

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

Structural Bioinformatics

Classical Physics and Chemistry (-1916)

Evolution of physical and chemical theories has always gone in parallel, classical chemistry being evolving in parallel to classical physics

The main features of Dalton's atomic theory are:

- 1 Elements are made of extremely small particles called atoms.
- 2a Atoms of a given element are identical in size, mass, and other properties
- 2b Atoms of different elements differ in size, mass, and other properties.
- 3 Atoms cannot be subdivided, created, or destroyed.
- 4 Atoms of different elements combine in simple whole-number ratios to form chemical compounds.
- 5 In chemical reactions, atoms are combined, separated, or rearranged. <-- **NO CHEMICAL BOND**

1850-: Black-body radiation (something that is heated emits energy)

1887-: Discovery of photoelectric effect (Hertz)

1896-: Discovery of radioactivity (Becquerel-P&M Curie)

1897-: Discovery of the electron (JJ Thomson)

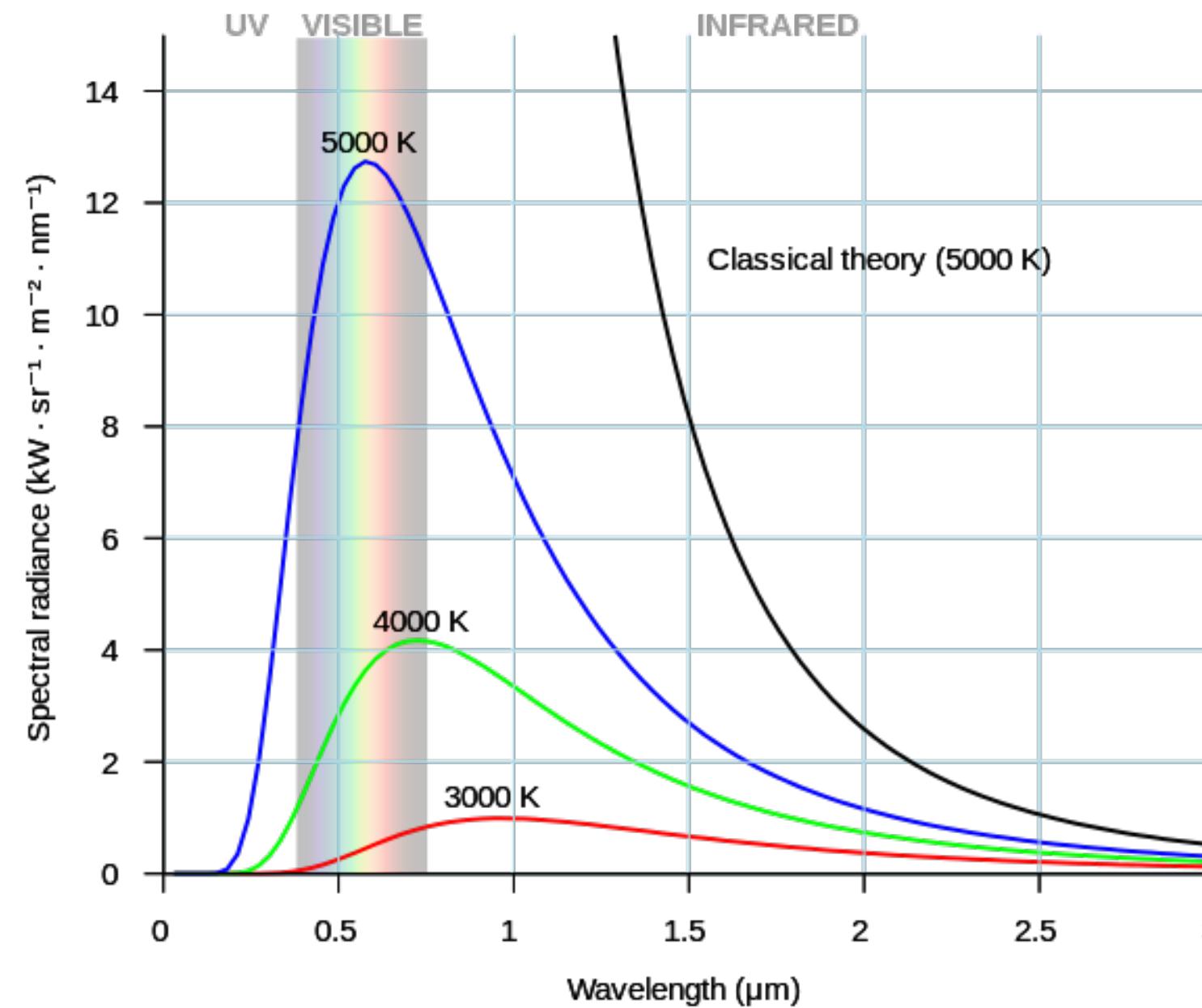
1909-: The electron is the smallest existing charged object (Millikan)

1911-: Discovery that atoms are made by a small positive nucleus surrounded by a negative region (Rutherford)

For all these, and other, experiments there were contradictory explanations.



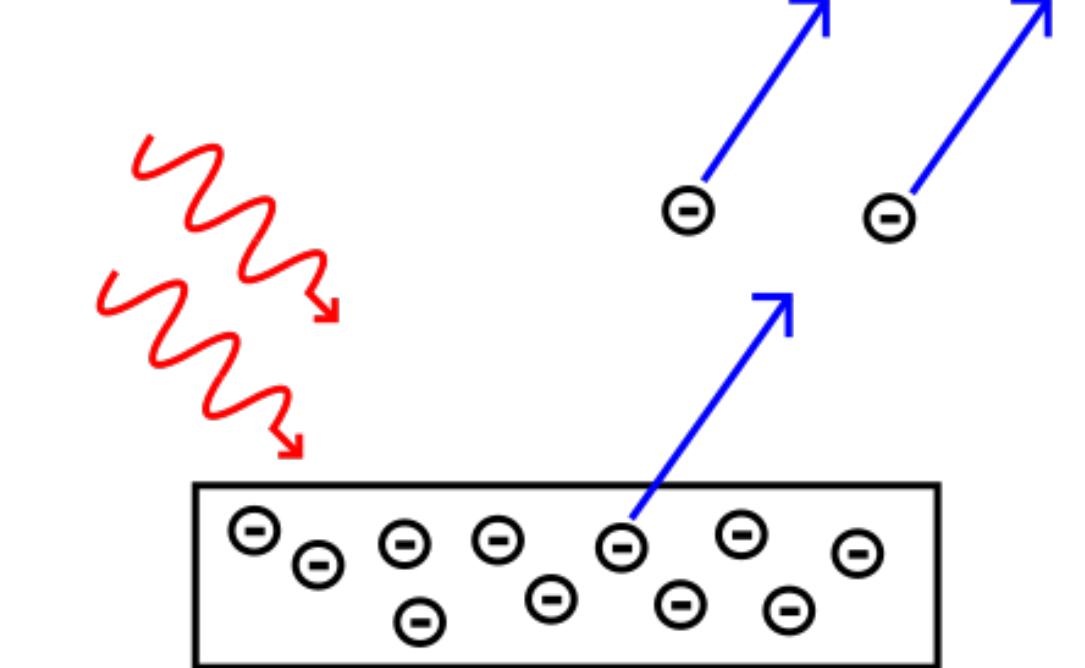
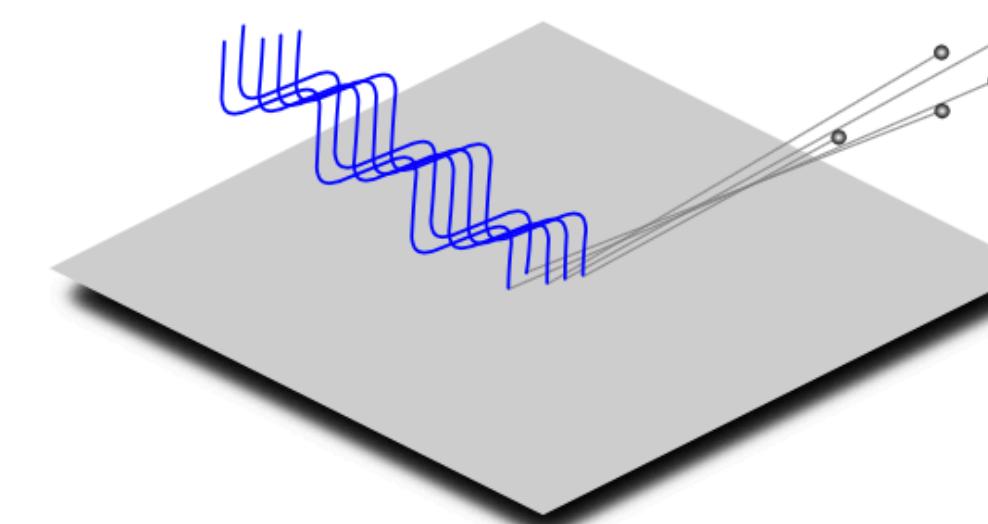
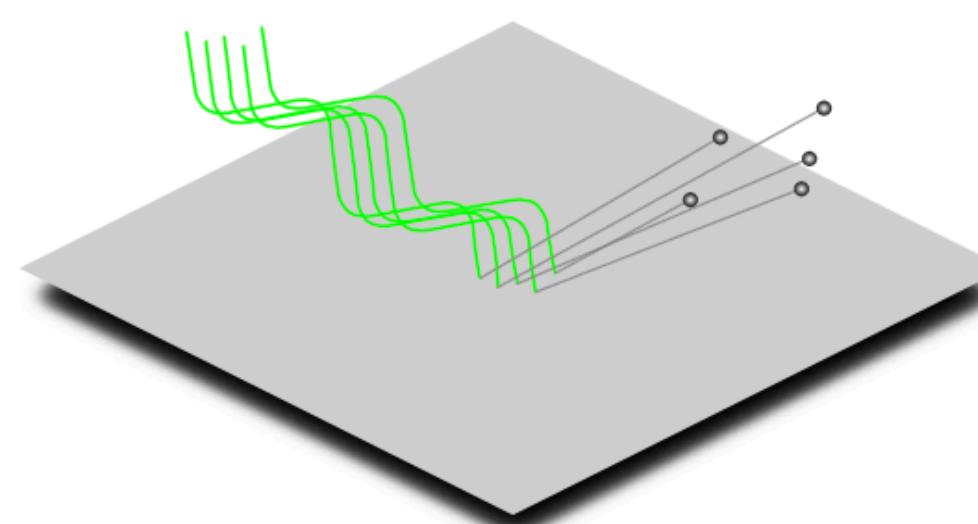
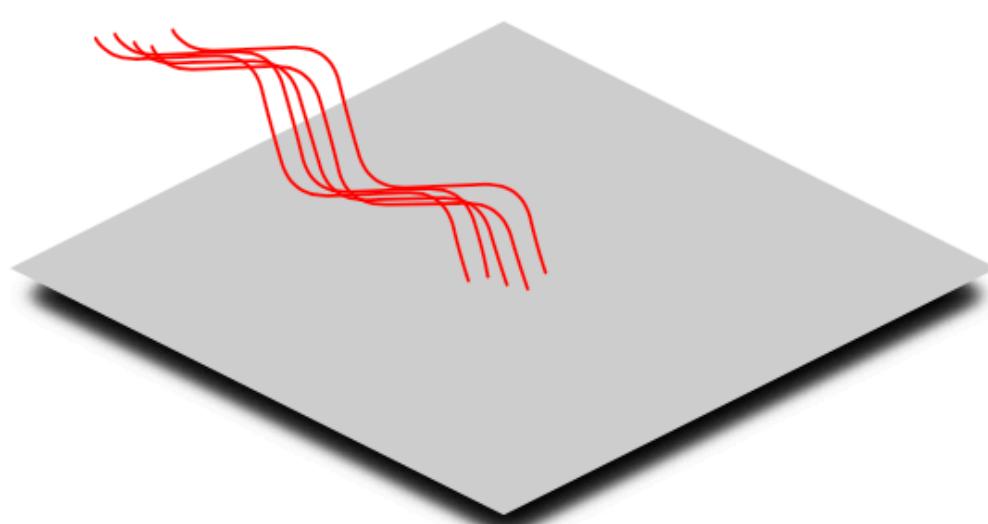
The Planck h constant: Black body radiation and Photoelectric effect



1902 - Planck: The radiance of a body is related to its temperature, that is from the color of an object emitting light we can estimate its temperature

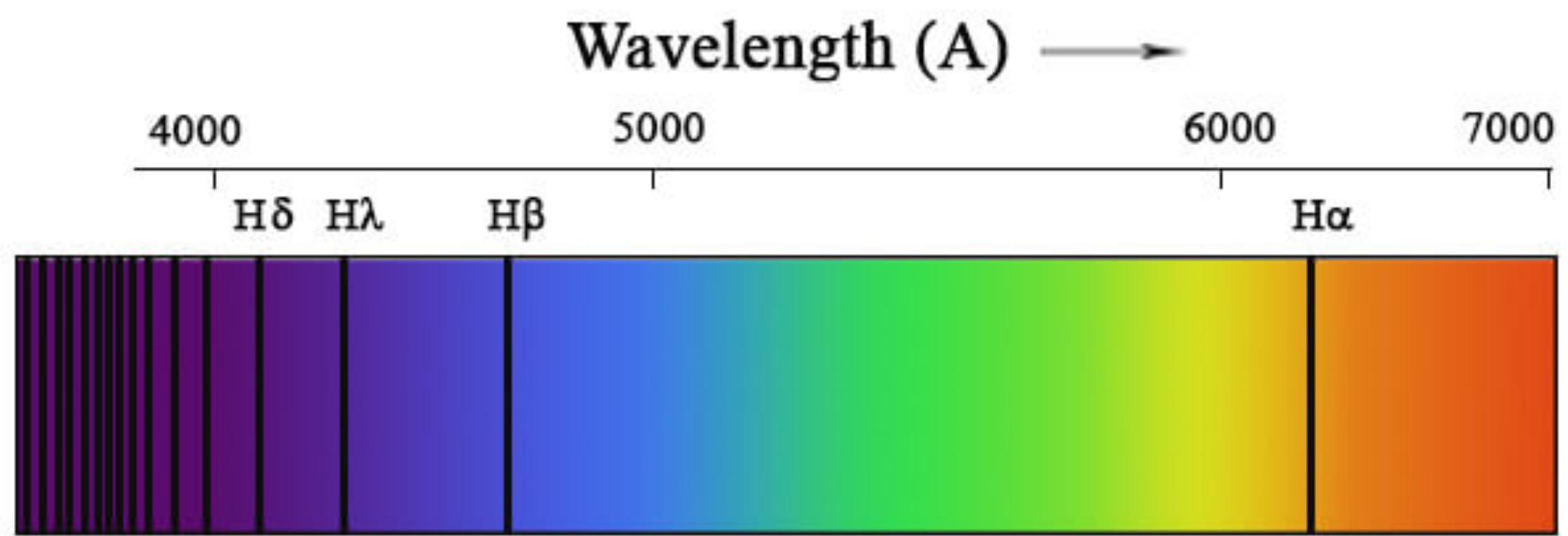
$$B_\nu(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1}, \quad \text{Fitting the experimental spectra you can measure } h$$

1905 - Einstein: Light is made of particles, called photons. Brightness is related to the number of photons. Frequency is related to the kinetic energy of the photons. $\epsilon = h\nu$

