

Complex and simple models: Quantum Chemistry, QM/MM and more

Structural Bioinformatics



# Quantum Mechanics for chemistry in brief

The electrons of an atom can only be described by a pdf:  $|\Psi(r, r_2, \dots, r_N)|^2$

psi is called the wave function. The norm of psi,  $|\psi|^2$  is the actual all-electron probability distribution: this is the probability of finding one electron in  $r_1$ , a second in  $r_2$  etc.

In QM, the problem is no longer how a point moves, but how a probability distribution evolves over time, which is obtained by solving the Schrödinger equation:

$$\hat{H} \Psi = E \Psi \quad \longleftrightarrow \quad \frac{-\hbar^2}{2m} \nabla^2 \Psi(r) + V(r) \Psi(r) = E \Psi(r)$$

This is only  
Coulomb

Kinetic Energy + Potential Energy = Total Energy

Hamiltonian Operator (Energy operator)

Energy eigenvalue

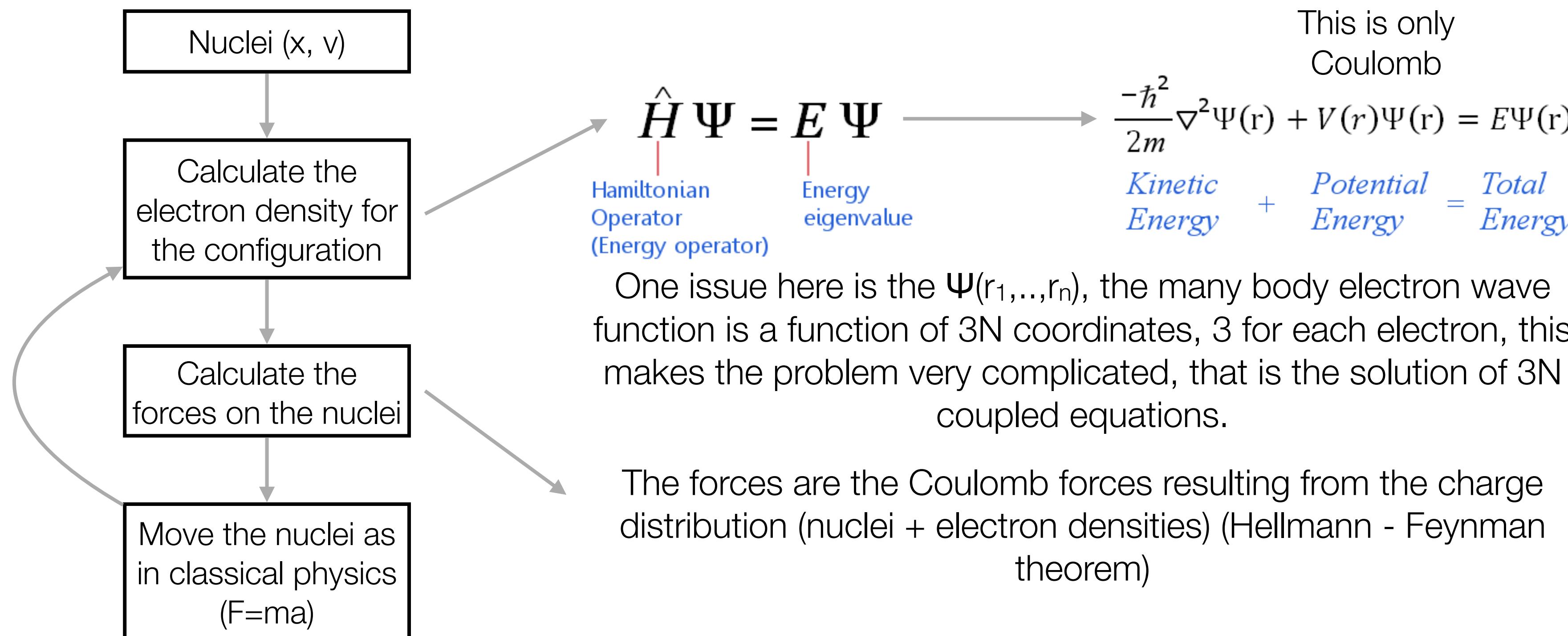
Solving this equation for objects more complex than the hydrogen atom is both analytically and computationally difficult.





# Quantum Chemistry: only electrons are treated as quantum particles.

Born-Oppenheimer approximation: atomic nuclei are considered to be 'classical particles' with a fixed charge.

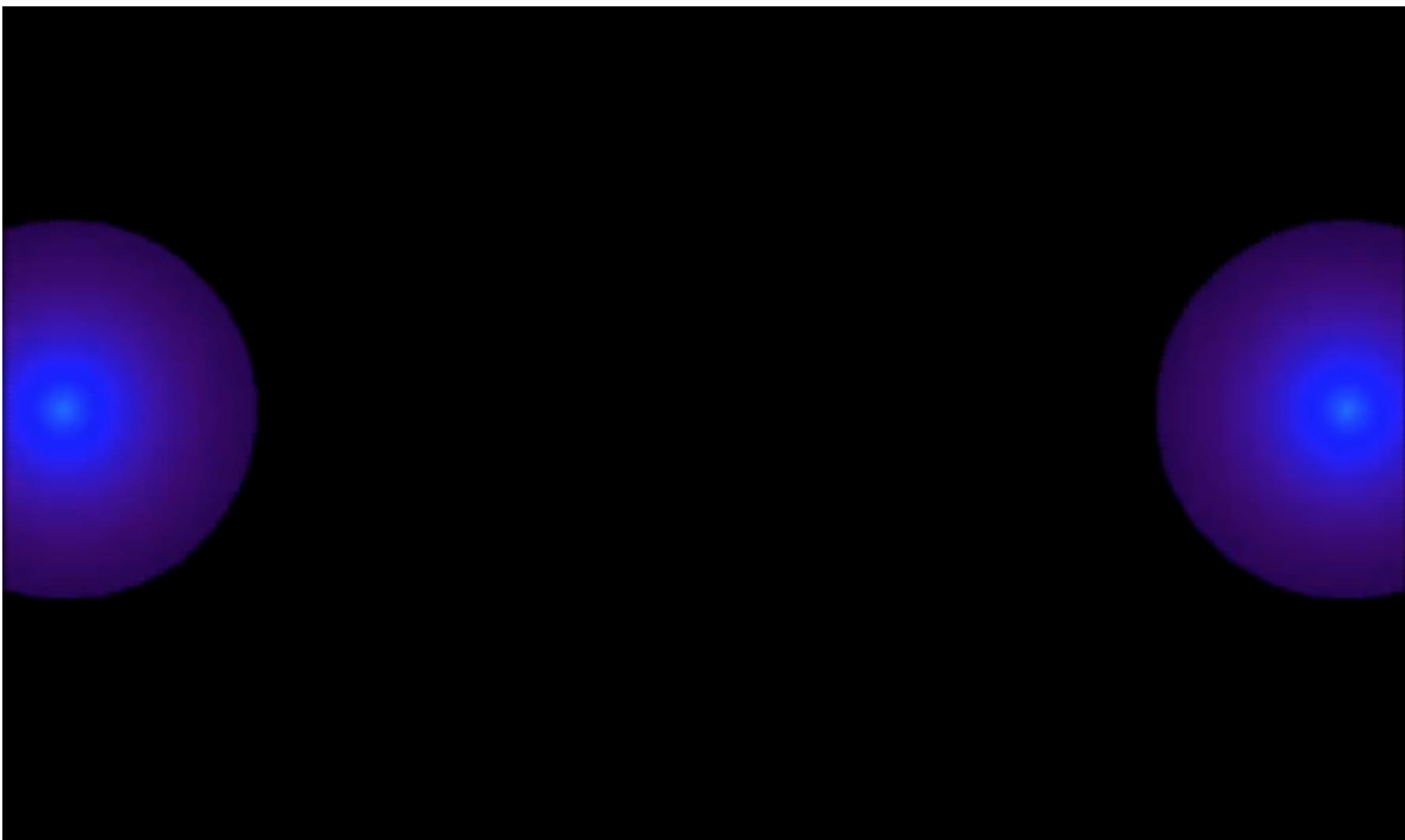


You may think as that each configuration has its own force-field determined by finding the all-electron density.



# Chemical Bonding

---



A chemical bond is the result of a redistribution of the electronic density.



UNIVERSITÀ  
DEGLI STUDI  
DI MILANO





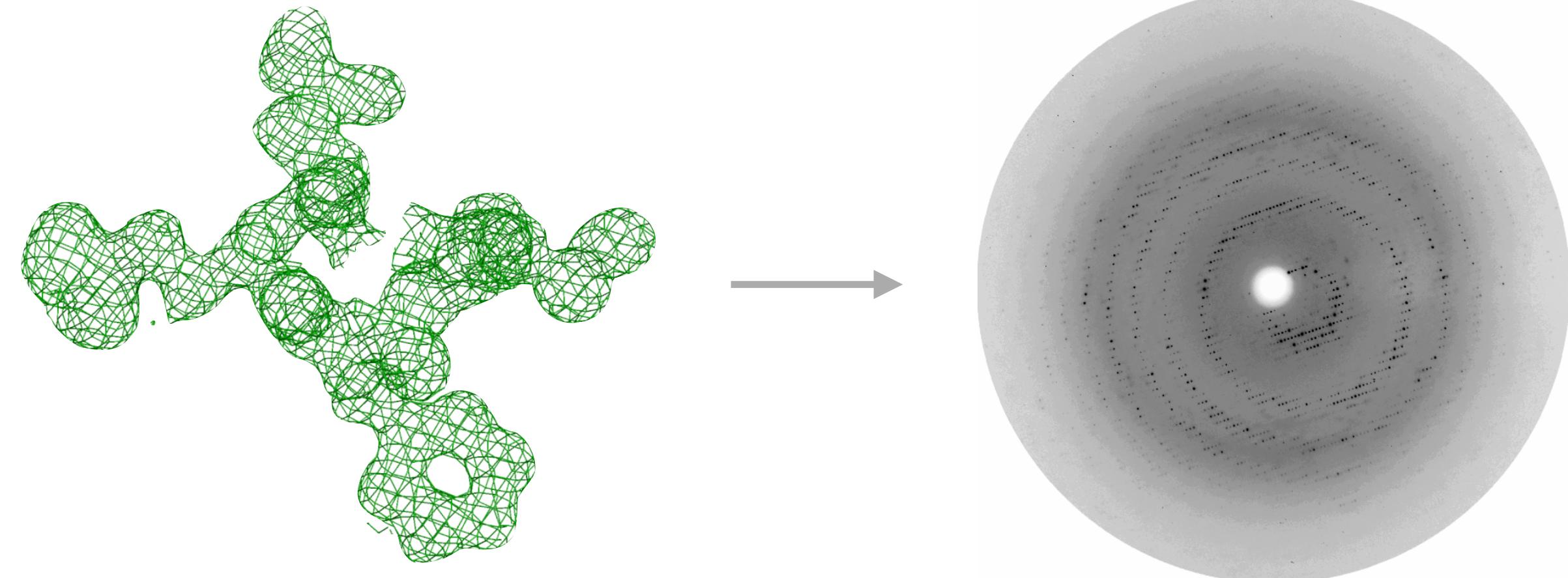
# Density functional theory (DFT)

To speed up the calculation of electronic forces on nuclei, a theory has been developed that shifts the focus from total electron density to electron density, that is, the density of electrons in a given region of space:

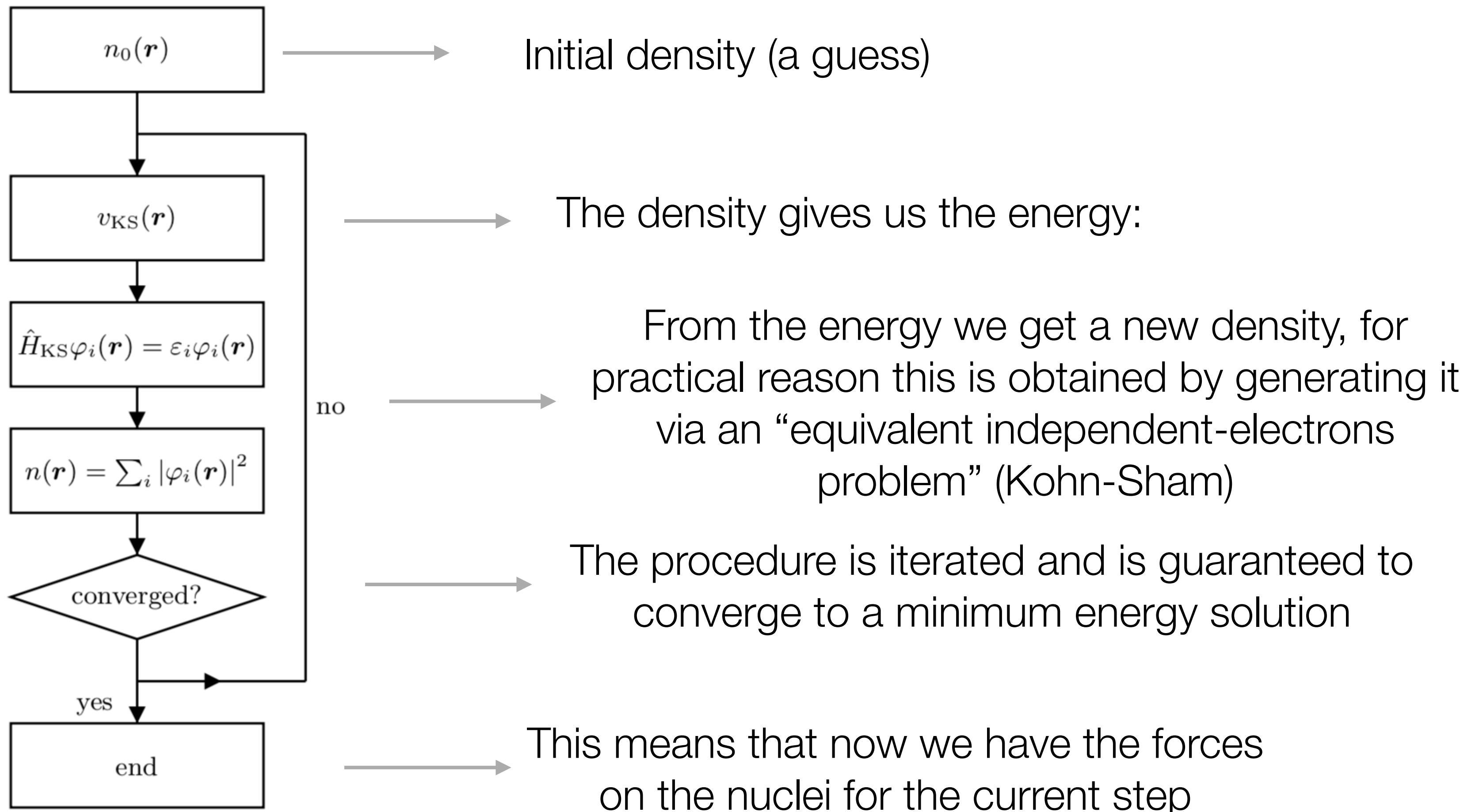
$$n(r) = N \int dr_2 \dots dr_N |\Psi(r, r_2, \dots, r_N)|^2$$

$n(r)dr$  is the number of electrons that could be found in a small sphere centred at position  $r$

**This is what X-ray crystallography observes**



# Density Functional Theory



# Density Functional Theory: key technical choices

---

The **XC** (exchange and correlation) potential: this is a key component of the DFT and is unknown. There are many different approximations, which can be divided into three classes

- LDA (local density approximation): Very fast and simple, not very accurate (PZ81, PW92, ...)
- GGA (Generalised Gradient Approximation): Accurate and fast (BLYP, PW91, ...)
- Hybrid GGA: slow, sometimes very accurate (B3LYP, PBE0, ...)

