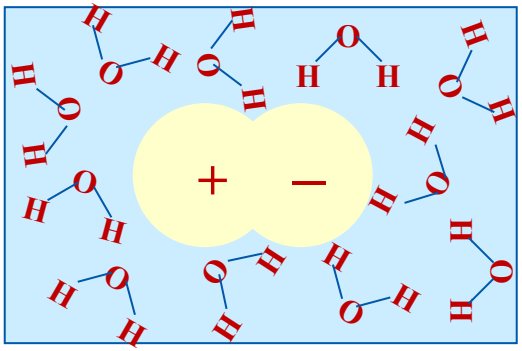
Solvate the system because many biological processes occur in aqueous solution, to take into account solvation effects, that play a crucial role in determining molecular conformation, electronic properties, binding energies, etc

Prepare the System



Prepare the System: Solvent

Explicit

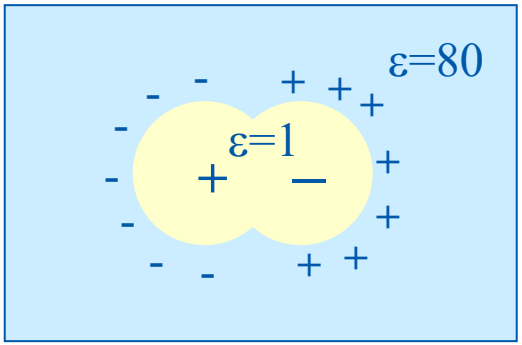


Solvent molecules are added to the molecular System,

WATER MODELS

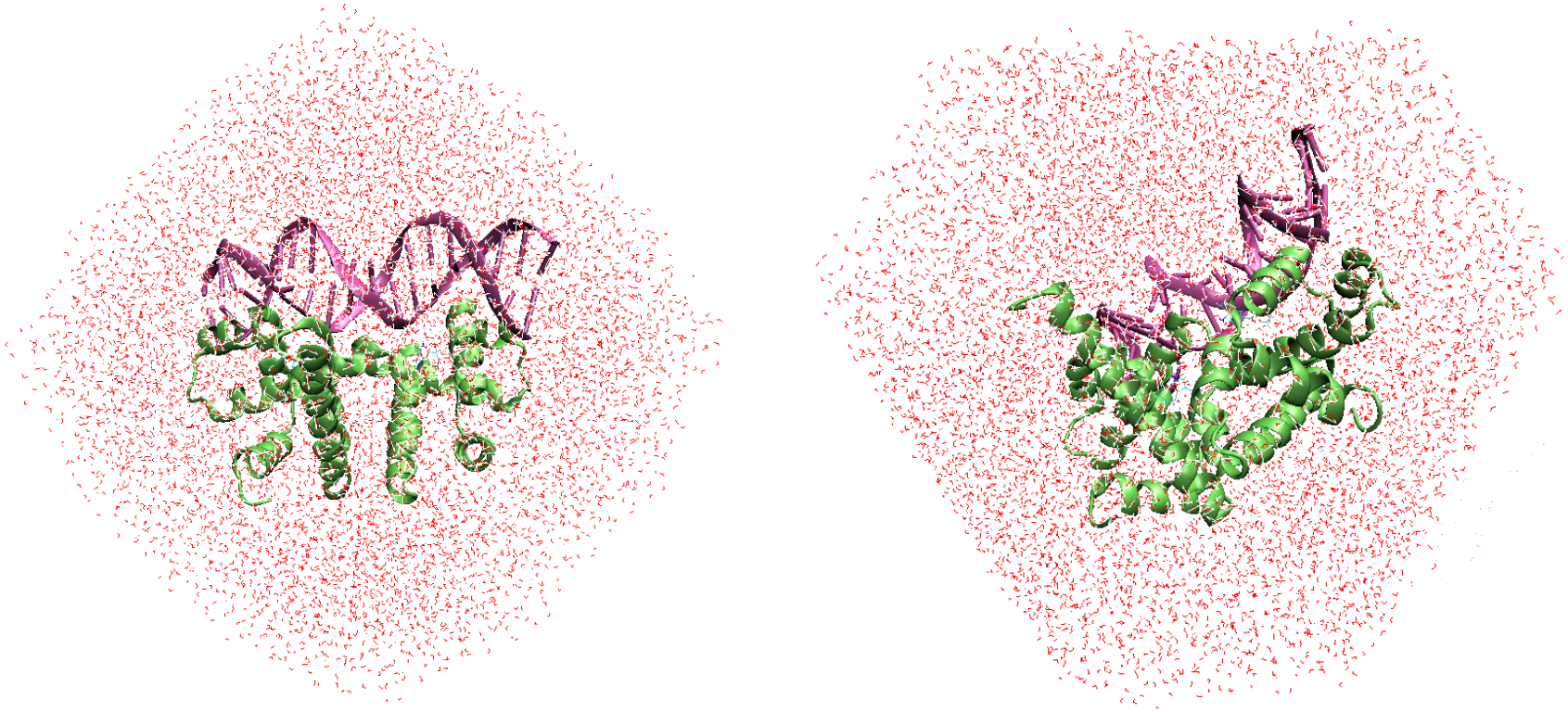
SPC, TIP3P, TIP4P, etc.