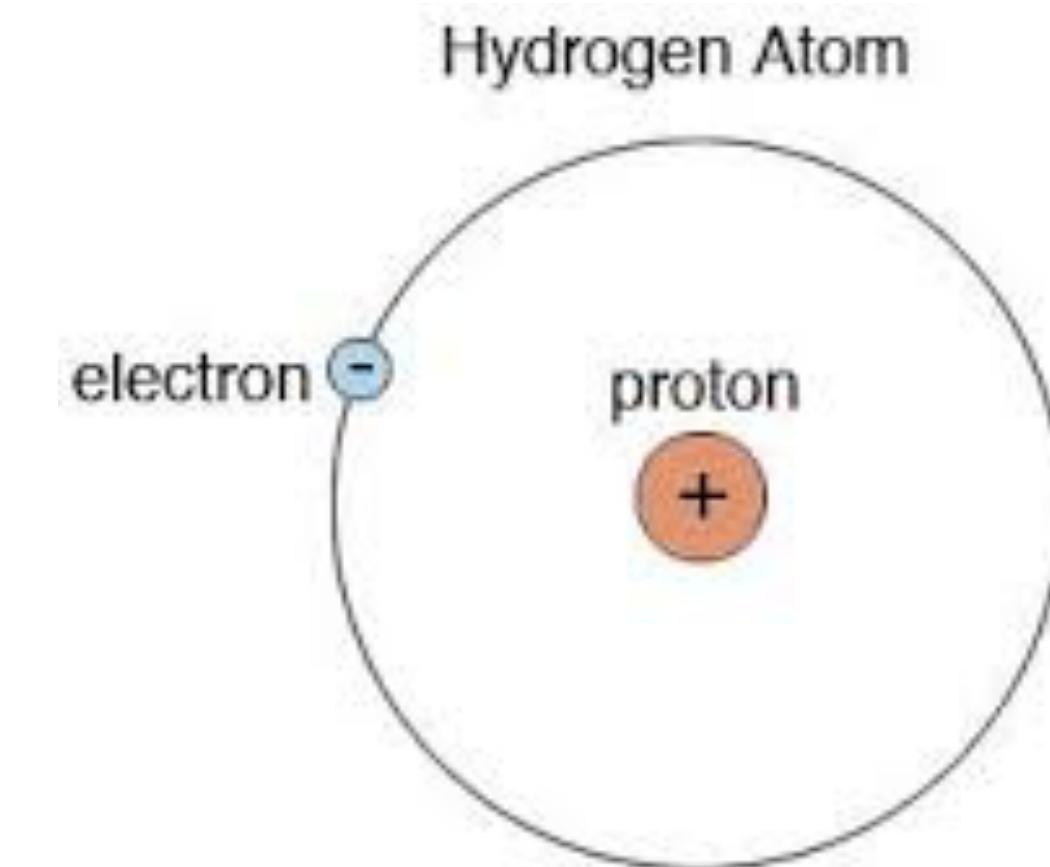
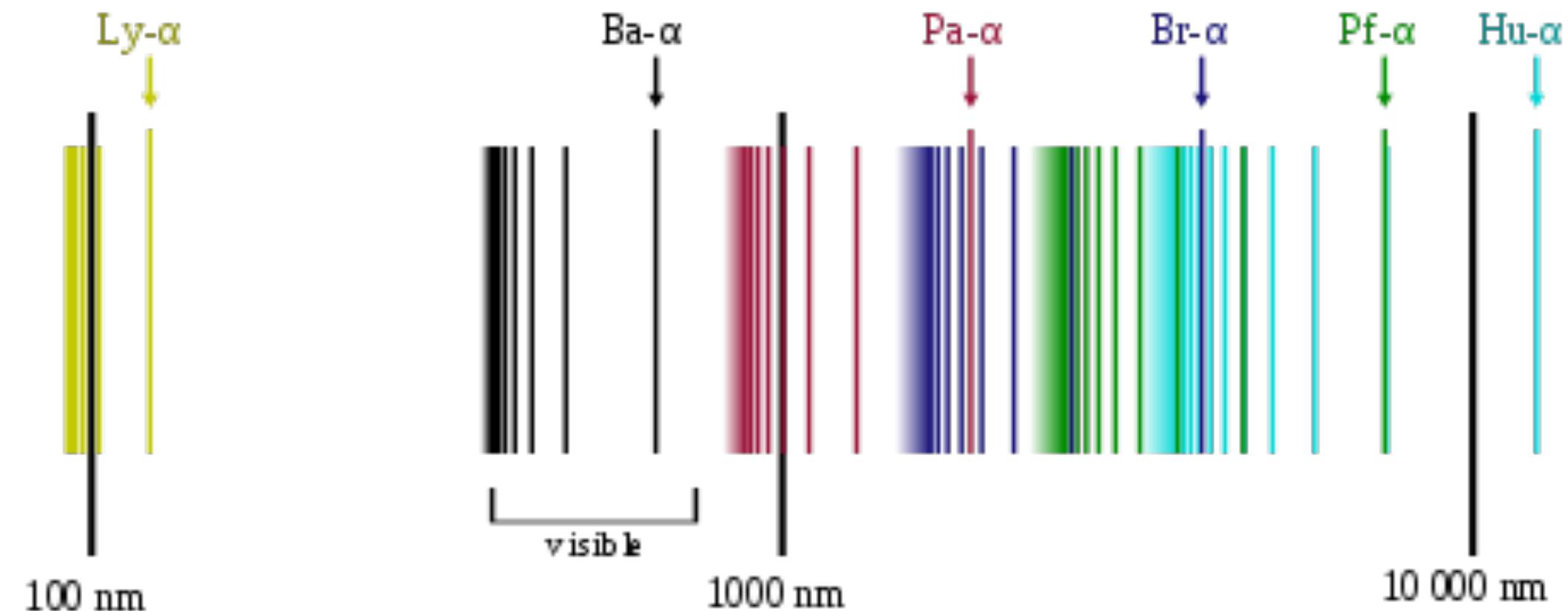


Atomic structure is studied by spectroscopy

Hydrogen absorption

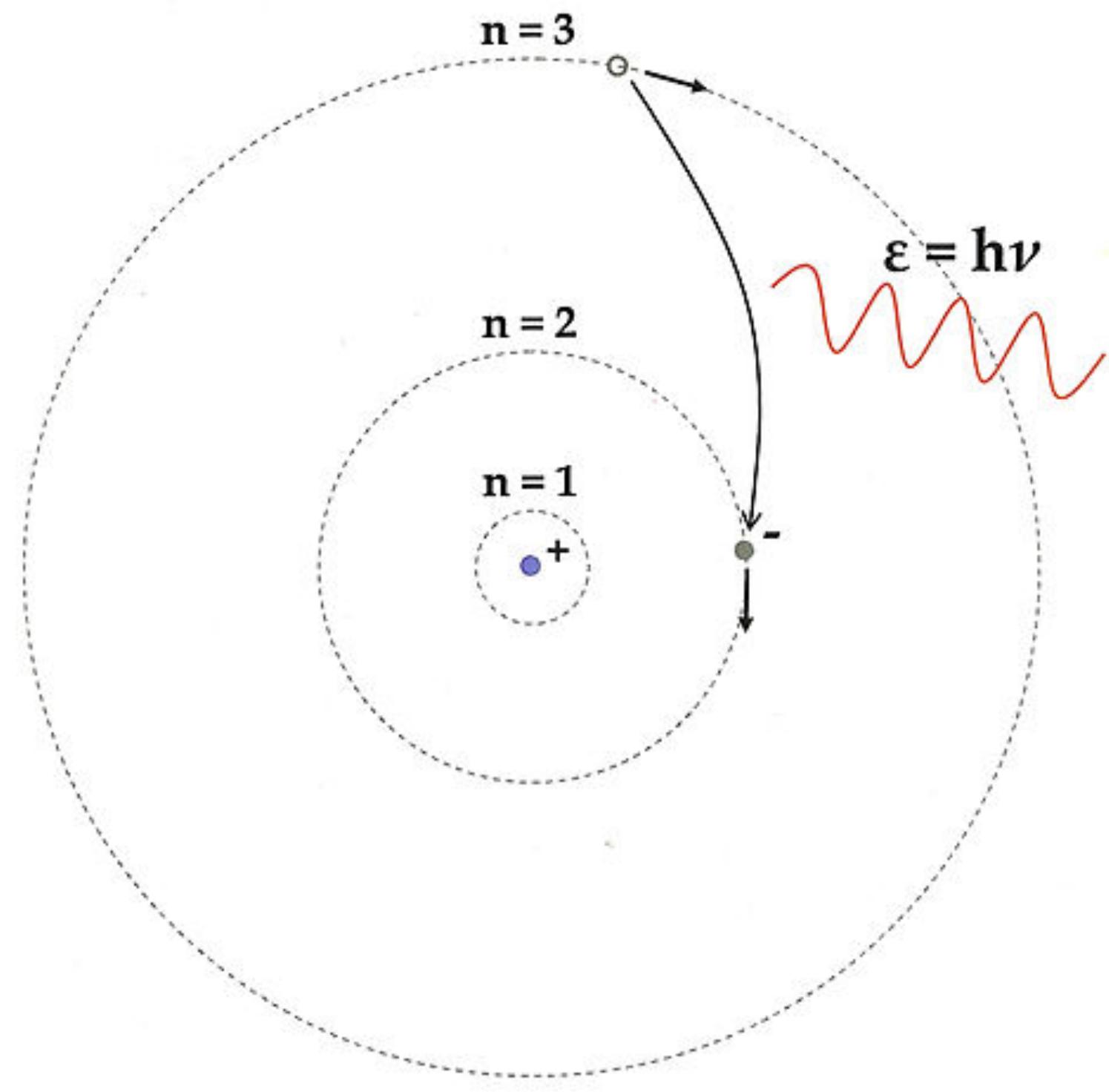


Hydrogen emission



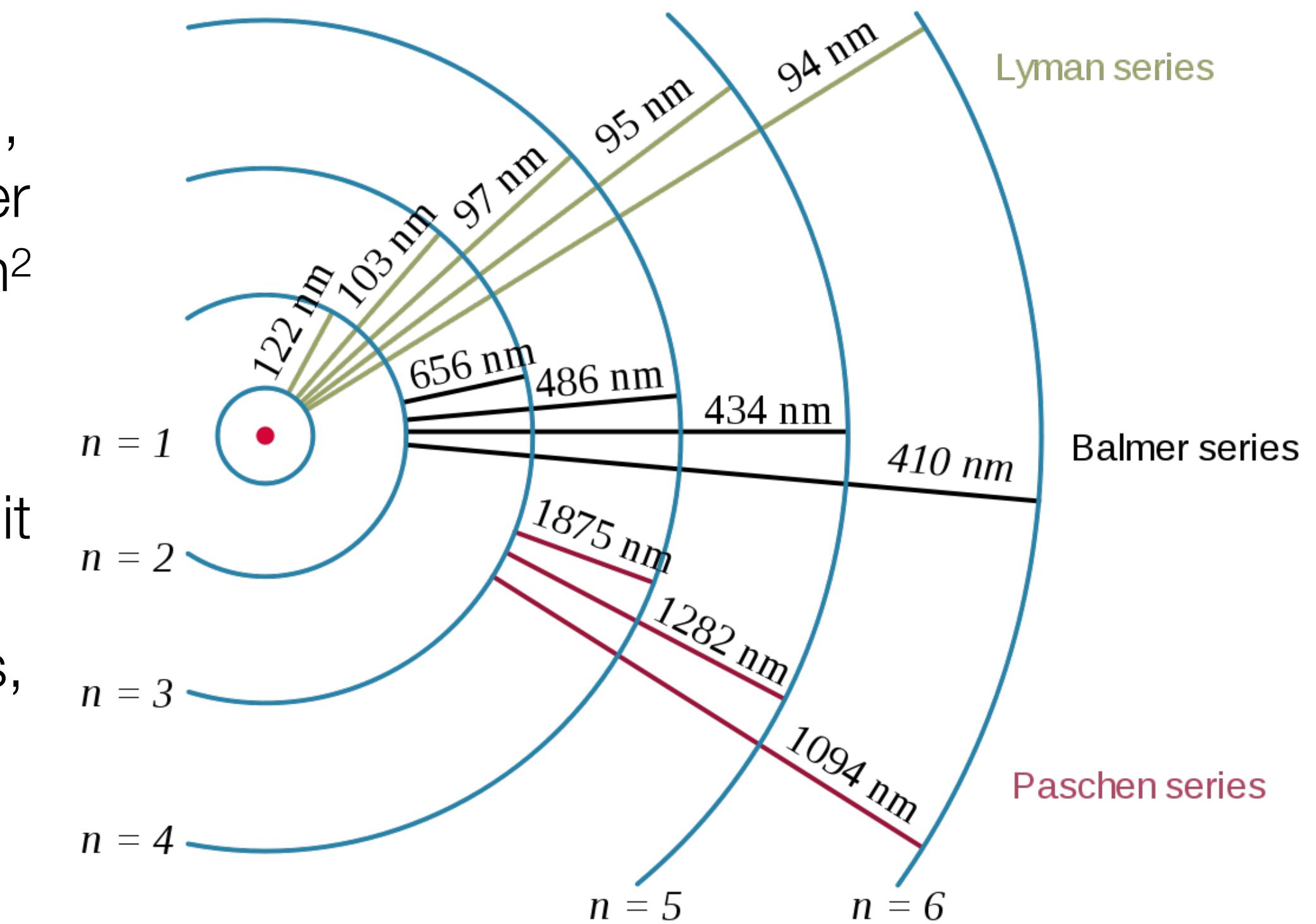
This idea of atom cannot be explained by classical physics because charges in motion should dissipate energy through electromagnetic radiation, so eventually the electron should collapse onto the nucleus. Furthermore in this model atoms can absorb and emit energy at any frequency.

Bohr's atomic model introduce the concept of orbitals



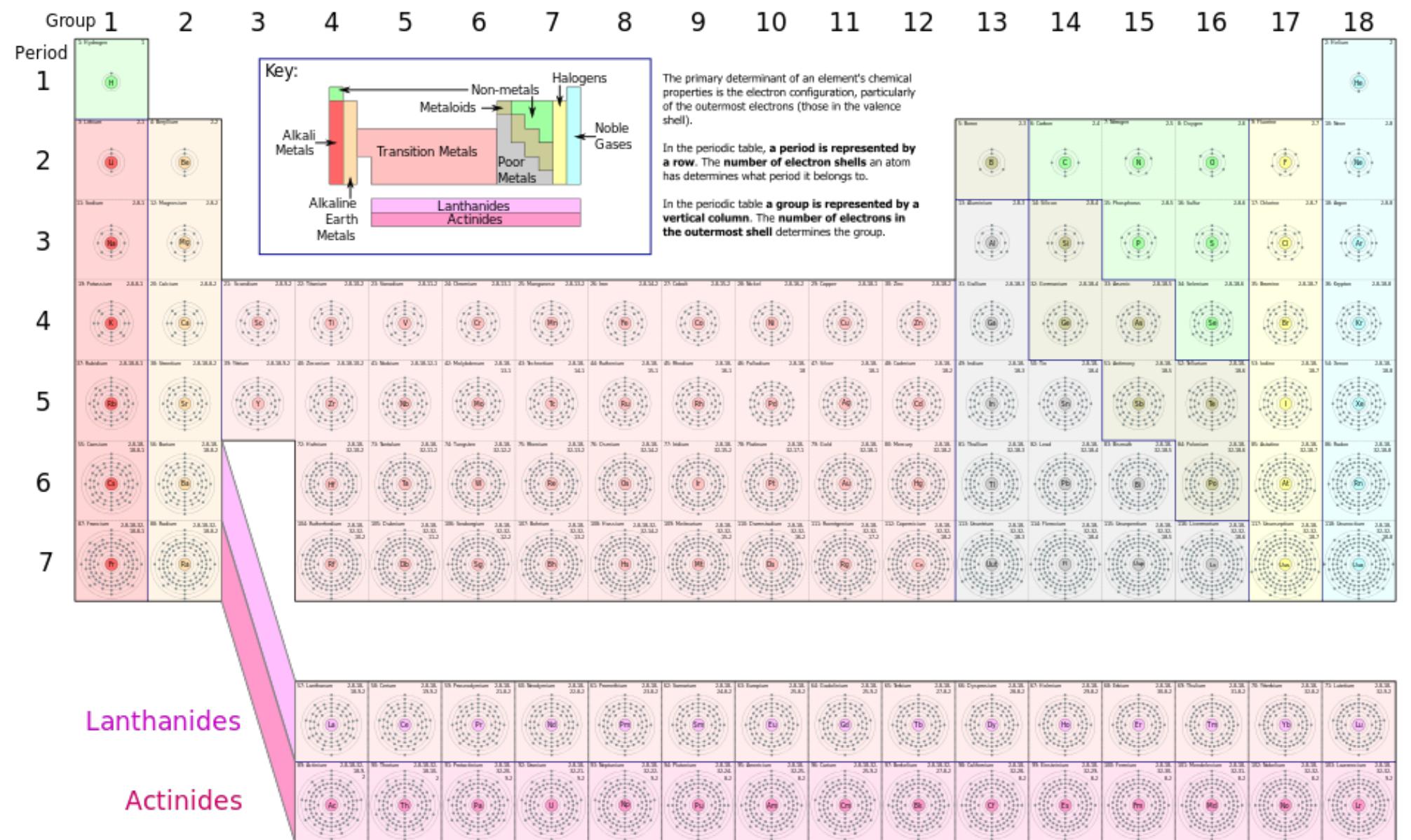
Only few orbits are possible, this are stable by assumption, generally speaking the number of electrons in each orbit is $2n^2$ (2, 8, 18, ...)

Light can be absorbed only if it corresponds to the energy difference between two orbits, and is emitted in the same frequencies.