The **PSEUDOPOTENTIAL**: The idea is that to reduce the number of electrons in the density calculation one can divide them into valence and non-valence electrons, the nucleus plus the non-valence electrons are described together by a PSEUDOPOTENTIAL.

DESCRIPTION OF THE **DENSITY**: The electron density can be described on a 3D grid, or using a basis set, and/or using simple waves.





## Summarising:

In principle, it is possible to simulate molecules without having to parameterise a force field as we saw before, but instead using the laws of physics. However, you still have to make approximations and choices, see the previous slide, which can affect the results you get.

In terms of performance, QM based approaches are extremely expensive, they get slower and slower the more ELECTRONS you have in your system (remember that in MD we were talking about ATOMS, not electrons).

Furthermore, while MD gets slower with  $N^2_{\text{atoms}}$ , QM gets slower with  $N^3_{\text{electrons}}$ .

So if in **MD** we are limited to simulating tens of millions of atoms on the microsecond timescale, in **QM** we are limited to simulating thousands of atoms on the nanosecond timescale.





# QM/MM: mixing QM and Classical MD simulations

## DFT Simulations

- Study small systems on short time scales.
- Study chemical reactions if they happen on the time scale of the simulation.

## MD Simulations

- Study relatively large systems on relatively long time scales.
- Cannot study chemical reactions.

Enzymes, which catalyse chemical reactions, are a large class of proteins whose function cannot be studied by either technique. A nice feature of enzymes is that they usually speed up chemical reactions dramatically, making them happen on time scales that can in some cases be compatible with DFT calculations, the only problem is size!

## How can we simulate enzymatic reactions?

Warshel & Levitt 1976



UNIVERSITÀ  
DEGLI STUDI  
DI MILANO

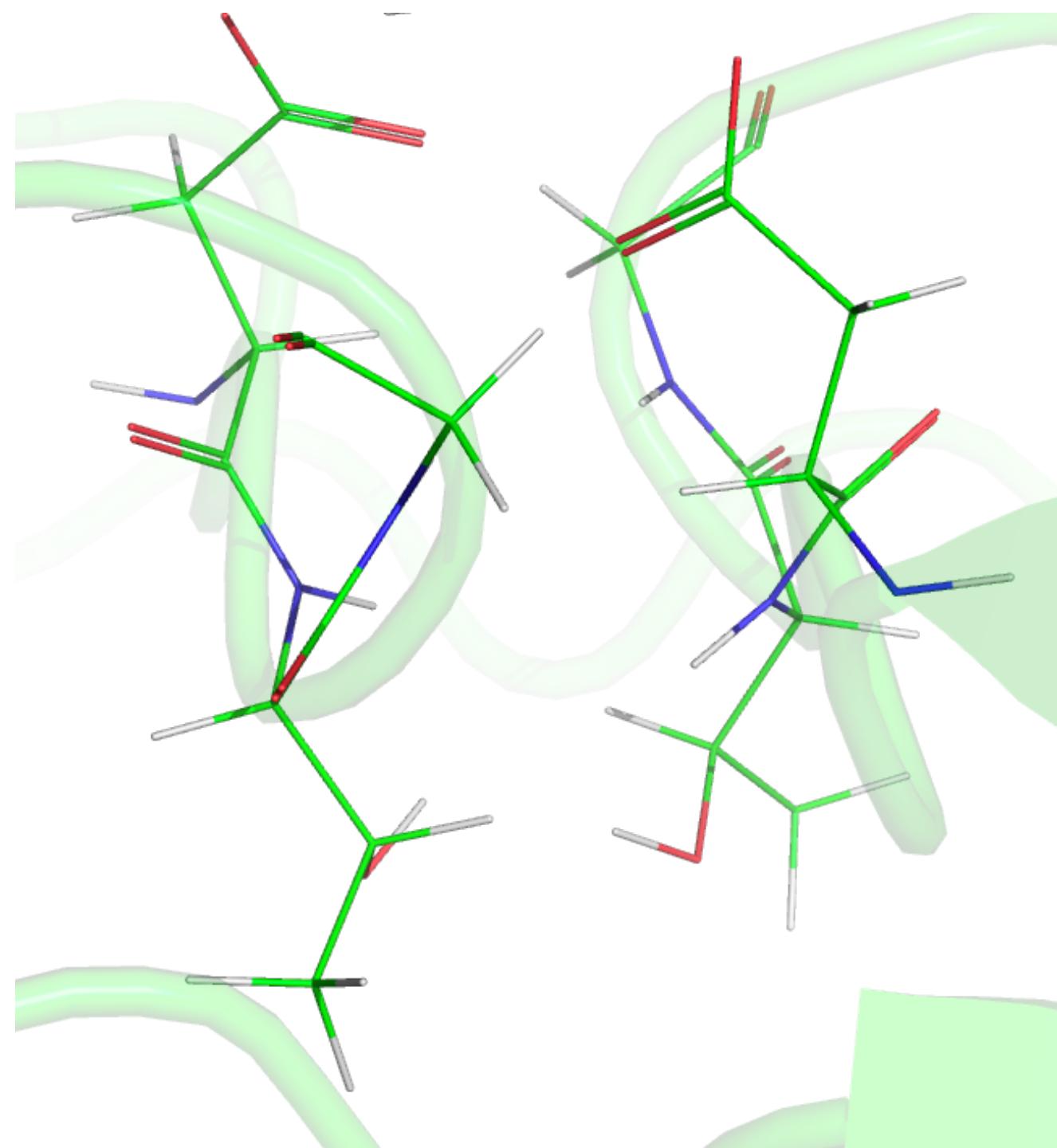


**Learn More:** Senn, H. M. & Thiel, W. QM/MM methods for biomolecular systems.  
*Angewandte Chemie Int Ed Engl* **48**, 1198–229 (2009).



# QM/MM: mixing QM and Classical MD simulations

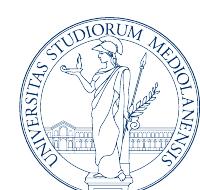
The idea of QM/MM is to simulate part of the system at the QM level, for example the active site, and the remainder using classical MD.



## **The main issues are:**

1. How to separate the QM and MM regions;
2. How to describe the interaction between the two parts (QM and MM).

To separate the QM and MM regions, amino acids are usually cleaved at C<sub>b</sub>, unless the backbone forms relevant interactions with the substrate. A number of hydrogens are added to saturate the bond (complete the valence shell).



# QM/MM interactions: the ONIOM subtractive scheme

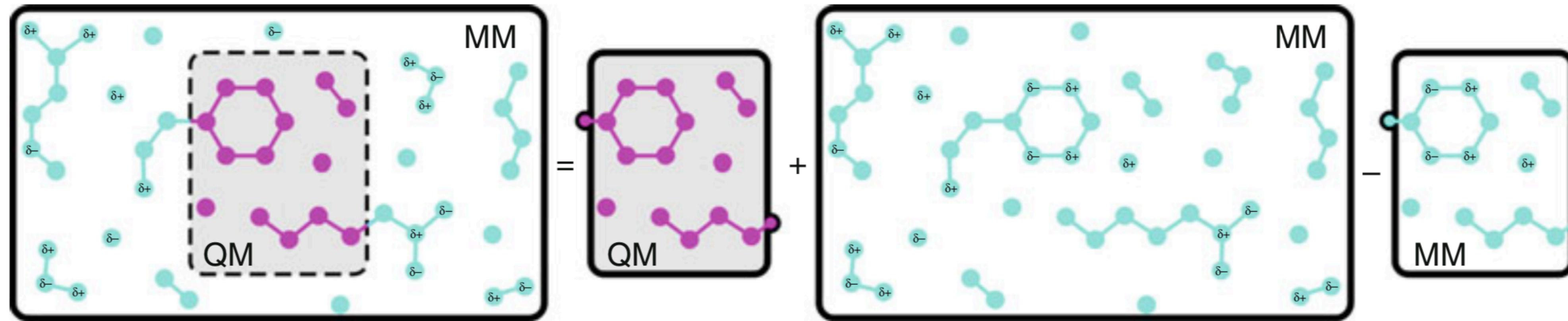


Fig. 2. Subtractive QM/MM coupling: The QM/MM energy of the total system (*left hand side of the equation*) is assumed to be equal to the energy of the isolated QM subsystem, evaluated at the QM level, plus the energy of the complete system evaluated at the MM level, minus the energy of the isolated MM subsystem, evaluated at the MM level. The last term is subtracted to correct for double counting of the contribution of the QM subsystem to the total energy. A prerequisite for the calculation is that a force field for the QM subsystem is available.

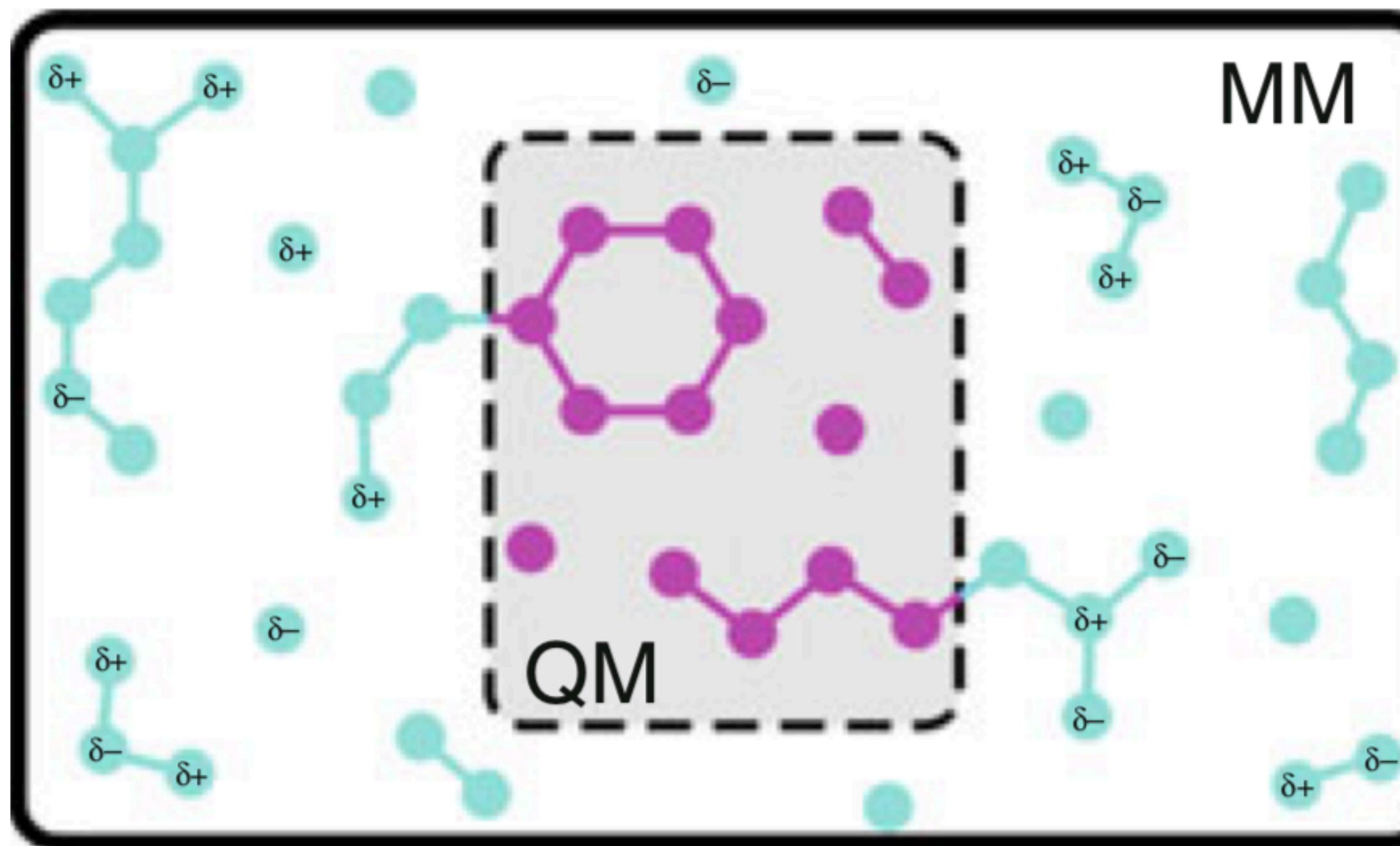
$$V_{\text{QM/MM}} = V_{\text{MM}}(\text{MM} + \text{QM}) + V_{\text{QM}}(\text{QM}) - V_{\text{MM}}(\text{QM}).$$

