Studying complex macromolecules

M2 SERP-Chem
Fabien Cailliez
LCP – Bât 349
fabien.cailliez@u-psud.fr

Outline

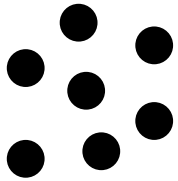
I. Classical forcefields

II. Molecular simulations and biomolecules

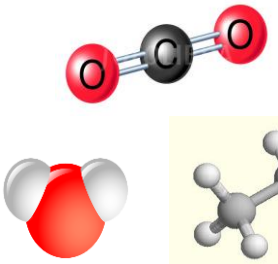
III. One typical example: Study of cadherins

IV. Some challenges in molecular modelling

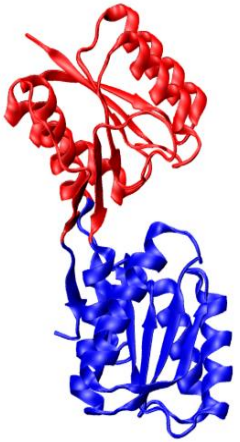
Complex molecules



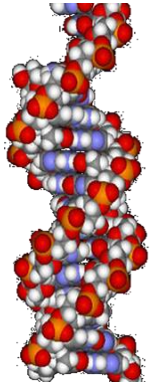
Lennard-Jones particles (rare gases)



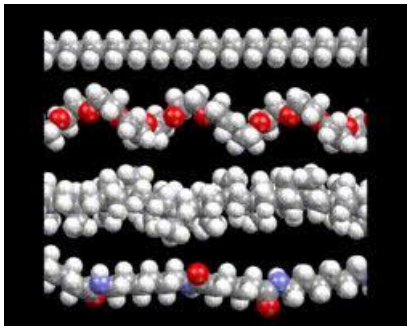
Small molecules (mainly rigid)

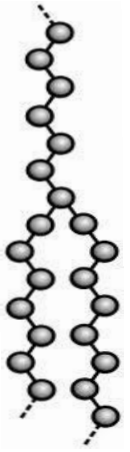


Proteins



DNA





Organic polymers

Increase in molecular complexity

The “problem” of complex molecules

- Large number of atoms
  - Time consuming
- Internal flexibility and many stable conformations
  - How to find the stable conformations?
  - Which is the most stable?
  - Are all the conformations important?
- Stability of the molecules strongly depend on solvation:
  - Use of solvent molecules in the simulation (more computation time)

Necessity of a cheap representation of energy

Energy representation

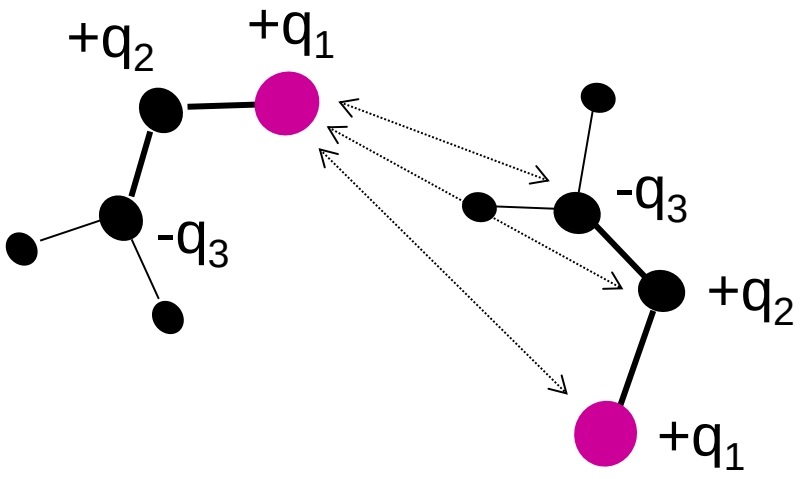
Energy representation	Representation level	System size	Time scale
Quantum Mechanics	Nuclei, electrons	N < 100 atoms	Up to few picoseconds
Classical Mechanics	Atoms	N < 10^6 atoms	Nanoseconds to microseconds

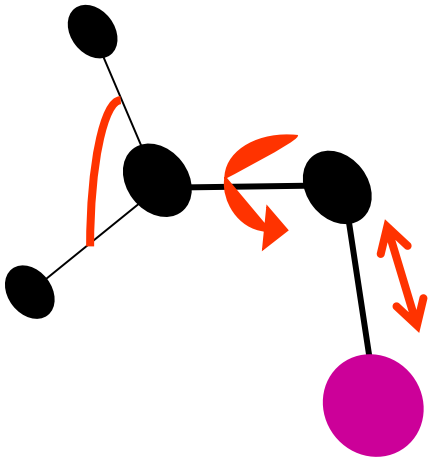
Necessity to use a classical representation of the energy: forcefields

Forcefield for complex molecules

Different contributions to the energy

Molecule = partial charges (atomes) interconnected (bonds)





$$E = E_{electrostatic} + E_{dispersion-repulsion} + E_{bond} + E_{angle} + E_{torsion}$$

« non-bonded » interactions
Intermolecular Forces

Deformation modes
Intramolecular energy

Intermolecular forces and «non-bonded» interactions

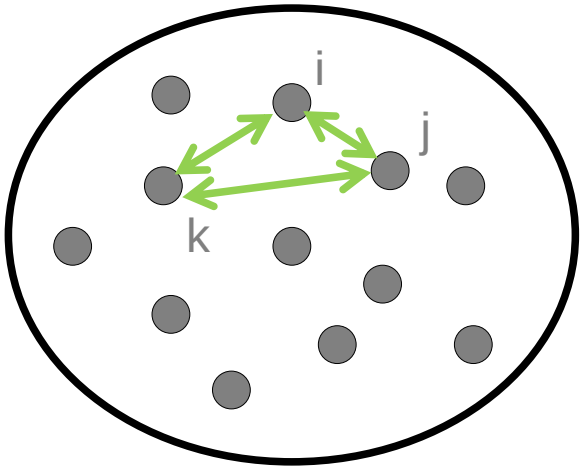
- Intermolecular foces are responsible for:
  - Cohesion of matter
  - Folding of complex macromolecules (polymers)

- Weak forces with respect to covalent interactions (energies ~ few $kJ.mol^{-1}$)

- Principal intermolecular forces:
  - Electrostatic interactions between charges
  - Hydrogen bonds
  - van der Waals interactions
  - Repulsive forces

Pair potentials for intermolecular forces

- Mathematical expression of the potential energy for one configuration:



$$E(ijk) = V_{ij}(q_i, q_j) + V_{ik}(q_i, q_k) + V_{jk}(q_j, q_k) + V_{ijk}(q_i, q_j, q_k)$$

$$E = \sum_{i<j} E_{ij} + \sum_{i<j<k} E_{ijk} + \ldots$$

Two-body interaction term
(~N^2 terms)

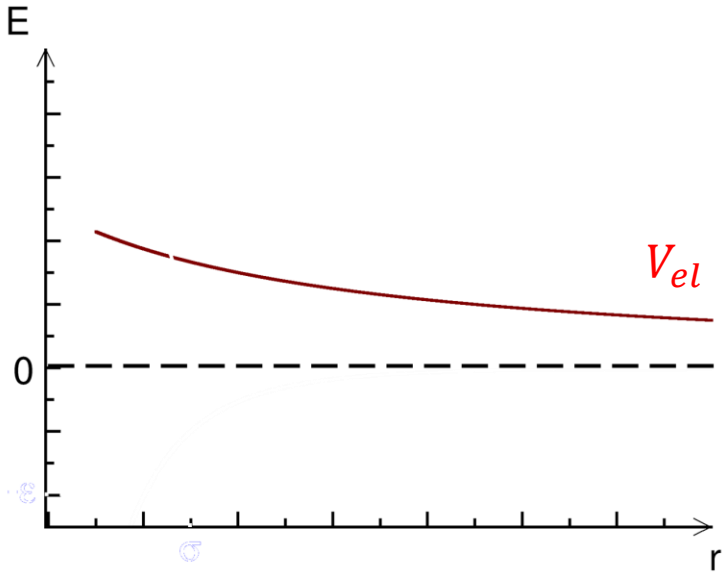
Three-body interaction terms
(~N^3 terms)

- The two-body terms have the greatest contribution

In the majority of forcefields, pair potentials are considered.
N-body interactions are introduced in an effective way

Non-bonded interactions

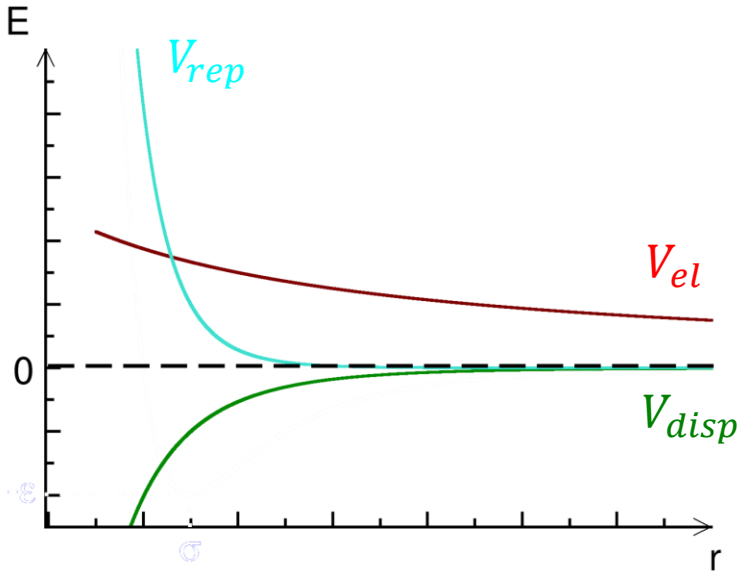
- Electrostatic interactions :
  - Electronic and nucleic density on atoms= partial charges
  - Coulombic potential: $V_{el}(r) = \frac{q_i q_j}{4\pi\varepsilon_0 r}$
  - Long-range force



- Hydrogen bonds:
  - Short-range directional interaction
  - Energy range: 15-20 kJ.mol^{-1}
  - Origin: electrostatic and orbital overlapping
  - Sometimes a specific term is added to the forcefield

Non-bonded interactions

- Dispersion interactions:
  - Attractive dipole-dipole interactions
  - One of the ven der Waals interactions:
    - Keesom
    - Debye
    - London (dispersion) $V_{disp}(r) = -\frac{C}{r^6}$
  - Short range: few Å
  - Energy range: ~1-10 kJ.mol^{-1}



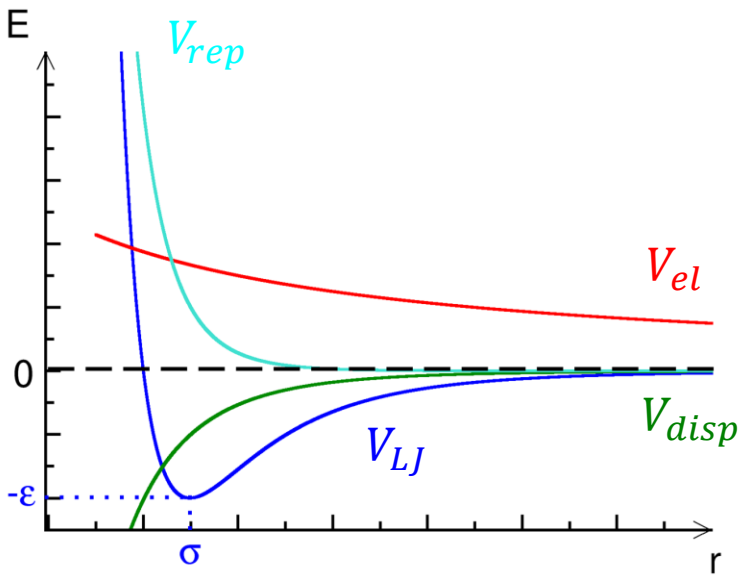
- Repulsive interactions:
  - Quantum origin
  - Very short range
  - Empirical expression: $V_{rep}(r) = \frac{A}{r^{12}}$ or $V_{rep}(r) = A\exp\left(-\frac{B}{r}\right)$

Non-bonded interactions

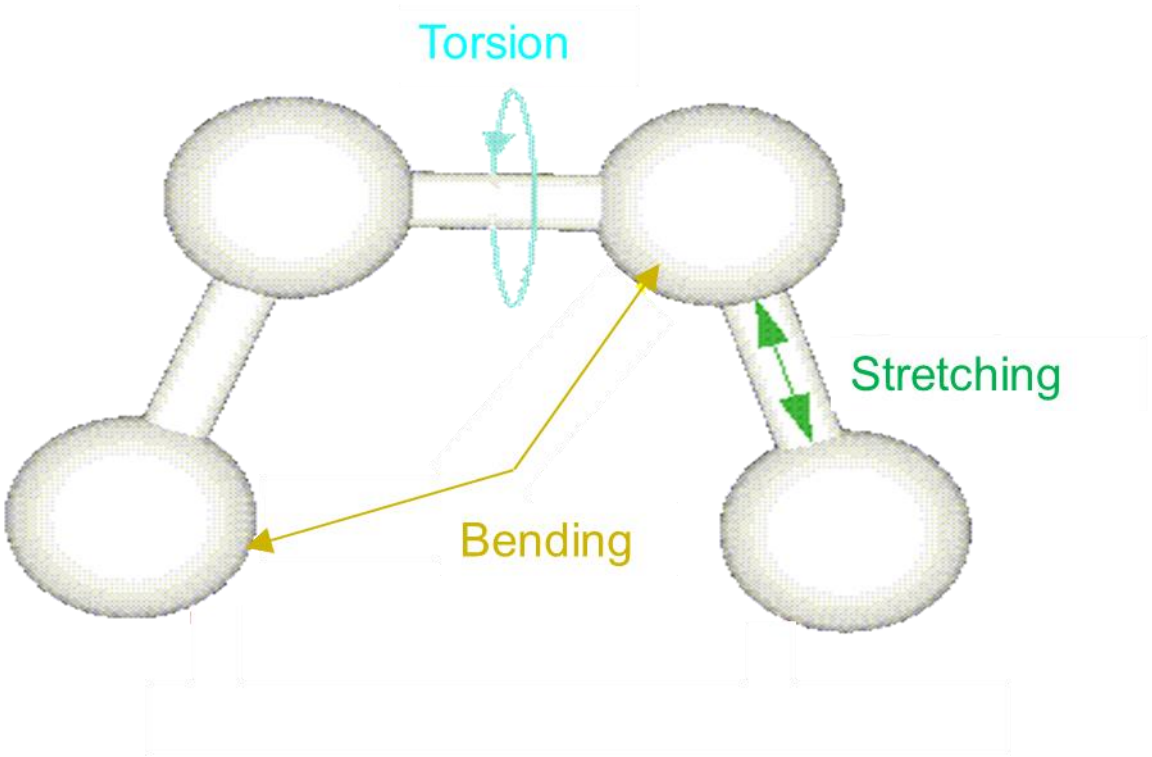
- Dispersion-repulsion interactions gathered in one expression:

$$V_{LJ}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

$$V_{Buck}(r) = \alpha \exp\left(-\frac{\beta}{r}\right) - \frac{\gamma}{r^6}$$

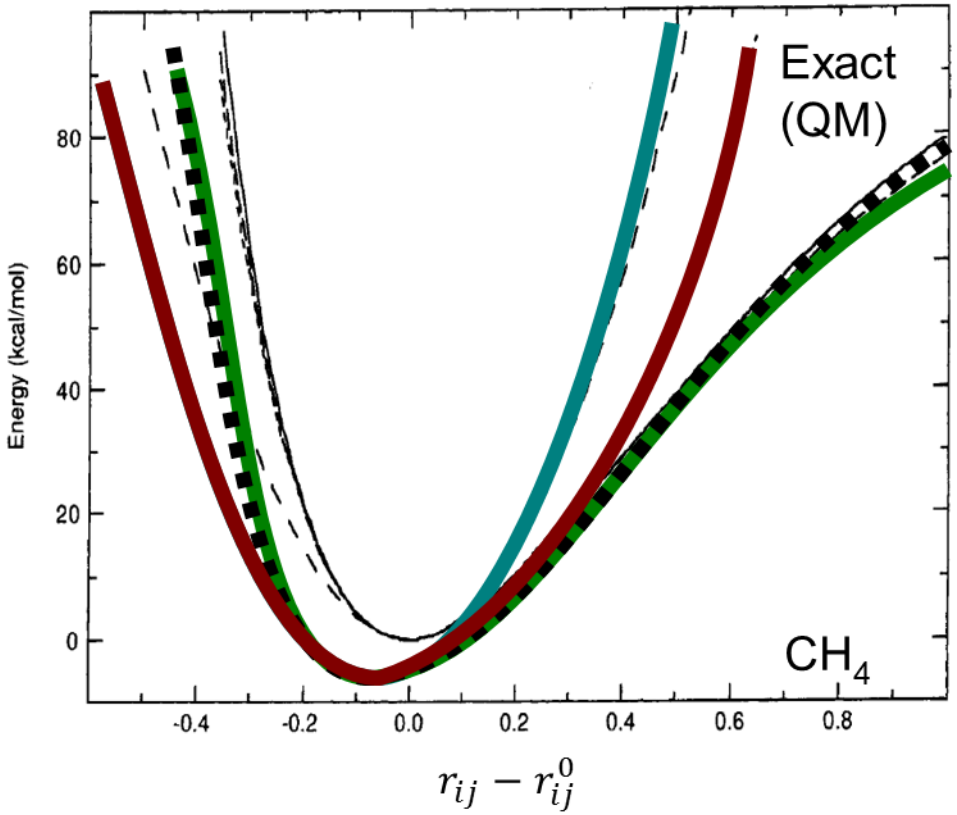


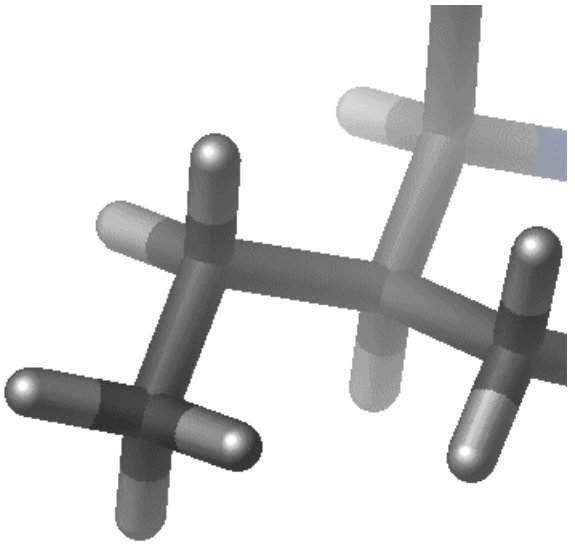
Intramolecular potentials



- Intramolecular potentials depend on functional groups
- Fundamental concept of forcefields: atom types
- Atom type:
  description with a certain parameter set of an atom included in a specific chemical environnement