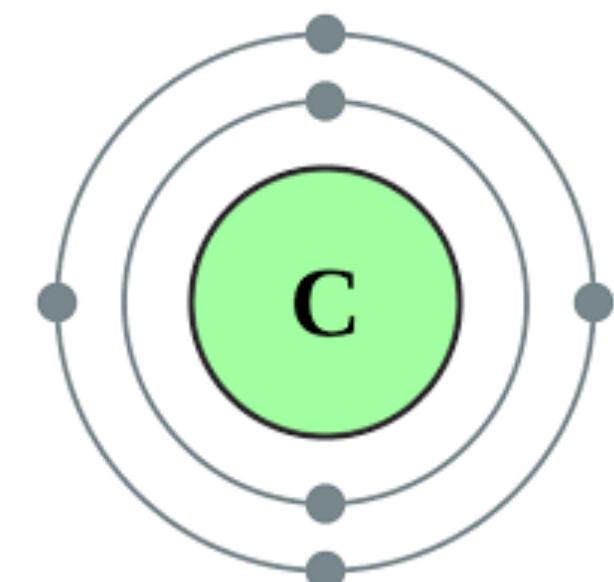


Bohr model explains the periodic table of elements (in part) and the chemical covalent bond

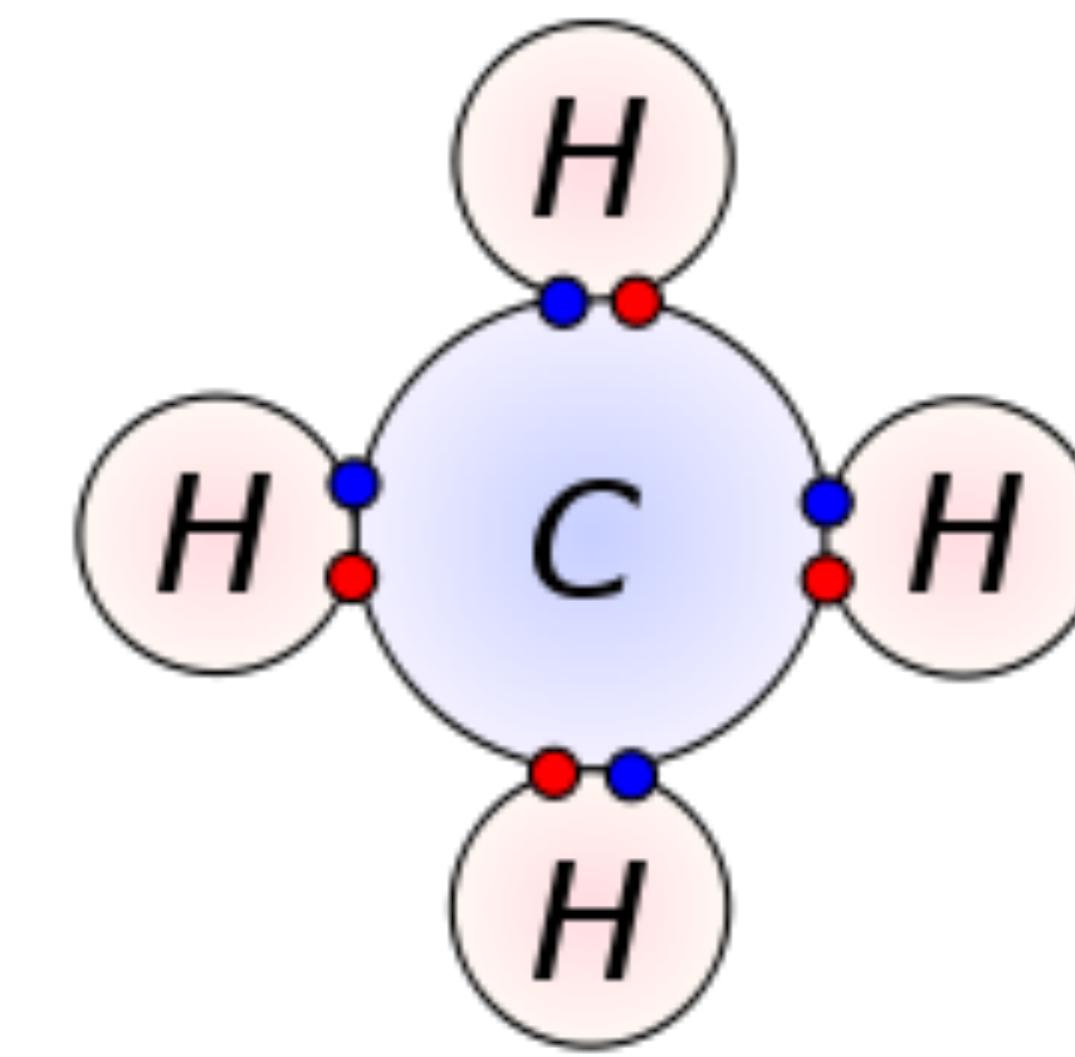
Periodic Table Of Elements Showing Electron Shells



Chemical Bond



Periodic Table:
Valence Electrons
and shell completion



- Electron from hydrogen
- Electron from carbon



UNIVERSITÀ
DEGLI STUDI
DI MILANO



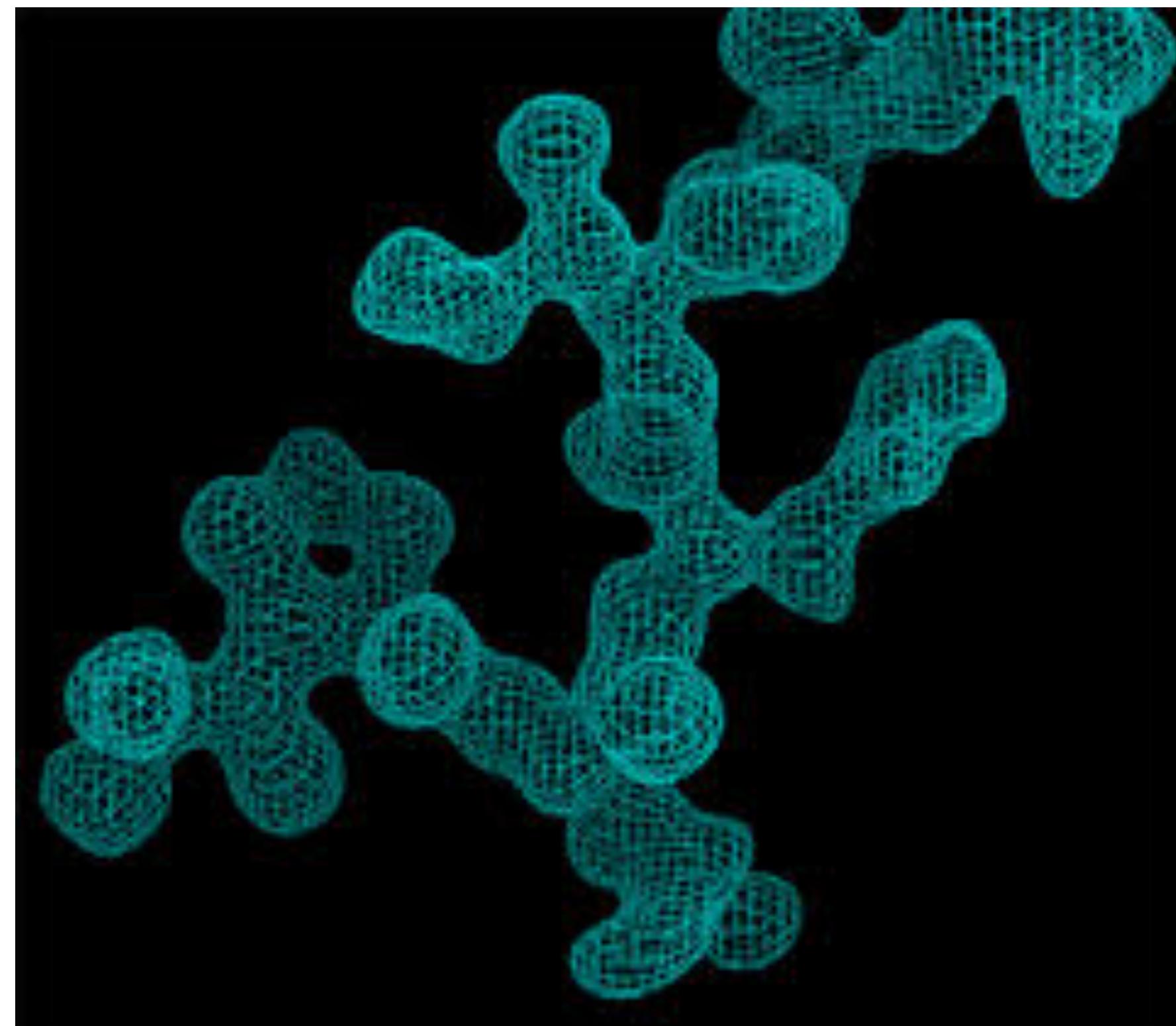
Quantum Mechanics (1925-): from energy levels to probability densities

All the models before are almost classical in the sense that particles have well defined properties.

With Quantum Mechanics there is a change of paradigm and a completely new theory.

At the microscopic level we cannot say anymore how things move from the perspective of a “cannon-ball”.

The object that evolves is a probability.



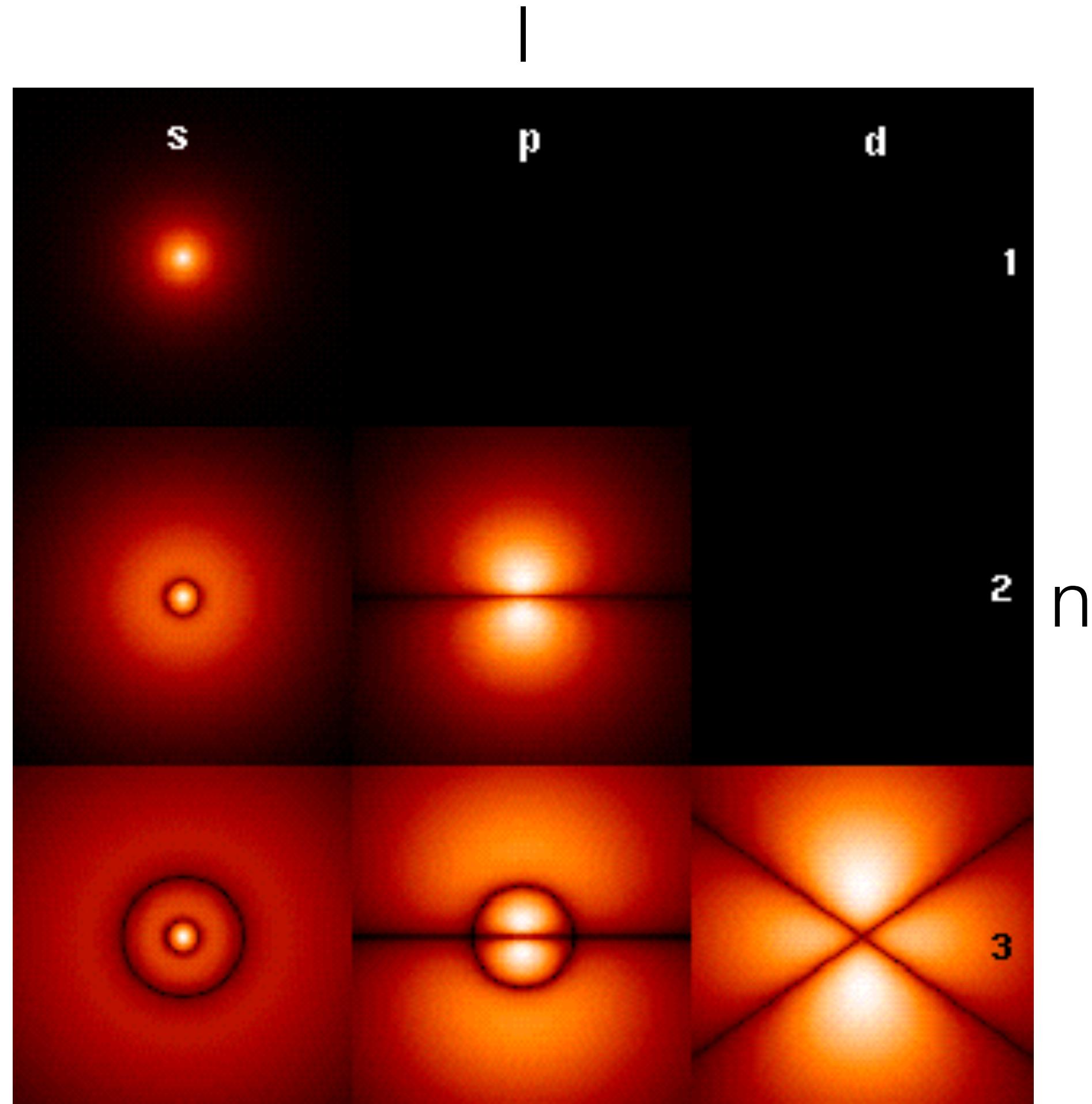
Quantum Mechanics (1925-): from energy levels to probability densities

An electron in an atom is defined by its probability density associated with the so called quantum numbers (n, l, m_l, m_s). For example in the case of the hydrogen one electron is determined by one density that corresponds to one energy.

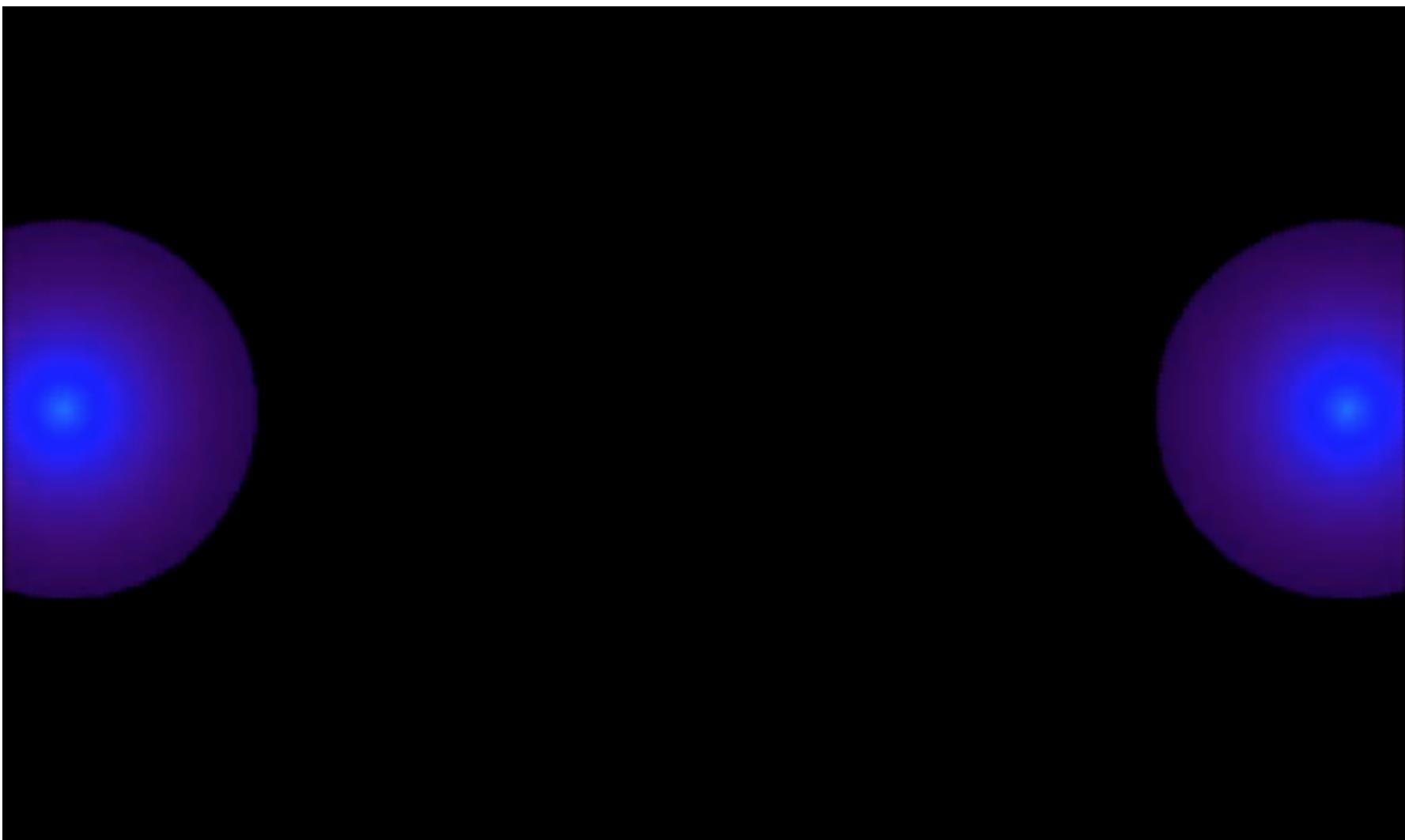
$$E_{1s} < E_{2s} < E_{2p}$$

The electron can be ‘excited’ to a higher energy by absorbing a photon or by temperature and so its density will change.