**In this scheme QM nuclei feel MM forces from the MM region but electrons are not affected.**



# QM/MM: and the additive scheme

The effect of the MM region is added to them QM by considering the electric field generated by the point charges, the QM electron distribution is used to generate additional forces on the MM region.



$$H = H_{DFT} + \sum_{MM} \frac{q}{r}$$

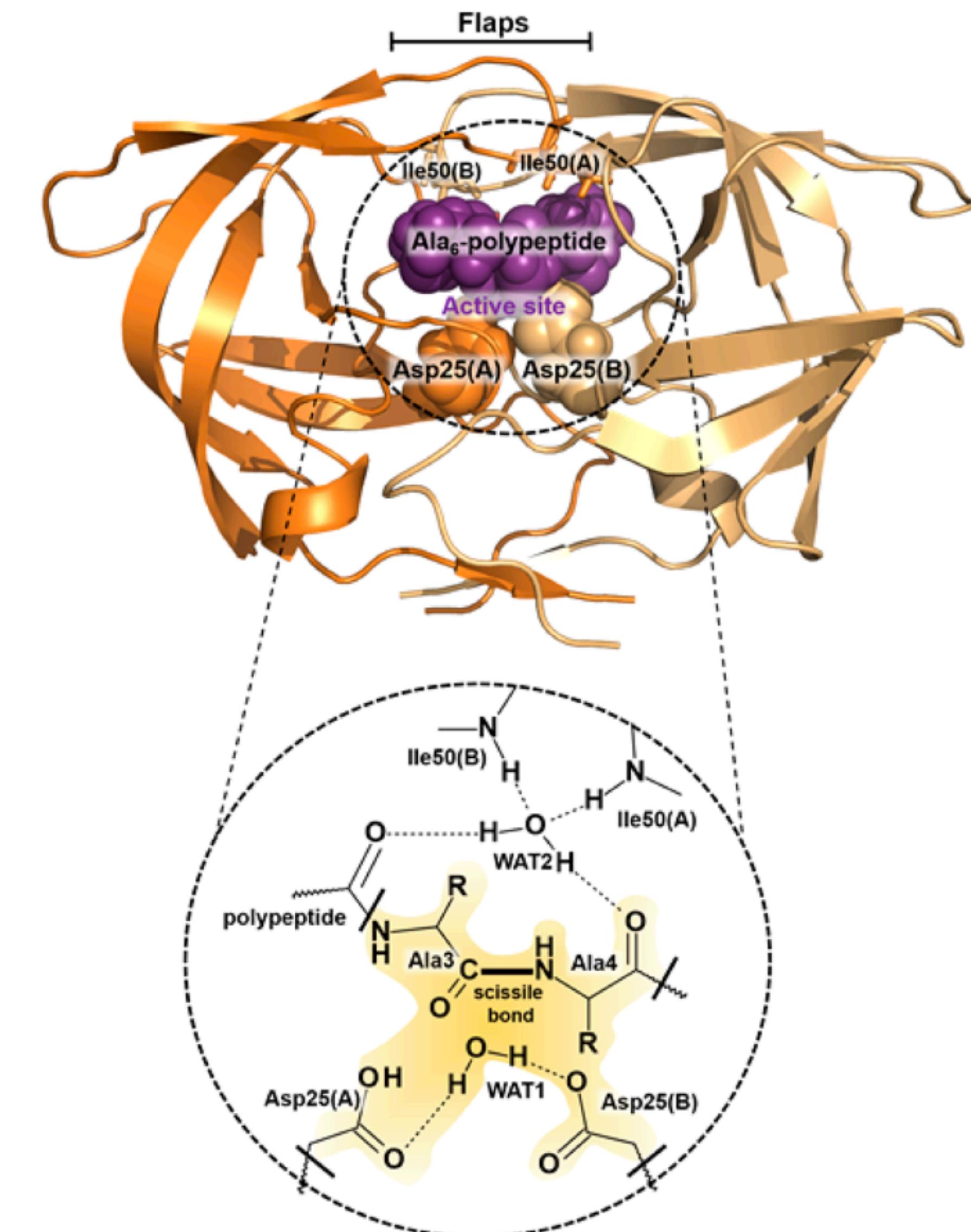
The energy is the DFT energy plus an additional external potential. This open a new issue for ‘classical charges’ that are very close to the border. This potential can be too attractive (take out electrons from the box) or too repulsive (push electrons away from the border), but there are corrections that can be added.



# Example: HIV-1 protease activity

How does a protease cleave a peptide bond?  
Catalytic residues are treated at the QM level, including a water molecule, the rest of the protein contributes with an electric field resulting from all the point charges derived from the force field.

QM/MM simulations allow the system to evolve and the reaction to be observed. Multiple pathways are observed and their relative energies can be calculated and compared with experiments.

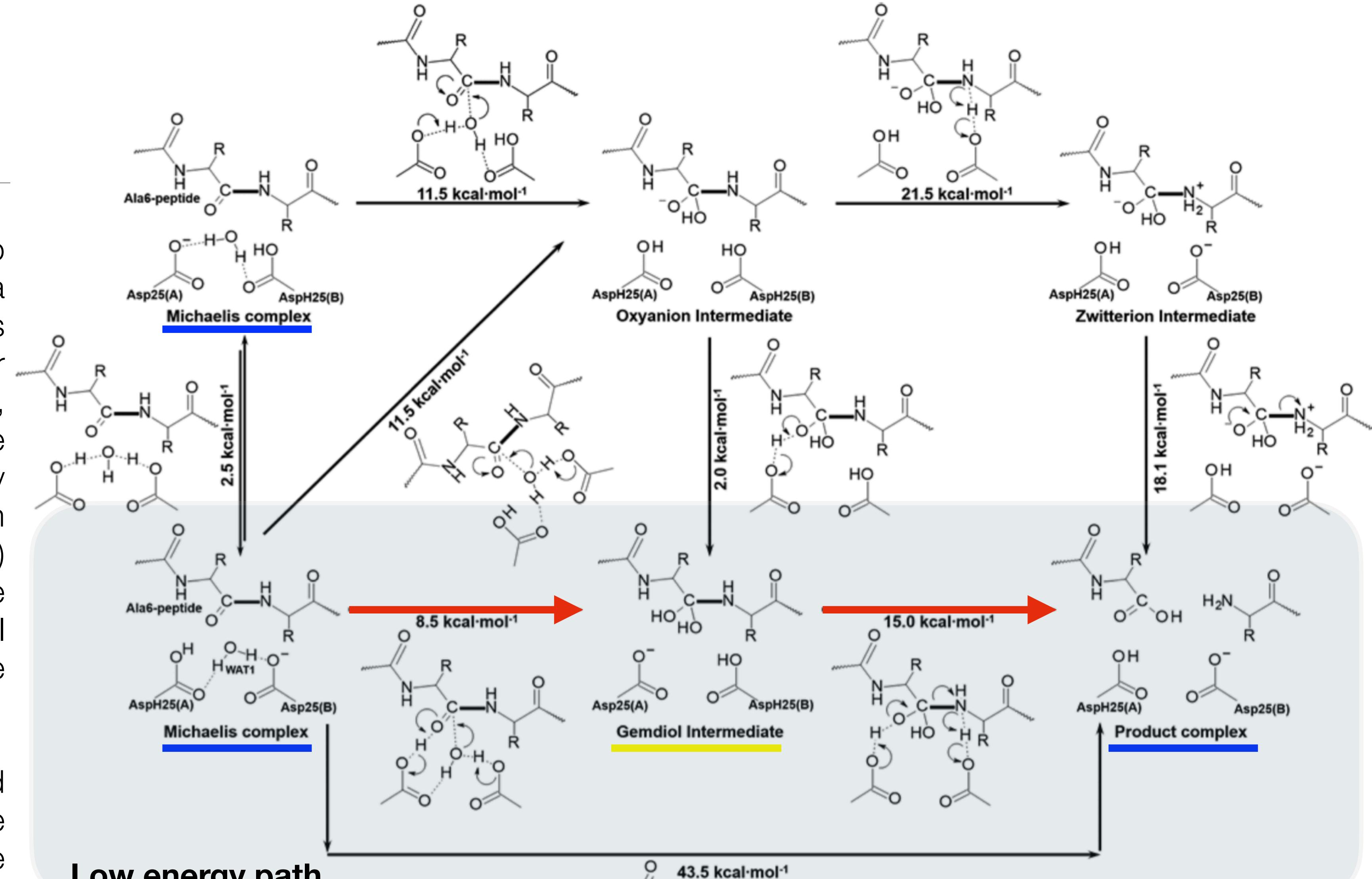


**Figure 1.** Structure of HIV-1 PR and detail of the active site, with protonated Asp-25(A), and Ala<sub>6</sub> peptide as substrate. Yellow area in the bottom panel contains the part of the system treated at QM level of theory during all the QM/MM calculations. Four link atoms are indicated as thick black lines.

# HIV-1 protease

The reaction pathway is a two-step process. First, a nucleophilic attack of a water molecule on the C of Ala3 is accompanied by a hydrogen transfer from this water molecule to Asp25(B), resulting in a gemdiol intermediate. The peptide bond is then concertedly cleaved by a double proton transfer from the oxygen of the protonated Asp25(B) to the nitrogen of the scissile peptide bond and from one of the hydroxyl groups of the carbon of the peptide bond to Asp25(A).

The rate-limiting step would correspond to the decomposition of gemdiol into the product complex, 15.0 kcal/mol, close to the experimental 15.1-17.9 kcal/mol.



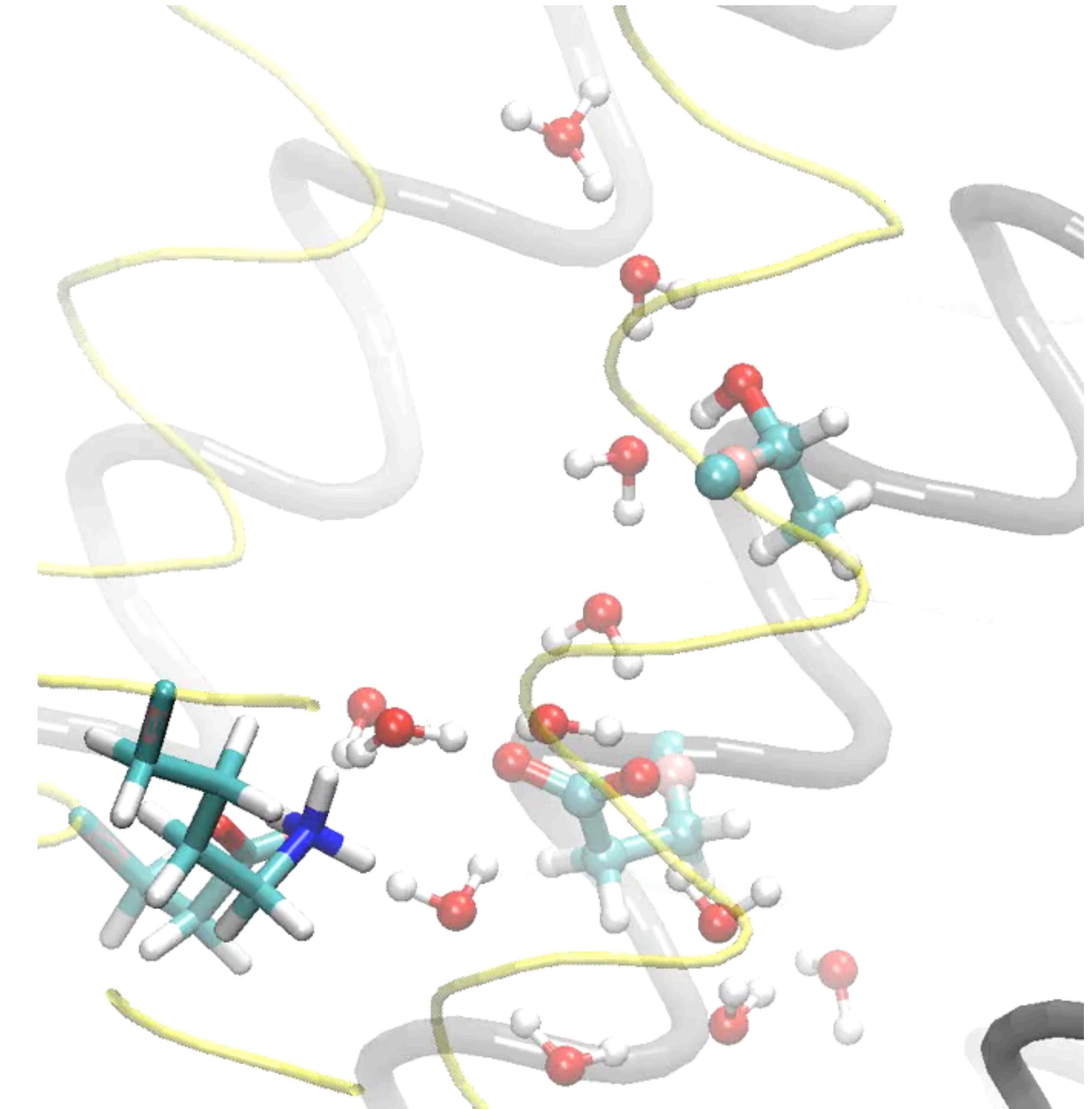
Krzemińska, A., Moliner, V. & Świderek, K. Dynamic and Electrostatic Effects on the Reaction Catalyzed by HIV-1 Protease. *J Am Chem Soc* **138**, 16283–16298 (2016).