Intramolecular potential: bond-stretching





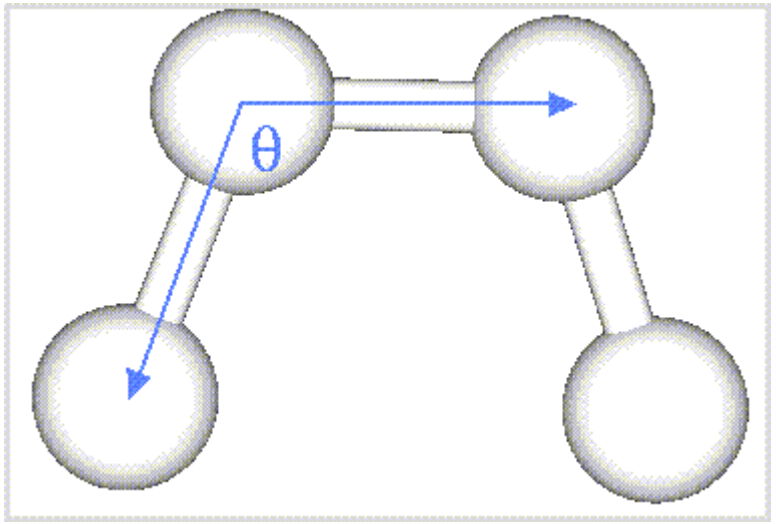
$$E_{Morse} = D\left[1 - \exp\left(\sqrt{\frac{k}{2D}}(r_{ij} - r_{ij}^0)\right)\right]$$

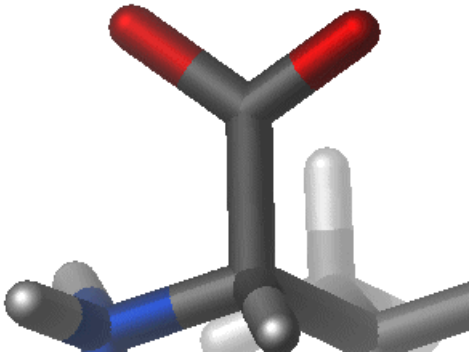
$$E_{P4} = k_{ij,2}(r_{ij} - r_{ij}^0)^2 + k_{ij,3}(r_{ij} - r_{ij}^0)^3 + k_{ij,4}(r_{ij} - r_{ij}^0)^4$$

$$E_{P2} = k_{ij}(r_{ij} - r_{ij}^0)^2 \longleftarrow E_{bond}$$

- Morse potential more precise, but:
  - Computationnaly more expensive
  - Energy minimum well described with a quadratic expression
- In practice: harmonic approximation
- Two parameters k_{ij} et r_{ij}^0 which depend on the bond type

Intramolecular potential: angle bending



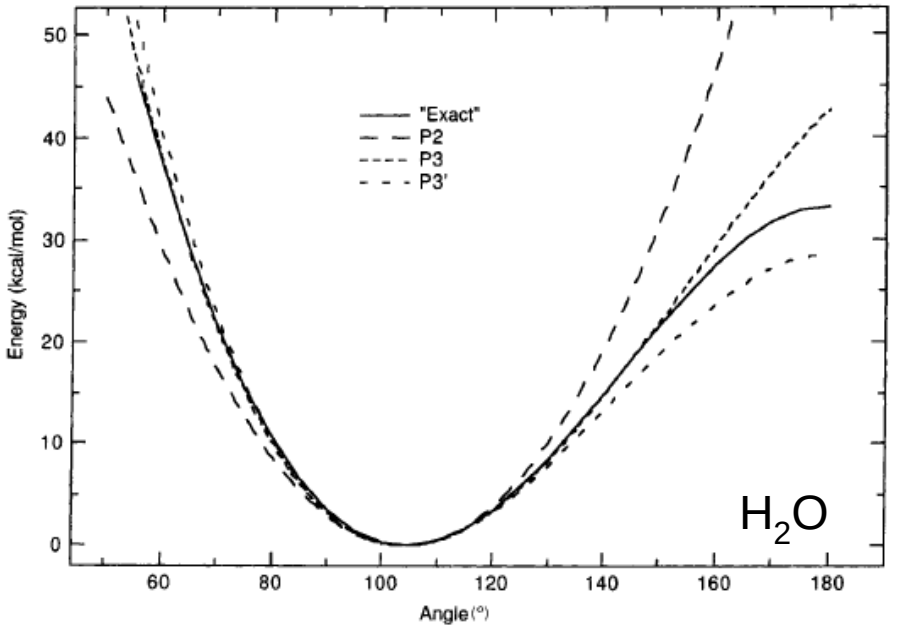


- Polynomail decomposition
- Harmonic potential:

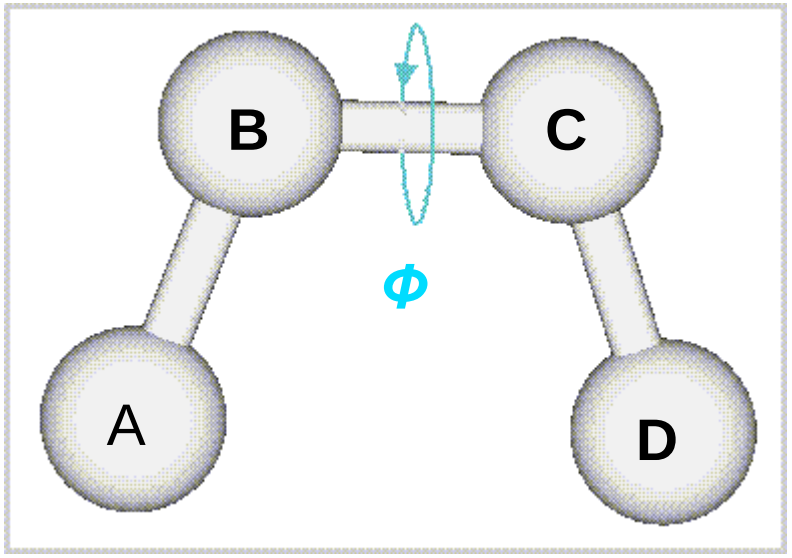
$$E_{bend} = k_{\theta,ij}\left(\theta_{ij} - \theta_{ij}^{0}\right)^2$$

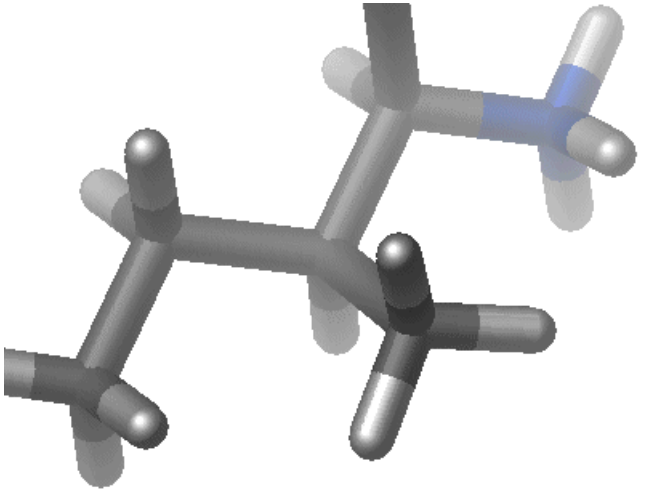
- Two parameters:

$$k\_(\theta, ij) \quad , \quad \theta\_ij\text{^}0$$



Intramolecular potential: torsional energy



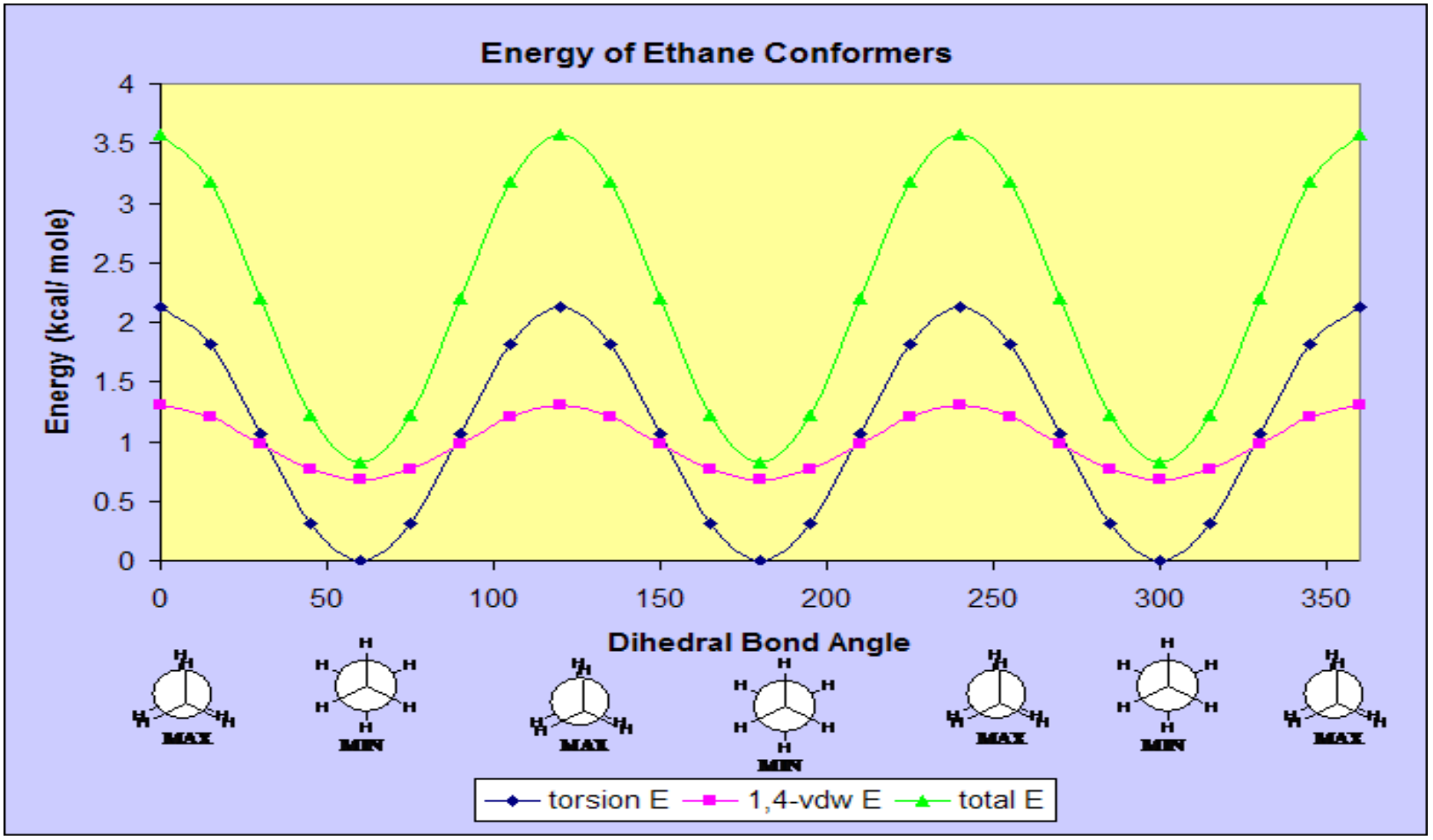


- Torsion energy: steric interaction resulting from rotation around the bond B-C in a sequence A-B-C-D.
- Periodic function → decomposition in sum of sinuosidal terms:

$$E_{torsion} = \sum_n \frac{V_n}{2}[1 + \cos(n\phi - \delta)]$$

- This is rather a correction to non-bonded interactions between A and D than a real physical interaction

Intramolecular potential: torsional energy



Molecular mechanics energy significance

$$E = E_{electrostatic} + E_{dispersion-repulsion} + E_{bond} + E_{angle} + E_{torsion}$$

- Value of the energy calculated with respect to a reference where all geometrical values are equal to their reference value
- Absolute MM energy has no significance (unlike quantum mechanics energy)
- Only differences of MM energies are relevant

Forcefield parameters calibration

- Forcefield are parameterised to reproduce:
  - Structural data (bond lengths, valence angles)
  - Thermodynamic data (densities, phase equilibriums, solvation enthalpies)
  - Results from quantum mechanics calculations (electrostatic potential map for example) or experimental results
  - Minimisation of a scoring function:

$$\chi^2 = \sum_i \frac{\left(Y_i - Y_{i,ref}\right)^2}{u_i^2 + u_{i,ref}^2}$$

$Y_{i,ref}$ Reference data

u_i , $u_{i,ref}$ Uncertainties

- Very laborious task because:
  - Many parameters that can interact
  - Necessary to make compromises

Calibration strategy for biomolecular forcefield

$$E = E_{electrostatic} + E_{dispersion-repulsion} + E_{bond} + E_{angle} + E_{torsion}$$

- Equilibrium value of bond lengths and angles fixed to crystallographic data
- Constant forces for bond and angle harmonic potentials fitted on IR spectroscopic data
- Partial atomic charges are fitted to reproduce electrostatic potential calculated with quantum mechanics
- van der Waals interaction parameters are fixed to “realistic” values
- Torsional parameters are fitted to reproduce quantum calculations
- At any stage, reevaluation of parameters may be done

Transferability of a forcefield

- Forcefields are based on the definition of atom types
  - One atom type = one set of parameters
  - Increasing the number of atom types allow to better reproduce a set of calibration properties
- Transferability:
  - Between molecules: ability to transfer parameters to different molecules
  - Between properties: ability of a forcefield to reproduce properties that were not used in the calibration procedure

Linked to the predictivity capacities of the forcefield

- Compromise to be found between number of atom types and transferability

Examples of force fields

Force field	Application	Software
MM2, MM3, MM4 (Allinger)	Purely steric Small molecules	Macromodel, Chem-3D, Hyperchem
Tripos	Adapted to large molecules	Alchemy, SYBYL
Amber	Proteins, Nucleic acids	Amber, Hyperchem, NAMD
Charmm	Proteins, Nucleic acids	Charmm, NAMD
Biosym	Proteins, Organic molecules	Discover
OPLS	Organic molecules	

Advantages and limitations of forcefields

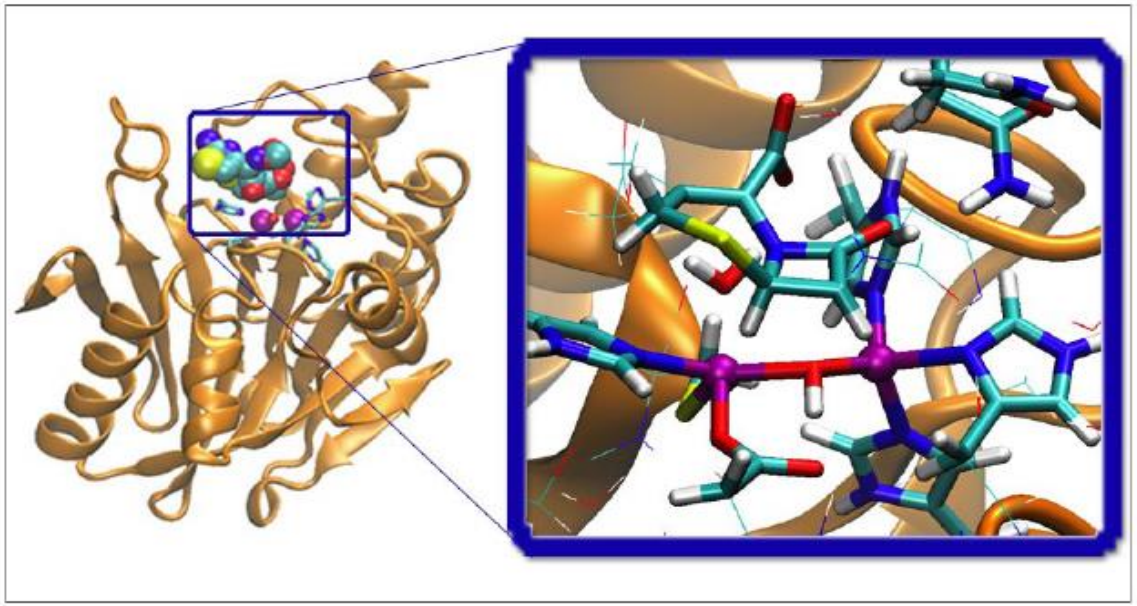
- Advantages:
  - Reduced computational time
  - Possibility to study complex systems
  - Time scale accessible more important than with quantum mechanics

- Limitations:
  - Electrons are only considered implicitly
  - No bond breaking or excited state studies
  - Polarisability not taken into account in the majority of forcefields

- How to circumvent the forcefields limitations?
  - Polarisable or reactive forcefields (computationaly more expensive)
  - Hybrid QM/MM approaches