Implicit/Continuum

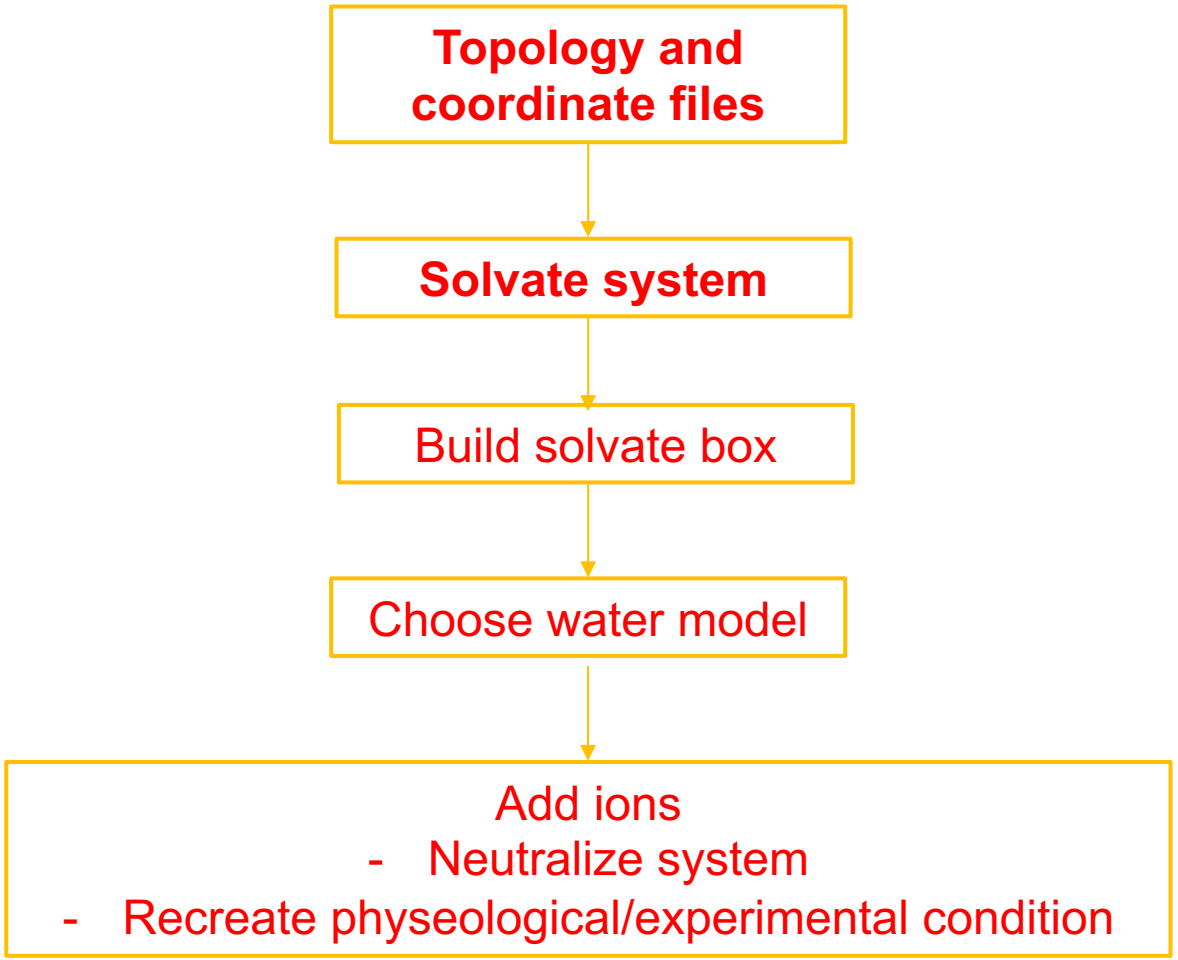


Solvent is modeled as a continuum dielectric

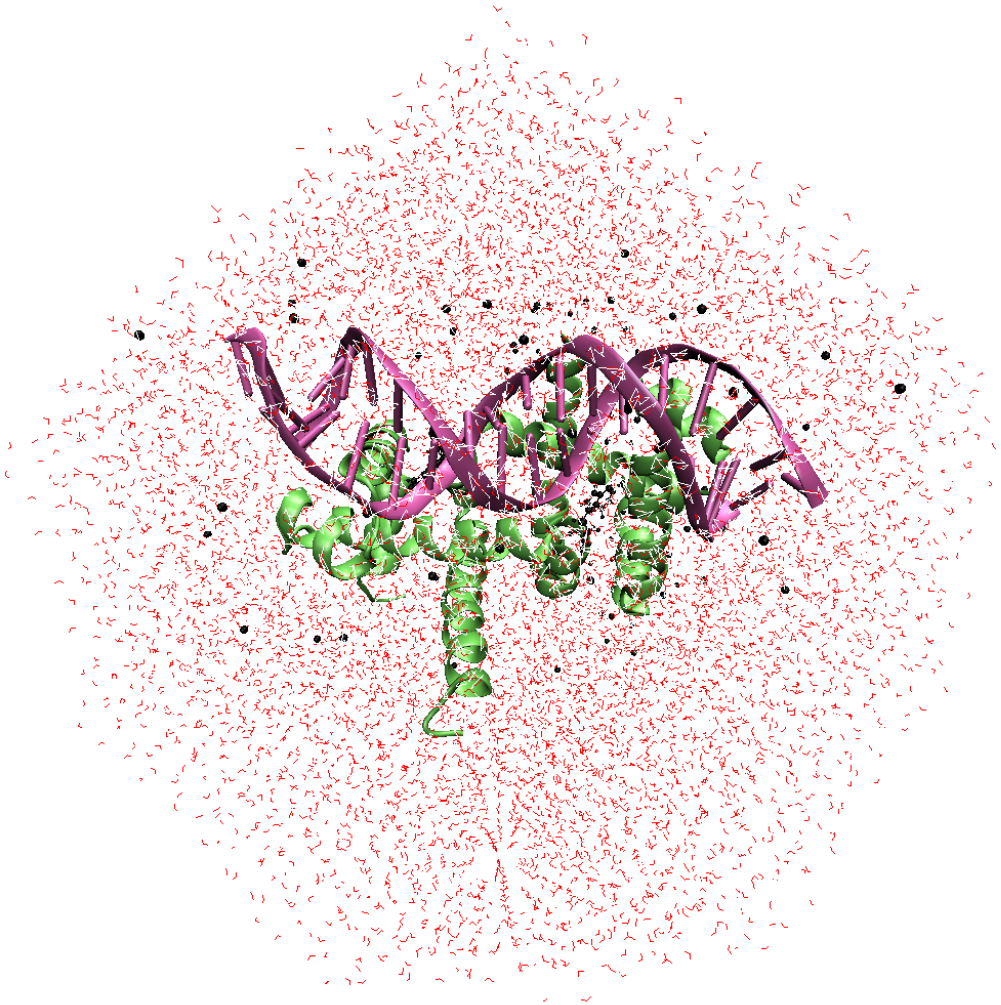
Prepare the System: Solvent



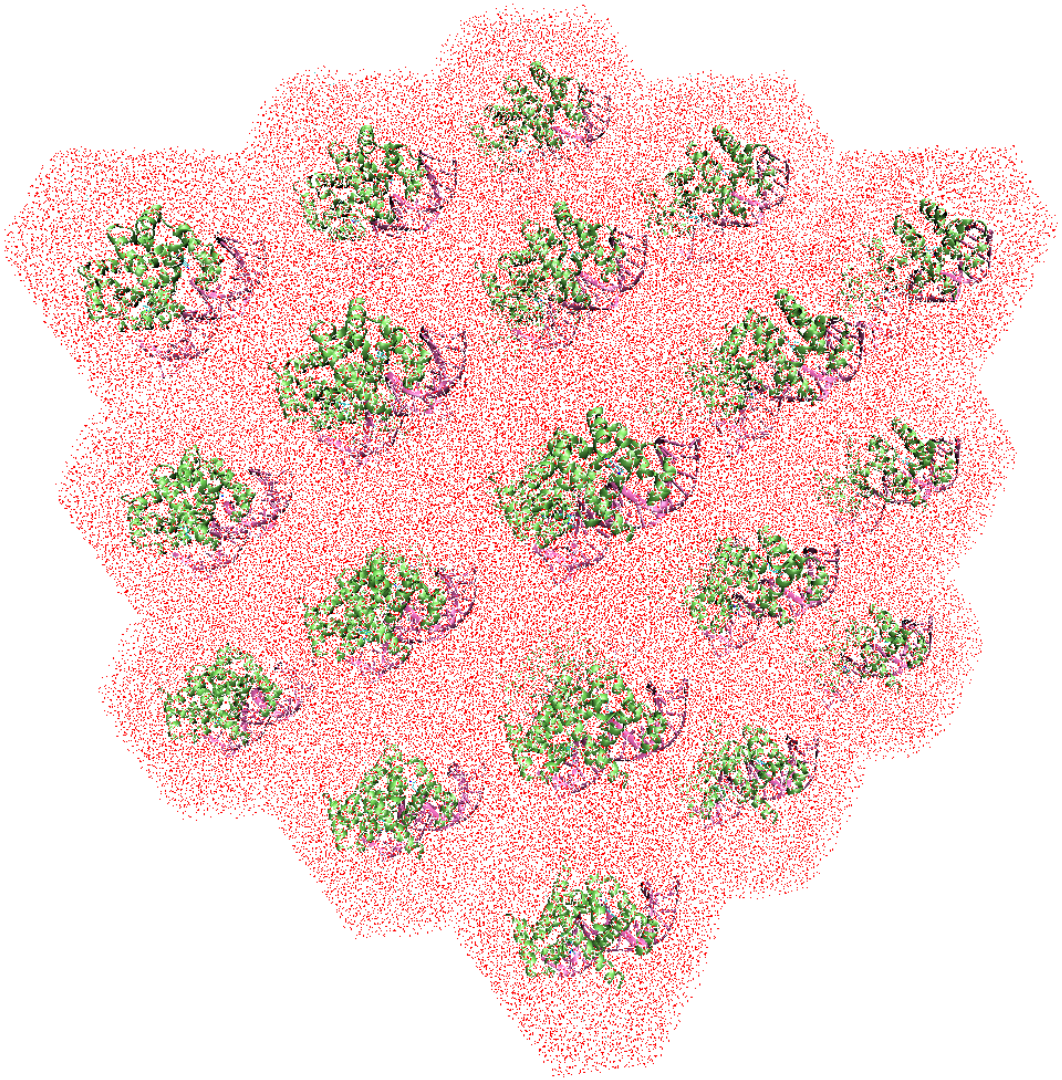
Prepare the System



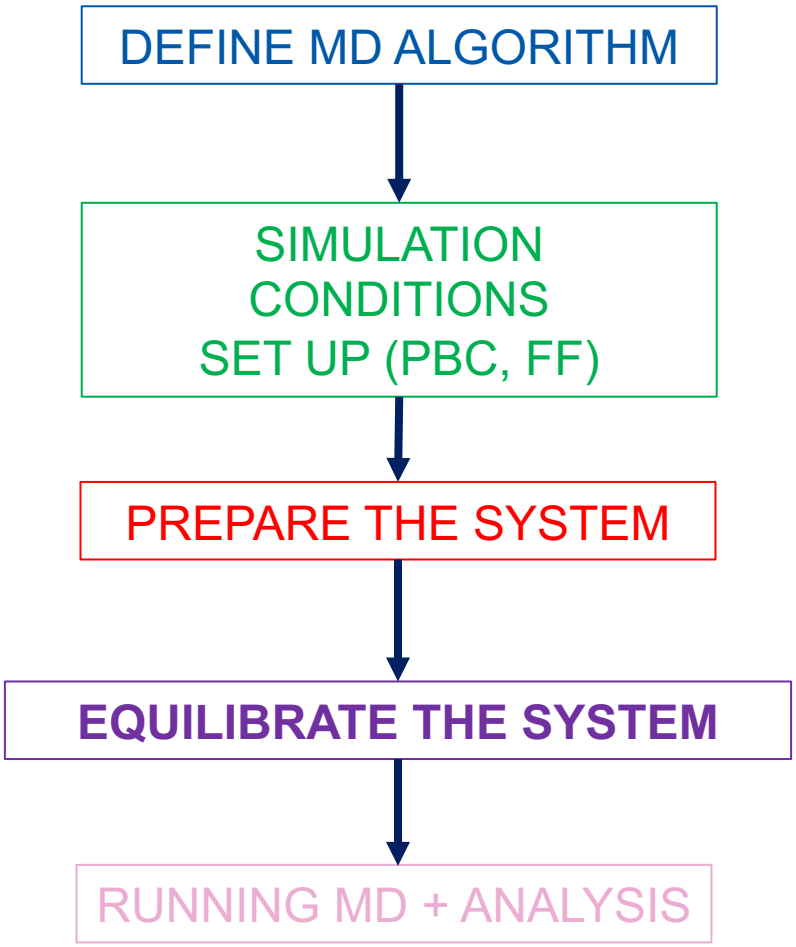
Prepare the System: Ions



Prepare the System