When more electrons are present there are joint probability densities.



Chemical Bonding



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



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Schrödinger equation

$$\hat{H} \Psi = E \Psi$$

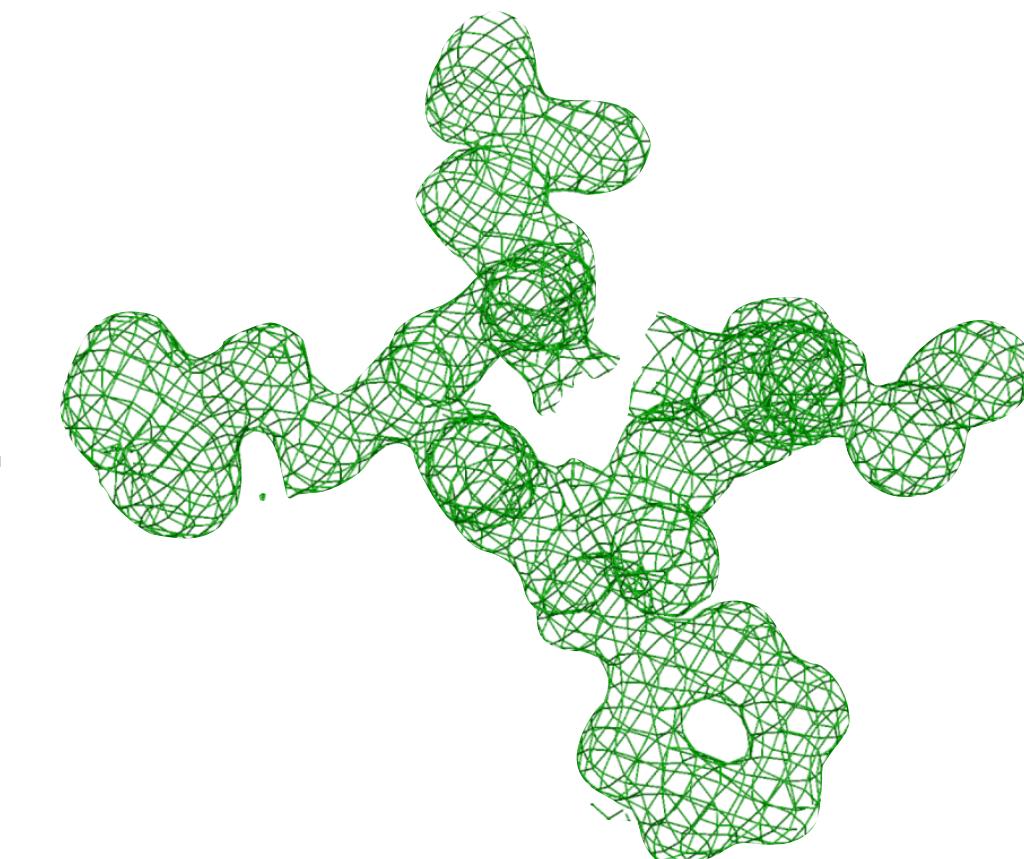
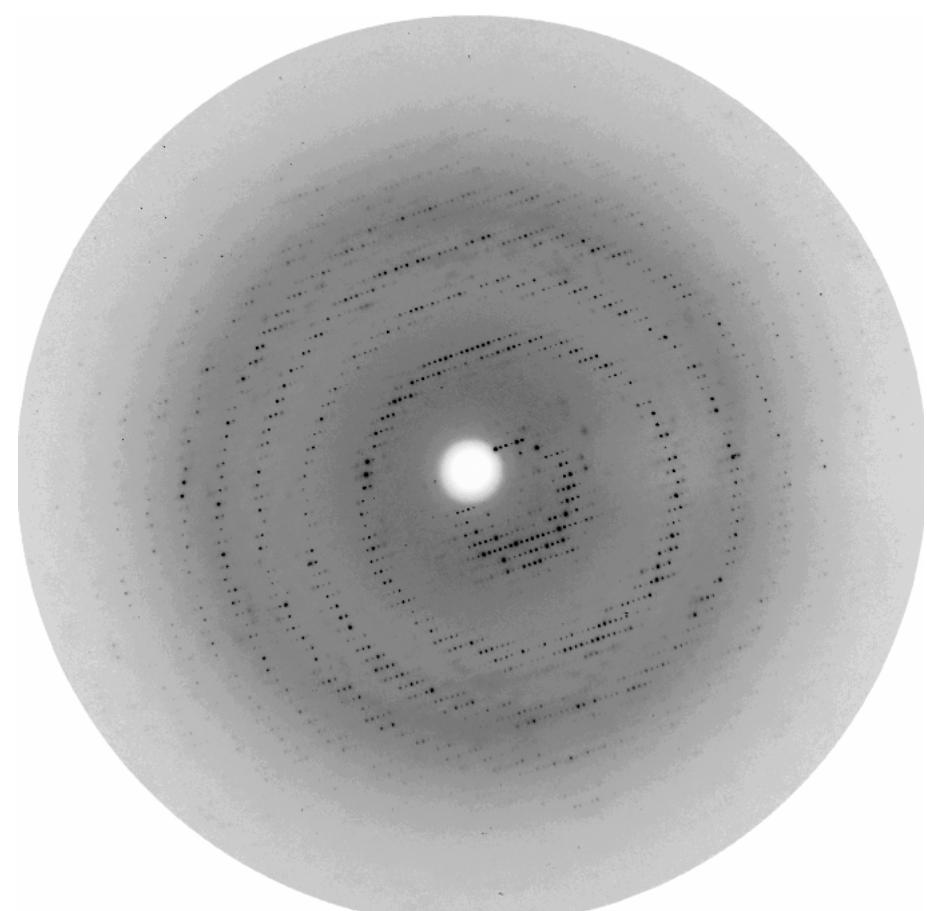
Hamiltonian Operator (Energy operator) Energy eigenvalue

This is only Coulomb

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

Kinetic Energy + *Potential Energy* = *Total Energy*

One issue here is the $\Psi(r_1, \dots, r_n)$, the many body electron wave function is a function of $3N$ coordinates, 3 for each electron, this makes the problem very complicated, that is the solution of $3N$ coupled equations.



Psi is called the wave function. The norm of psi is the actual all electrons probability distribution:

$$|\Psi(r, r_2, \dots, r_N)|^2$$

This is the probability of finding one electron in $r_1=x_1, y_1, z_1$ another in $r_2=x_2, y_2, z_2$, ecc

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

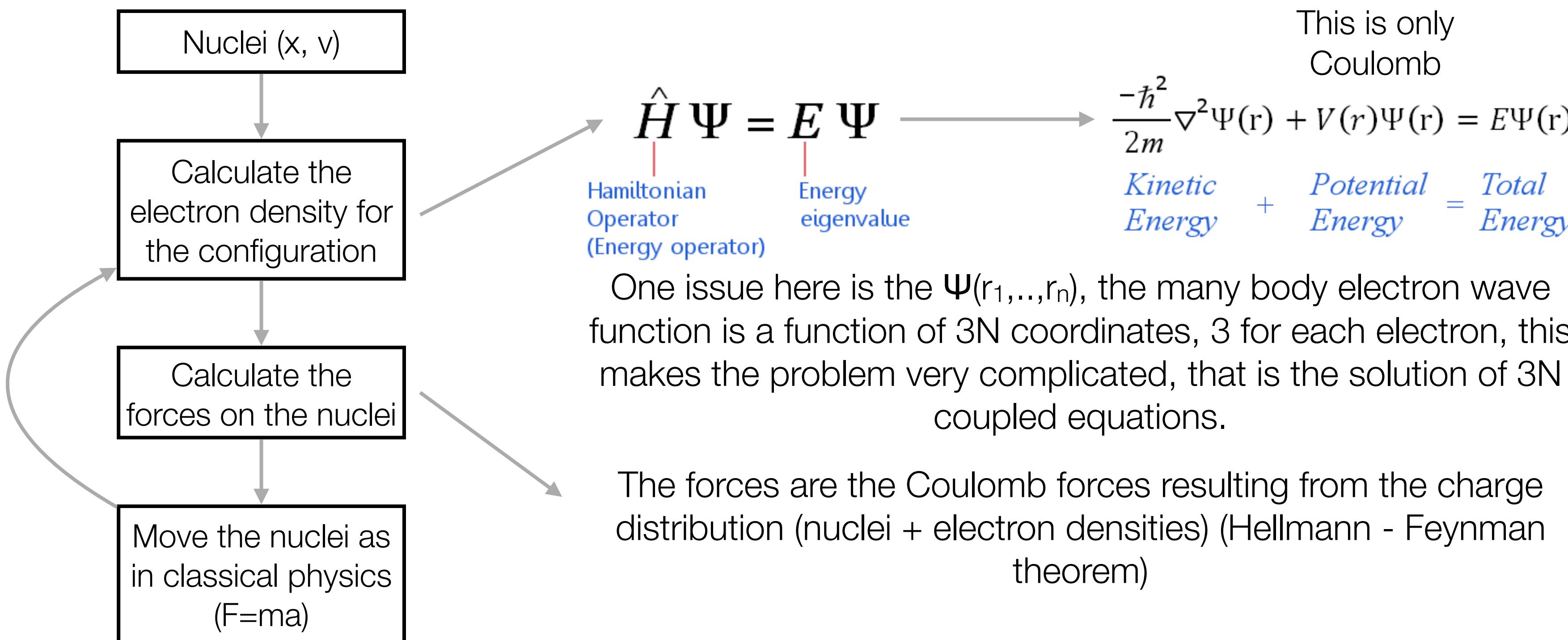
This is the probability density function for N electrons, so $n(r)dr$ is the number of electrons that could be found in that region

That you observe by X-Ray crystallography



Quantum Chemistry: only electrons are treated as quantum particles.

Born-Oppenheimer approximation: atomic nuclei are considered as ‘classical particles’ so the density becomes a ‘sphere’.



Density-Functional Theory

In most of the cases has we have already seen the only quantity that one can observe experimentally is the overall electron density, so not the density of a single electron but all of them together.

This is so true and important that it is possible to rewrite all QM theory based on the overall electron density instead than for each single electron, this is the so called Density Functional Theory

DFT is based on the two Hohenberg-Kohn theorems:

1. The total energy of a system is a unique functional of the electron density
2. The density that minimises the total energy is the exact ground state density.

So instead of $\Psi(r_1, \dots, r_N)$ that is a function $3N$ coordinates, it is enough to study $n(r)$ that is a function of only 3 coordinates.

Function: given a number 'x' you get a number $y=f(x)$.

$$y = \sin(x)$$

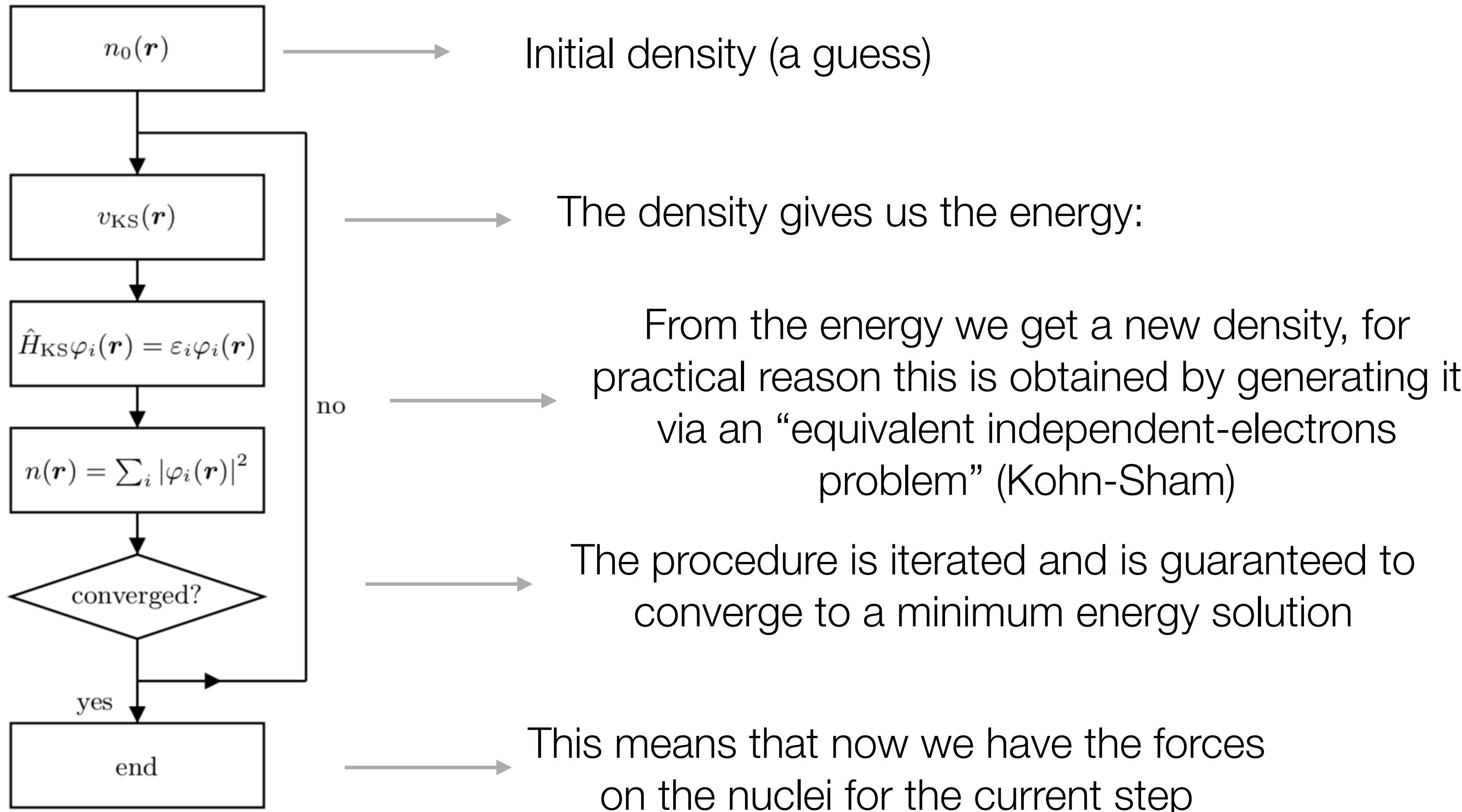
Functional: given a function 'f' you get a number 'y'

$$y = \int_a^b f(x)dx$$

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



Density Functional Theory

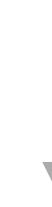


DFT Key Choices: Exchange and Correlation Potential

Density Functional Theory is exact but unfortunately its practical implementation needs the knowledge of an unknown “Exchange and Correlation” potential.

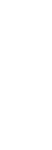
There are three classes of such potentials:

Local Density Approximation
(LDA)



Very fast and simple,
not very accurate
(PZ81, PW92, ...)

Generalised Gradient
Approximation
(GGA)



Accurate and fast
(BLYP, PW91, ...)