Hybrid QM/MM approaches

- In a single simulation:
  - A small part of the system is treated with quantum mechanics
  - The rest of the system is treated with a classical forcefield

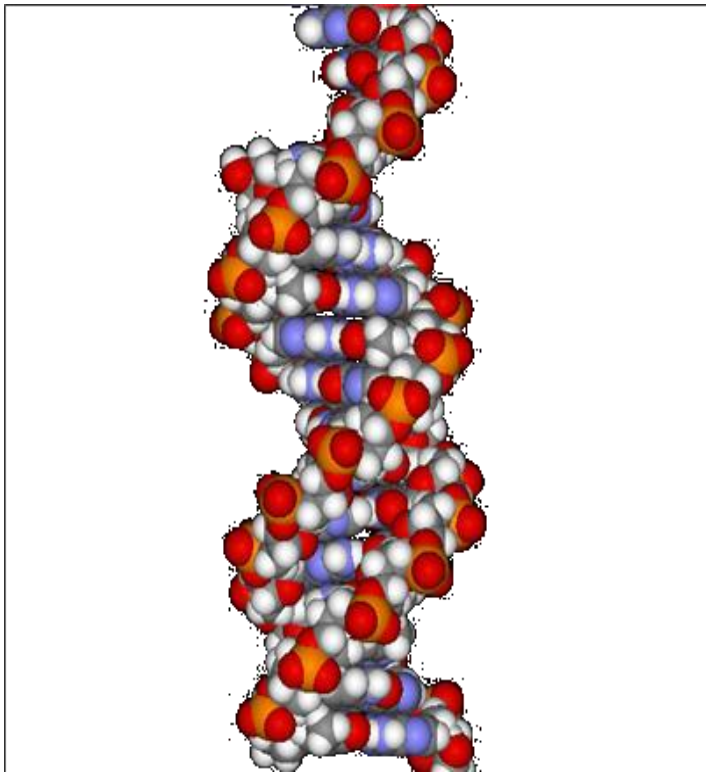


- Bond breaking/formation is possible in the QM part
- The global dynamics of the system and molecular environment of the QM part is preserved thanks to the MM part

Molecular simulations and biomolecules

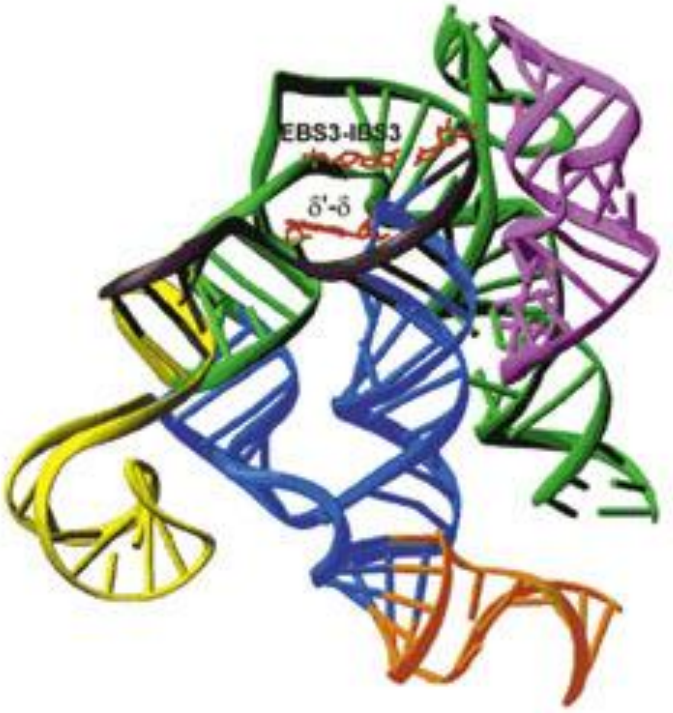
Biomolecules: nucleic acids

DNA



- A, T, C, G
- Double-stranded structure

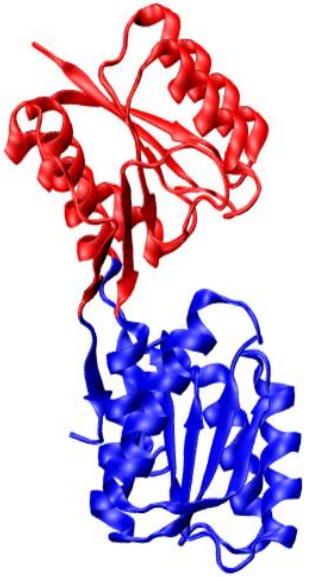
RNA



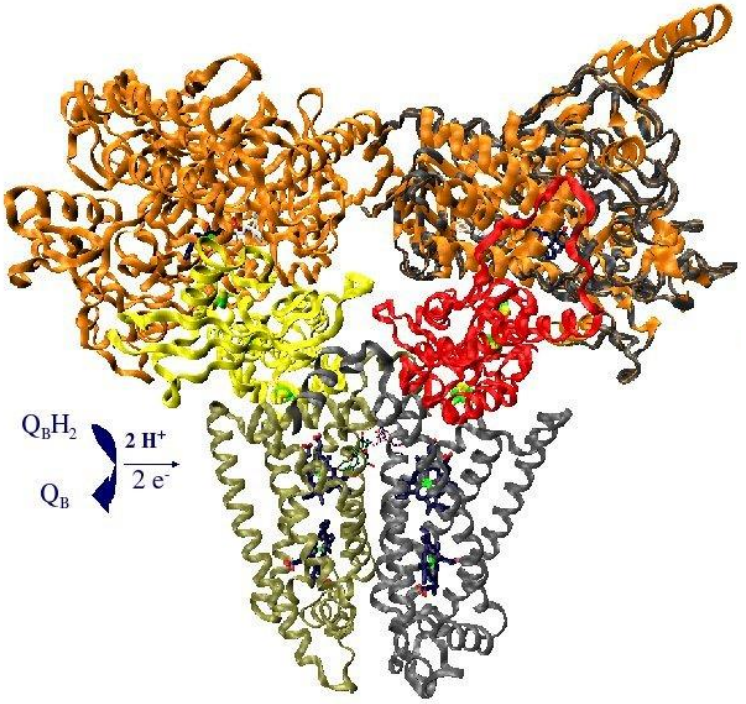
- A, U, C, G
- Single-strand
- Complex folding

Biomolecules: proteins

Fibronectin



D-ribose Binding Protein



Fumarate reductase

- 20 amino acids
- Complex folding
- Secondary structures (helices, strands,...)

PDB : http://www.rcsb.org/pdb/home/home.do

Molecular modeling in biology

- Refinement of experimental structural data
  - Conformational exploration
- Study of systems at equilibrium (structural fluctuations, thermodynamic properties)
  - Relative flexibility of the different parts of the molecule (enzyme active sites for example)
  - Computation of binding affinities (docking)
- Study of the dynamics of the system evolution (folding, movement linked to function)
  - Temporal and/or directional moves

Few numerical data

- System sizes:
  - Proteins: 50 to thousands amino acids (~500 to 10000 atoms)
  - DNA: ~800 atoms for a decamer (10 successive bases)

  Extensive quantum mechanics calculations untractable

- Timescales of biological phenomena:
  - Atomic fluctuations: fs-ps
  - Ligand biding: ns – ms
  - Movement of large proteic domains: ms – s
  - Protein folding: s-h

  Timescales often incompatible with available computational power

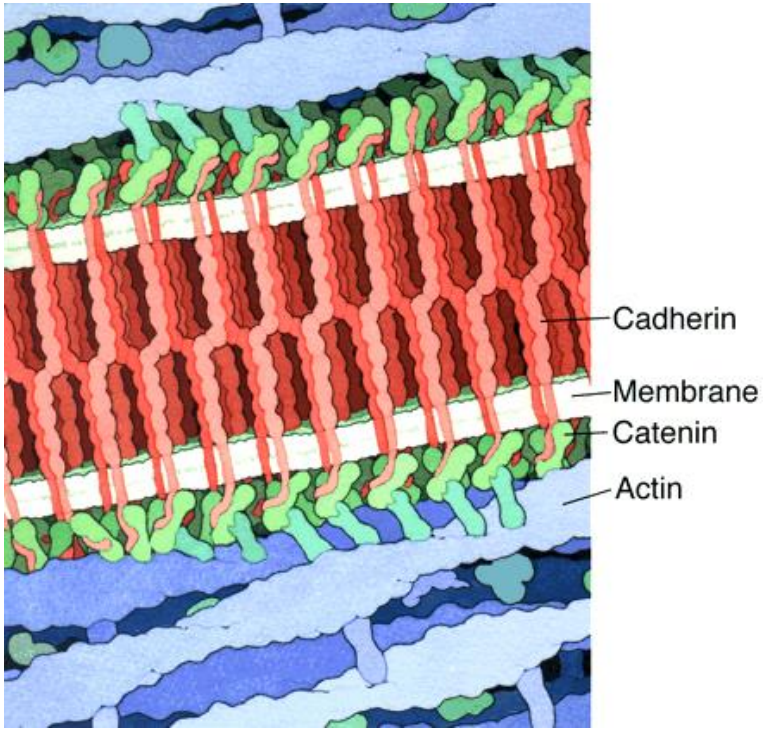
Simulation methods

- Energy minimisation (Molecular Mechanics MM)
  - Determination of stable conformations
  - Preparation of the system for further simulations
- Molecular Dynamics (MD)
  - Conformational exploration
  - Dynamics of protein moves
- Monte Carlo (MC)
  - Conformational exploration
- Normal Mode Analysis (NMA)
  - Collective motion of protein atoms

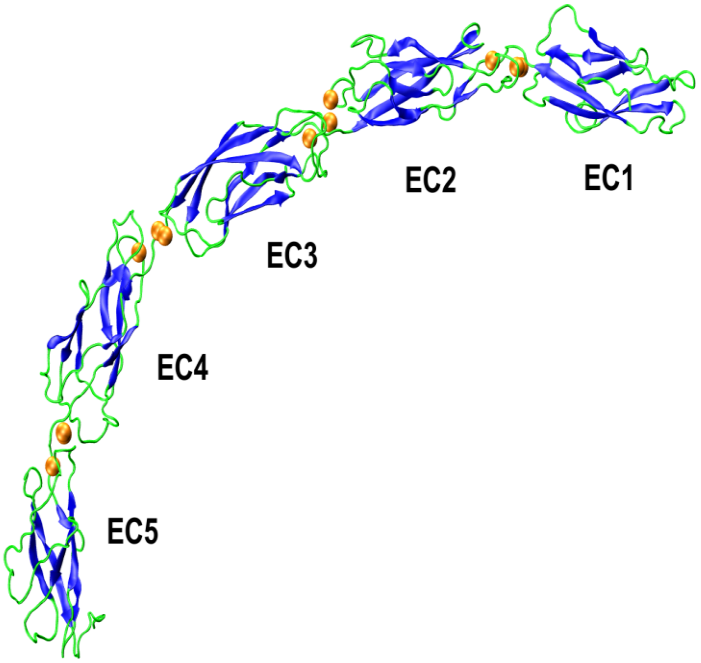
One typical example: Study of cadherins

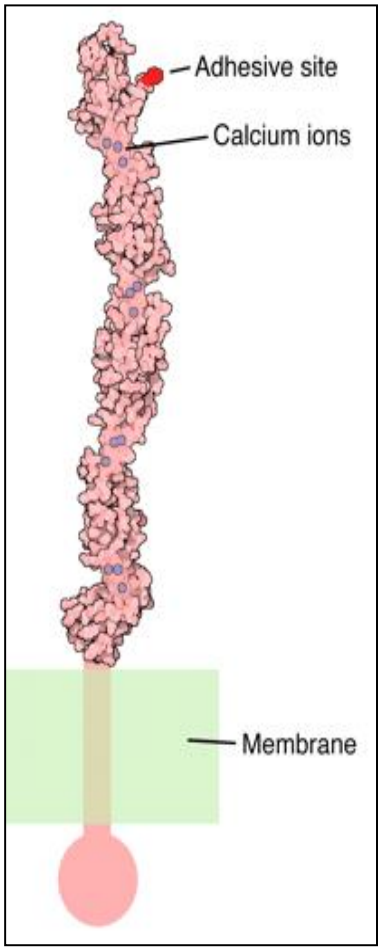
Example of a MD study: cadherins

- Transmembrane proteins implied in adhesion phenomena between cells
- Activity regulated by the presence of calcium ions
- 3 ions bound between two successive EC domains in X-ray structures



Goodsell, D.S. – The oncologist 2002; 7 : 467-468

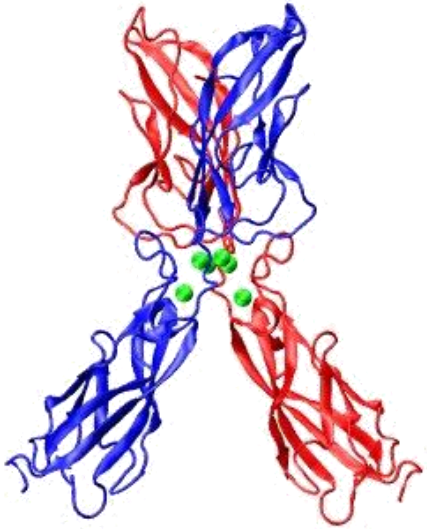




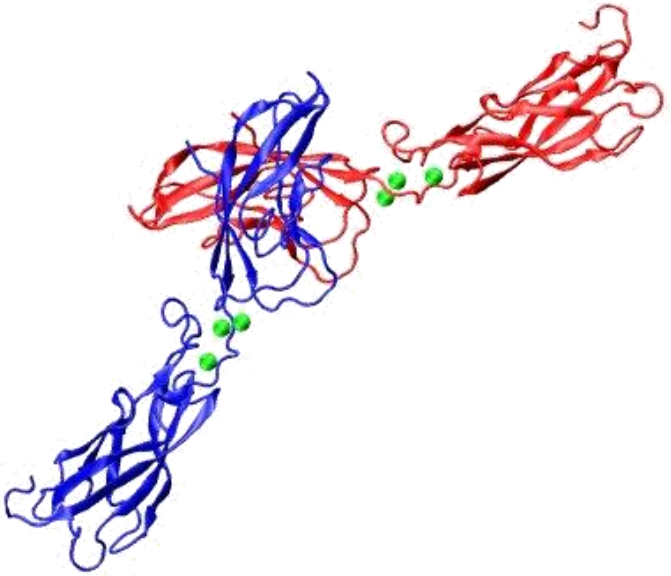
Goodsell, D.S.
The oncologist
2002; 7 : 467-468

Questions and aims of the study

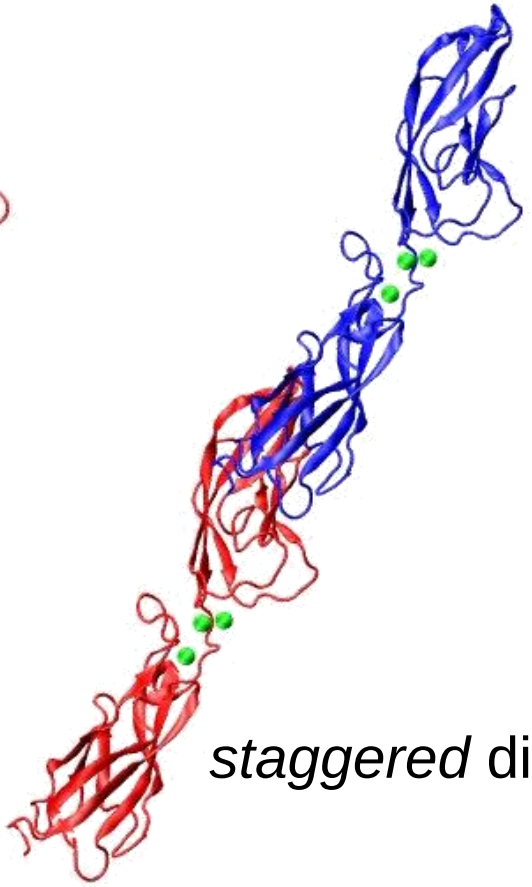
- Understanding the role of the calcium ions in the structure of the cadherin molecule
- What is the structure of the cadherin dimer implied in the adhesion process?



cis-dimer



trans-dimer



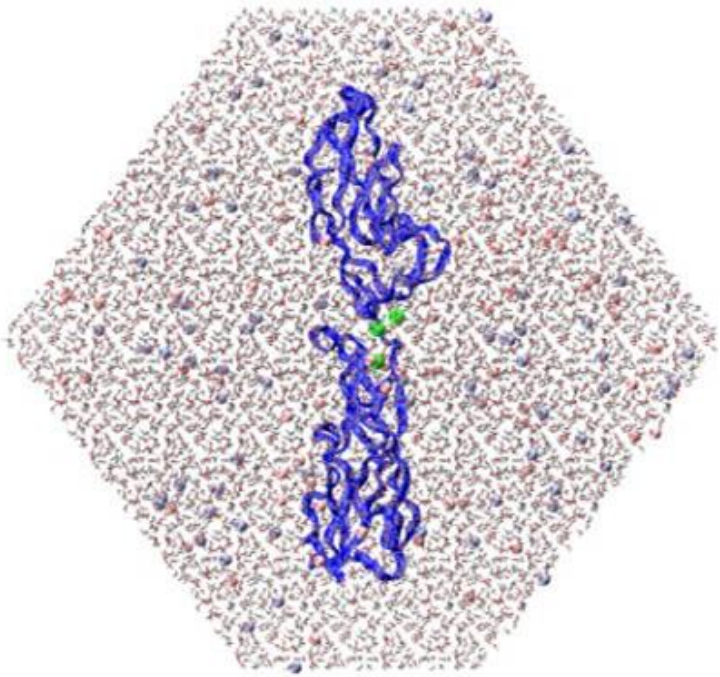
staggered dimer

Protocol of the study

- Setting up the system under study
- Equilibration and production phase
- Analysis and results