# Respiratory complex I

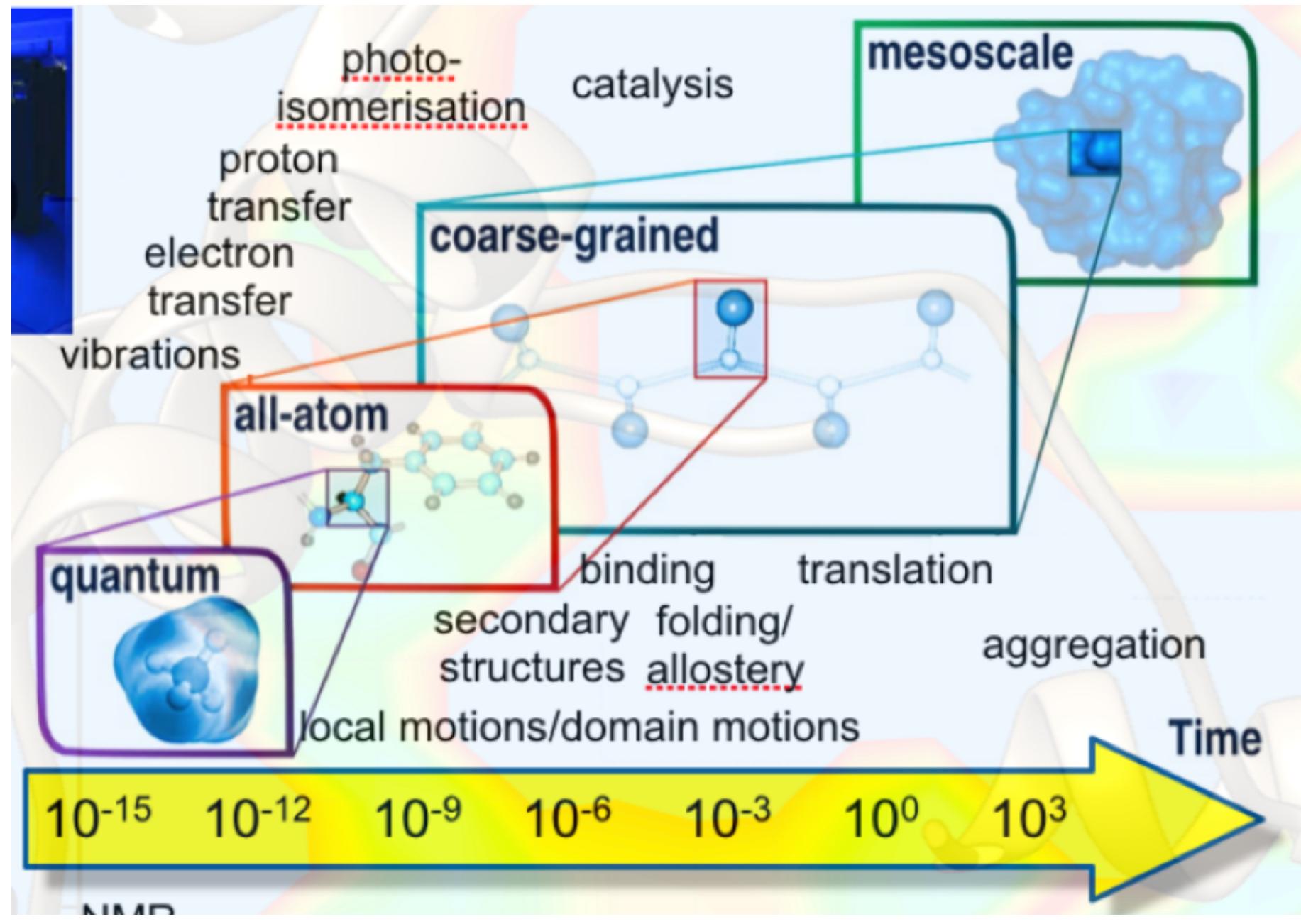
Electron transfer by metal ions. Reaction with quinone, which force a conformational change in the transmembrane part change the charge state as a result of which a proton can pass through.

- Standard MD suggests the hydration of some channel starting from the X-Ray structure.
- Modifying the protonation state of charged residues in the channel result in a dehydration.
- This suggest that a conformational change is not needed.
- Is the hydrated channel enough to transfer protons?
- QM simulation with water and charged residues in the QM region (~5 ps longs) show that a proton can be transferred through the wire in less than 1 ps.





# Simple and complex models



Complexity can take into account both the accuracy of the interactions and the spatial resolution (i.e. electrons/atoms/amino-acids/etc).

Different processes may require different simulation techniques.

Size and time scale are two key issues related to the sampling problem. The larger the system, including solvent if applicable, the slower the simulation.

Long time scales can be inaccessible even for small systems given the short time scale of the time step (fs).





# Simplified Models

The aim of simplifying models in simulations is both to extend the accessible time scale and to simulate ever larger systems.

Simplification of a model can be achieved, for example, by

- Reducing the resolution, in the same way that we have reduced the resolution of electrons/nuclei by using classical atoms.
- Removing part of the system (for example, removing the solvent and somehow implicitly accounting for that in some other way)
- Change the way the atoms/beads interact with each other, for example in classical MD simulations electrostatic interactions are the most computationally expensive part.





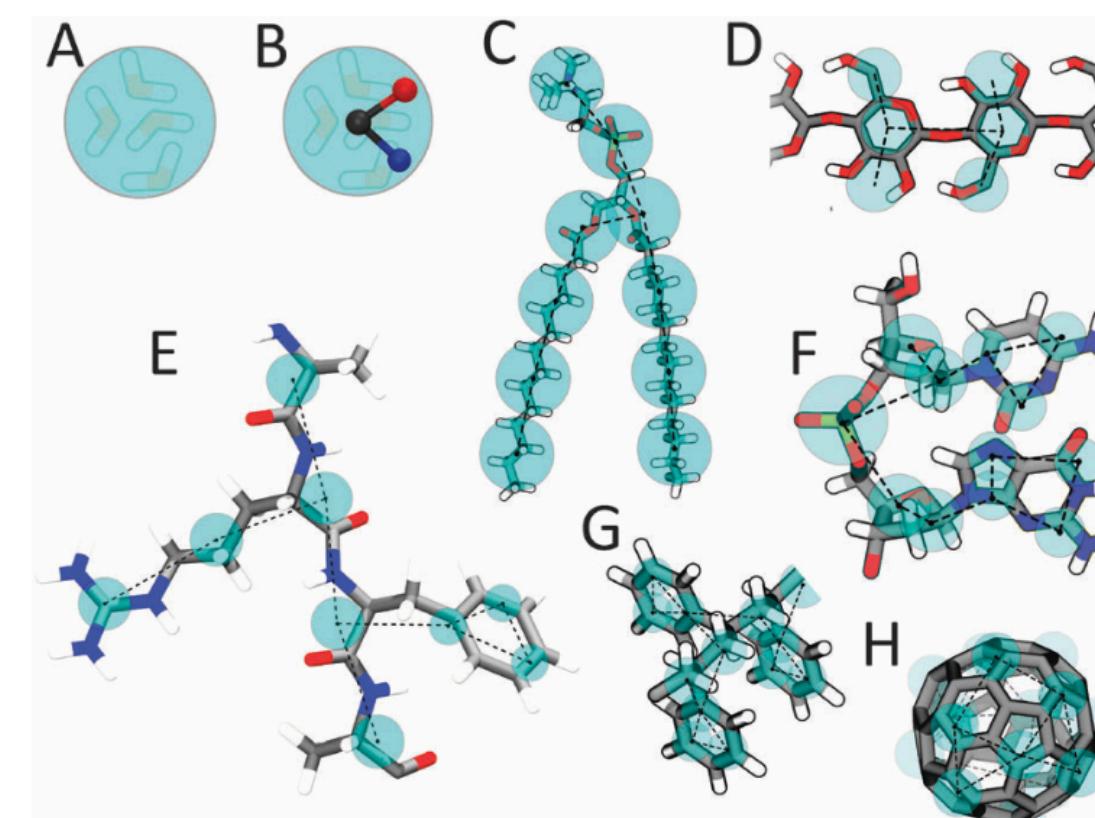
# Simplified Models

The aim of simplifying models in simulations is both to extend the accessible time scale and to simulate ever larger systems.

**There are two main strategies for simplifying a model:**

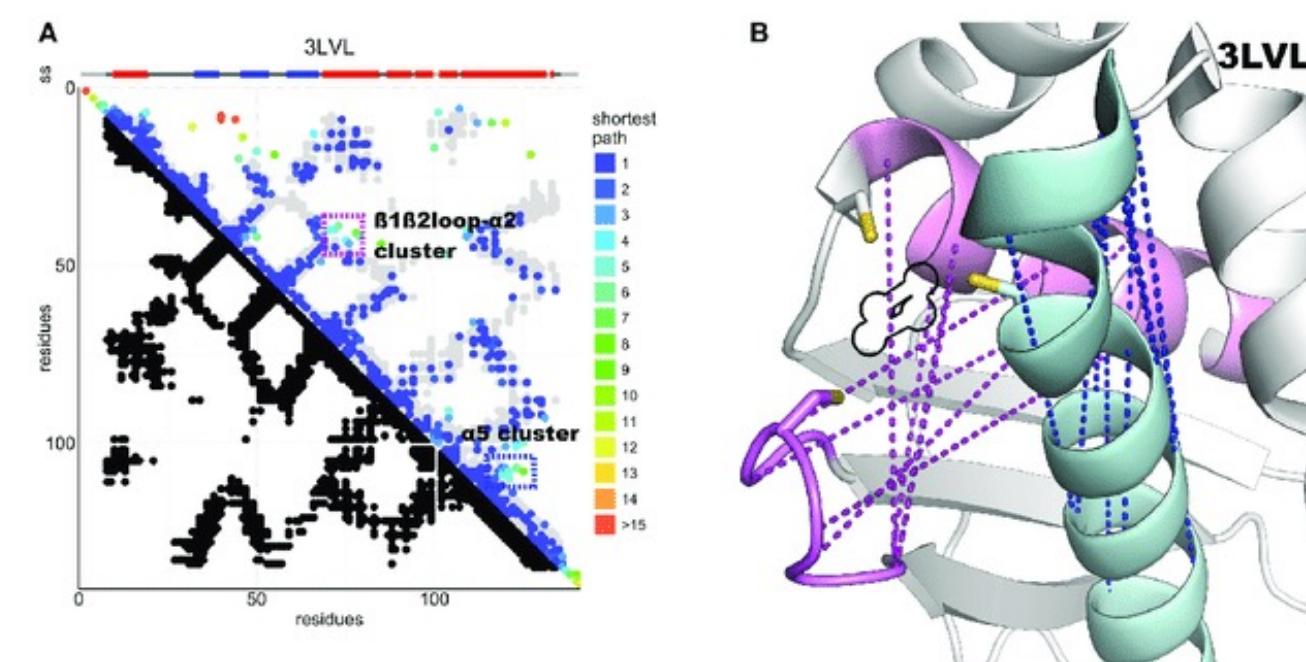
## Phys/Chem Based

In this case the interactions try to preserve relevant physico-chemical properties such as the polar/hydrophobic/charged nature of amino acids. An example is the MARTINI force field



## Knowledge Based

In this case, the interactions attempt to reproduce other sources of knowledge, e.g. sequence conservation profiles, co-evolutionary analysis, structural knowledge. An example is Go models (also known as structure-based models).