Hybrid GGA
(hGGA)

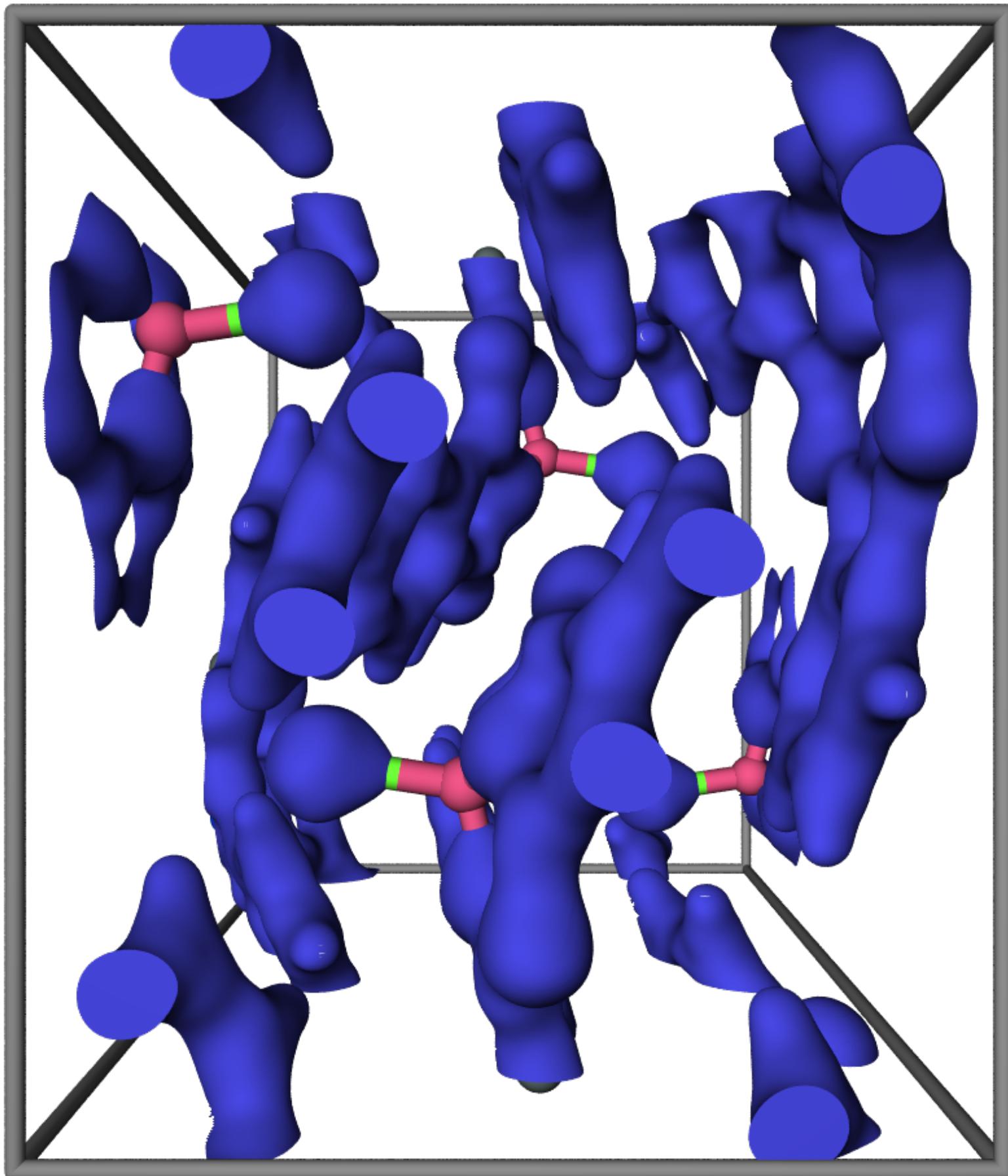


Slow, sometimes
very accurate
(B3LYP, PBE0)

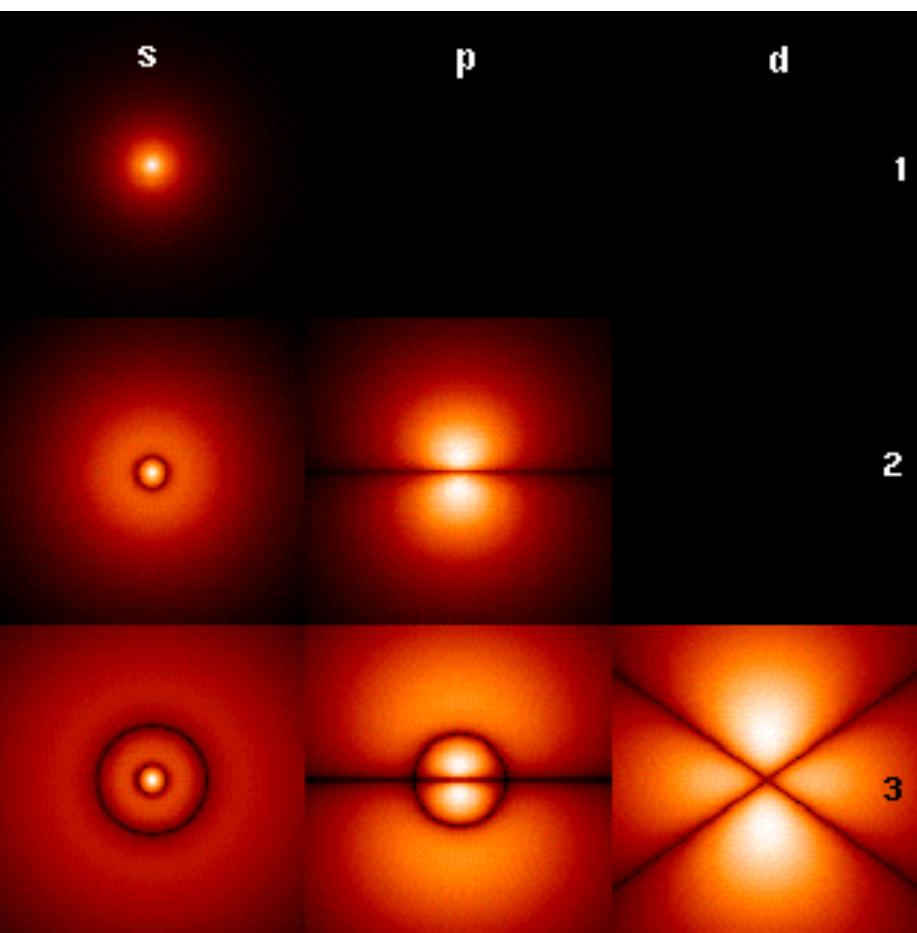


DFT Key Choices: describe the density

GRID



BASIS SET



Each orbital is a mathematical function, we can write the density as a combination of the orbitals (in practice as combination of Gaussians)

$$\sum_k a_k f_k(r)$$

- Examples:
- STO
 - 3-21G
 - cc-pVQZ

PLAIN WAVES

A free electron in vacuum is described by a plane-wave, these are also good for periodic systems (Bloch's theorem)

$$\sum_k a_k e^{ikr}$$

Hybrid methods allow using a combination of basis set that are good at describing the behaviour of electrons close to their nuclei and plane waves that are good for their behaviour far away from them. Grid are then added to calculate electrostatic interactions.



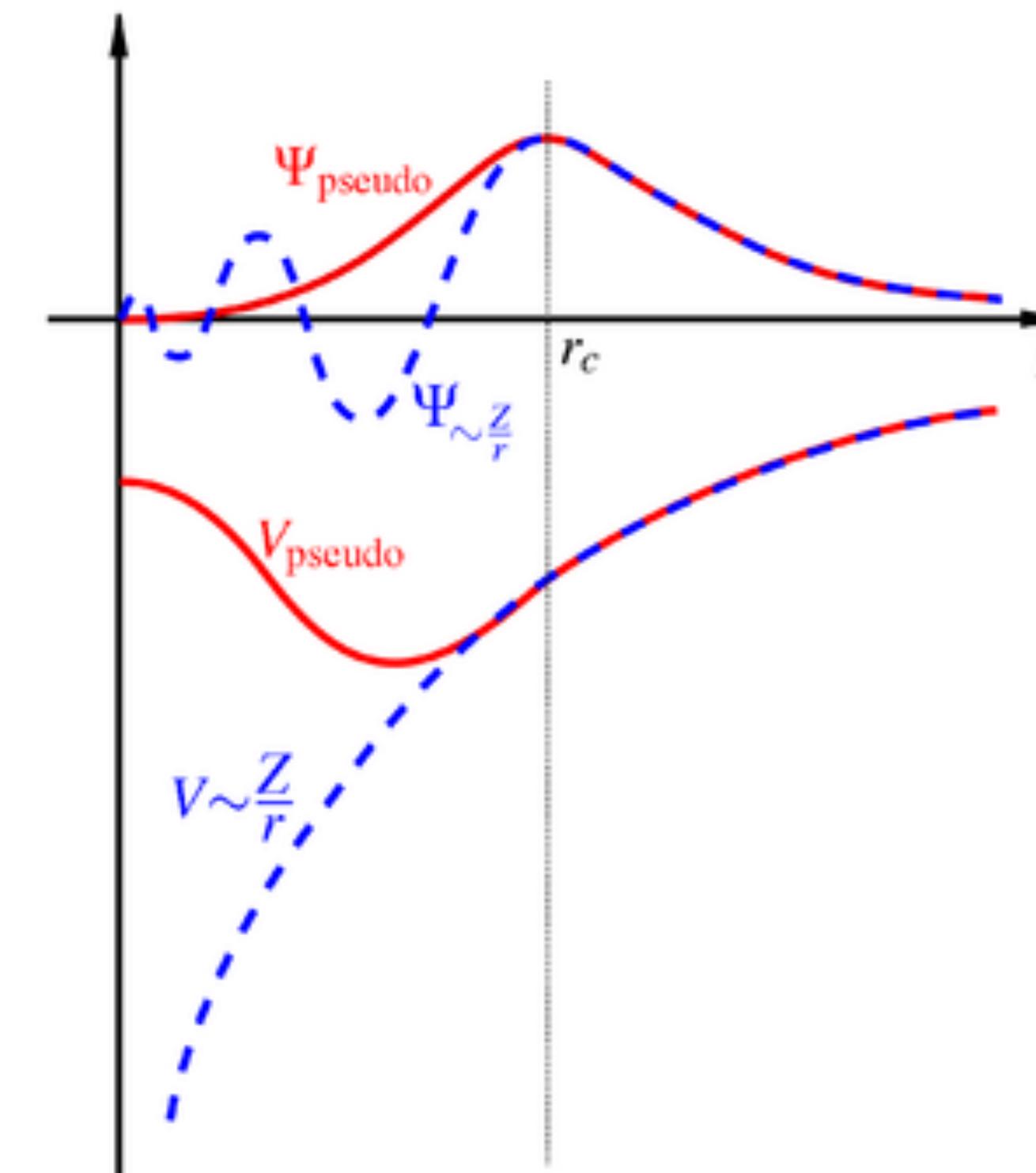
DFT Key Choices: Pseudopotentials

Quantum Chemistry simulations are computationally very demanding. The number of electrons is what determines the calculation time. How can we reduce the number of electrons in the calculation?

The concept of valence electrons:

The chemical property of an atom are essentially determined by its valence electrons

We could build pseudo nuclei that are the sum of the nucleus+core electrons and simulate only the valence electrons.



QM/MM mixing QM and Classical MD simulations

Warshel & Levitt 1976

DFT Simulations

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

MD Simulations

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

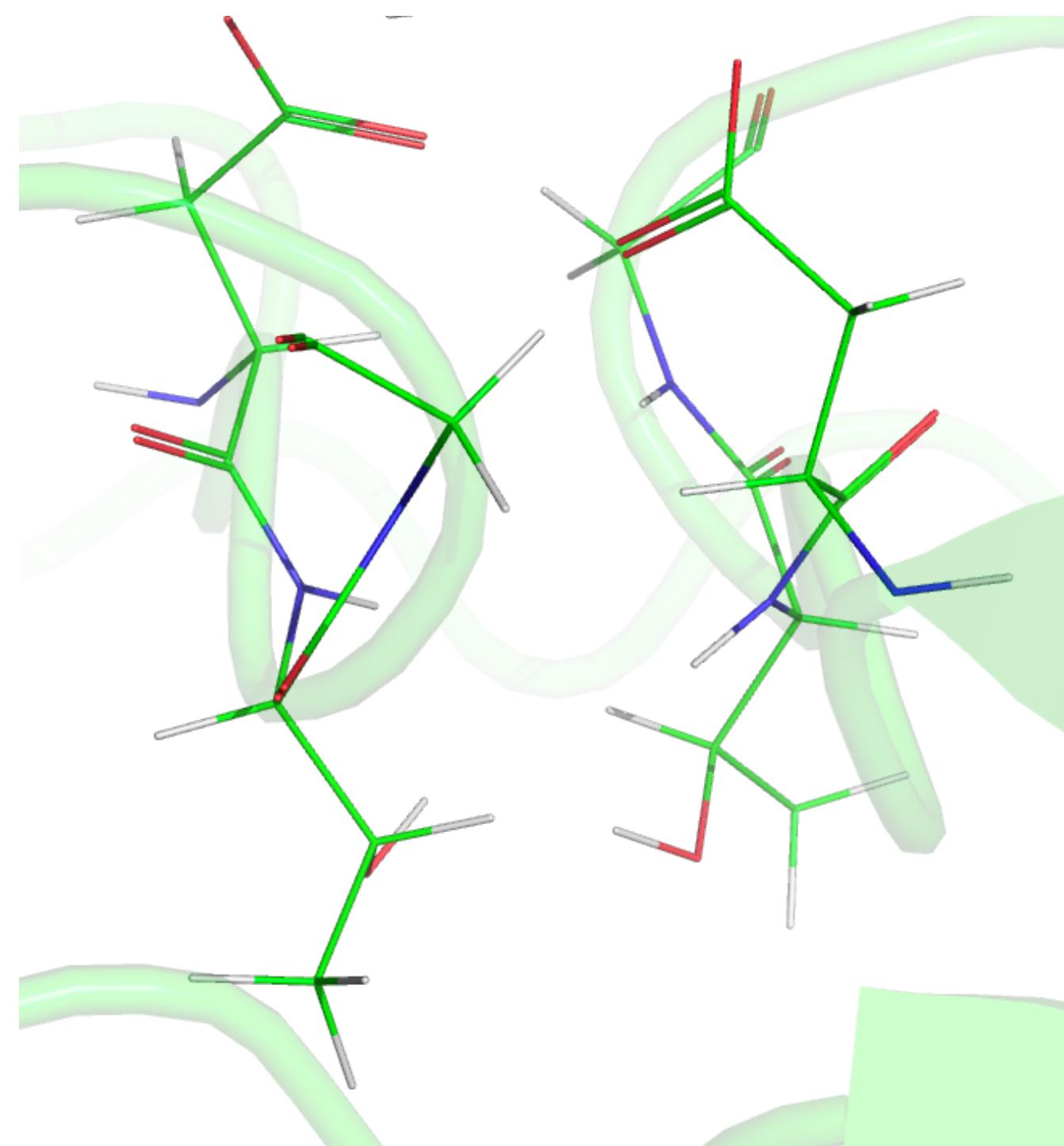
Enzymes catalysing chemical reactions is a large class of proteins whose function cannot be study by neither techniques. A nice feature of enzyme is that usually they can speed up chemical reactions dramatically, making them happen on time scales that can be compatible with DFT calculations in some cases, the only issue is the size!

How can we make simulations of enzymatic reactions?



QM/MM: mixing QM and Classical MD simulations

The first possibility is to cut the system, for example only the active site:



The main issues here are:

1. We are cutting some covalent bonds
2. We are neglecting the flexibility of the active site
3. We are neglecting the interaction of remainder of the environment with the active site

Amino acids are usually cut at the C_b unless the backbone forms relevant interactions with the substrate. A number of hydrogens is added to saturate the bond (complete the valence shell). This should solve the first issue.



QM/MM: including flexibility

Subtractive scheme: ONIOM

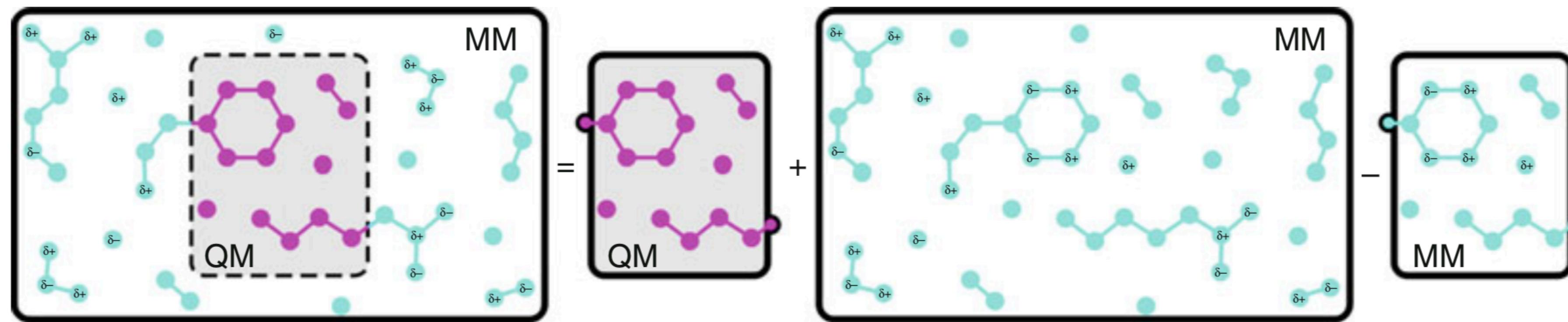


Fig. 2. Subtractive QM/MM coupling: The QM/MM energy of the total system (*left hand side of the equation*) is assumed to be equal to the energy of the isolated QM subsystem, evaluated at the QM level, plus the energy of the complete system evaluated at the MM level, minus the energy of the isolated QM 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.