Setting up of the system

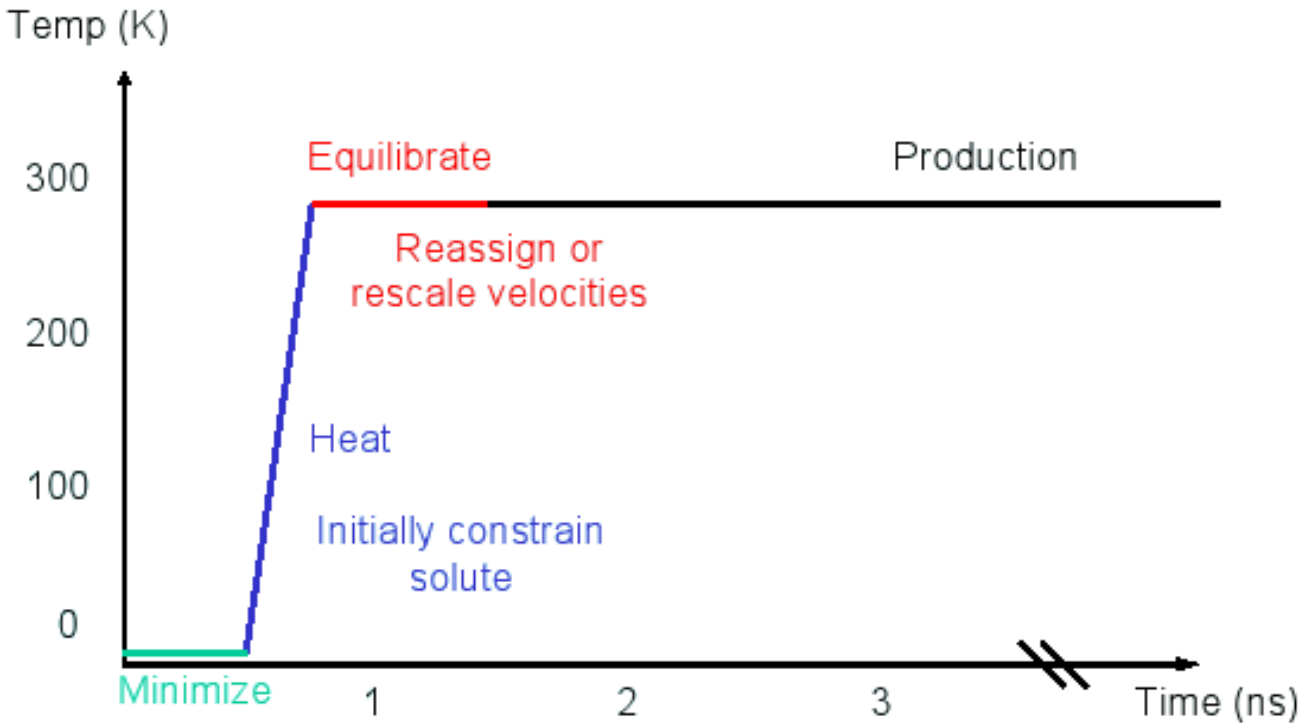
- Initial structure
  - X-ray or NMR structure (from the PDB)
  - Computation of the protonation state
  - ~ 3000 atoms



- Creation of a surrounding water box
  - adapted to the global shape of the protein
  - ~ 80000 atoms
- Adding counter-ions to neutralise the protein charge
- Choice of the forcefield and of the simulation conditions (Temperature, Pressure, cutoffs,...)

Equilibration phase

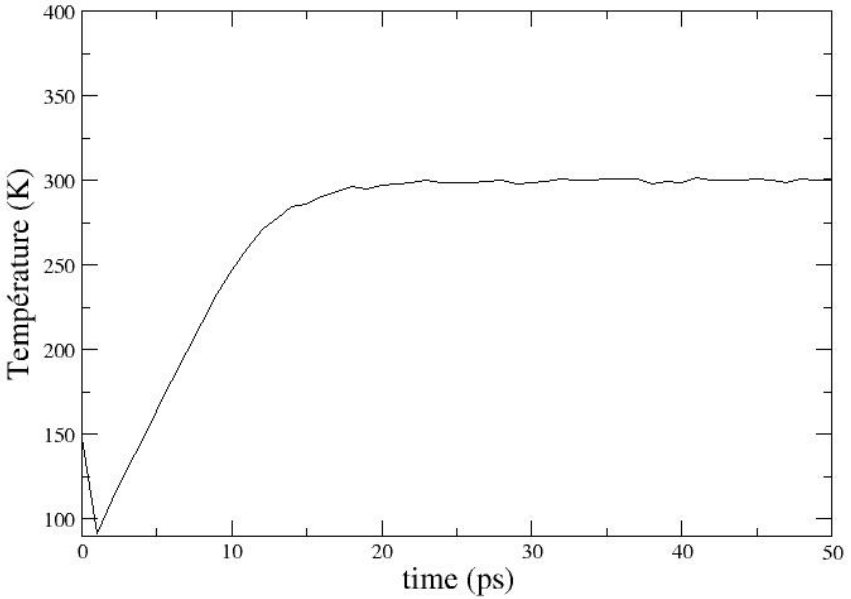
- Adaptation of the solvent molecules to the protein (energy minimisation under constraint)
- Relaxing the constraints on the protein atoms and progressive increase of the temperature

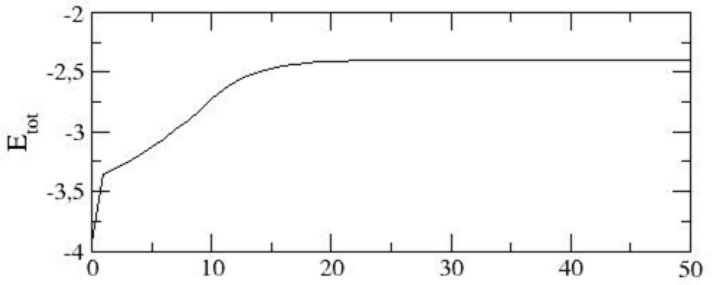


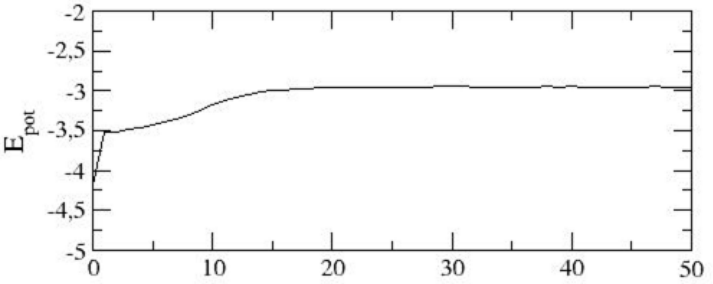
Equilibration phase – Analysis

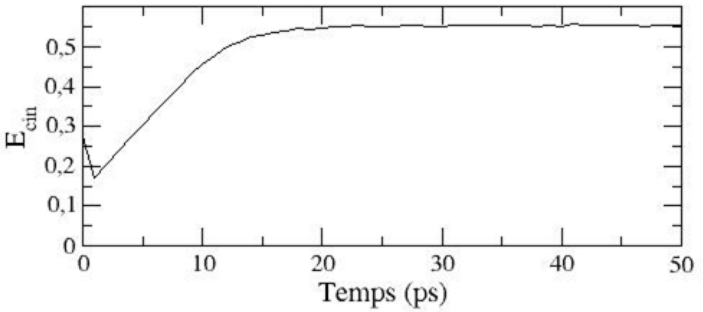
- Verification of the “good” behaviour of the simulation:
  - Temperature, pressure, volume stables
  - Energies

Heating and equilibration

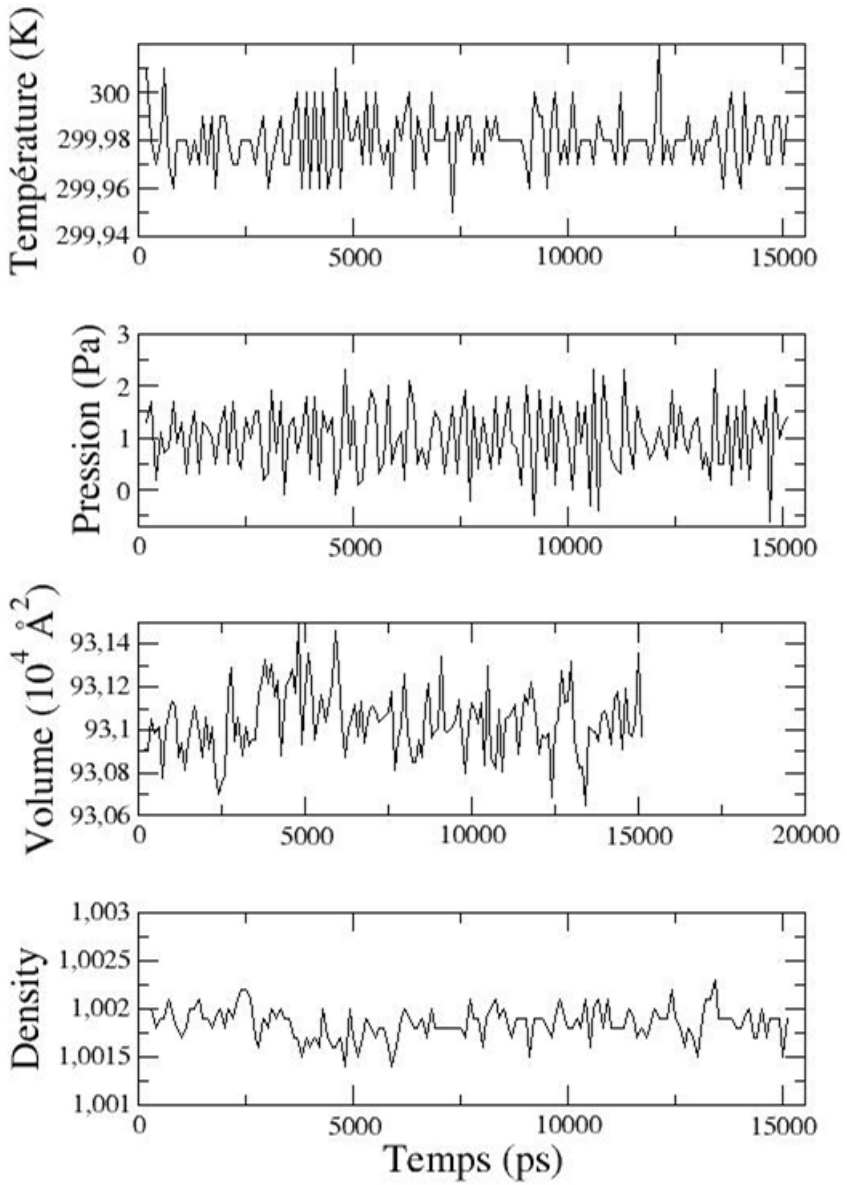


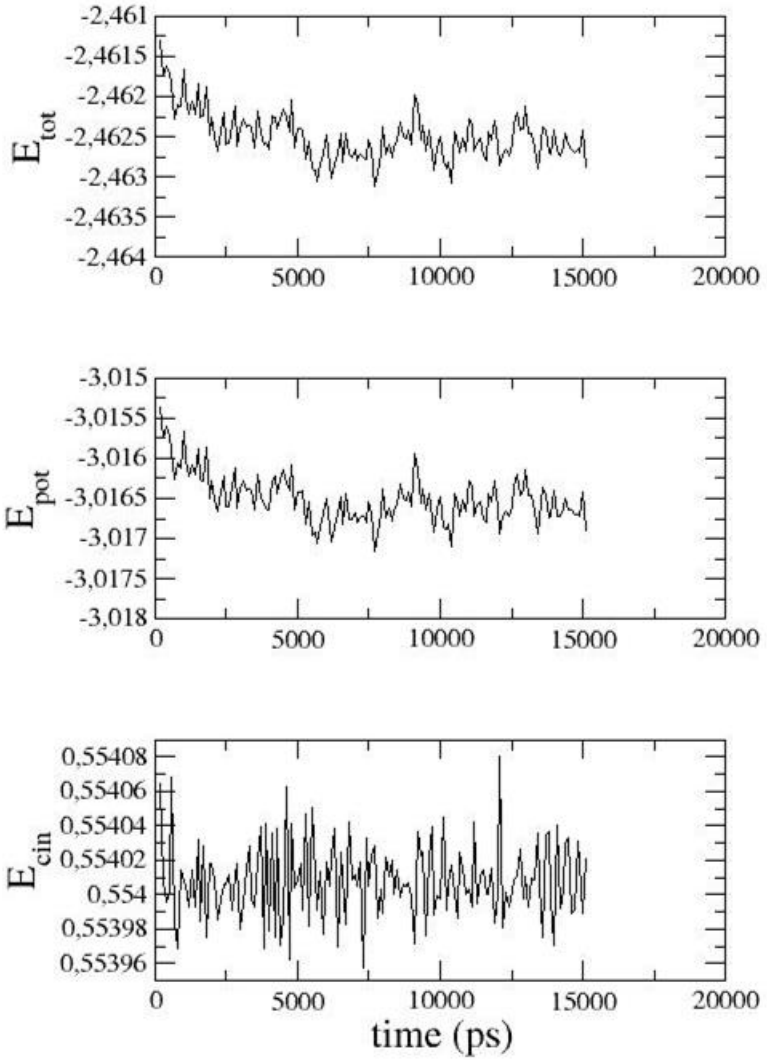






Production phase





Production phase – Analysis

- Verification of the “good” behaviour of the simulation:
  - Temperature, pressure, volume
  - Energies
- Analyses:
  - Structural parameters (RMSD, RMSF, interesting distances,...)
  - Principal component analysis
  - Calculation of thermodynamics quantities
  - ...

Classical analysis metrics

- Root Mean Squared Deviation (measure of the structure deformation)

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\vec{r_i}(t) - \vec{r_i}(0)\right)^2}$$

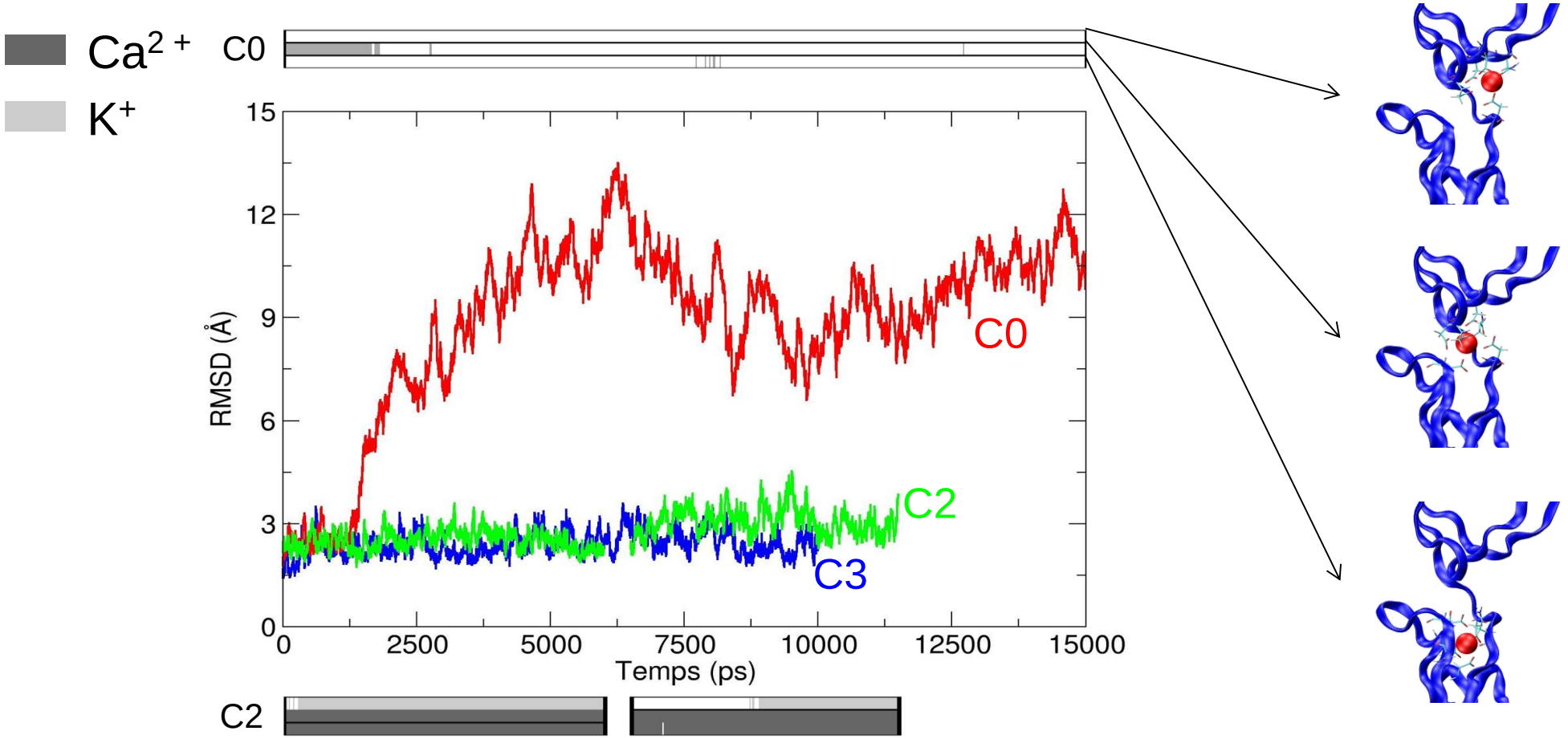
- Gyration radius (compacity of the molecule)

$$R_{gir}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\vec{r_i}(t) - \overrightarrow{r_{CM}}\right)^2}$$