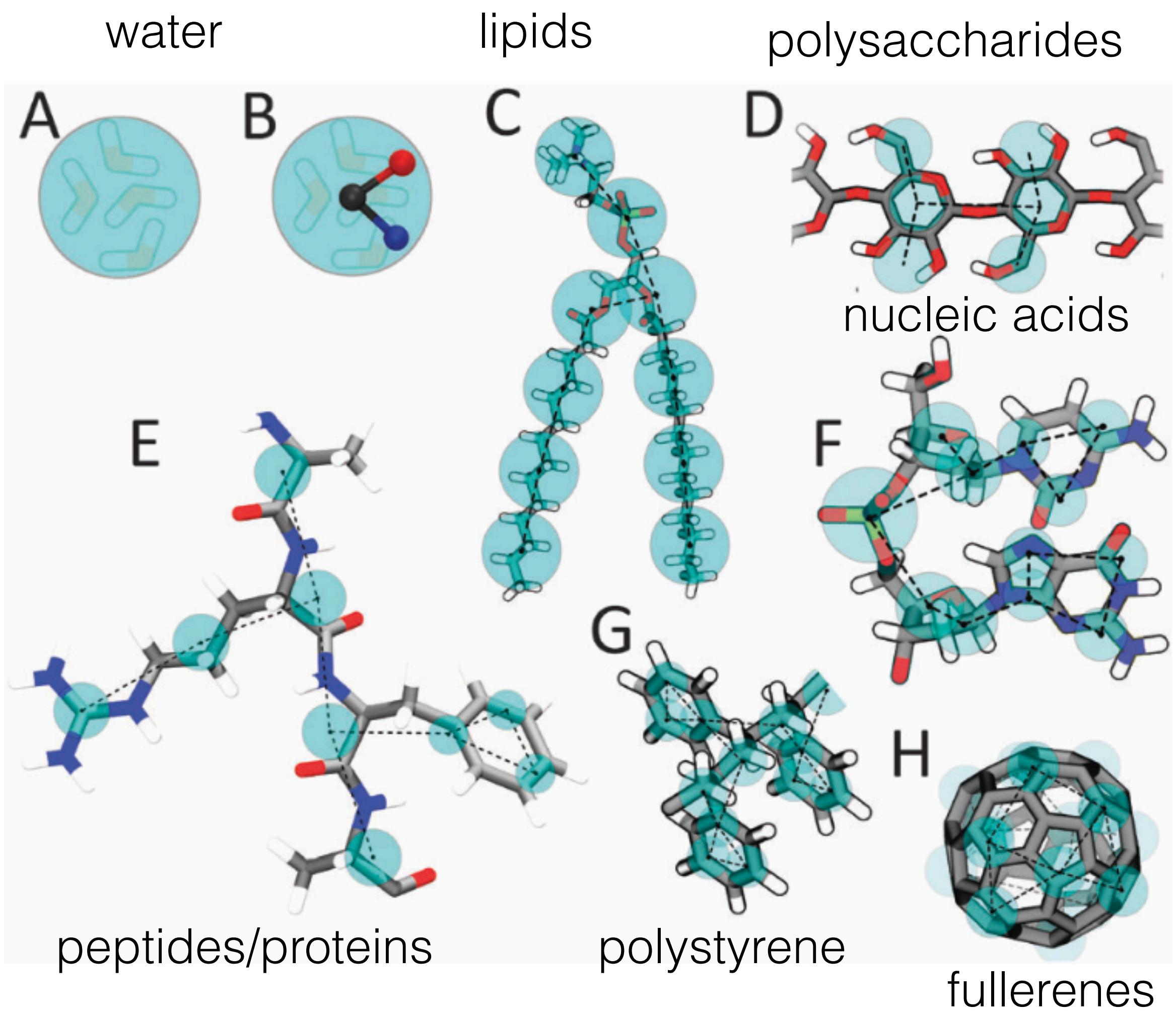




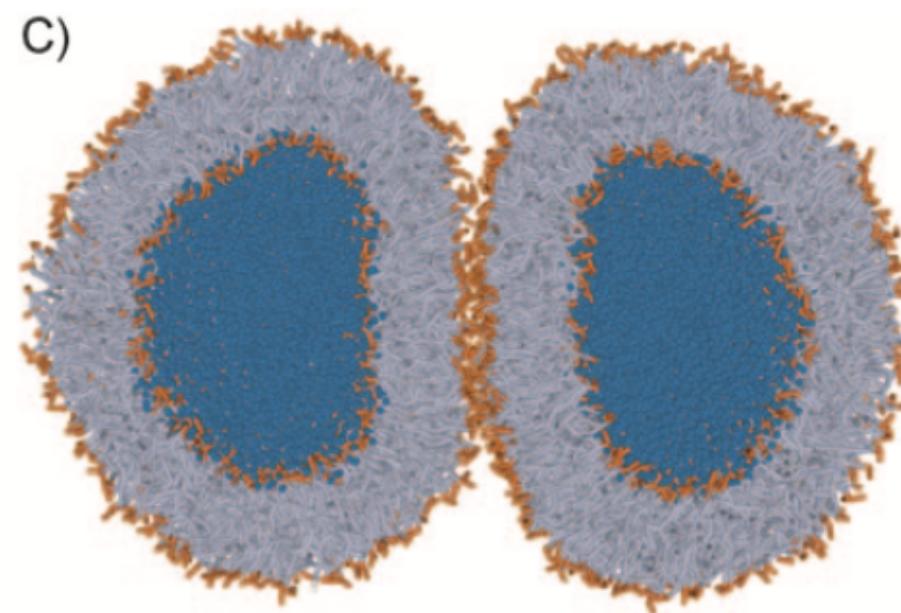
# Martini: a transferable Coarse Grain model

**Aim:** a coarse-grained model with a transferable, physically based potential, capable of simulating very large systems with quasi-chemical resolution.

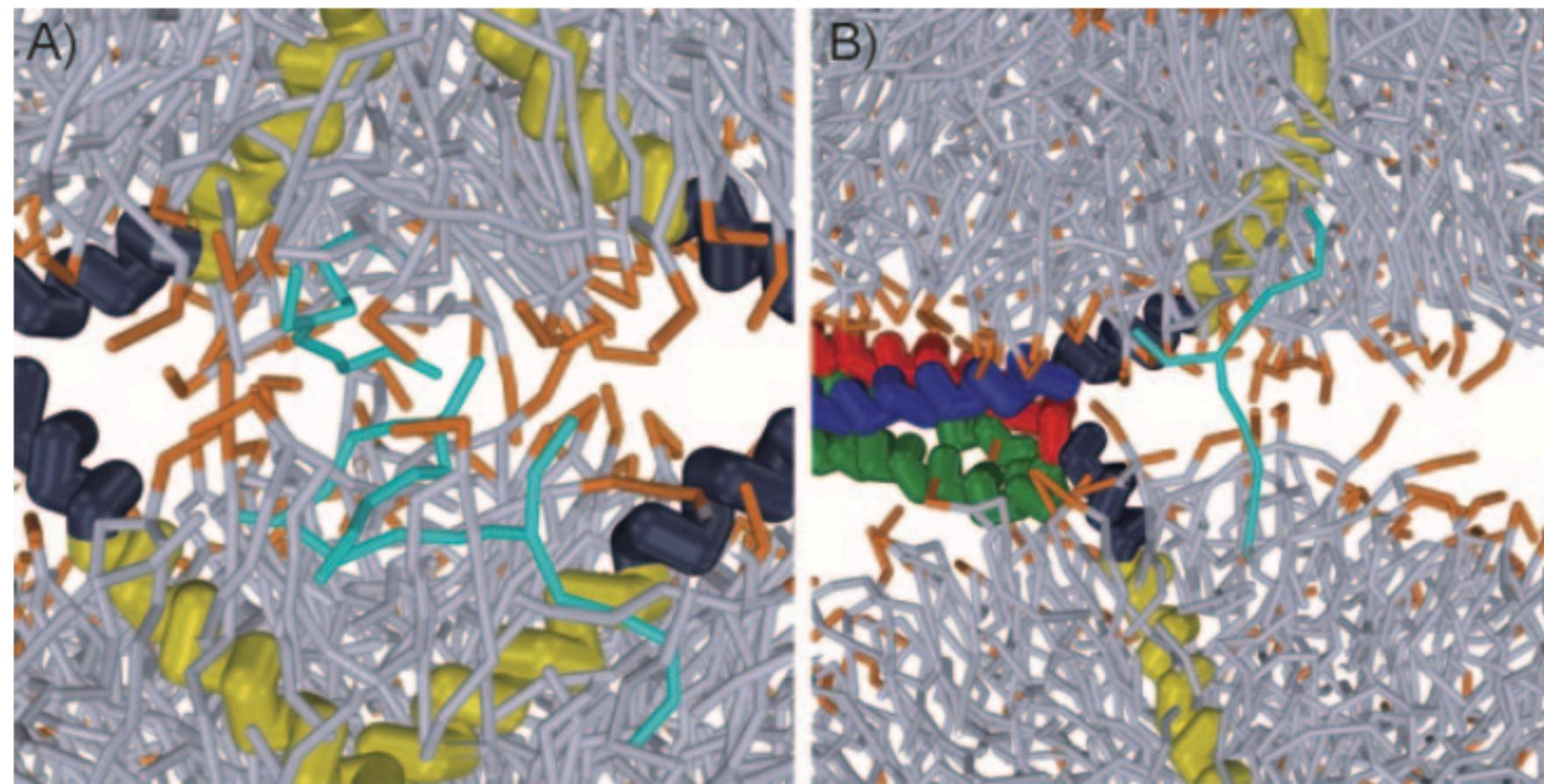
The main idea is to use a 4 to 1 mapping (4 heavy atoms and their hydrogens in a single particle), except for rings where the mapping is 2 to 1. Then define a set of building blocks and fit the force field to reproduce thermodynamic behaviour.



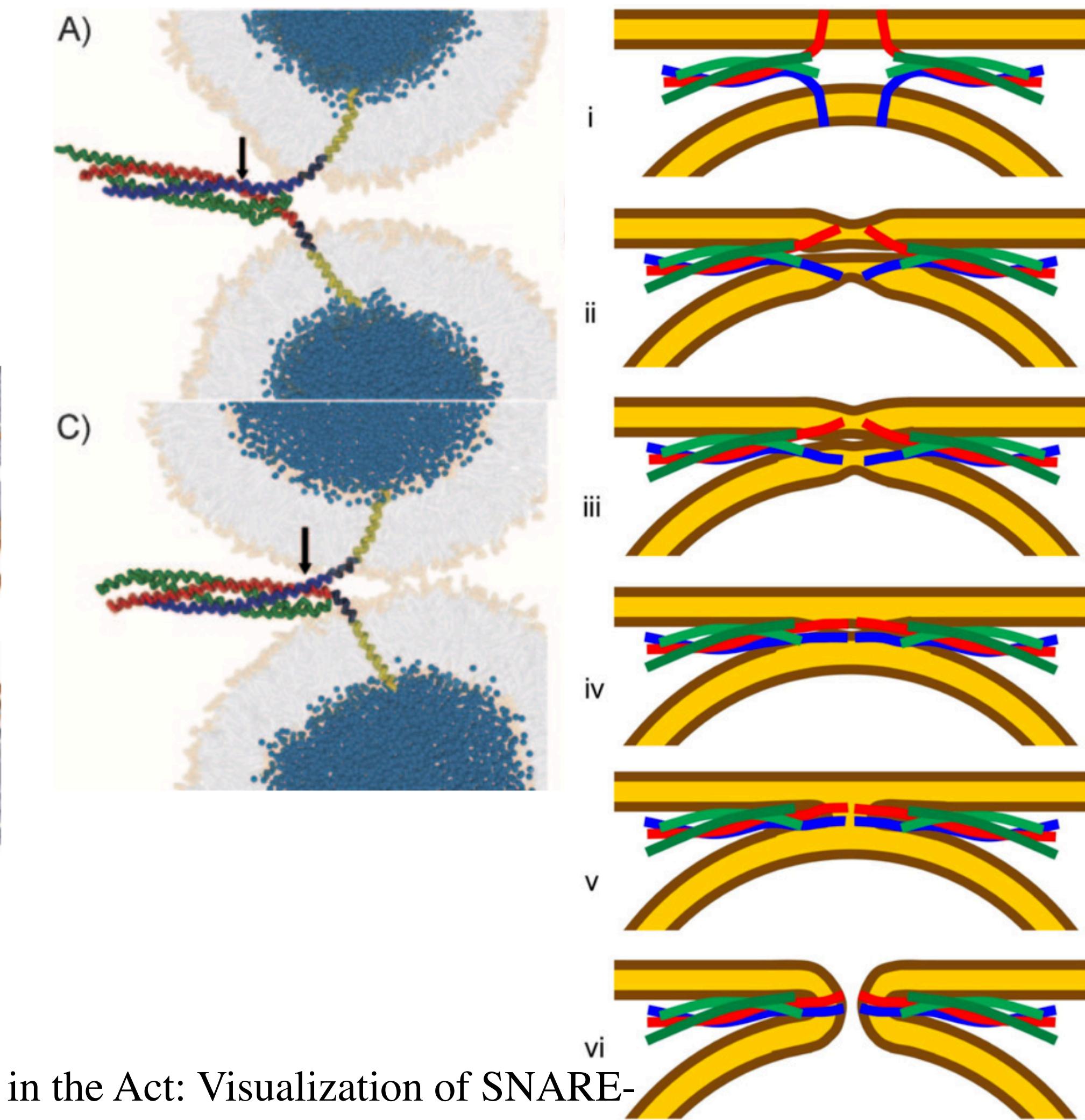
# Martini examples: Vesicular Fusion



3 millions beads ~  
30 millions atoms  
4 us of simulation



Vesicular fusion is a key transport process inside and among cells. While it can happen spontaneously it is often regulated by proteins. In particular the SNARE complex is a key element



Risselada, H. J., Kutzner, C. & Grubmüller, H. Caught in the Act: Visualization of SNARE-Mediated Fusion Events in Molecular Detail. *Chembiochem* **12**, 1049–1055 (2011).

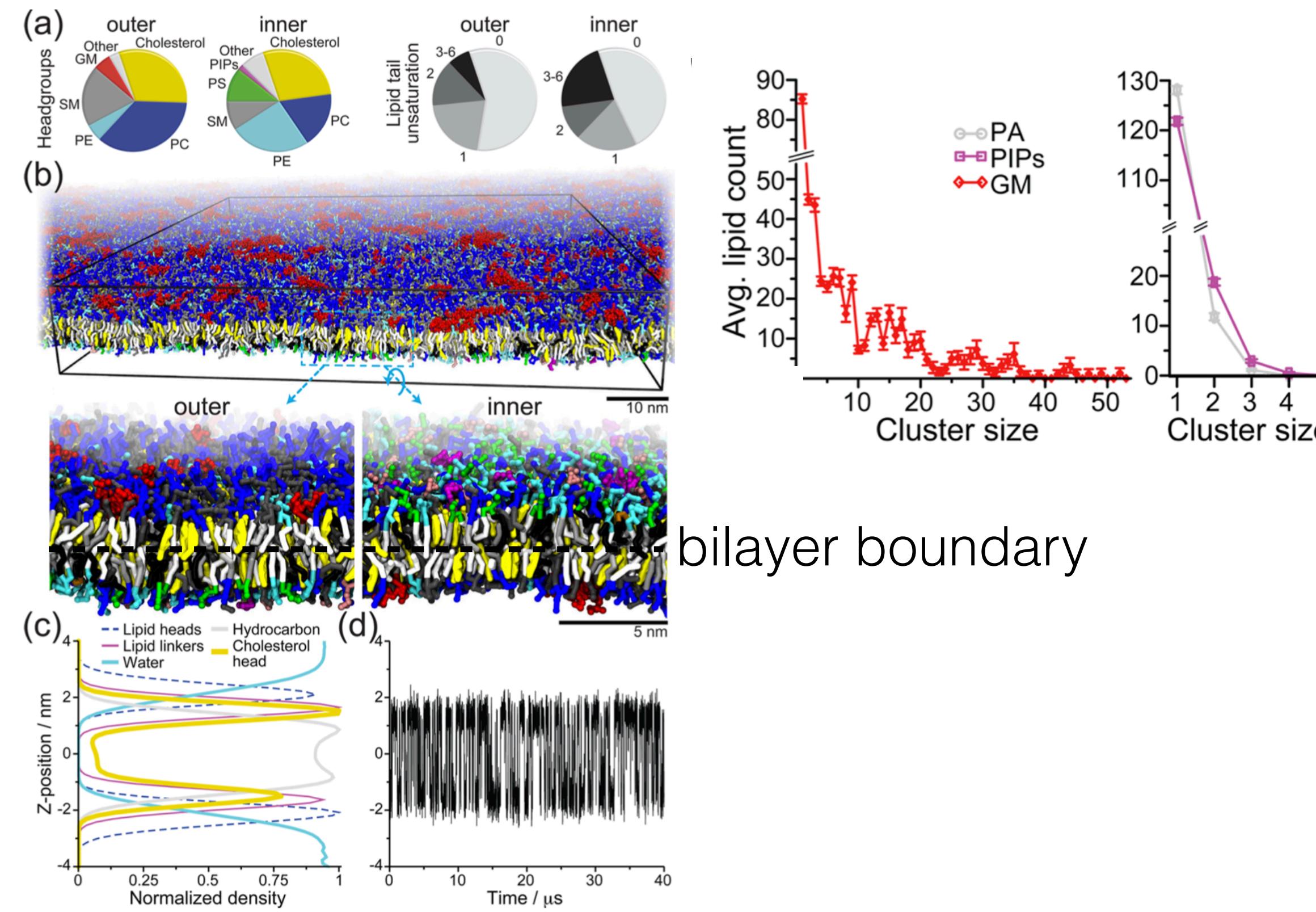


UNIVERSITÀ  
DEGLI STUDI  
DI MILANO



# Martini Examples: Visualising a Plasma Membrane

63 types of lipids with charged species in the inner membrane and glycolipids in the outer membrane, equal distribution of cholesterol in both.