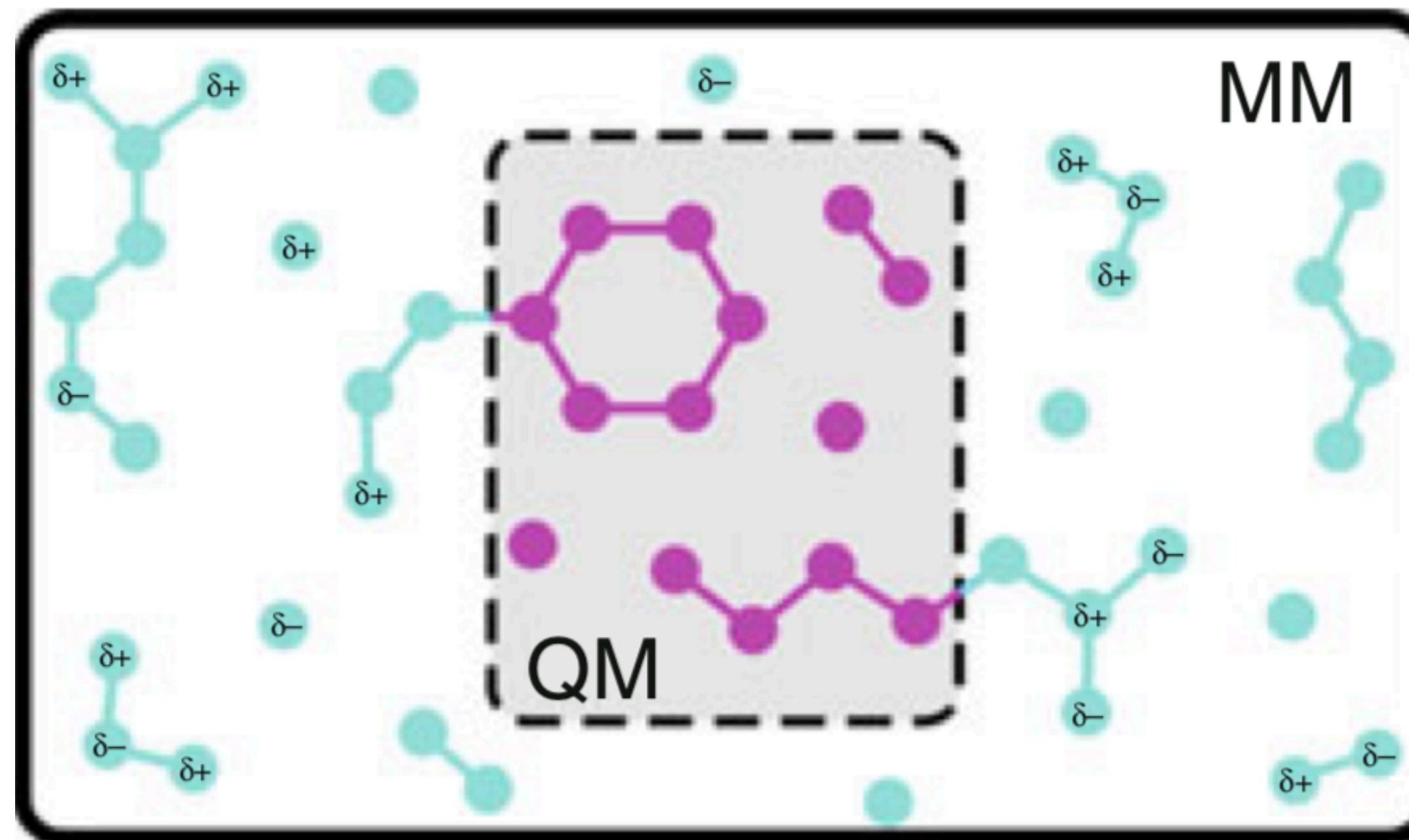
$$V_{\text{QM/MM}} = V_{\text{MM}}(\text{MM} + \text{QM}) + V_{\text{QM}}(\text{QM}) - V_{\text{MM}}(\text{QM}).$$

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



QM/MM: and the environment

The effect of the removed environment can be added back by considering the electric field generated by the point charges of everything left out, these can be taken from a force field:



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

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.



HIV-1 protease

How does a protease cleaves a peptide bond?

Catalytic residues are treated at QM level, including a water molecule, the remainder of the protein contributes with an electric field resulting from all the force-field derived point charges.

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

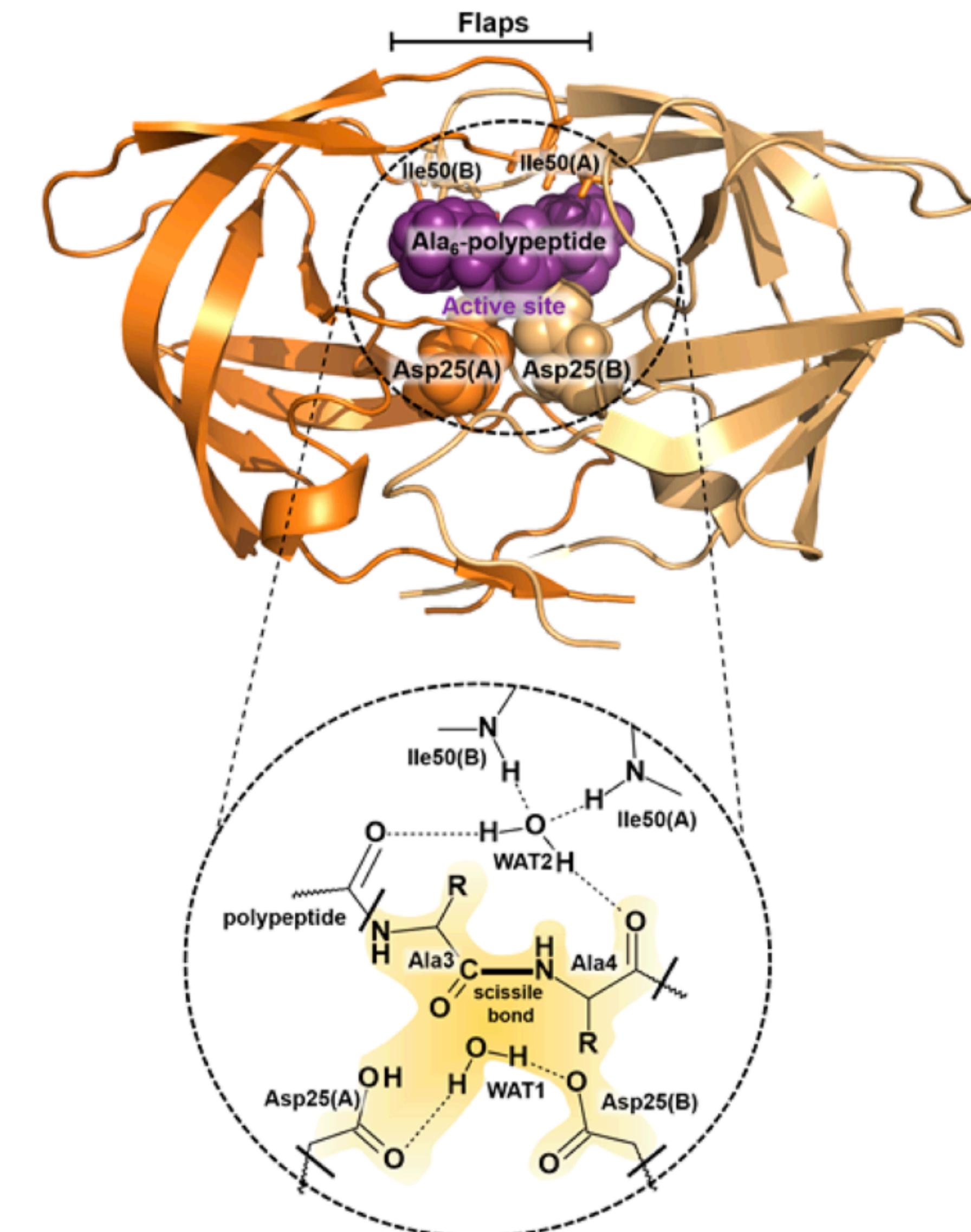
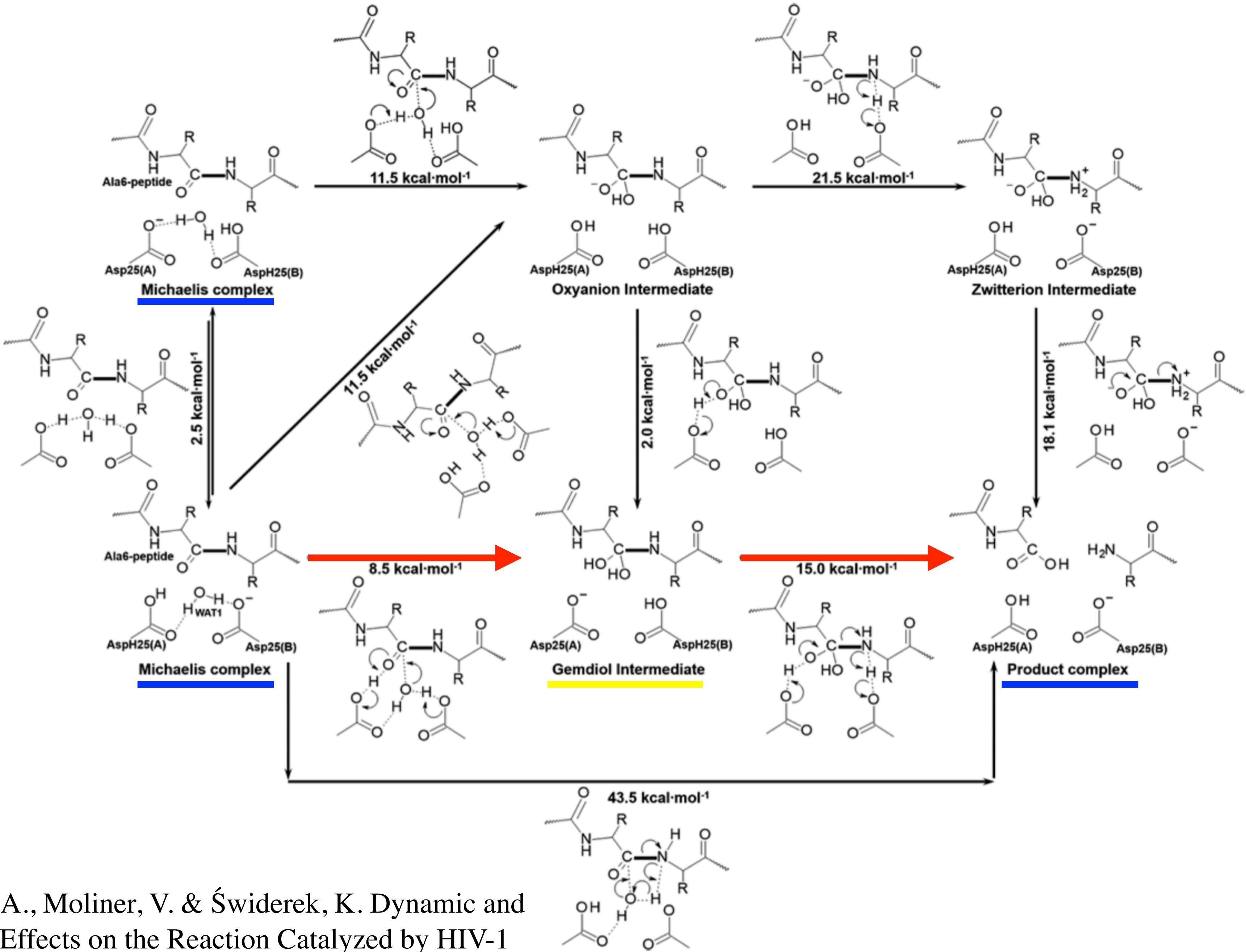


Figure 1. Structure of HIV-1 PR and detail of the active site, with protonated Asp-25(A), and Ala_6 peptide as substrate. Yellow area in the bottom panel contains the part of the system treated at QM level of theory during all the QM/MM calculations. Four link atoms are indicated as thick black lines.

HIV-1 protease

The reaction pathway proceeds as 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. Subsequently, the peptide bond is broken concerted with a double proton transfer, from the oxygen of the protonated Asp25(B) to the nitrogen atom of the scissile peptide bond and from one of the hydroxyl groups of the carbon atom of the peptide bond to the Asp25(A).

The rate-limiting step would correspond to the gemdiol decomposition into the products complex, 15.0 kcal/mol, close to the experiments 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 through metal ions.
Reaction with quinone that could:

- force a conformational change in the transmembrane part
- Modify the charge state
- As a result proton can go through.

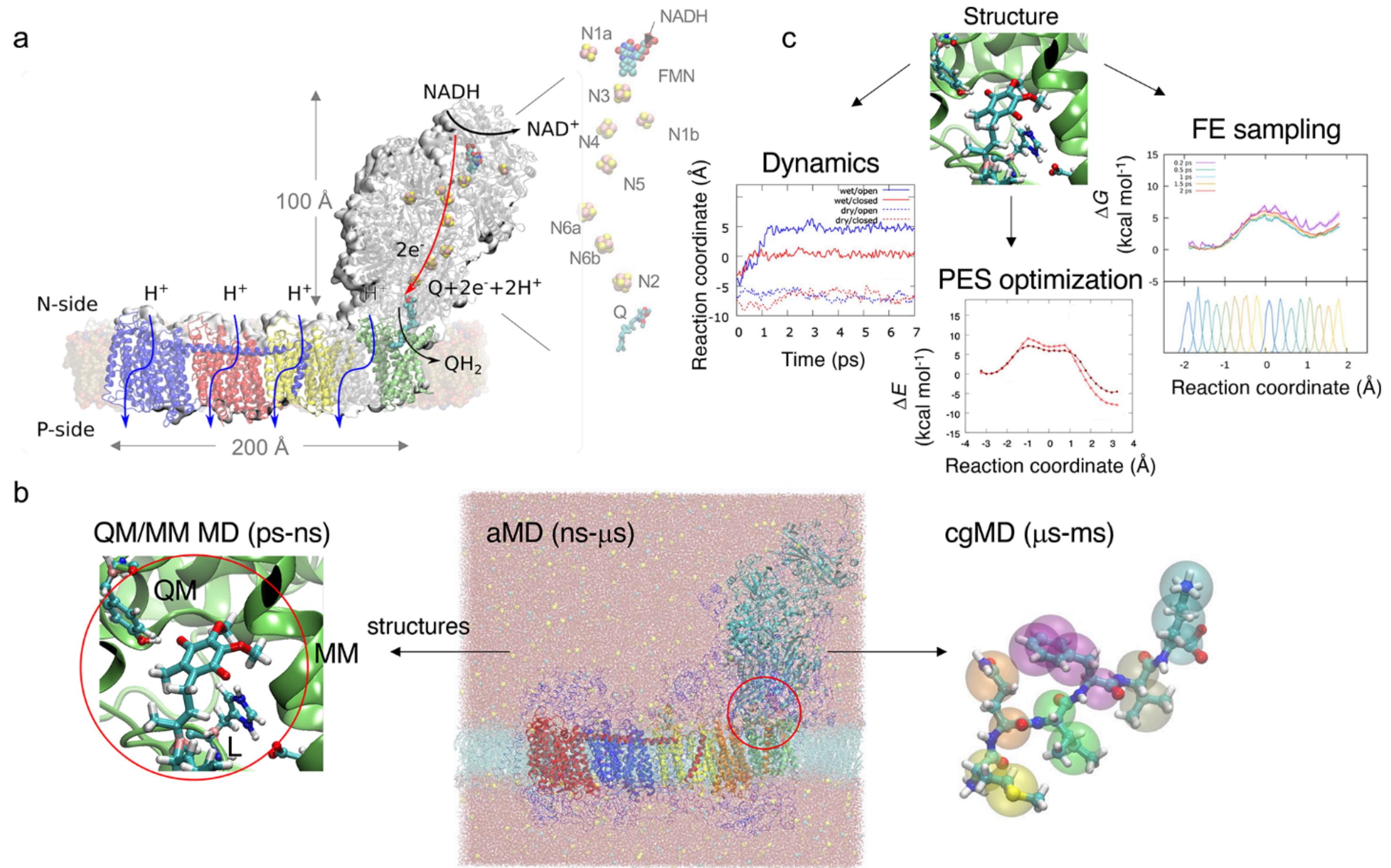
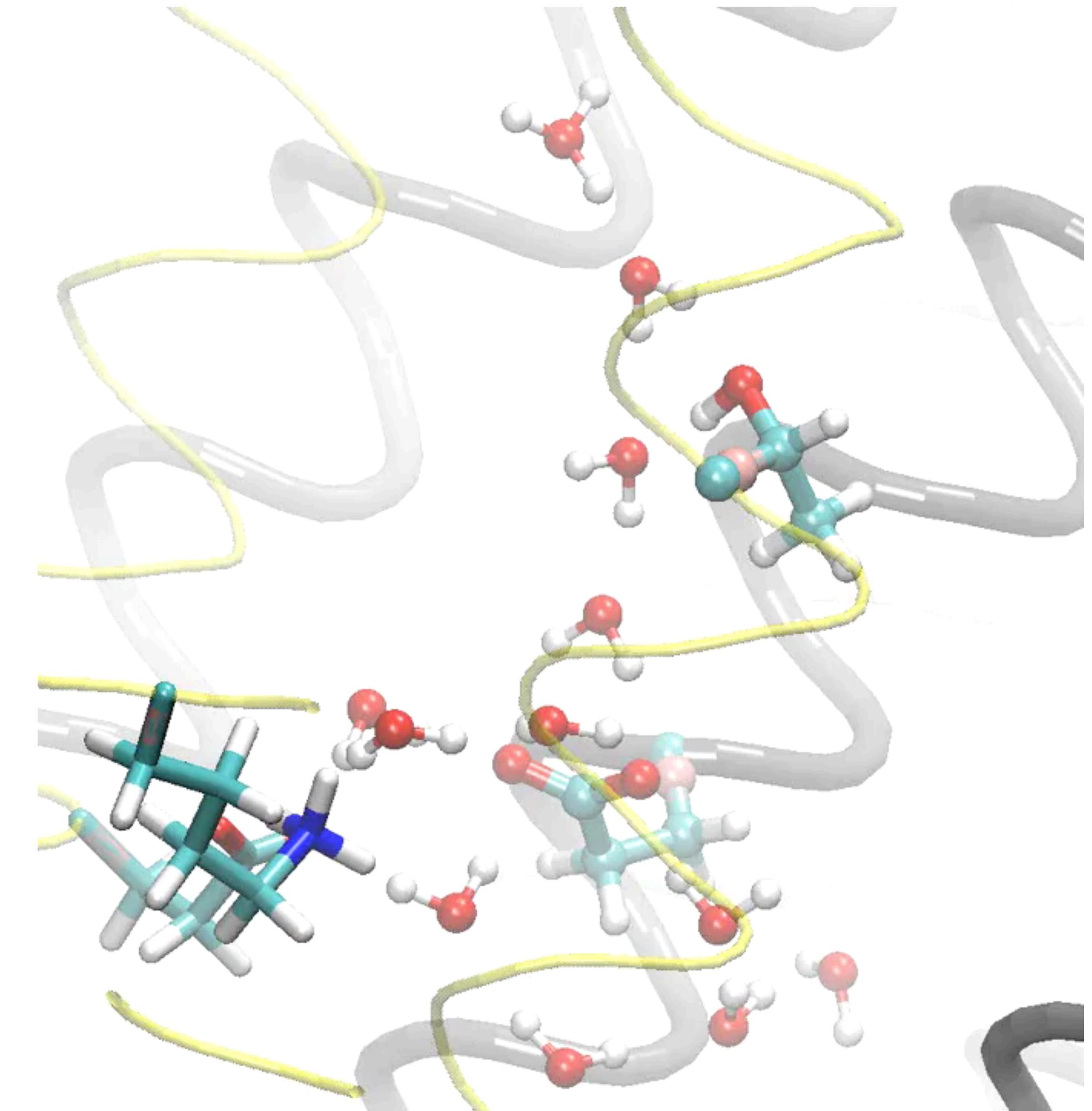