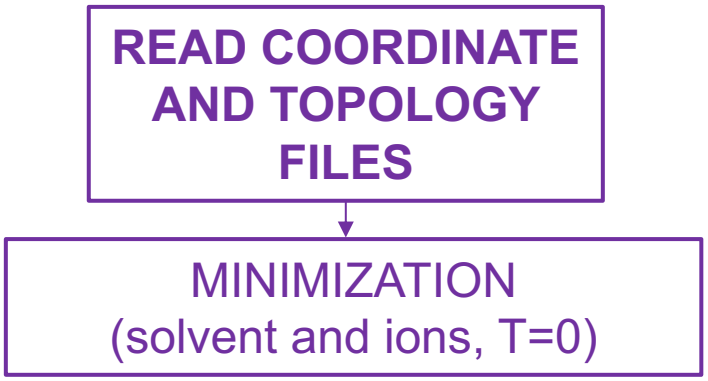
MD Simulation Step by Step



Minimize and Equilibrate the System

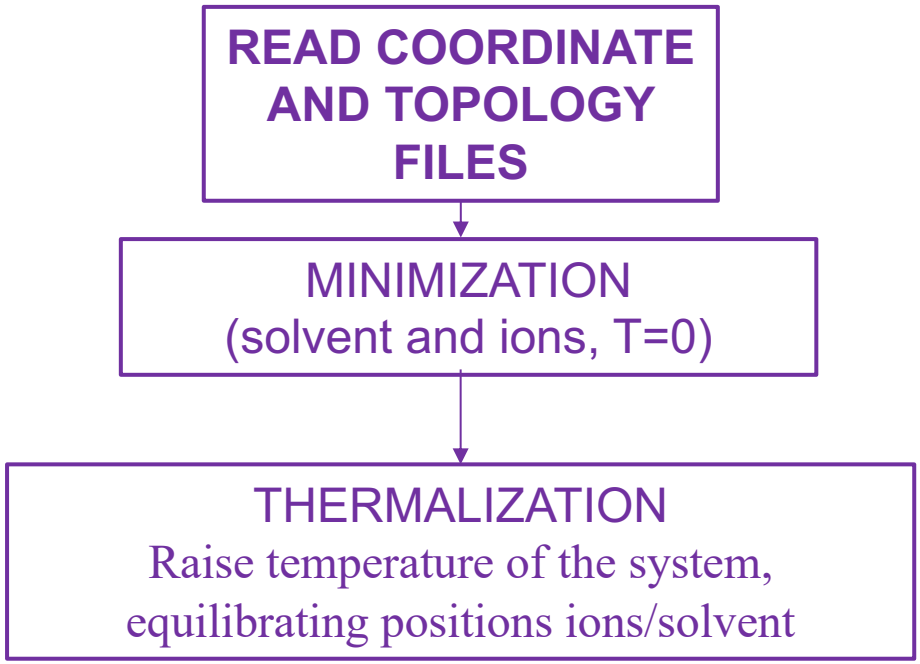
READ COORDINATE AND TOPOLOGY FILES

Minimize and Equilibrate the System

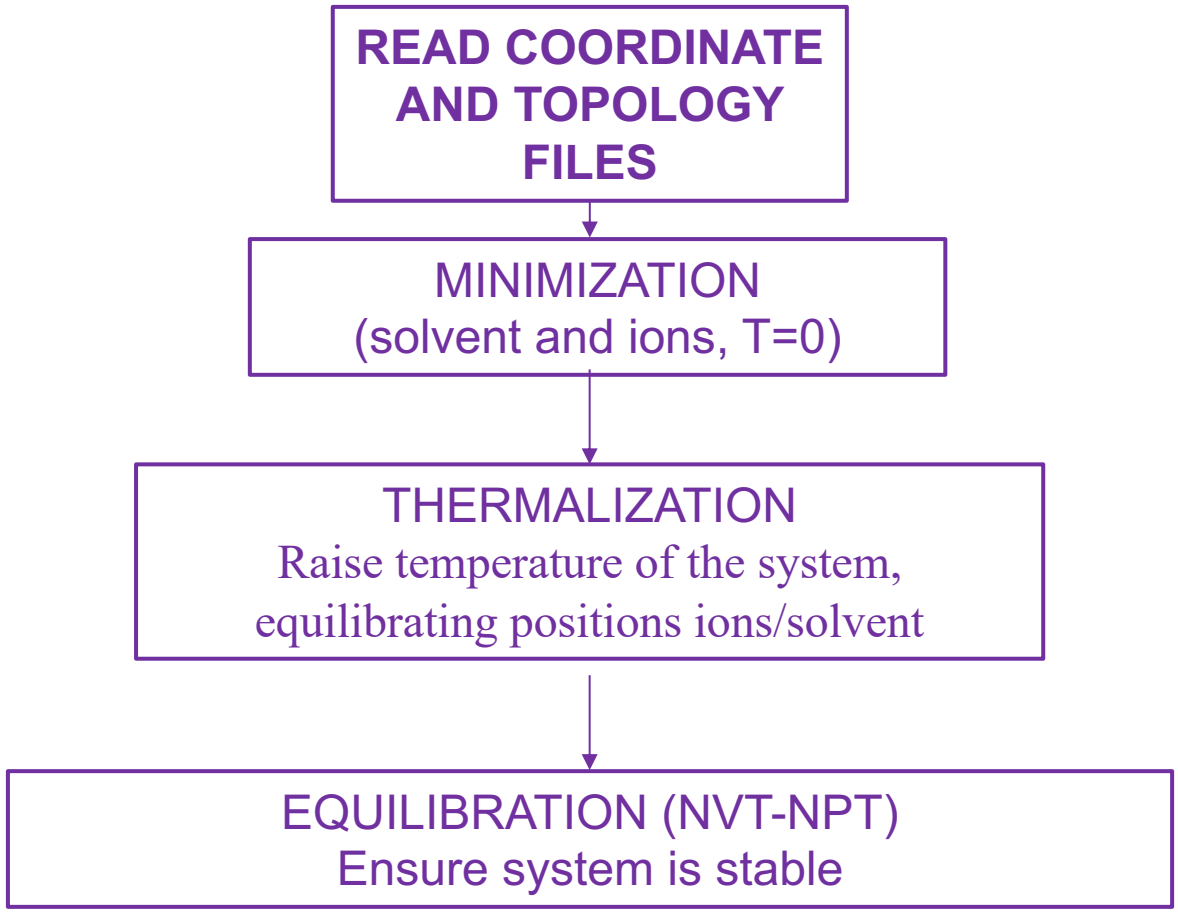


Steepest descent – relates to the gradient of the potential energy surface:
$F(r) = -dV(r)\ dr$