## Observations

1. GlycoLipids (GMs) cluster more than PhosphoLipids (PA/PIPs).
2. Cholesterol diffuses very rapidly between inner and outer leaflet
3. Cholesterol is more concentrated in the outer leaflet than in the inner leaflet (54:46).





# Structure-Based models

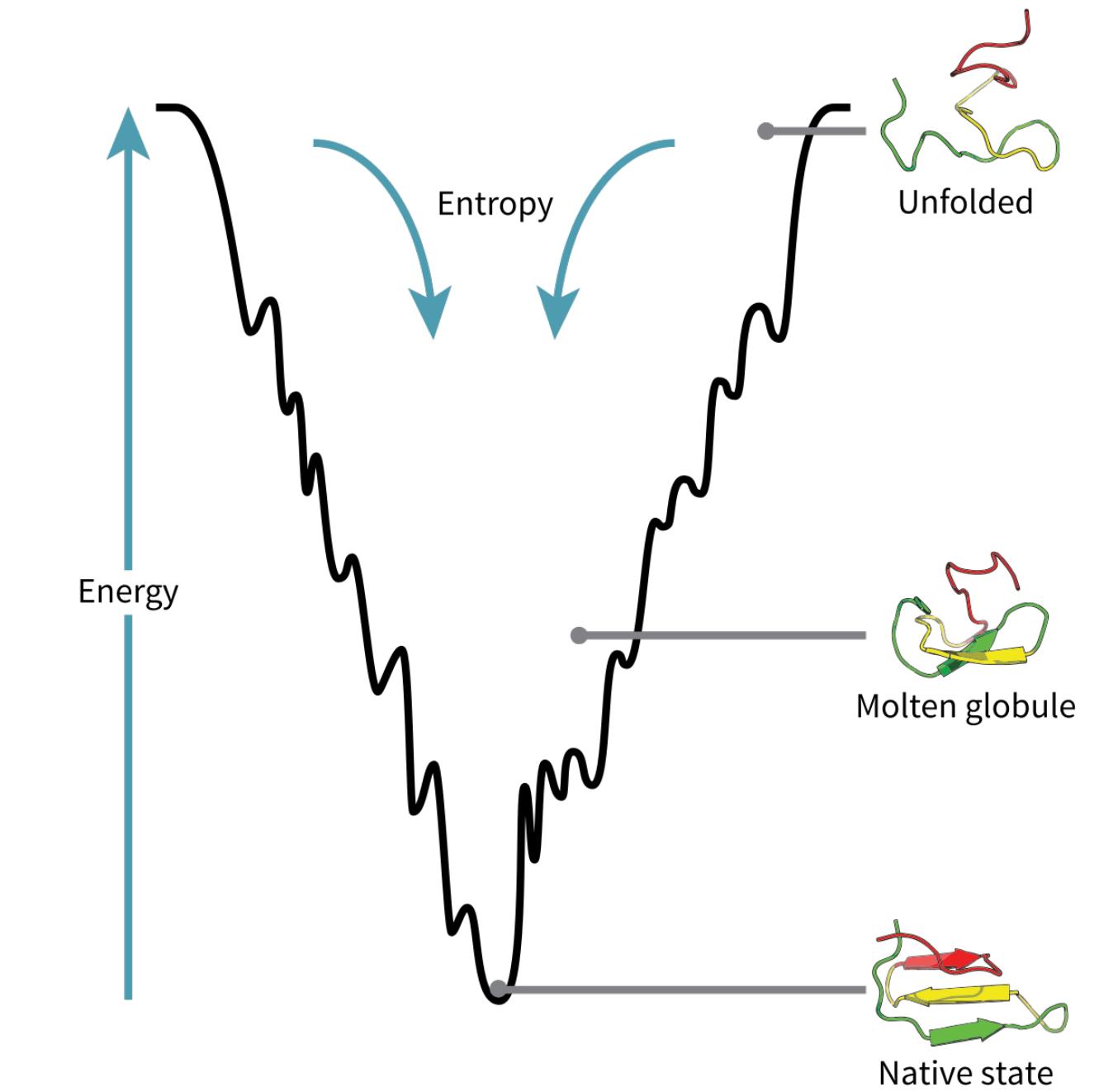
Structure-based models for proteins originated from the observation that

- folded proteins have a very well defined 3D structure
- protein 3D structure is very stable, so there must be a free energy and a potential energy minimum for that sequence of amino acids.

As a consequence, the spatial organisation of the atoms in 3D should reflect near-optimal interactions between the atoms.

So given a protein PDB structure we can parameterise a force-field based only on the geometry of that structure:

$$H(x; x_m, X_a) = \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_q (\theta - \theta_0)^2 + \sum_{improper} K_\phi [1 + \cos(n\phi - \phi_0)] \\ + \sum_{dihedrals} K_\psi [1 + \cos(n\psi - \psi_0)] + \sum_{native} \varepsilon \left[ \left( \frac{r_{ij,m,a}}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij,m,a}}{r_{ij}} \right)^6 \right] + \sum_{others} \frac{c_{ij}^{(12)}}{r_{ij}^{12}},$$

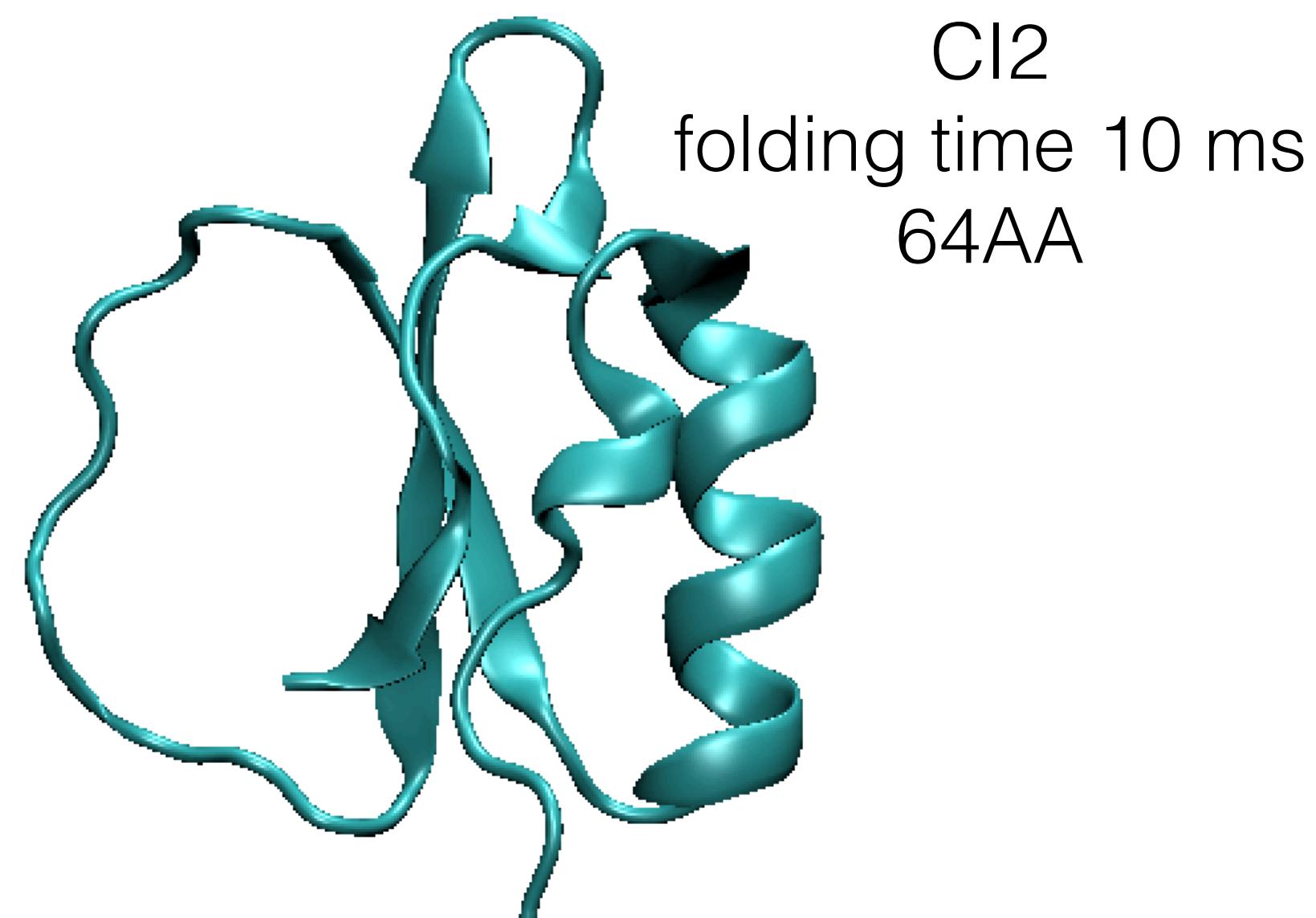


Instead of having parameters valid in general, we have parameters valid for the specific protein structure, in this way the most stable configuration is by definition the protein structure.