- Atomic fluctuations (local flexibility of the molecule)

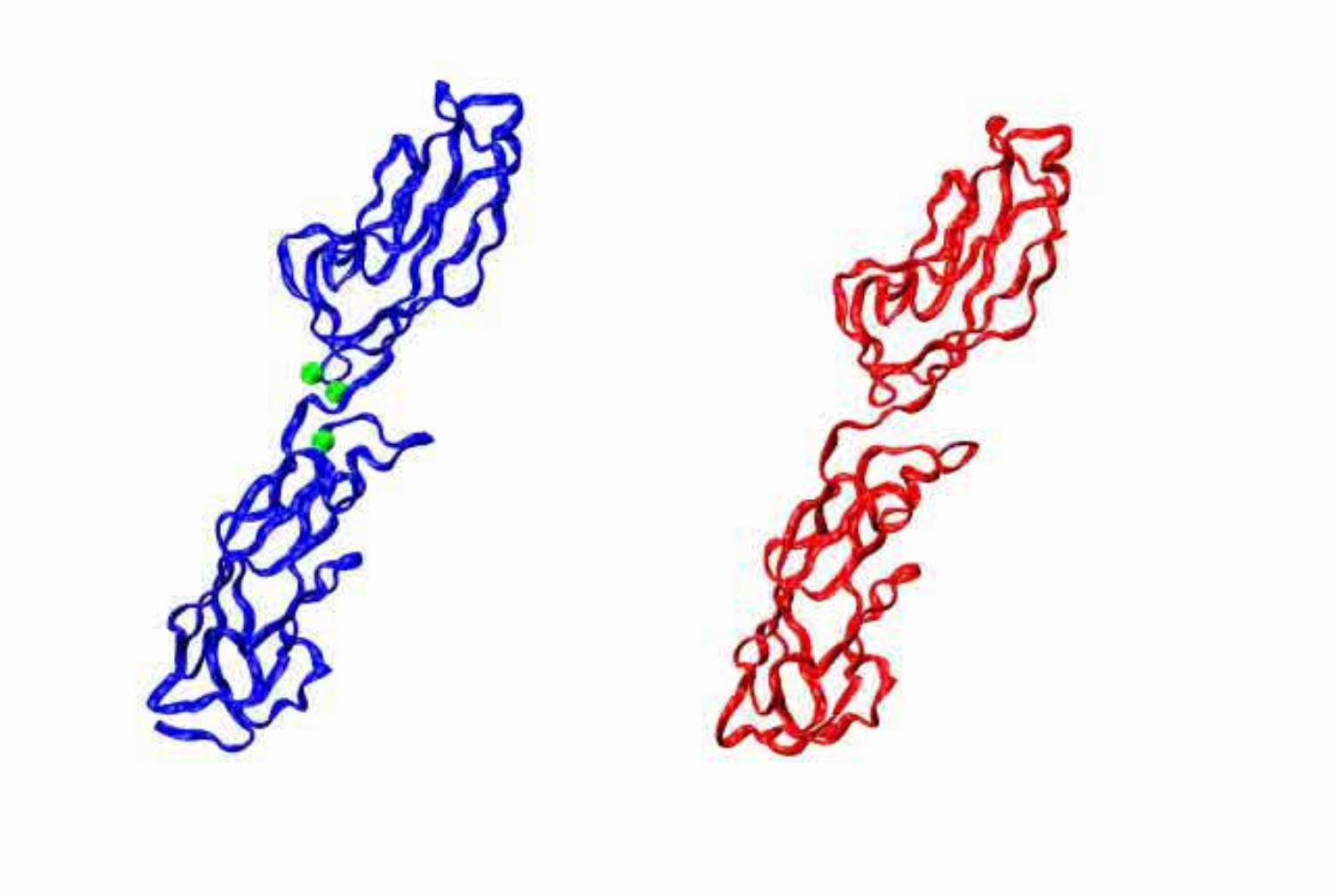
$$RMSF(i) = \langle\left(\vec{r_i}(t) - \langle\vec{r_i}\rangle\right)^2\rangle_t$$

Monomer structure and calcium ions

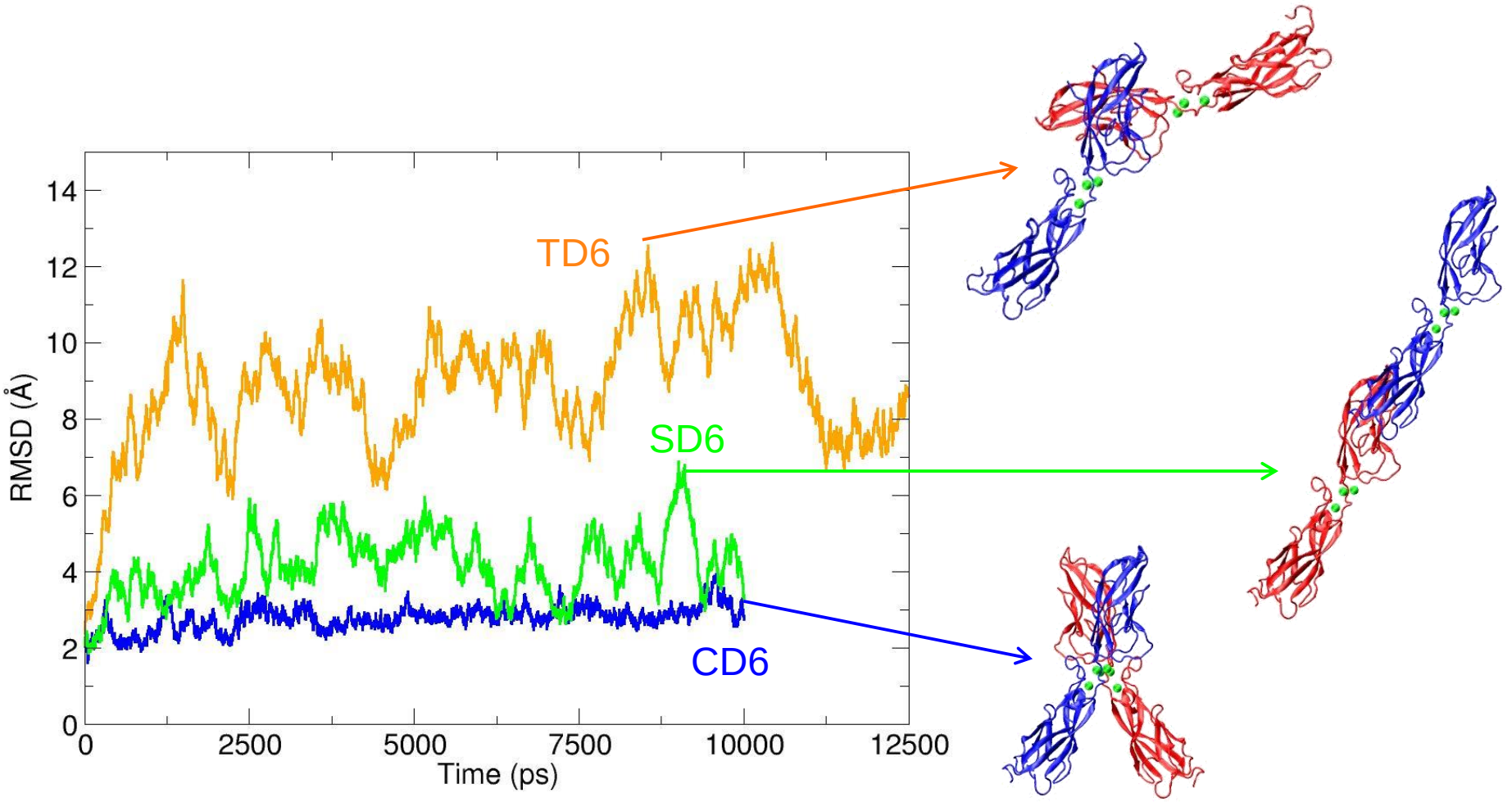


- RMSD : measure of the stability of the structure with respect to a reference structure (initial structure or mean equilibrium structure
- Measure of distances to determine the occupation of the binding sites of calcium ions

Monomer structure dynamics

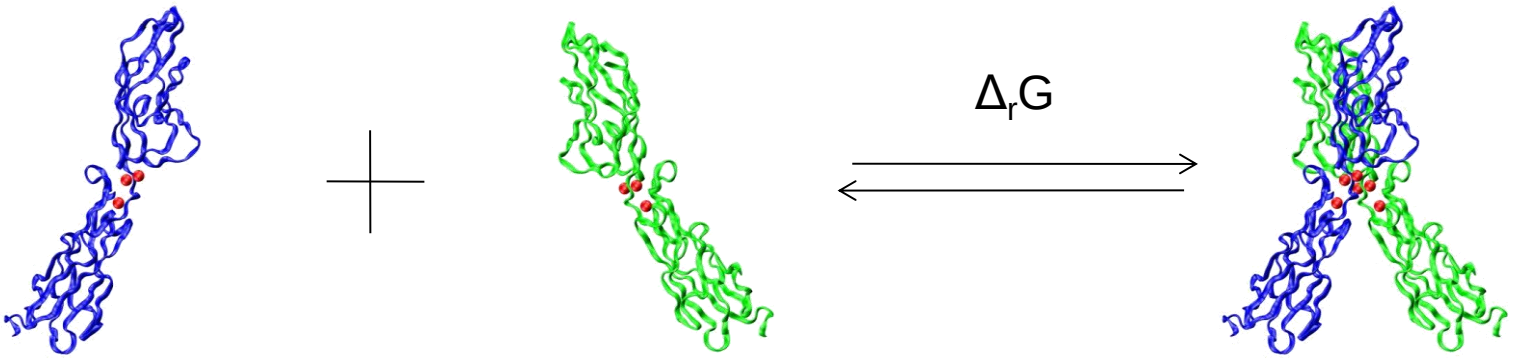


Dimers - Structure



- Important structural modification of TD6: crystallographic structure unstable
- RMSD: structural measure but no thermodynamics information

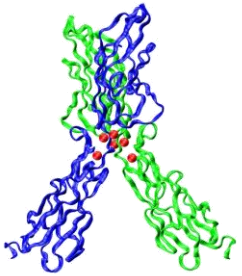
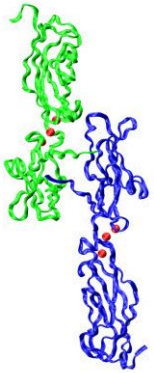
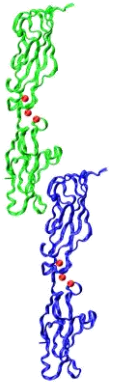
Thermodynamical analysis



- Extraction of a sample of structures after equilibration
- Calculation of mean enthalpy
  - From the forcefield energies
- Calculation of the mean entropy
  - Calculation of vibration modes (NMA)
  - Use of statistical thermodynamics formulae
- Finally:

$$\Delta_r G = (H_{dim} - TS_{dim}) - 2 \times (H_{mon} - TS_{mon})$$

Thermodynamical analysis

	cis-dimer CD6 	*trans*-dimer TD6 	*Staggered* dimer SD6 
$\Delta_r H$ (kcal.mol^{-1})	-84	-74	-44
$-T\Delta_r S$ (kcal.mol^{-1})	55	34	38
$\Delta_r G$ (kcal.mol^{-1})	-29	-40	-6

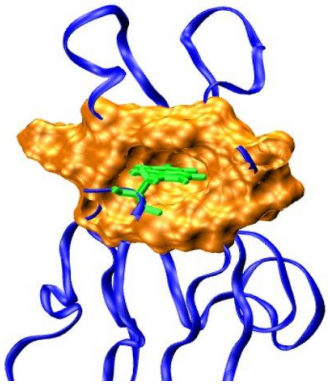
The *trans*-dimer is the most stable and thus may be the more able to support the adhesive function of cadherins

We got great information from “standard” MD, but...

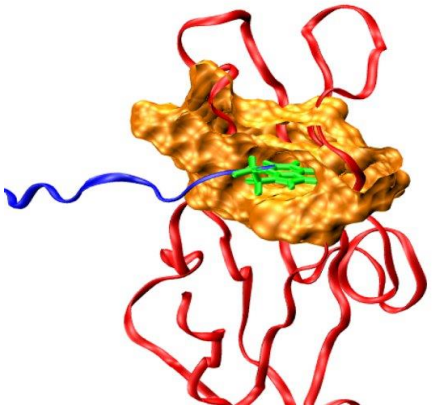
- System size important
  - *trans*-dimer: ~300000 atoms
  - 10ns: few weeks of calculation (at that time)

Reducing the dimensionality of the system

- Strand-exchange mechanism
  - Necessary for the *trans*-dimer formation
  - Rare event: impossible to see it in a “free” MD



Monomeric structure



trans-dimer

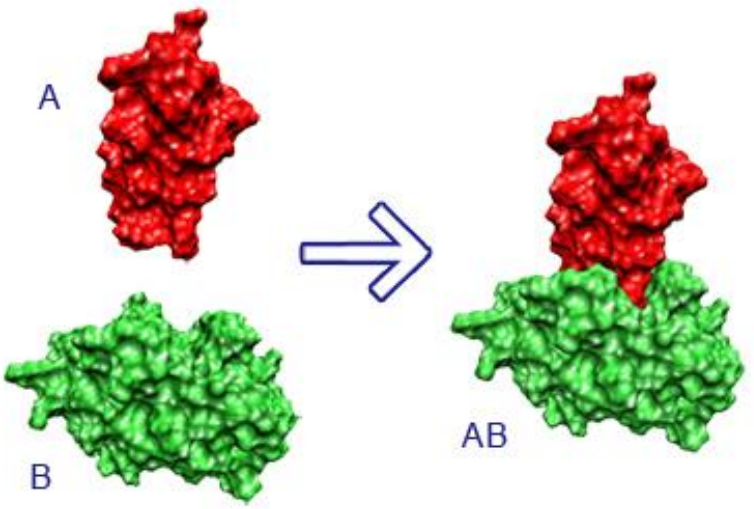
Biased-molecular dynamics

A short introduction to molecular docking

Definition of molecular docking

- Molecular docking:
  computational prediction of the structure of the complex formed by two molecules

- Various types of molecular complexes:
  - Protein/protein complexes
  - Protein/Nucleic acids complexes
  - Protein/small molecule complexes
  - DNA/small molecule complexes



Receptor-ligand complexes:
Pharmaceutical interest

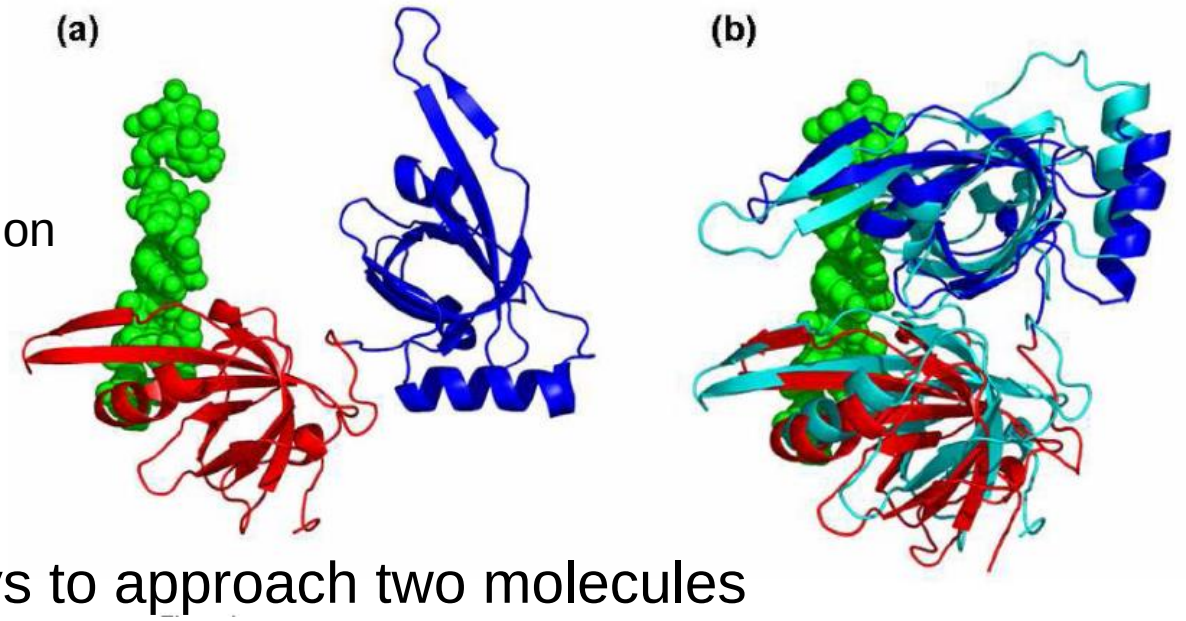
Mapping interactions between molecules in cells (interactomes)

Difficulties of molecular docking

- Molecular flexibility:
  Conformation of docked molecules may not be similar to that of free molecules

Replication Protein A / DNA complex

Free conformation of the protein



Docked conformation of the protein (cyan: prediction, blue and red: RX structure)

- Many ways to approach two molecules

Astronomical number of possible complexes to be tested

- How to know which is the best complex?

Definition of an accurate scoring function

Formulation of docking problem

- Given data: initial conformation of two molecules in their free form
- Question:
  - Do the two molecules bind?
  - What is the best complex between the two molecules?