towards the local energy minima.

Minimize and Equilibrate the System

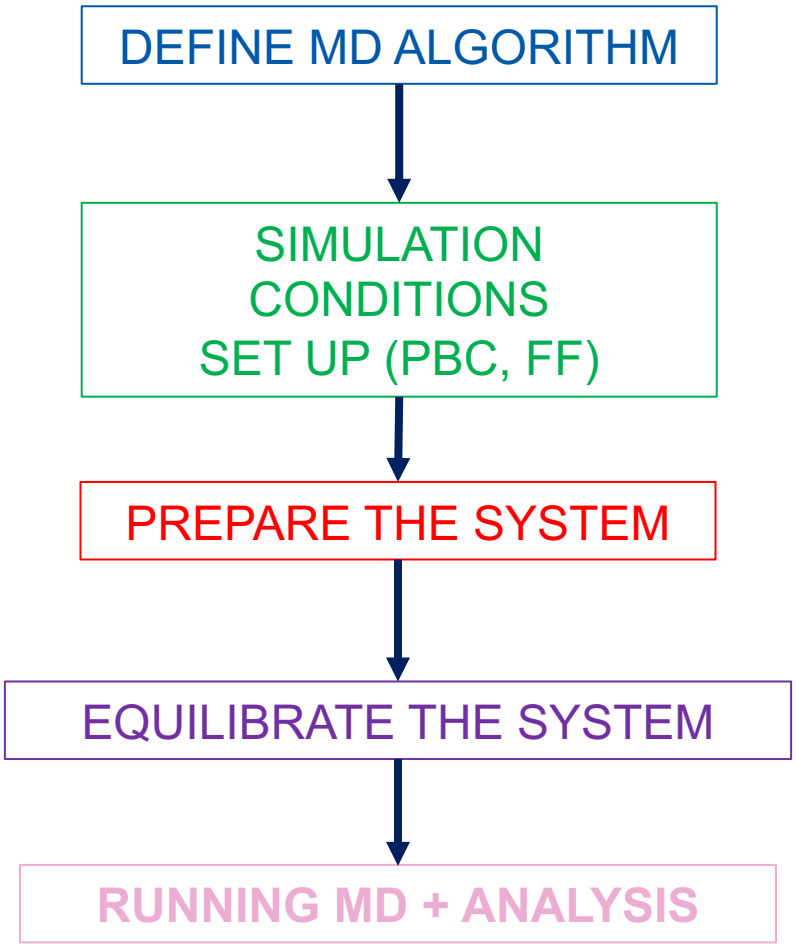


Minimize and Equilibrate the System



Several properties are monitored: pressure, temperature and energy
The purpose of this phase is to run the simulation until these properties become stable with time.

MD Simulation Step by Step



MD Production + Analysis

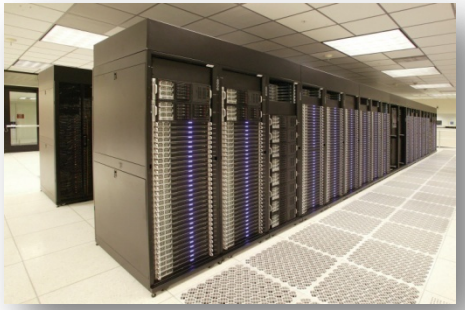
Production Time:

- Simulate under desired conditions (NVE, NpT, etc) for a time t
- Collect data (trajectory files with new positions, velocities and energy at each step)

Final Analysis:

- Evaluate observables (macroscopic properties, structural, energetics….)