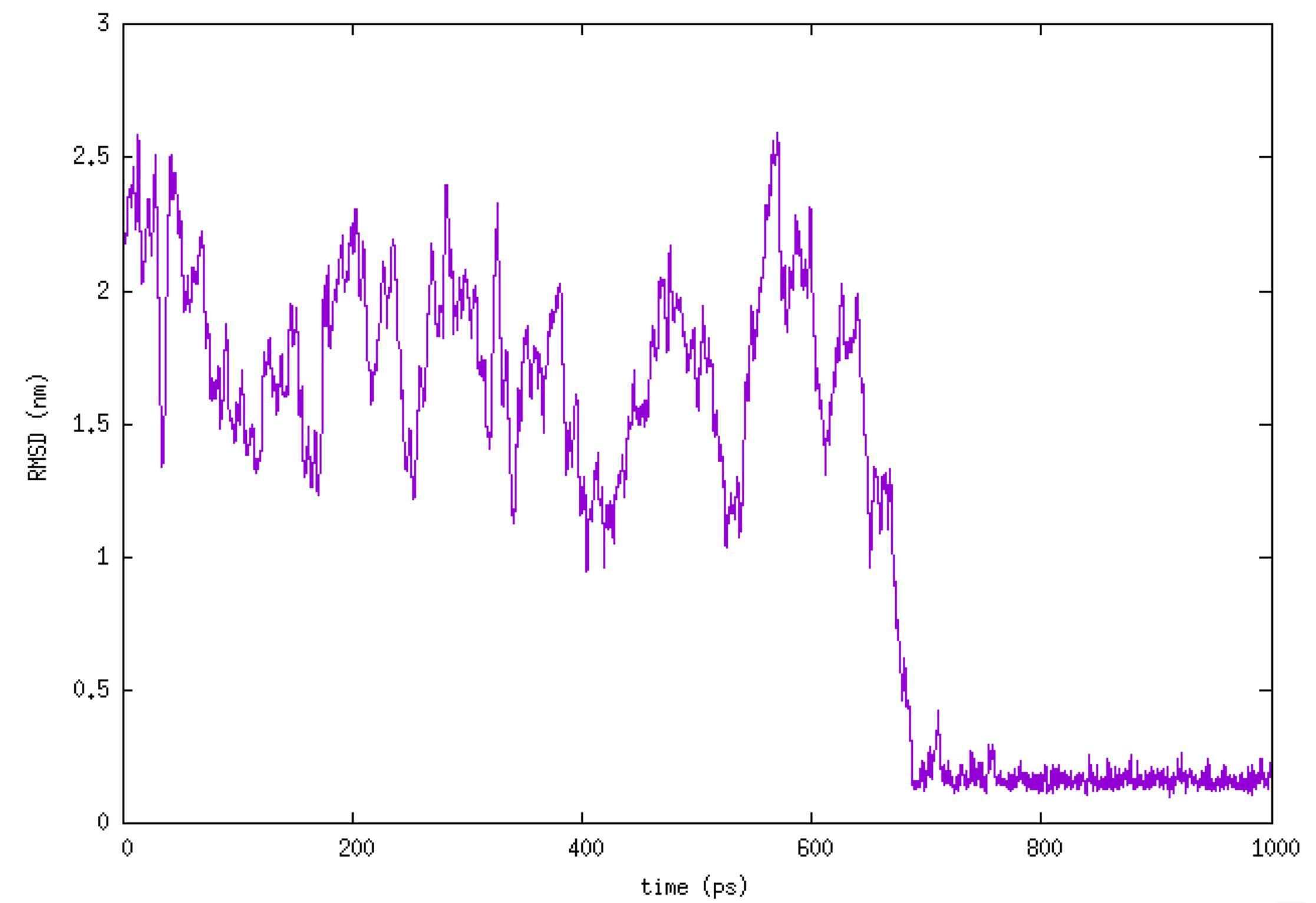
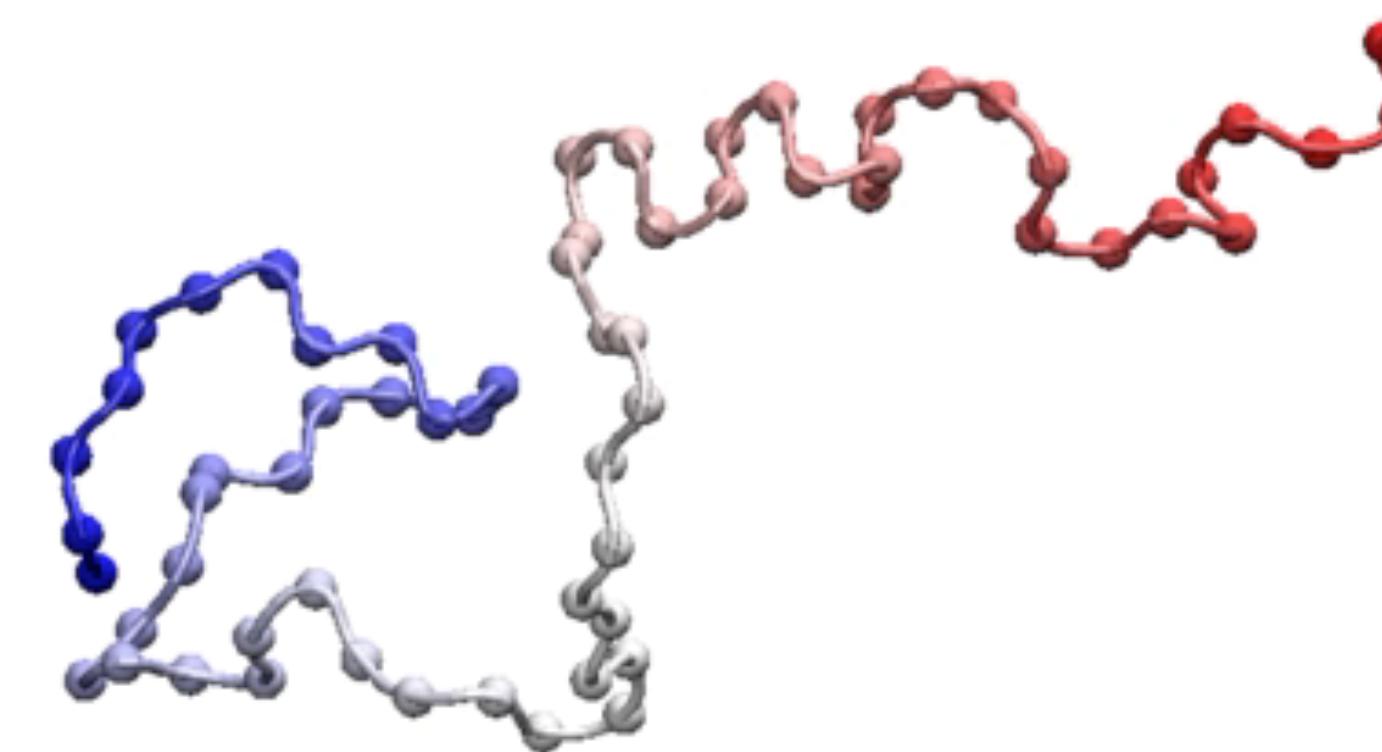


# Structure-Based models

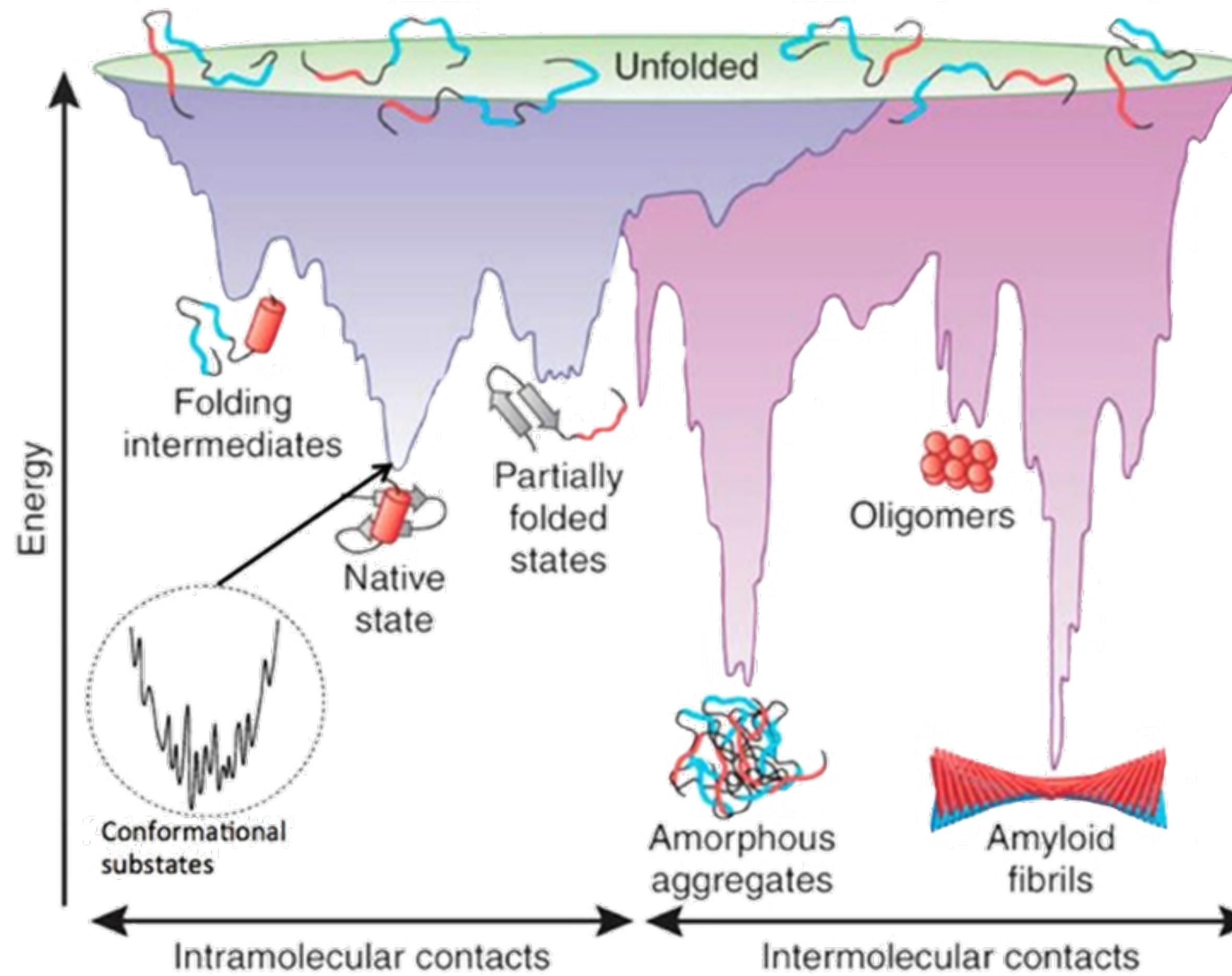
In the '90-early 2000 this approach was used to study the mechanism of protein folding, nowadays it is used (not very often) to study the motion of very large proteins



UNIVERSITÀ  
DEGLI STUDI  
DI MILANO



# Structure-Based models for protein aggregation and more



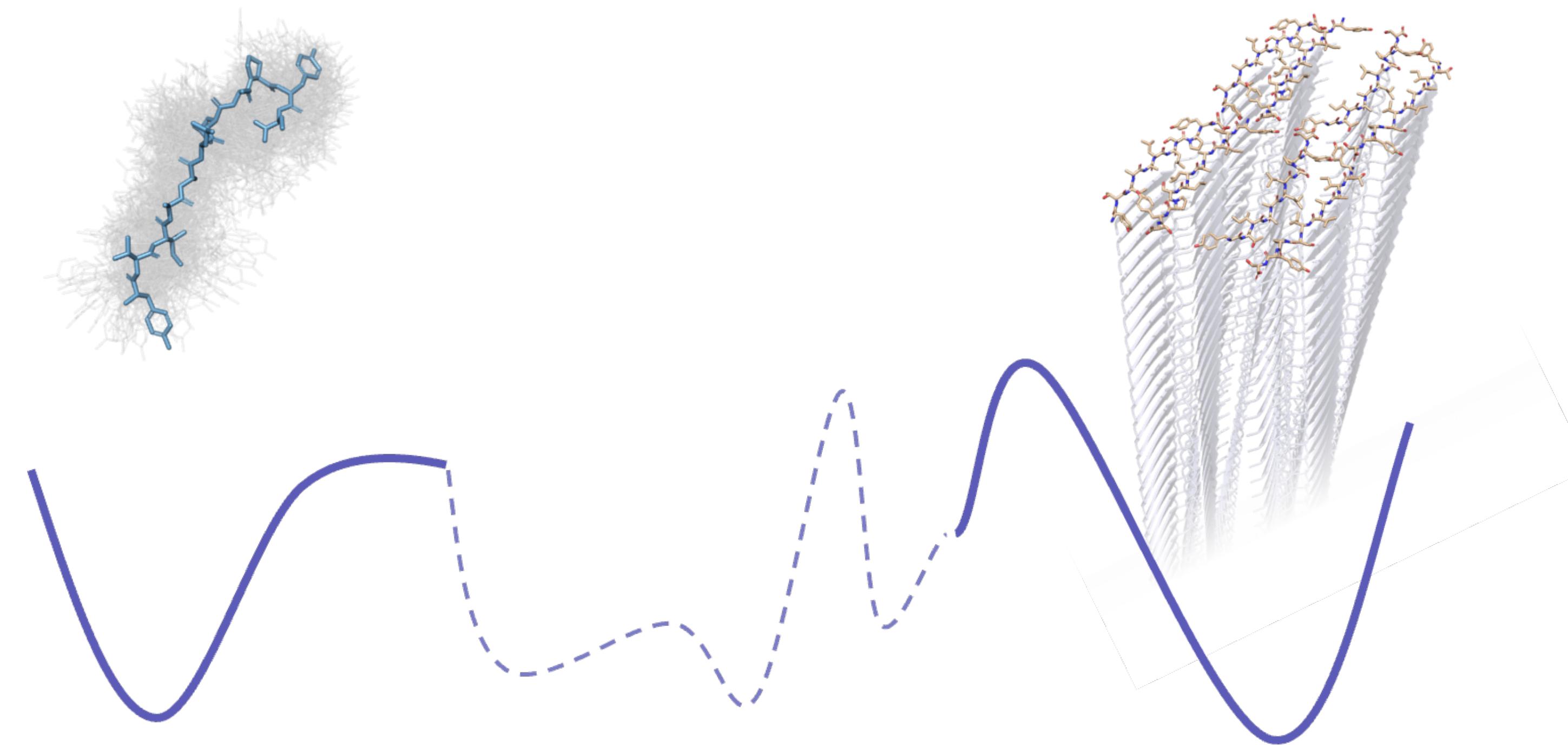
The general understanding of the behaviour of a protein in solution is that its native and amyloid states are the only free energy minima characterised by a well defined structure.

We can use this information to build a multi-Structures-Based model to simulate protein folding/unfolding/aggregation

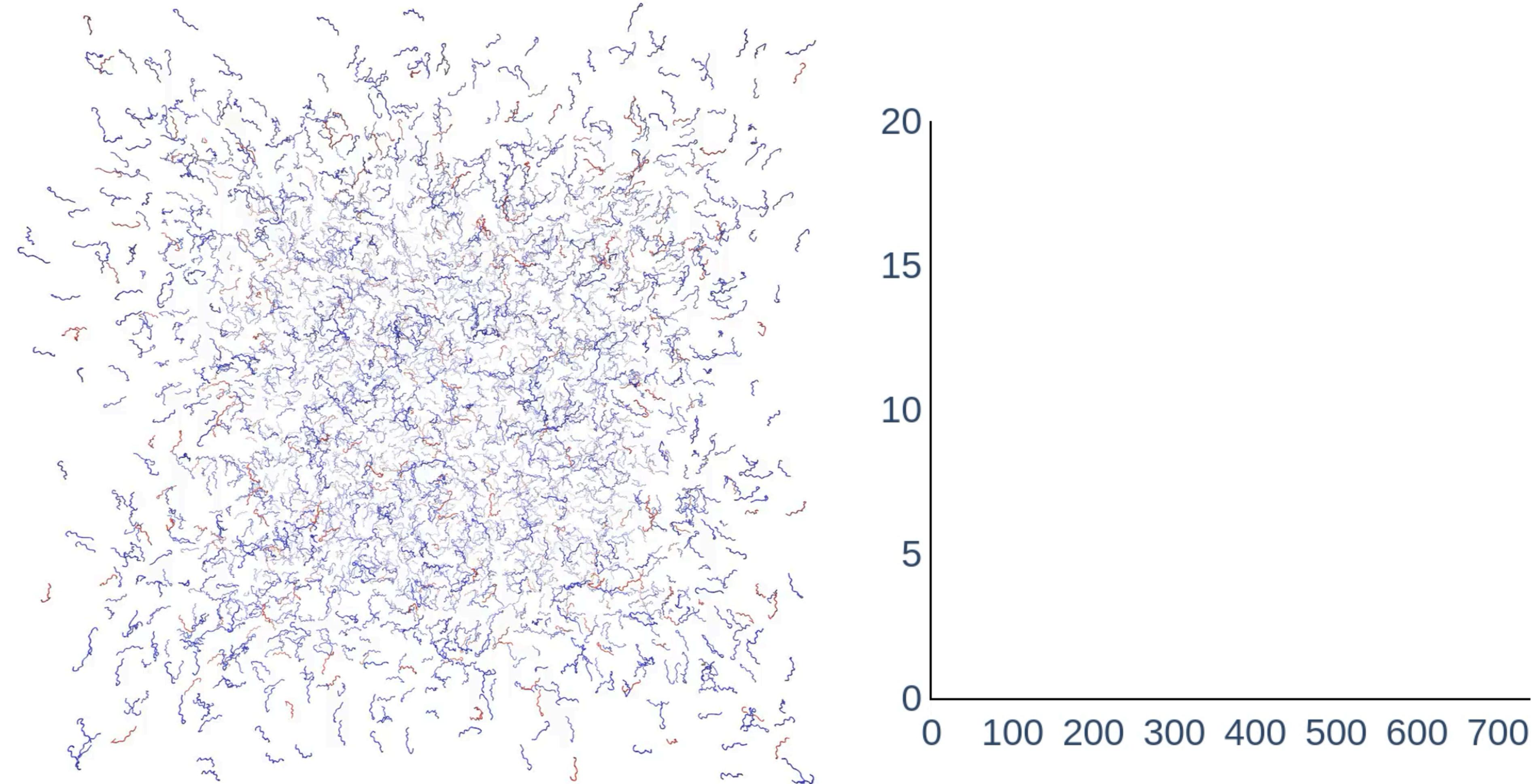


# Structure-Based models for protein aggregation and more

As a further step instead than learning from a native structure, we can learn from a simulation of the native state, thus enabling also the study of disordered proteins