Figure 1. (a) Structure and function of Complex I. Reduced NADH donates electrons to the 100 Å long FeS chain that transfers them to quinone (Q). Q is reduced to quinol (QH_2), which triggers the transfer of four protons across the 200 Å long membrane domain. Point mutations of residues in the proton pumping subunits (shown in blue, red, yellow, gray/green) lead to inhibition of the Q oxidoreduction activity. (b) Multiscale simulation approaches can be used for probing the structure, function, and dynamics of PCET mechanisms in Complex I and other energy transducing enzymes. QM/MM models (left) allow exploring the local electronic structure, energetics, and dynamics on picosecond time scales (QM, QM region; MM, MM region; L, link atom in pink); atomistic MD (aMD) simulations (middle) enable sampling of the microsecond dynamics in a model of the biochemical environment; whereas coarse-grained models (cgMD) (right, showing a 1:4 mapping of beads:heavy atoms) allow exploring the micro- to millisecond time scale, but with loss of atomic detail. (c) The systems can be explored by unbiased MD simulations, potential energy surface (PES) scans, or free energy sampling methods at the different theory levels. The MD simulations allow probing, e.g., the dynamics of a reaction coordinate over time (here proton transfer, PT), whereas PES scan or FE sampling allows computing free energy/energy profiles along a reaction coordinate of interest (here a PT reaction).

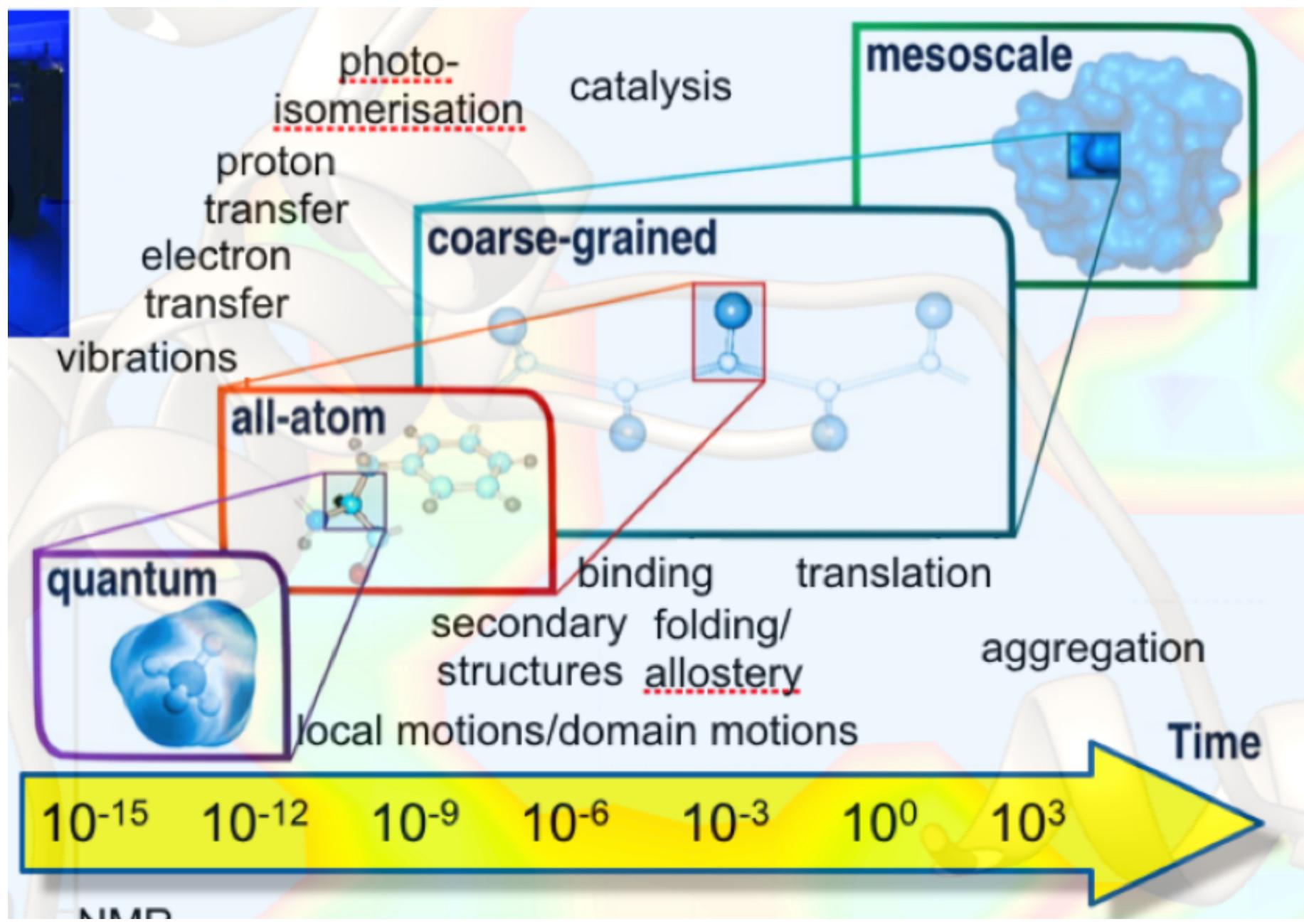


Respiratory complex I

- 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 1ps.



Simple and complex models



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

Different processes can require different simulation techniques.

Size and time scale are two key issues related together to the problem of Sampling. The larger is the system, including solvent if the case, the slower is the simulation.

Long time scales can be unaccessible also for small systems given the short time scale of the time step (fs)

Simplified Models

The goal of simplifying models when doing simulations is to both extend the accessible time scale as well as to simulate larger and larger systems

Simplifying a model can be achieved for example by:

- Decreasing the resolution, in the same manner by using atoms/beads we were decreasing the resolution from electrons/nuclei
- Removing part of the systems (for example removing the solvent and somehow implicitly taking that into account in some other way)
- Modifying the way atoms/beads interact one with the other, for example in classical MD simulations electrostatic interactions are the most computational expensive part.



Simplified Models

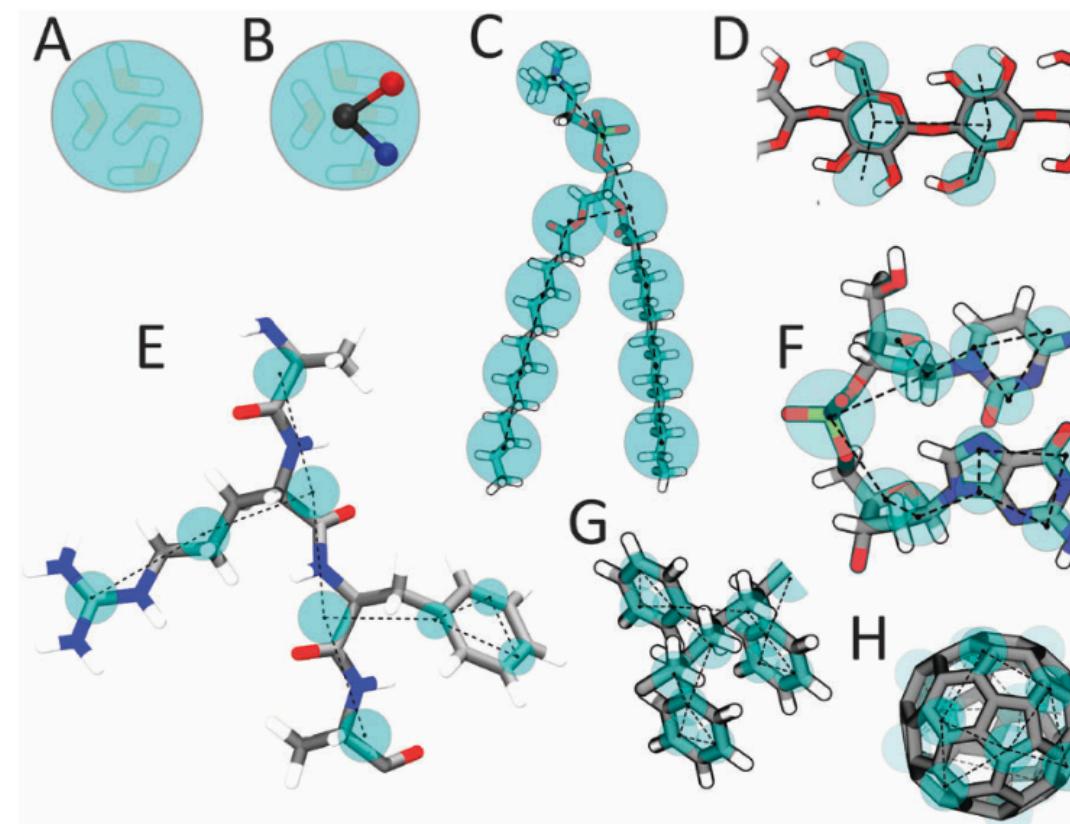
The goal of simplifying models when doing simulations is to both extend the accessible time scale as well as to simulate larger and larger systems

There are two main strategies to simplify a model:

Phys/Chem Based

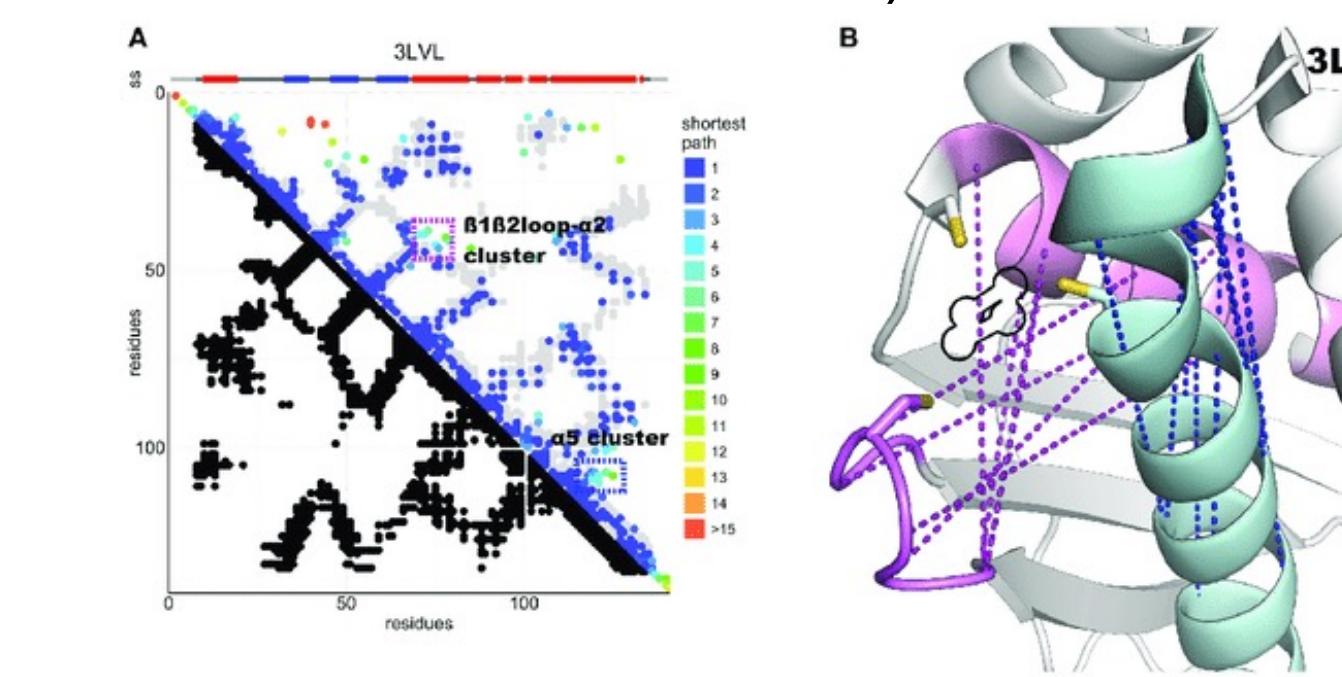
In this case the interactions try to conserve relevant phys/chem properties like the polar/hydrophobic/charged nature of amino acids.