- Tasks of docking:
  - Generating multiple initial structures
  - Improve the docking from these initial structures
  - Select the most probable complexes : ΔG_{bind} minimum

$$\Delta G_{bind} = \Delta G_{RL,gas} + \Delta G_{solv}$$

$$\Delta G_{bind} = \left(\Delta H_{RL,gas} - T\Delta S_{RL,gas}\right) + \Delta G_{solv}$$

Selecting the complexes: the scoring function

- Impossibility to compute exactly ΔG_{bind}
- Construction of a simple scoring function with desired properties:
  - Lowest value for the “natural” complex
  - Capable of distinguishing between correctly and incorrectly docked structures
  - Fast to compute
- Often a scoring function is almost uniquely energetic (no entropy): strong hypothesis !
- Nature of the scoring function:
  - Ab initio
  - Molecular mechanics
  - Empirical (learnt over a training set of complexes)

Increase in accuracy
Increase in cost

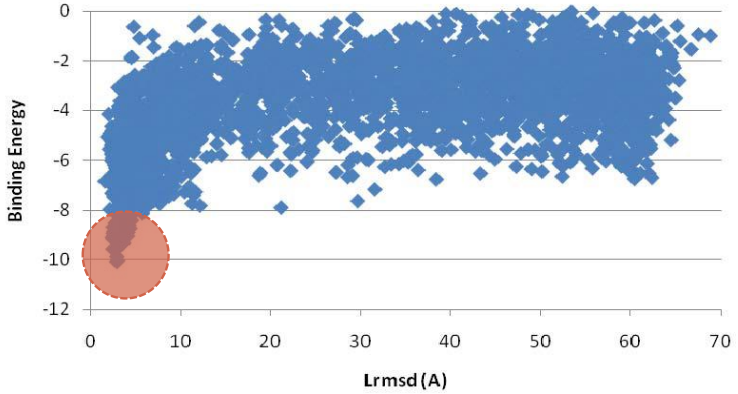
Selecting the complexes: the scoring function

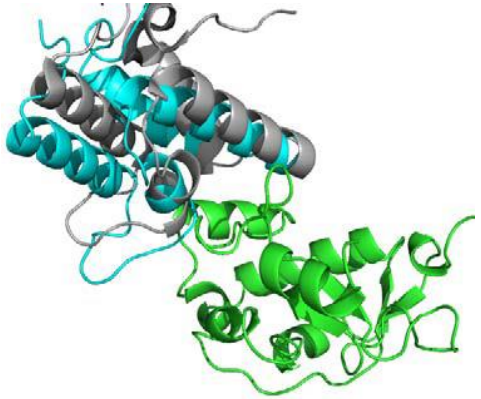
- Example of the scoring function in AutoDock:

$$V = W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec}\sum_{i,j}\frac{q_i q_j}{e(r_{ij})r_{ij}} + W_{sol}\sum_{i,j}\left(S_i V_j + S_j V_i\right)e^{(-r_{ij}^2/2\sigma^2)}$$

van der Waals | H-bond term | Electrostatic | Solvation term

- Testing a score function:





Superposition of crystallographic structure and best prediction

- Select good candidates rather than give a unique prediction

Types of docking

- Protein/Protein docking or DNA/Protein docking:
  - Rigid-body docking:
    limited to systems where unbound and bound molecules are identical
  - Flexible docking:
    using MD, MC, energy minimisation, NMA, rotamer libraries
    - Side-chain flexibility: more expensive
    - Backbone flexibility: much more expensive !

- Receptor-ligand docking:
  - Rigid receptor/Rigid ligand
  - Rigid receptor/Flexible ligand
  - Flexible receptor/Flexible ligand

Increasing computational cost

Illustration of various types of docking

- Test-case: Acetylcholinesterase – fasciculin 2 complex

- 4 docking strategies tested:
  - Key-Lock model (KL): no backbone flexibility
  - Conformation Selection (CS): backbone flexibility sampled by generating previous to docking various initial models differing by backbone conformation
  - Induced-Fit (IF): backbone flexibility sampled during the docking procedure (by MD, energy minimisation, MC,…)
  - Combined CS/IF

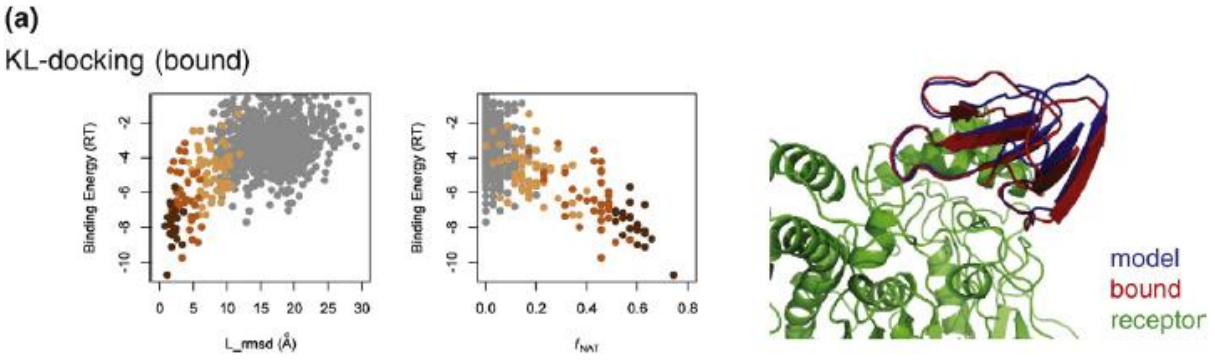
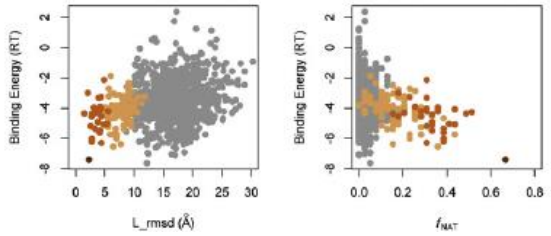
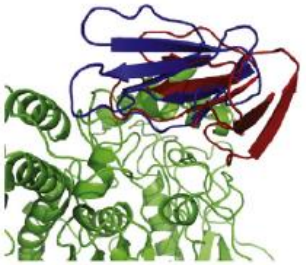


Illustration of various types of docking

(b)

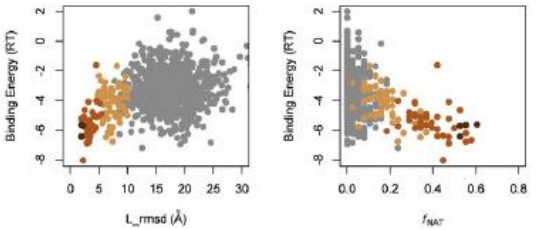
KL-docking

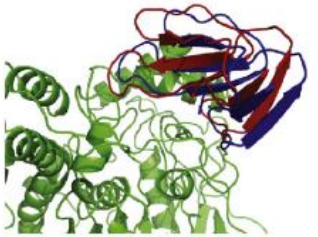




(c)

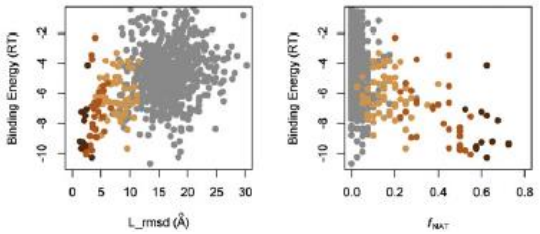
CS-docking

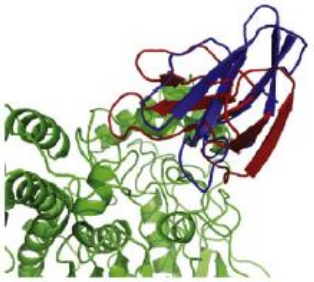




(d)

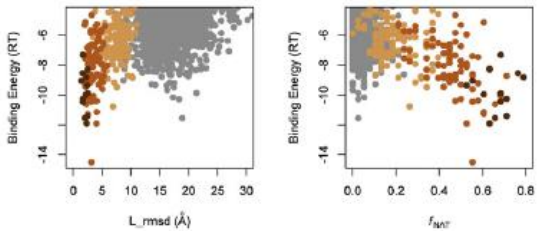
IF-docking

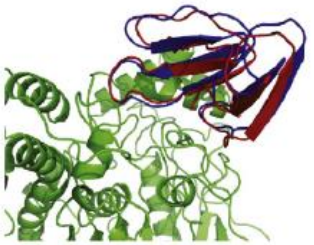




(e)

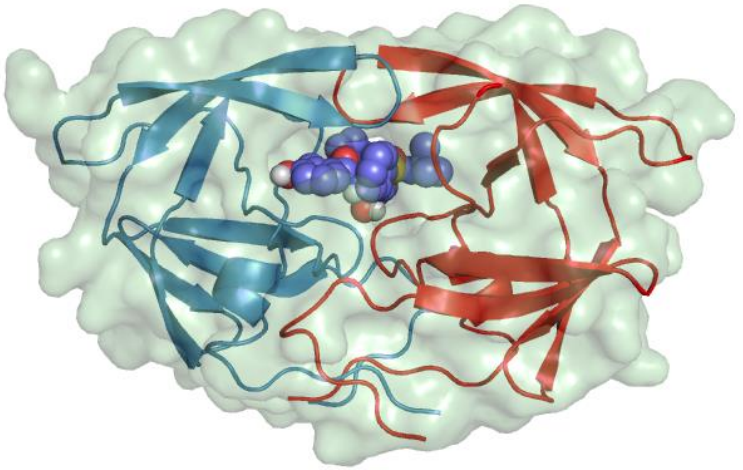
CS/IF-docking





Successes of docking in drug design

- Drugs against HIV:
  - HIV protease inhibitor amprenavir from Vertex and GlaxoSmithKline (Kim et al. 1995)
  - Nelfinavir from Pfizer (Greer et al. 1994)



- Influenza neuraminidase inhibitor zanamivir (Relenza) by GSK (Schindler 2000)

Docking has become an important tool in early stage of drug development to select potential candidates

Available softwares

- DOCK (Kuntz *et al.* 1982, Ewing & Kuntz 2001)
- FlexX (Rarey *et al.* 1996)
- Hammerhead (Welch *et al.* 1996)
- Surflex (Jain 2003)
- SLIDE (Kuhn *et al.* 2002)
- AutoDock (Olson *et al.* 1990, Morris *et al.* 1998)
- AutoDock Vina (Trott *et al.* 2010)
- Haddock (Bonvin *et al.*, 2003)
- Attract (Zacharias, 2005)
- … and many others

Some challenges in molecular modelling

Reducing the dimensionality of the system

- *trans*-dimer structure:
  - ~ 230000 atoms
  - ~ 74000 water molecules
- But we are interested only in the behaviour of the protein...
- Differences between the possibility of atomic molecular simulations and real biological systems:
  - One order of magnitude in length-scale
  - 4 to 6 orders of magnitude in time-scale

How to reduce the dimensionality?

- Use of implicit solvent (GB, PB equations):
  - No more solvent molecules
  - Fewer atoms
- Use of a simplified representation of the protein:
  - Fewer atoms for the protein
  - Simplification of the forcefield
  - In very simple representations, no more solvent molecules is needed

Implicit representation of the solvent

- Why is solvent important?
  - Biomolecules contain lots of charged groups
  - Often unstable in non-aqueous media (*a fortiori* in vacuum)

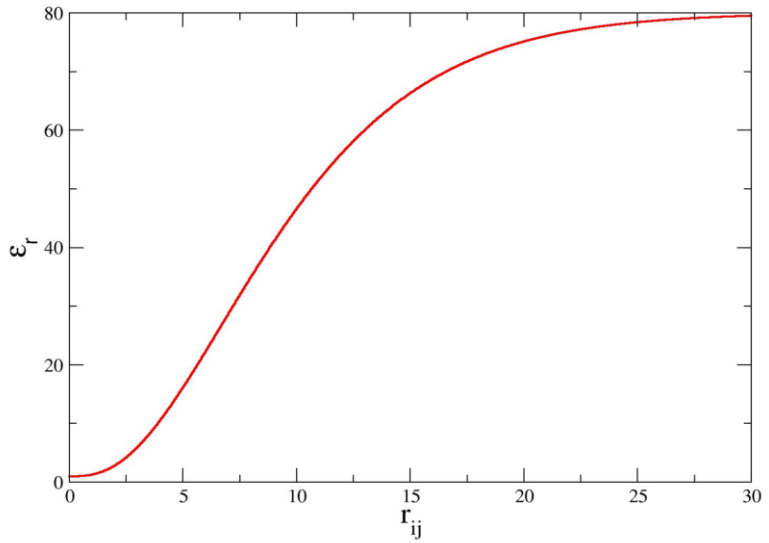
- Role of the solvent:
  - Attenuation of electrostatic interactions = vision of the solvent as a medium with a relative dielectric constant ε_r

$$V_{ij} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}$$

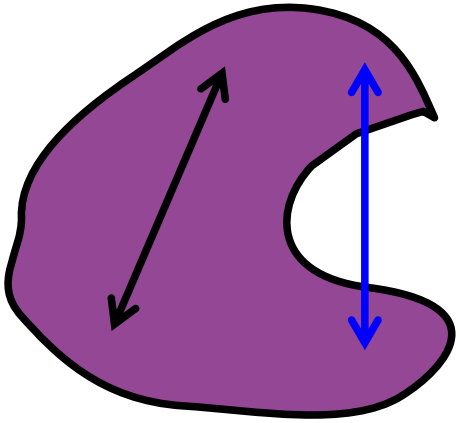
- All the interactions are not reduced to the same extent

Modification of electrostatic forces

- Modification of the dielectric constant as a function of the distance r_{ij} between two atoms:
  - ε_r proportional to r_{ij}
  - Sigmoïdal function of r_{ij}



- Problem: protein are not homogeneous



More complex implicit solvents

- Addition of one term in the forcefield that models the interaction of atoms with the solvent:

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{vdW} + \Delta G_{cav}$$

  - ΔG_{elec} : electrostatic interactions
  - ΔG_{vdW} : van der Waals solute/solvent interactions
  - ΔG_{cav} : creation of a cavity in the solvent (surface tension and reorganisation)

$$\Delta G_{vdW} + \Delta G_{cav} = \gamma A + b$$

$$\Delta G_{elec} = ?$$

Generalized Born model

- Solvation of an ion with a radius a:
  - Exact work necessary to transfer an ion from vacuum to a medium with dielectric constant ε_r

$$\Delta G_{elec} = \frac{-q^2}{2a}\left(1 - \frac{1}{\varepsilon_r}\right)$$

- Generalisation for a molecule:
  - Close partial charges → Screening of electrostatic interactions

$$\Delta G_{GB} = \frac{-1}{2}\left(1 - \frac{1}{\varepsilon_r}\right)\sum_{i,j}\frac{q_i q_j}{f_{GB}(r_{ij}, a_i, a_j)}$$

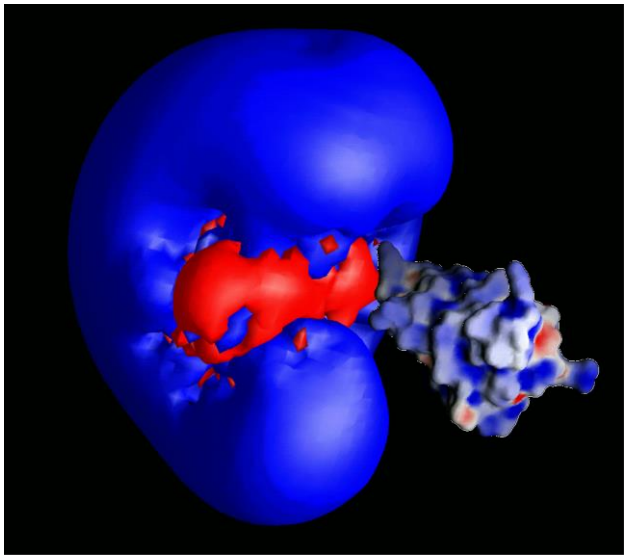
  - i=j, f_{GB} → Solvation of an ion
  - r_{ij} small, f_{GB} → Solvation of a dipole
  - r_{ij} big, f_{GB} → solvatation of the two ions + screening of electrostatic interactions

Poisson-Boltzmann model

- Poisson model
  - Continuous dielectric medium with permittivity ε
  - Resolution of the equation $\nabla\left(\varepsilon(r)\nabla\Phi(r)\right) = -4\pi\rho(r)$
  - Determination of Φ in each point of the system (numerical resolution)

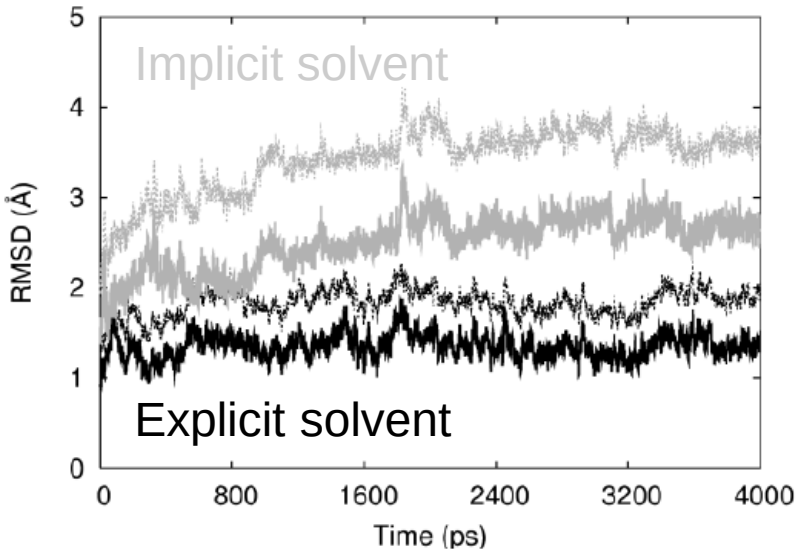
$$\Delta G_{PB} = \frac{1}{2}\sum_i q_i \left(\Phi_i - \Phi_i^{vac}\right)$$