# Structure-Based models for protein aggregation and more

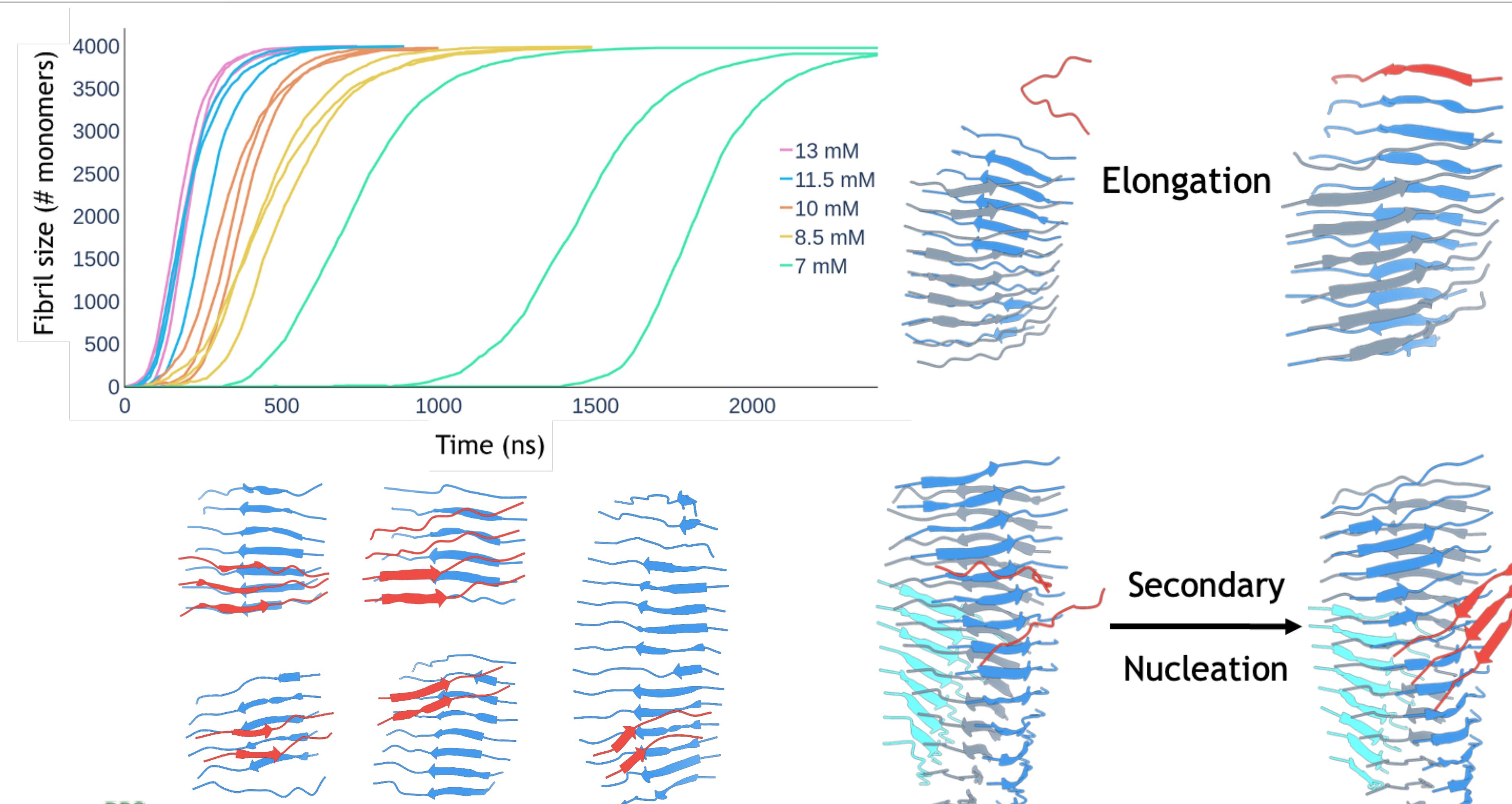


UNIVERSITÀ  
DEGLI STUDI  
DI MILANO

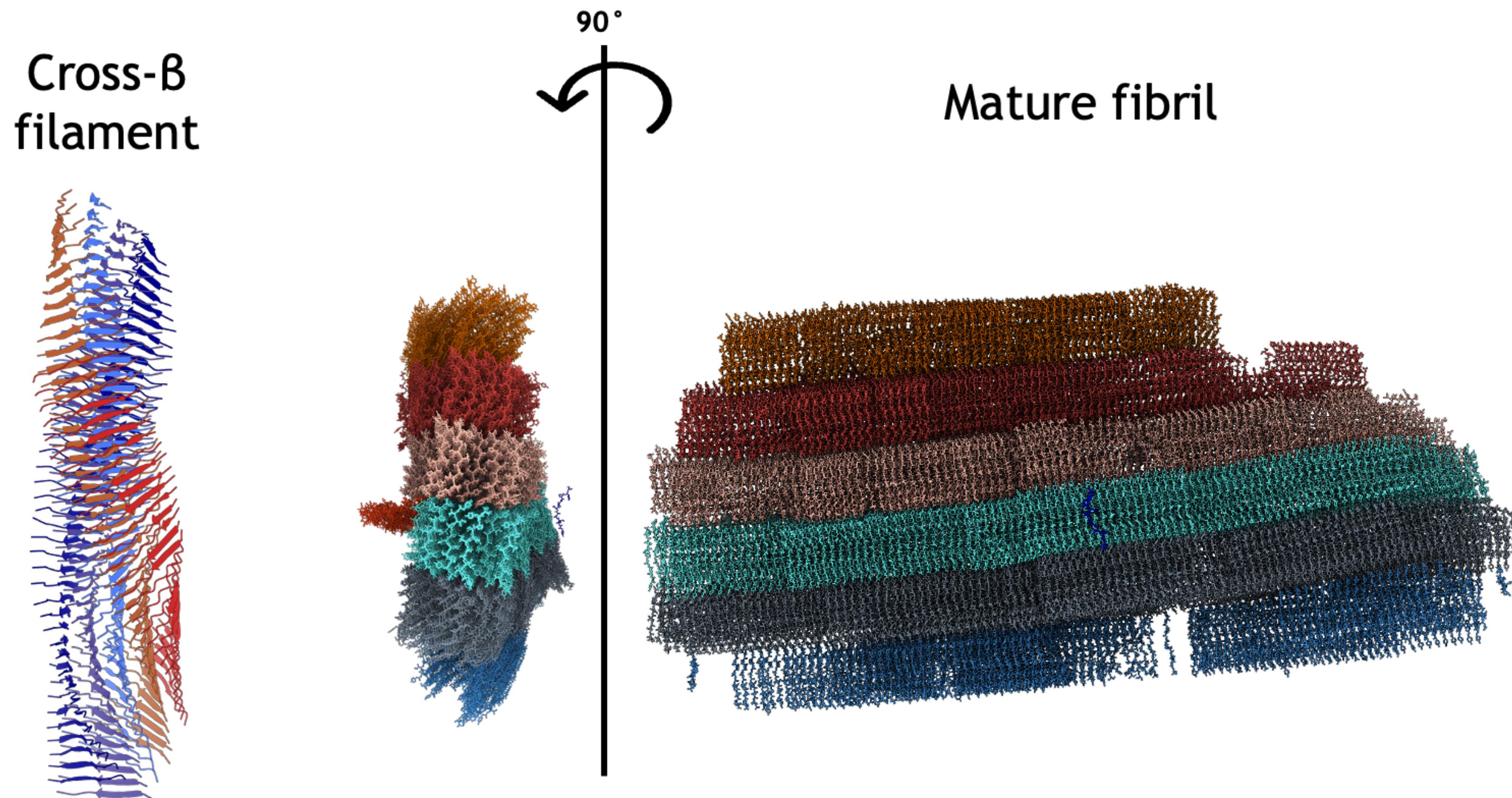


Scalone, E. *et al.* Multi-eGO: An in silico lens to look into protein aggregation kinetics at atomic resolution. *Proc National Acad Sci* **119**, e2203181119 (2022).

# Structure-Based models for protein aggregation and more



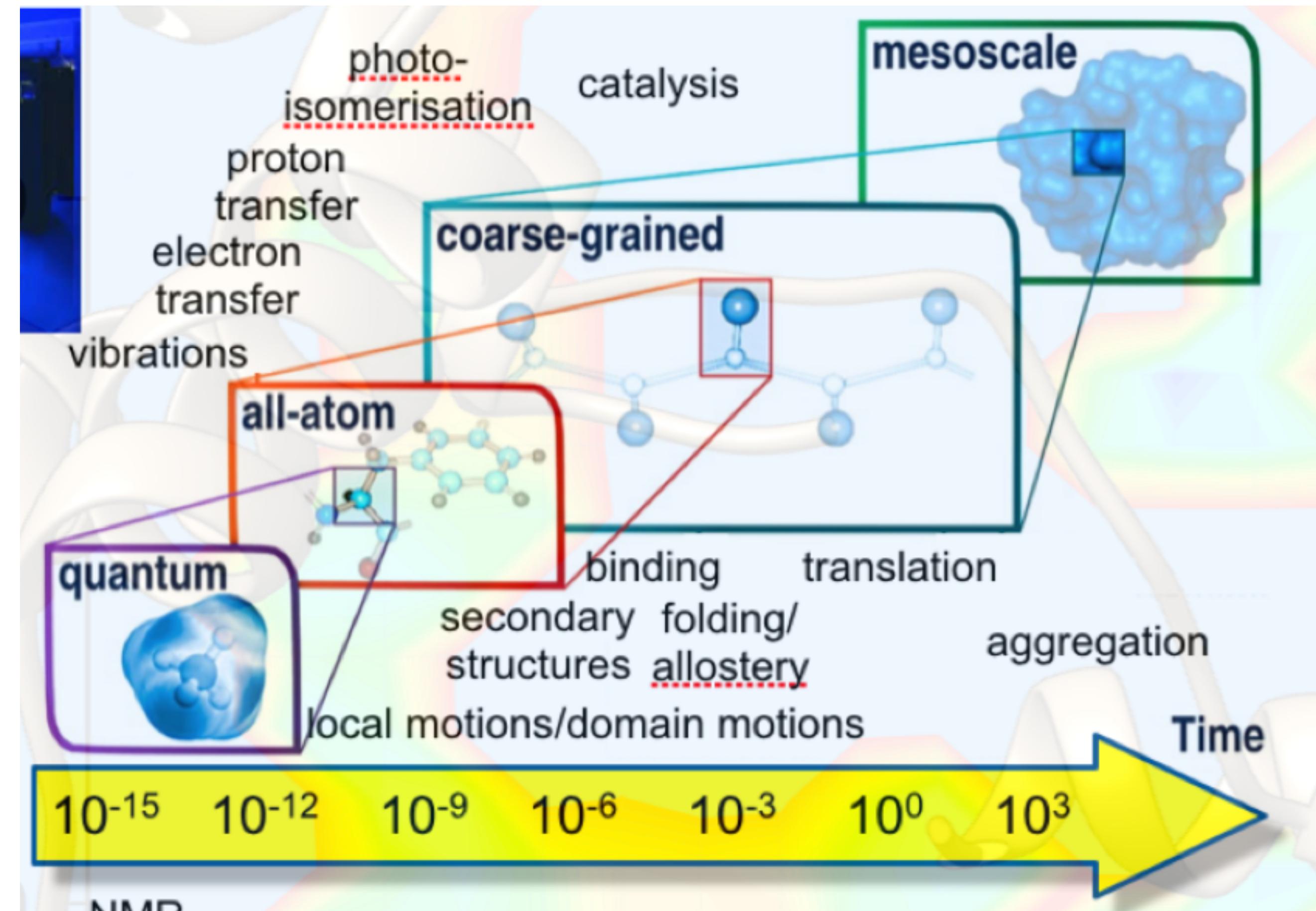
# Structure-Based models for protein aggregation and more



UNIVERSITÀ  
DEGLI STUDI  
DI MILANO



# Select the most appropriate simulation technique(s)



UNIVERSITÀ  
DEGLI STUDI  
DI MILANO



# Questions:

---

- What is the difference between the all-electron and the electron probability density function?
- What is the Born-Oppenheimer approximation?
- How does the performance of a classical simulation compare to a QM simulation?
- What are QM/MM simulations and what kind of problems can they solve?
- What is the difference between knowledge based and chem/phys based simplified models?
- What are the key features of the Martini model?
- What is a structure based model?