MD Limitations



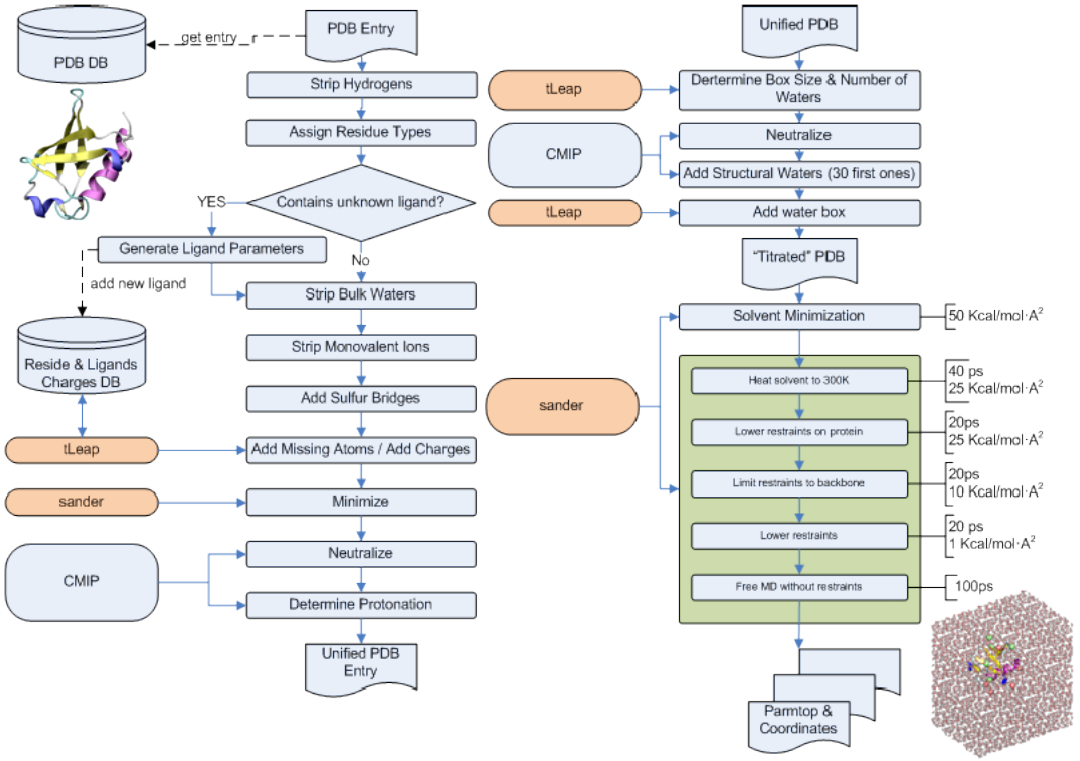
Large Computational Resources

$$V(\vec{\mathbf{r}}) = \sum_{\text{bonds}} {}^{(1)}K_b(b-b_0)^2 + \sum_{\text{angles}} {}^{(2)}K_\theta(\theta-\theta_o)^2 + \sum_{\text{dihedrals}} {}^{(3)}K_\chi(1+\cos(n\chi-\delta)) + \sum_{\text{nonbonded-pairs},i,j} \left[{}^{(4)}\frac{q_i q_j}{4\pi\varepsilon_o r_{ij}} - \varepsilon_{ij}\left\{ \left(\frac{R_{\min ij}}{r_{ij}}\right)^{12} - {}^{(5)}2\left(\frac{R_{\min ij}}{r_{ij}}\right)^{6} \right\} \right]$$

Energy dependencies on:

1. Bond length
2. Bond valence angle
3. Bond dihedral angle
4. Non-bonded electrostatic interactions
5. Non-bonded van-der Waals interactions

Force Fields Uncertainties



High level of expertise needed

Hospital et al, Adv Appl Bioinform Chem, 2015, 10:37-47 / Hospital A., Gelpí J.L. Wiley Interdisciplinary Reviews: Computational Molecular Science 2013, 3:364-377

MD Limitations:computational Resources

BSC Marenostrum Supercomputer

Graphical Processing Units (GPUs)