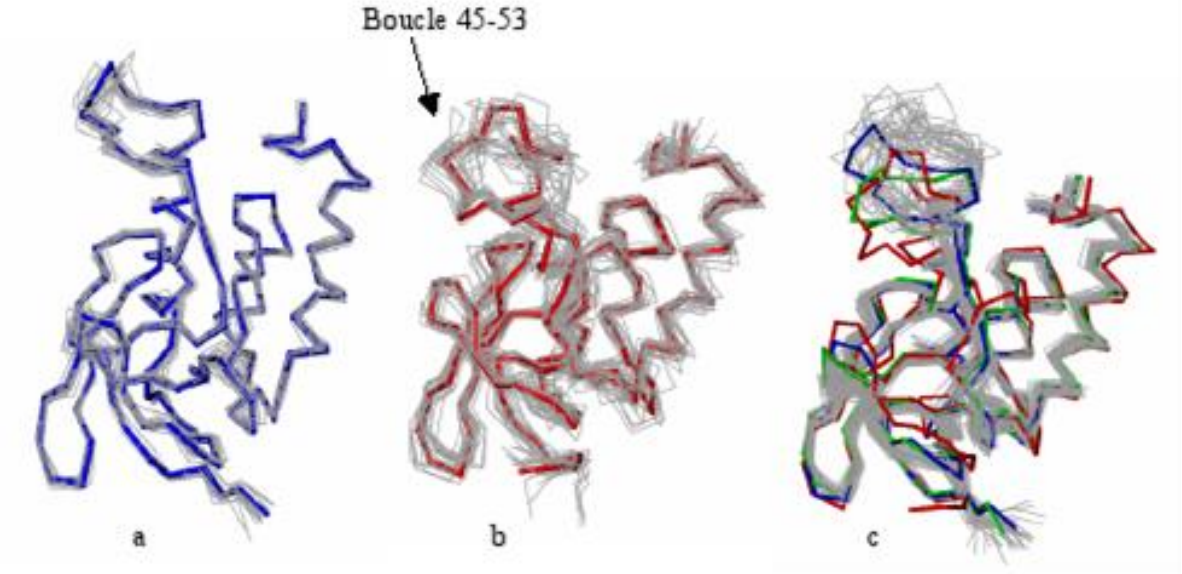
- GB or PB : Advantages / Drawbacks
  - Hard to differentiate ΔG_{PB} → limitation for MD
  - GB more empirical...
  - Calculation of electrostatic potential with PB equation
    → Potential maps / complementarity (docking)



Trypsin and one inhibitor

Implicit representation of the solvent: application





Explicit solvent

Implicit solvent

NMR structures

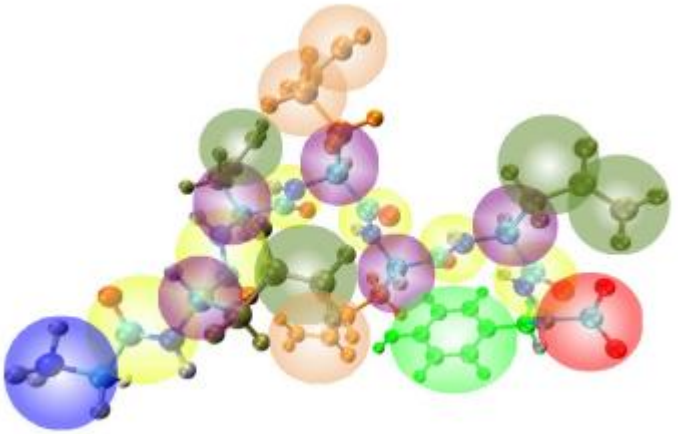
- Advantages:
  - Simulations quicker (> 30%)
  - No need to equilibrate the solvent
  - No friction with solvent molecules → seems more flexible (realistic?)
- Drawbacks:
  - No direct interactions with water molecules that may be needed for stabilisation of particular parts of the protein
  - Does not work well with elongated molecules (DNA for example)

Applications of implicit solvents

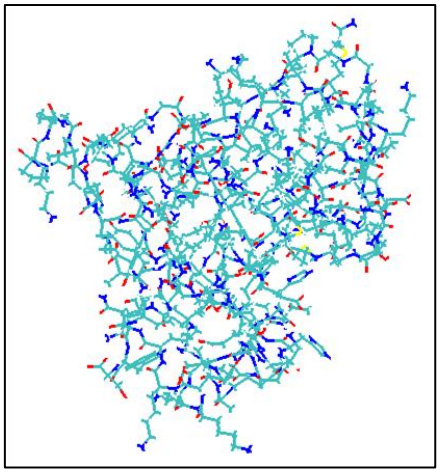
- Used in scoring functions for protein structure prediction
- Binding energy calculations (docking)
- Molecular Dynamics of big systems
- Conformational sampling:
  - Protein folding
  - Structure refinement

Simplified protein representation

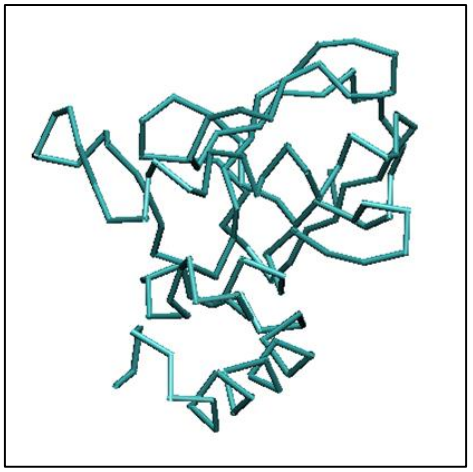
- Precise atomic description is not always necessary
- “United atoms” forcefields:
  - “Removal” of hydrogen atoms gathered with "heavy" atoms
  - Modification of vdW and electrostatic parameters
  - Gain of ~50% in atom number
- Coarse-graining:
  - Several atoms are gathered in a single pseudo-atom
  - One amino-acid = one or two atoms for the backbone + 0 to 3 atoms for the sidechain
  - Bonded and non-bonded interactions between pseudo-atoms: similar mathematical expression than all-atom forcefields, but parameters need to be refined



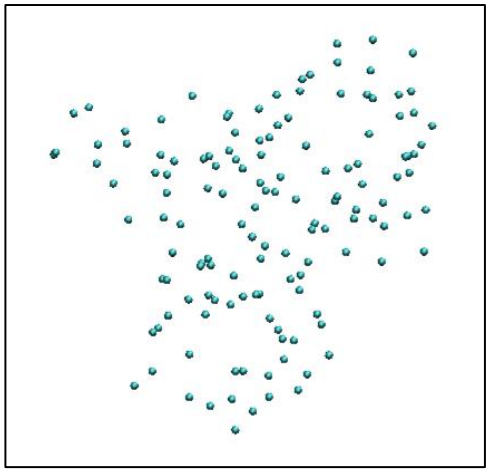
“Extreme” coarse-graining: elastic network



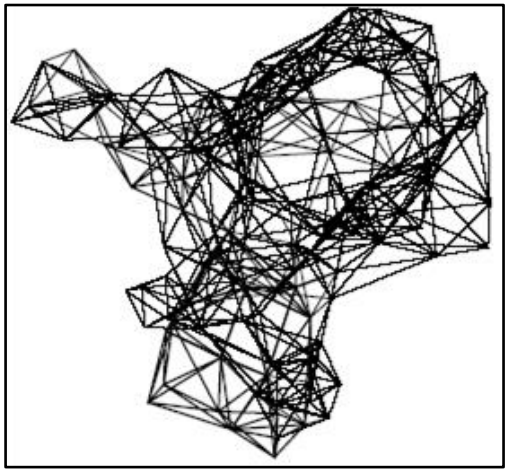
All atom



Trace

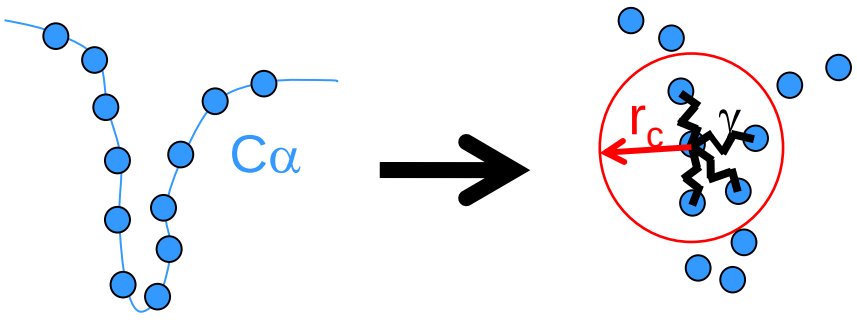


Cα = sites



Elastic network

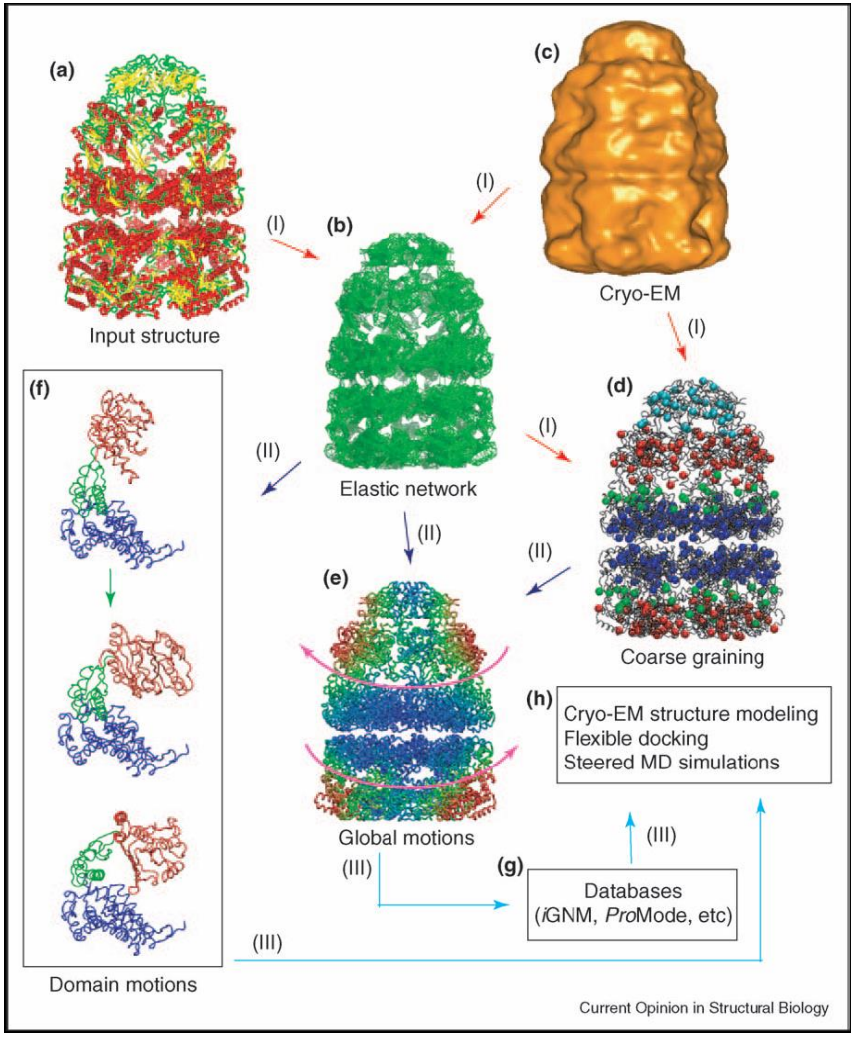
- One centrer of force per amino acid
  → Reduction of the number of atoms by a factor of 10
- Interactions between sites = harmonic springs
- Possibility to increase the graining for very big systems (1 bead for 5-10 amino acids)



Normal Mode Analysis with an elastic network

- Simplified representation are frequently used to study large motions in proteins in combination with Normal Mode Analysis (see course of D. Perahia)

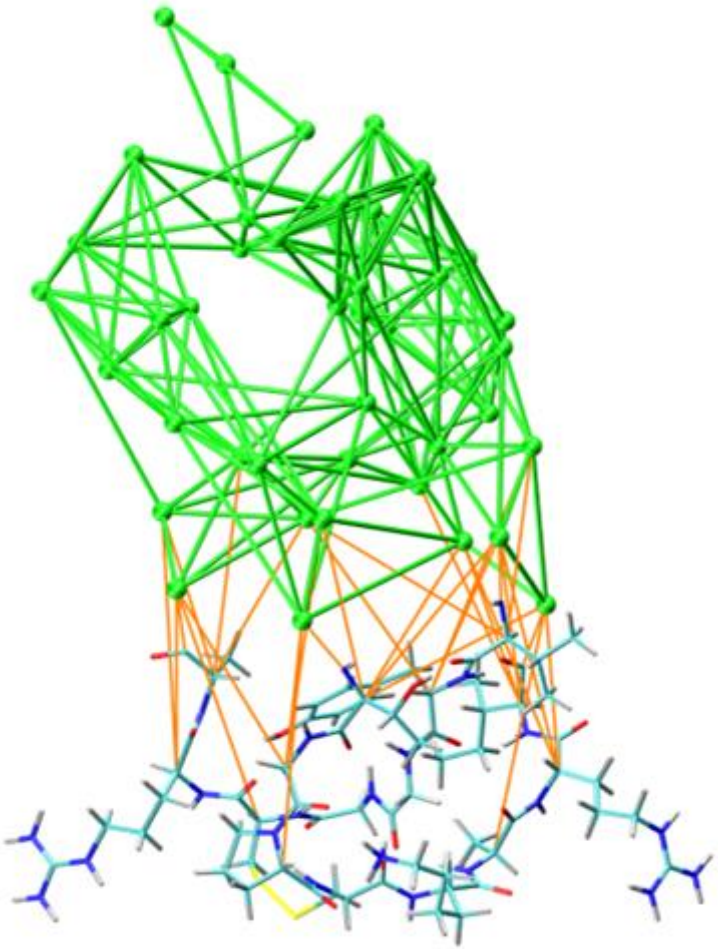
- Advantages of the simplified representation:
  - NMA necessitates the use of big matrices (high memory needs)
  - Offers the possibility to study bigger systems
  - Huge decrease of computational time
  - Reasonable agreement with all-atoms calculations



Complex of chaperone protein GroEL-GroES (Dimer of heptamer)

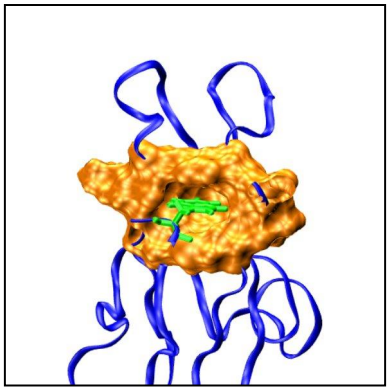
Limitations of reduced representations

- Loss of (atomic) information
- In the case of elastic network, the moves are constrained around the equilibrium structure
- There is a need to define intermediate (hybrid) representation
- Multi-scale problem:
  - Quantum Mechanics
  - Molecular Mechanics
  - Coarse Graining

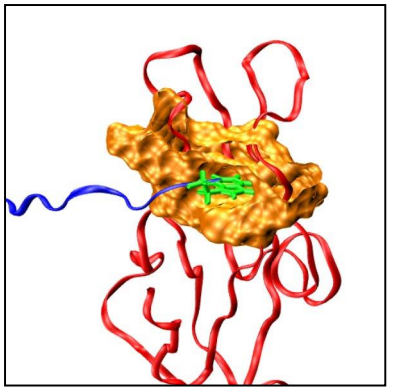


Biases in simulation

- Time-scales accessible for all-atom simulations:
  - Tens to hundreds of nanoseconds for intermediate size/big systems
  - Up to the microsecond scale for small systems (50-100 residues) with massively parallelized computation is used
- Many phenomena, especially in biology, take place on a much longer time scale (between ms and s)
  - Big moves between large domains
  - Moves with high energy barrier
  - Nucleation processes
- Example: strand-exchange mechanism in the formation of the *trans*-dimer of cadherin
- Moves can be forced with the use of biases (in MD, energy minimisations, or MC)