ANTON Supercomputer

GPUGRID.net

Folding@home distributed computing

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

EXCELENCIA SEVERO OCHOA

PRACE

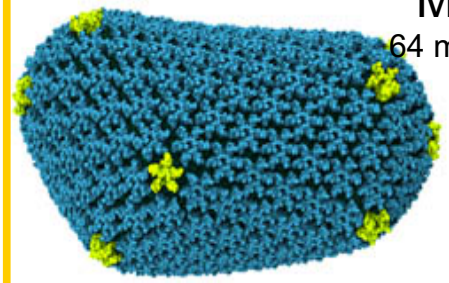
bioexcel

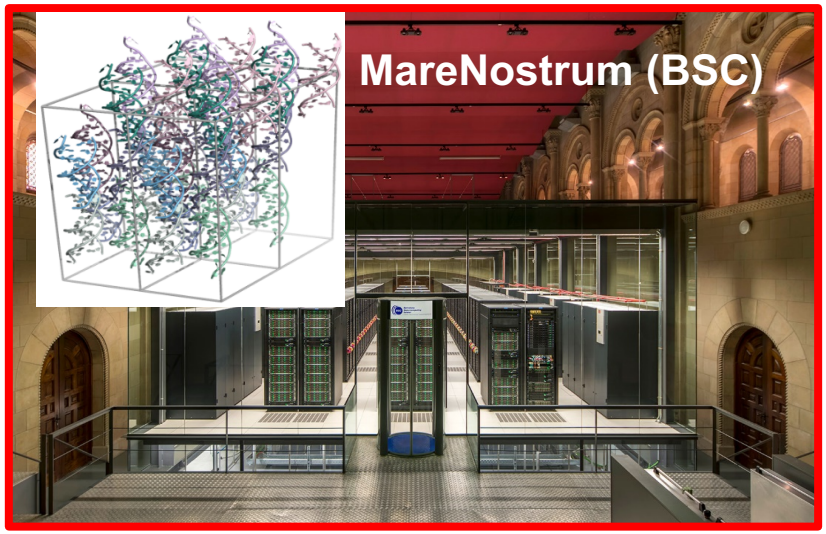
INB

MD Limitations:computational Resources

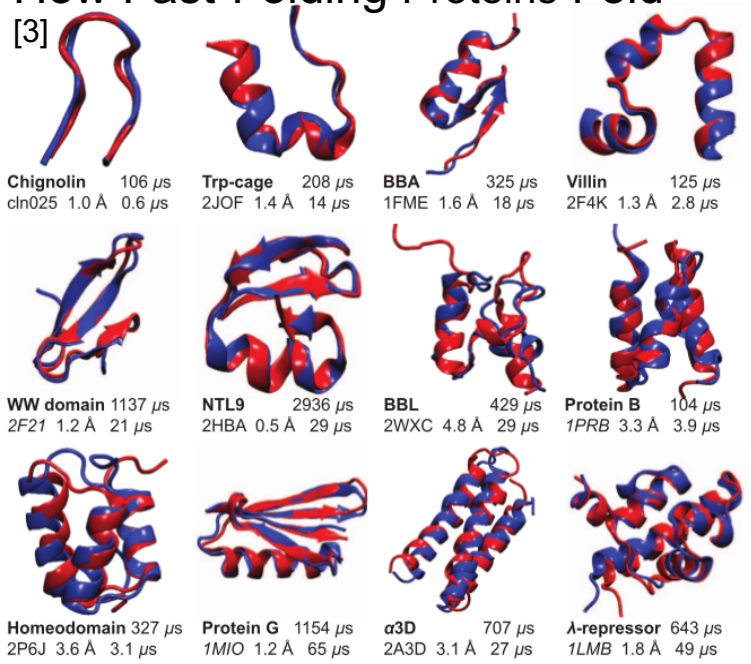
Mature HIV-1 capsid [1,2]

64 million atoms x 200ns / 35K CPU





How Fast-Folding Proteins Fold [3]



■ MD
■ Exp

[1] Zhao G,..., Schulten K,... *Nature.* **2013**,497:643-646. [2] Perilla JR, Schulten K. *Nat. Comm.* **2017** [3] Lindorff-Larsen K,...,Shaw DE. *Science.* **2011**,334:517-520.

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

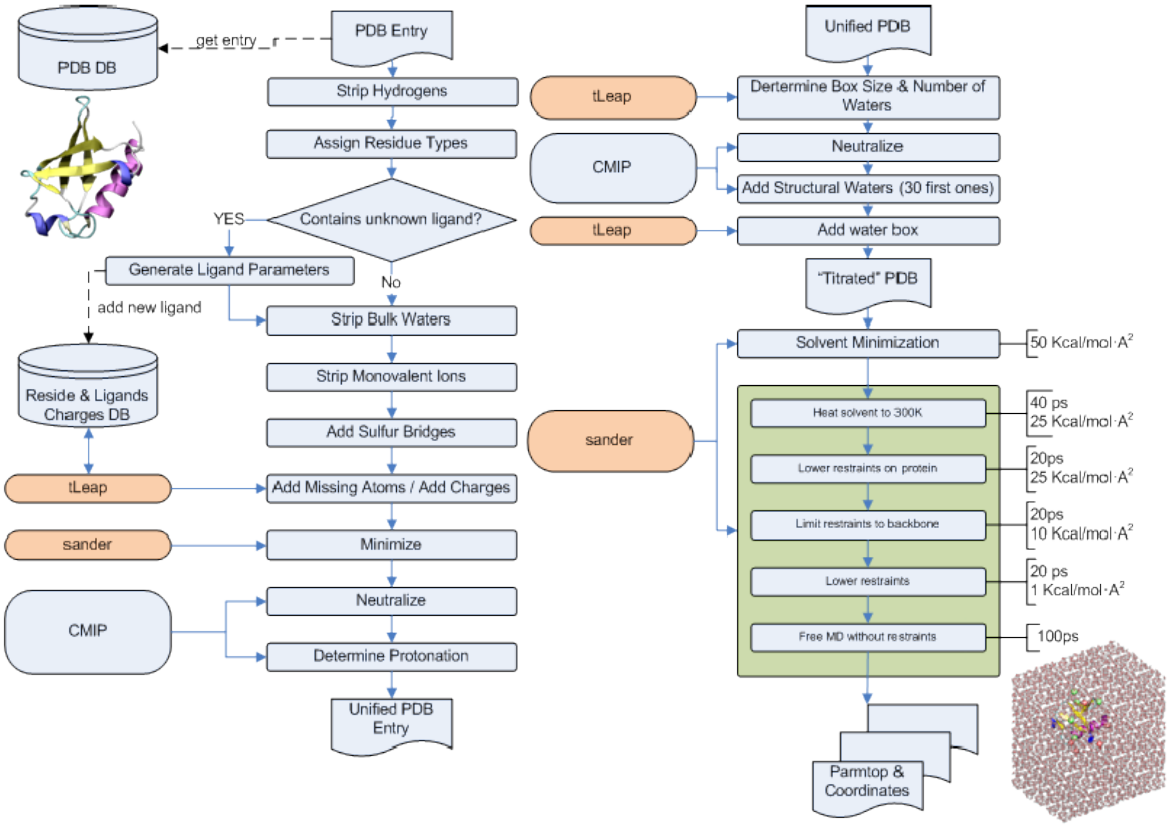
EXCELENCIA SEVERO OCHOA

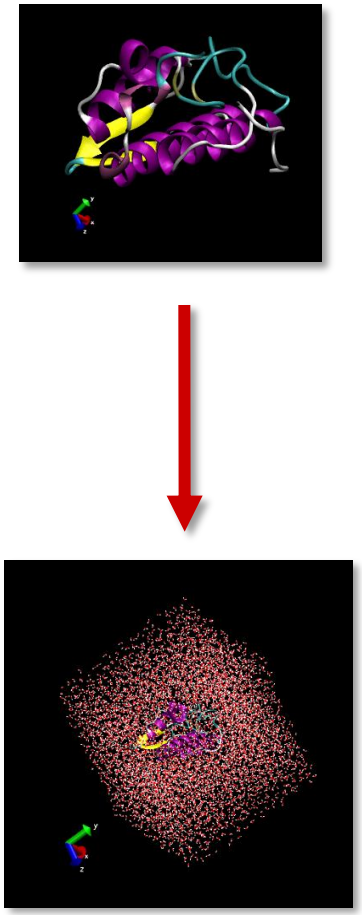
PRACE

bioexcel

INB

MD Limitations: Level of Expertise





A. Hospital et al, Bioinformatics 2012, 28 (9), 1278-1279.

MD Limitations: FF uncertainties

- **Reparameterization** of amber **force-field**.
  - Keeping α/γ **ParmBSC0** (2007) corrections.
  - Reparameterizing **sugar puckering, ε, ζ and χ torsions**.

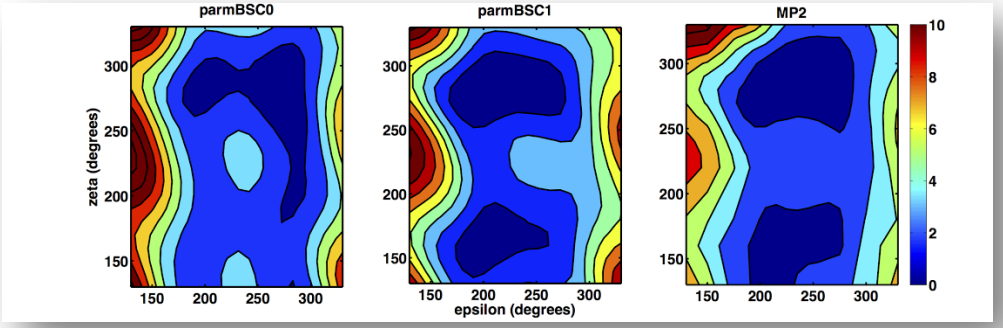
$$E_{bond} = \sum_{bonds} K_r (r - r_0)^2$$

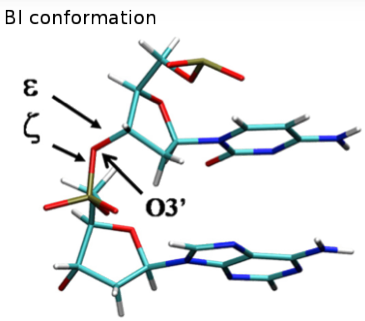
$$E_{angle} = \sum_{angles} K_\theta (\theta - \theta_0)^2$$

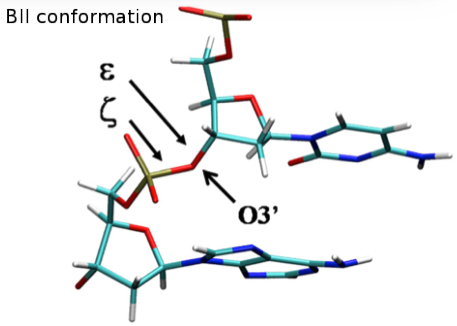
$$E_{dihedral} = \sum_{diheds} \sum_{n} \frac{V_n}{2} (1 + \cos(n\Phi - \gamma))$$

$$E_{vdw} = \sum_{i} \sum_{j>i} \left(A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-6} \right)$$

$$E_{elec} = \sum_{i} \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$







ParmBSC0: A. Pérez et al. Biophys J. (2007) 92 (11), 3817-3829.

ParmBSC1: I.Ivani et al. Nature Methods (2016) , 13 (1), 55-58.

Useful links: Simulation Setup - Tutorials

Amber

http://sf.anu.edu.au/collaborations/amber_on_fujitsu/amber-12/tutorial/standard-setup/index.html

http://ambermd.org/tutorials/

NAMD
Scalable Molecular Dynamics

http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-win-html/index.html

http://www.ks.uiuc.edu/Training/Tutorials/

GROMACS
Groningen Machine for Chemical Simulations

http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/lysozyme/index.html

http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/

http://www.gromacs.org/Documentation/Tutorials

Useful links: http://ask.bioexcel.eu/

bioexcel

all categories | Categories | Latest | New | Unread | Top | + New Topic

Category	Latest	Topics
HADDOCK The HADDOCK category is meant for discussing any topic related to the use of the HADDOCK software, either as a local installation or via the HADDOCK web portal. For details about HADDOCK please refer to http://www.bonvinlab.org/software/haddock2.2 Feel free to create new topics related to your quest…	Protein exceeds max number of atoms 10d Secondary structure prediction 23d Scoring interface histidine 23d Multibody docking active sites? 23d Violations analysis 24d Different running time using same partners 24d Different violation results in ana_noe_viol_all.list and ana_noe_viol_ambig.lis 27d Active residues 27d BioExcel webinar: Robust solutions for cryoEM fitting and visualisation of interaction space Feb 7 DNA three letter code Feb 3	6 / month 75 / year
GROMACS GROMACS is a free and open-source software suite for high-performance molecular dynamics and output analysis. It primarily targets bio-molecular simulations of thousands to millions of atoms, described by particle-based classical mechanics force fields. These simulations integrate the Newtonian equa…	Moving atoms between repartition and force evaluation 2d BioExcel Webinar #2: Performance tuning and optimization of GROMACS webinar May '16 AMBER force field Sep '16 Regarding minimization of edited structures Jul '16	1 / week 1 / month
CPMD The CPMD category is meant for discussing topics related to the use of the CPMD software in biological oriented applications, and in particular its QM/MM interface	QMMM with other solvent like alcohol Jul '16	1 / year
DISVIS This is the help forum for DISVIS, a software and web portal that allows you to visualizes and quantify the information content of distance restraints between macromolecular complexes	BioExcel Webinar: Robust solutions for cryoEM fitting and visualisation of interaction space (2017-02-15) Feb 7 DisVis web portal 2.0 Dec '16 Release of the DISVIS web portal Sep '16	3 / year