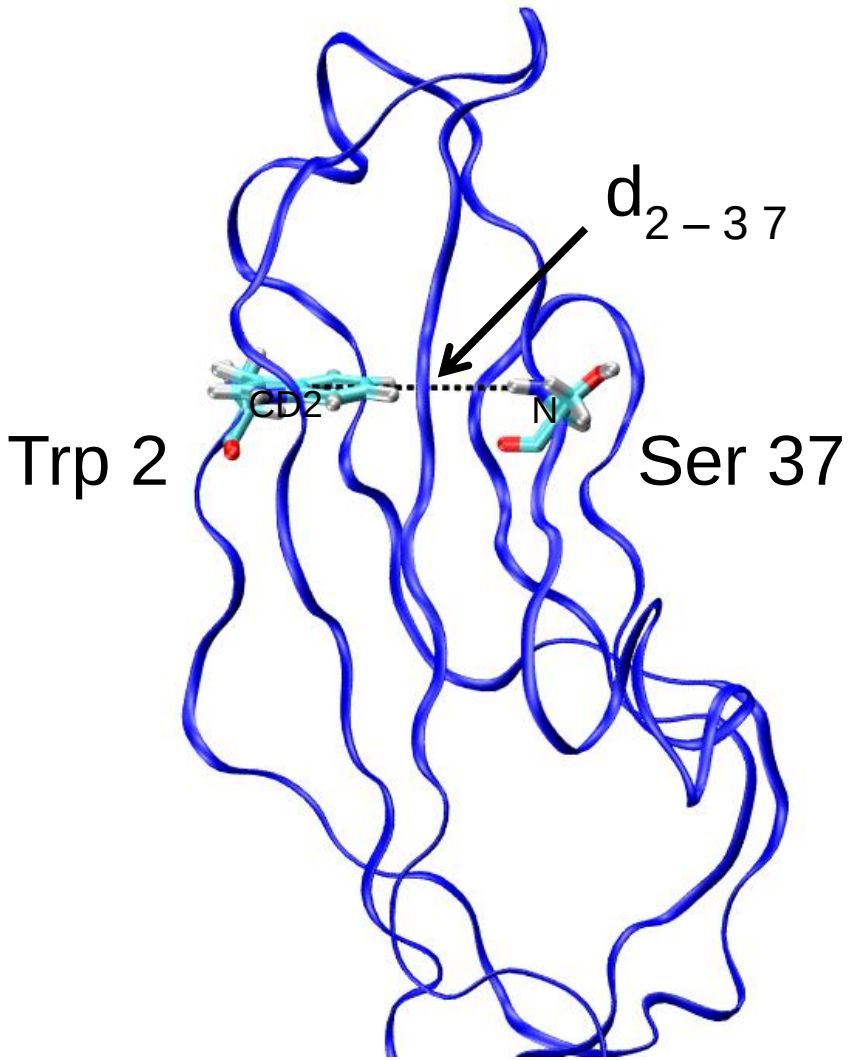


monomer



trans-dimer

Biased simulations - Example

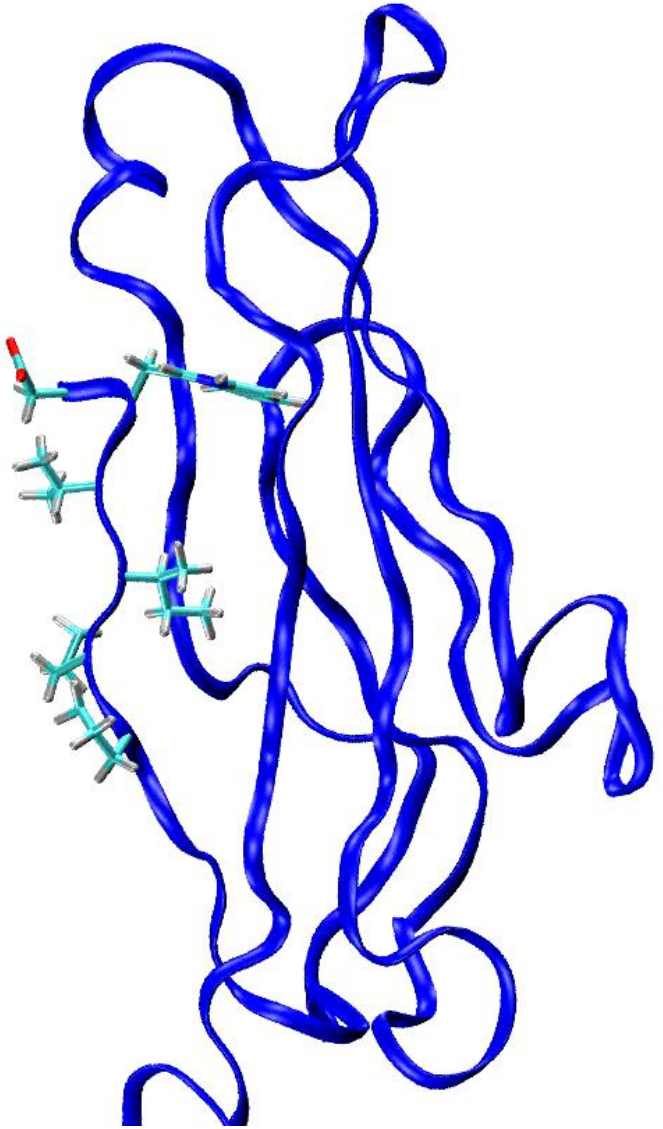


- Choice of a “reaction coordinate”:
  d_{2-37}
- Increase of d_{2-37} in MD simulations “step by step”
- Modification of the forcefield, adding an harmonic constraint:

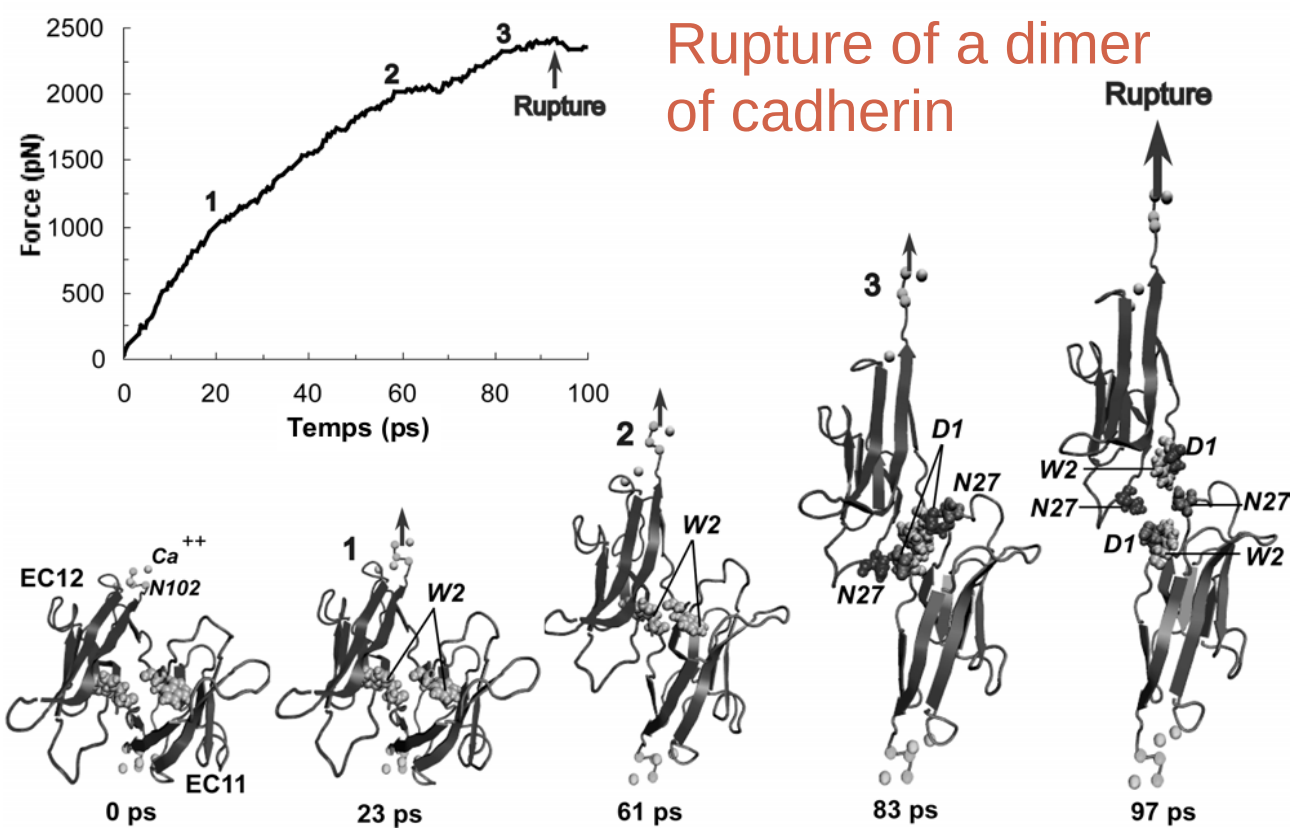
$$E_{cont} = k(d_{2-37} - d^0_{2-37})^2$$

- Possibility to determine free energy profiles

Example in « motion »



Stretching molecules: Steered Molecular Dynamics



- Some proteins undergo important mechanical stress
- This stress can be modelled with an additional force: steered molecular dynamics
- May lead to observation of intermediate structures (“transition structures”)
- Non-equilibrium simulations

Biased simulations – Summary

- Advantages:
  - Have access to “rare events”
  - New strutural, kinetic, or thermodynamical information
- Drawbacks:
  - Choice of the reaction coordinates not necessary obvious
  - Velocity of the deformation or forces introduced in the simulation
    - Sometimes “non-realistic” values
    - What is the influence on the observations?

Studying reactions in biomolecules

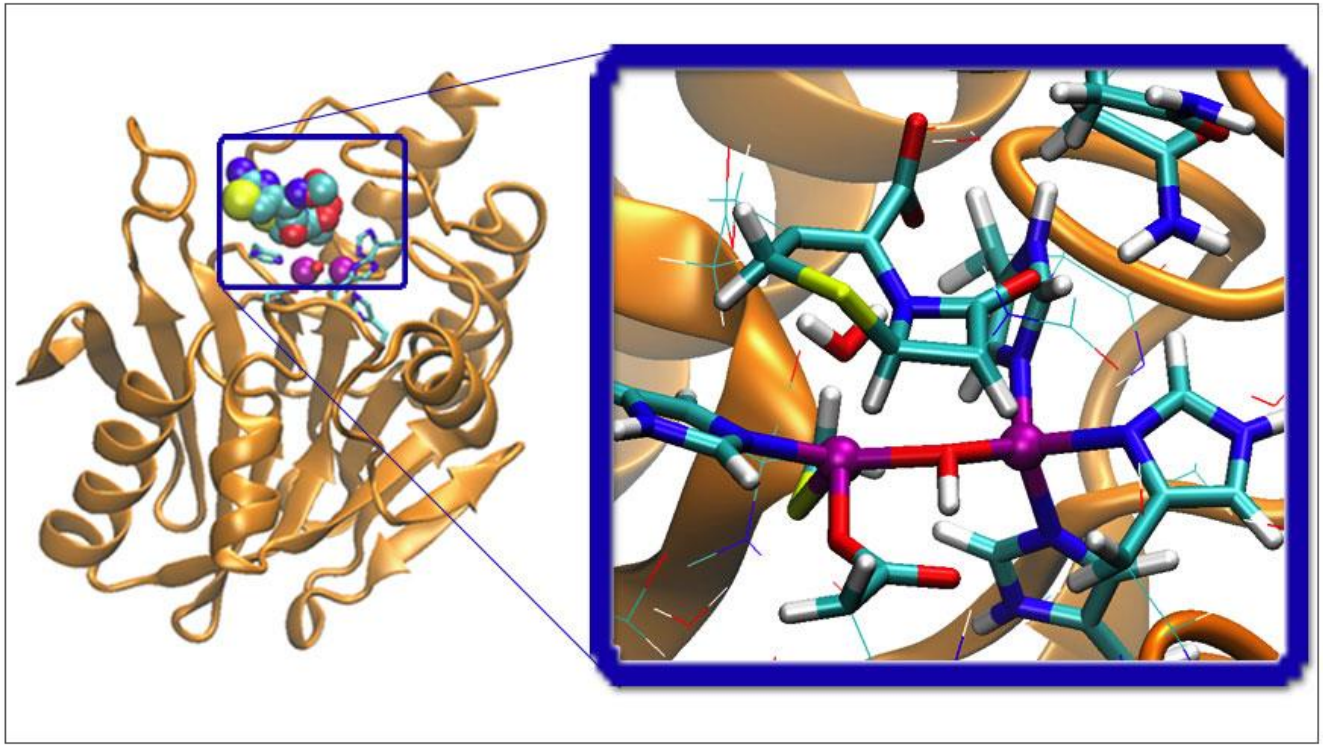
- Protein functions:
  - Structural function / Molecular motors
  - Enzymes: chemistry (bond formation/rupture, electron or proton transfer)

Need of quantum mechanics

- But system size prevents the use of quantum mechanics

- Make use of hybrid QM/MM representations

QM/MM – Example

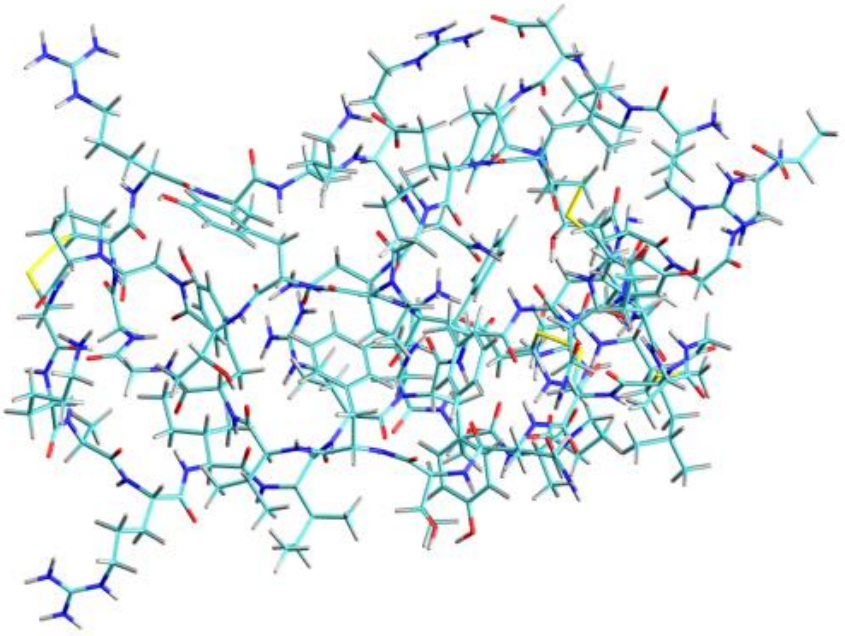


- 2 parts for the molecule treated with QM or MM

$$E = E_{QM} + E_{MM} + E_{QM/MM}$$

- Example:
  Zinc β-lactamase à Zinc: QM/MM simulations allowed to understand the mechanism of the hydrolysis of β-lactame

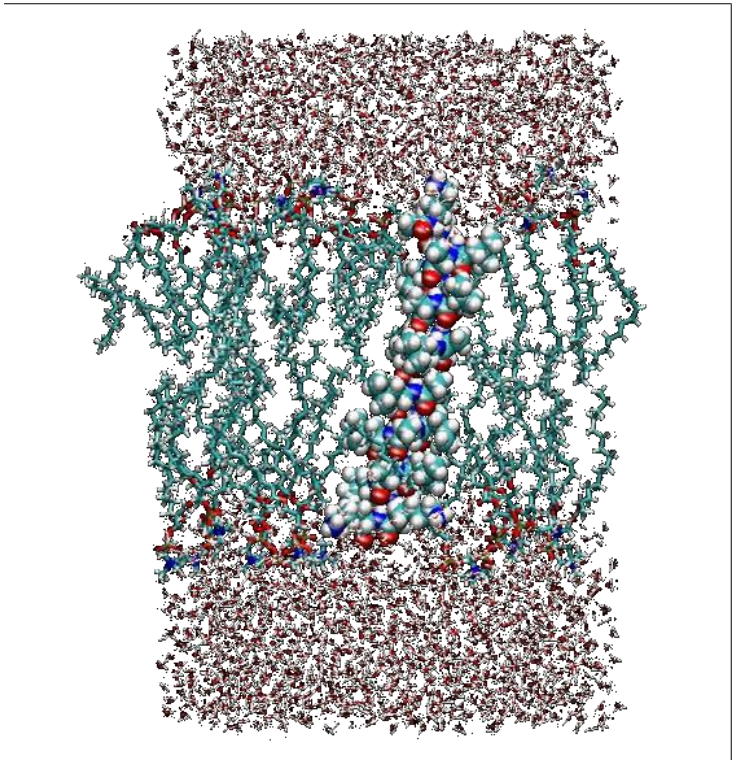
Biomolecule simulations: from prehistory...



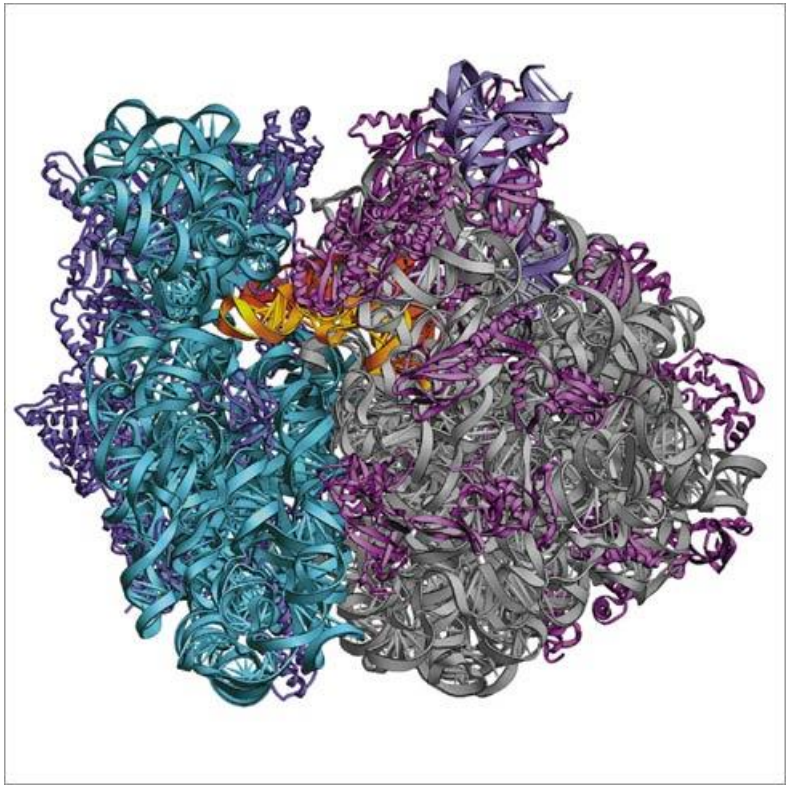
BPTI

- Nature 1977 : "Dynamics of folded proteins" :
- First molecular dynamics simulation of a protein
  - BPTI (58 amino-acids – 900 atoms)
  - No solvent molecules
  - 8.8ps of simulation

...to today and tomorrow



Membrane proteins :
100000 to billions of atoms



Ribosome (structure resolved in 2001): more than 50 proteins and DNA molecules