An example is the MARTINI force-field



Knowledge Based

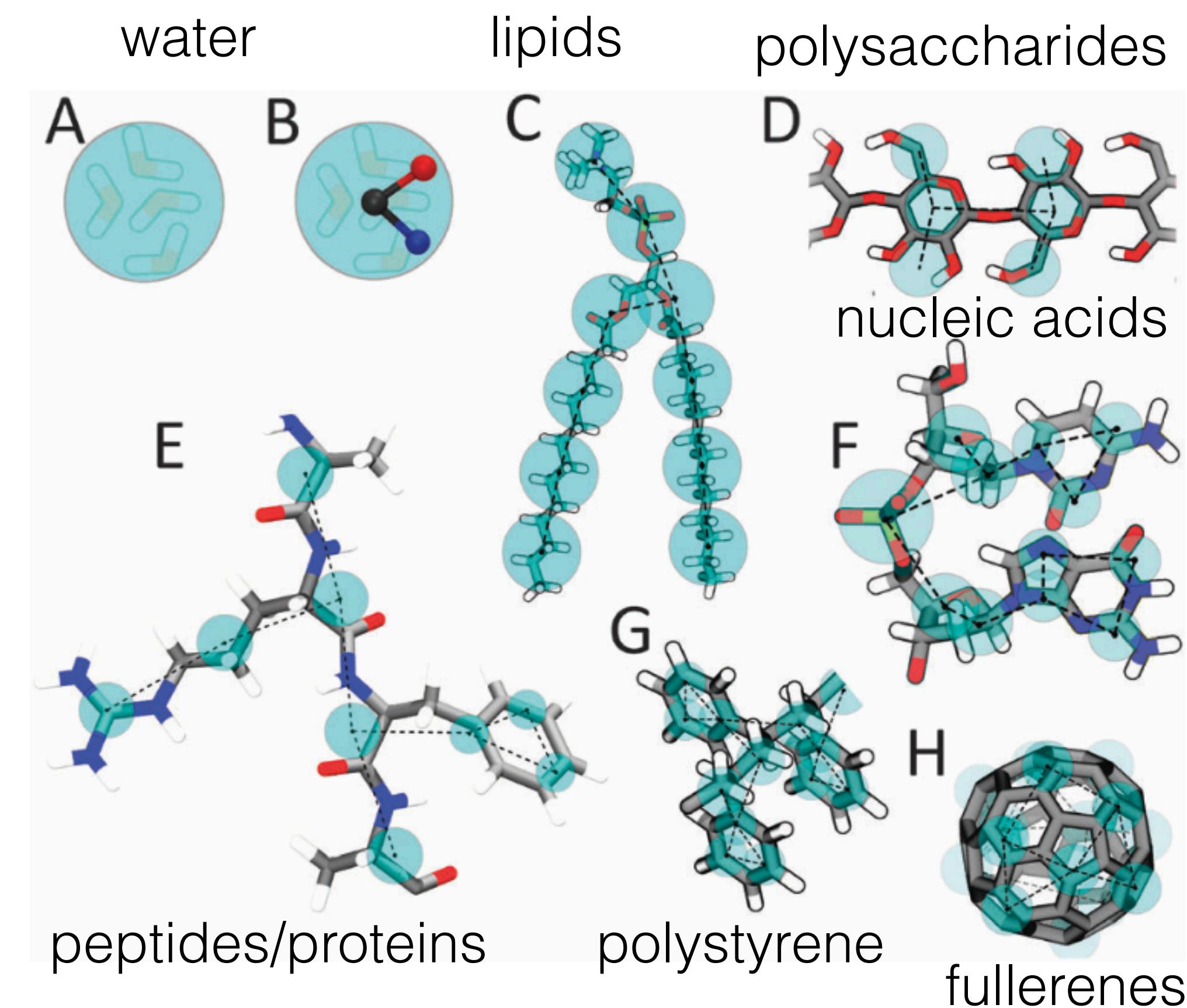
In this case the interactions try to reproduce other source of knowledge, for example sequence conservation profiles, co-evolution analysis, structural knowledge. An example are Go-Model (also known as structure-based models)



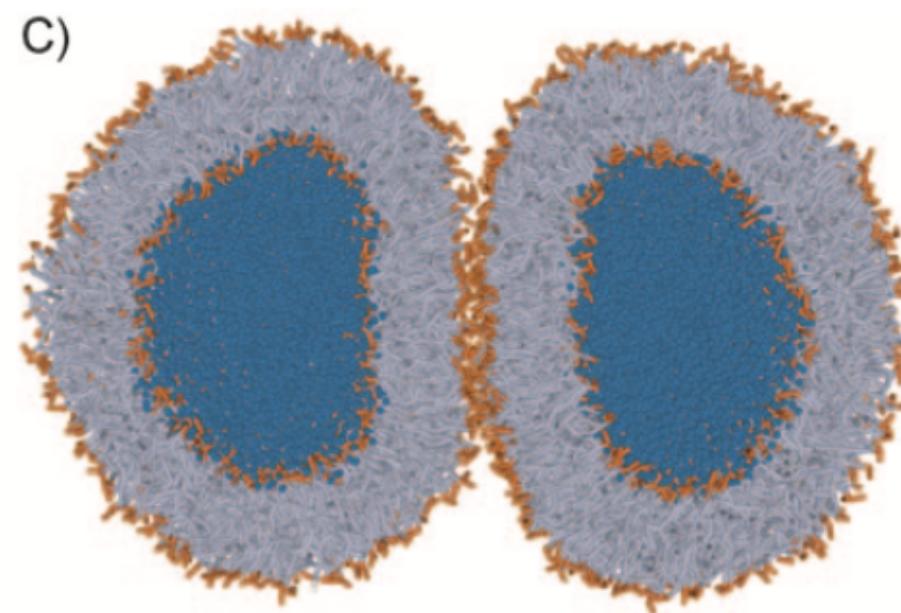
Martini: a transferable Coarse Grain model

Goal: a coarse grain model with a transferable, physics based potential, able to simulate very large systems at quasi chemical resolution.

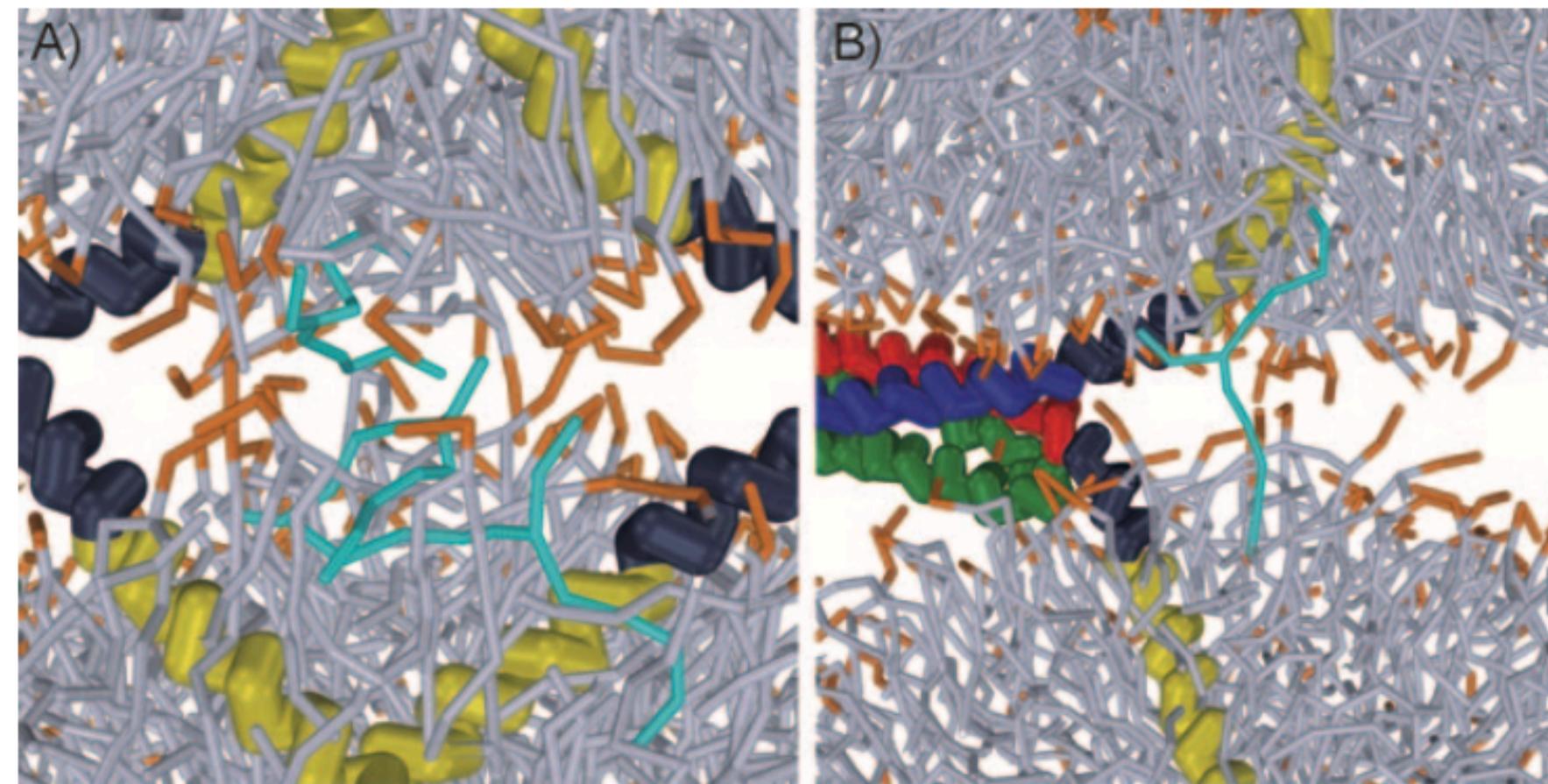
The main idea is to use a 4 to 1 mapping (4 heavy atoms and their hydrogens into a single particle) with the exception of rings for which the mapping is 2 to 1. Then define a number of building blocks and fit the force-field in order to reproduce thermodynamic behaviours.



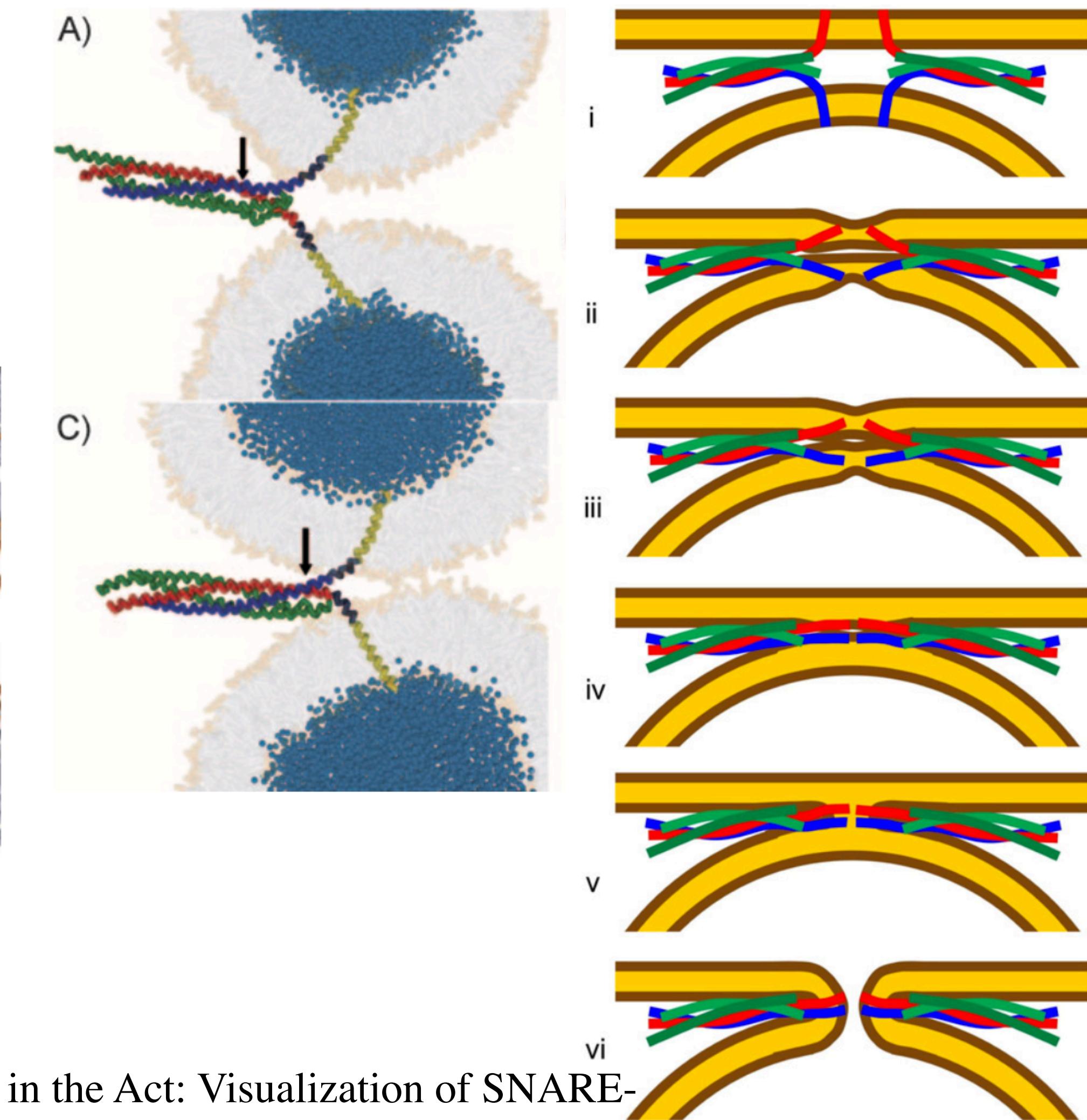
Martini examples: Vesicular Fusion



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

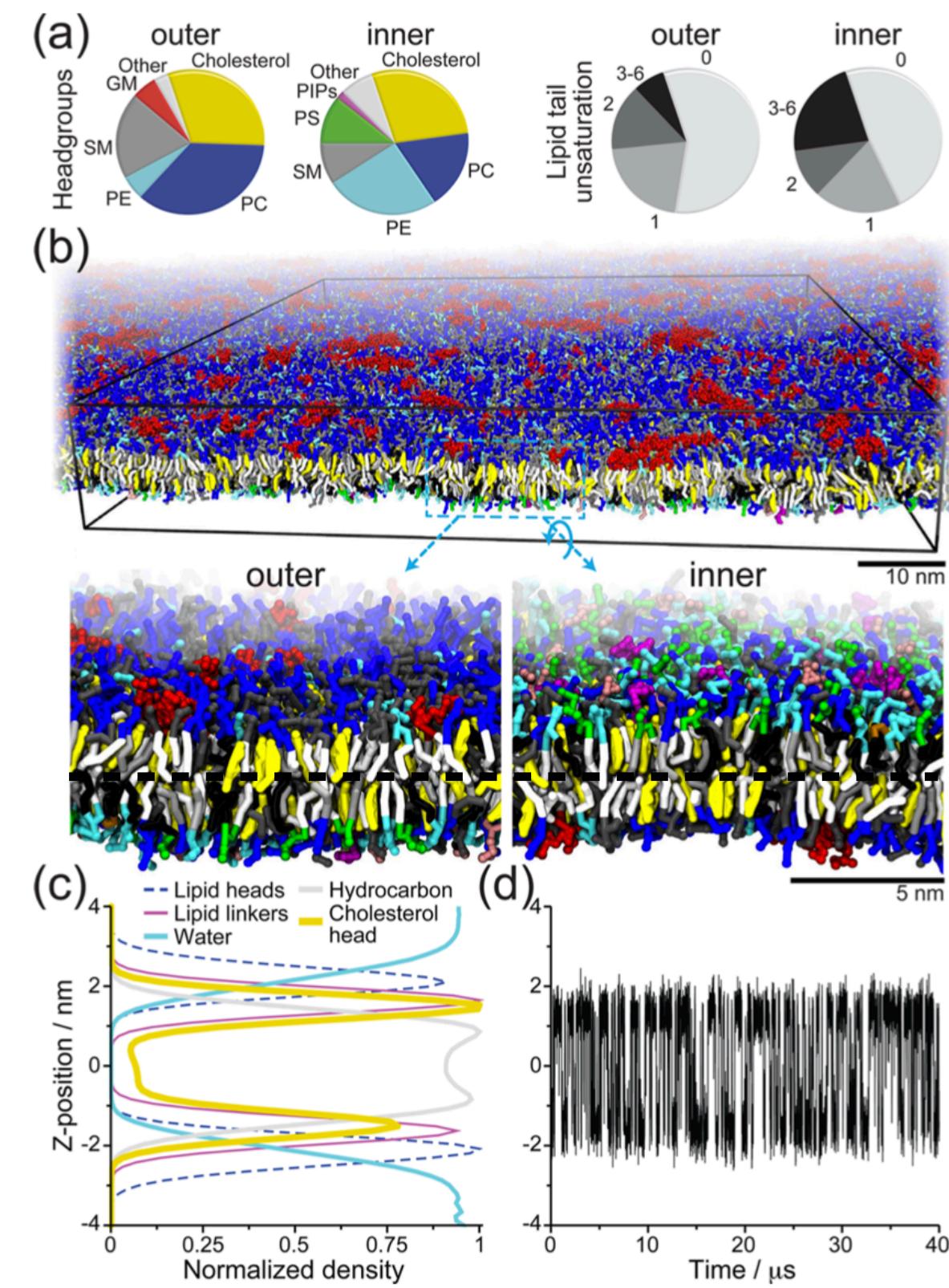


Risselada, H. J., Kutzner, C. & Grubmüller, H. Caught in the Act: Visualization of SNARE-Mediated Fusion Events in Molecular Detail. *Chembiochem* **12**, 1049–1055 (2011).

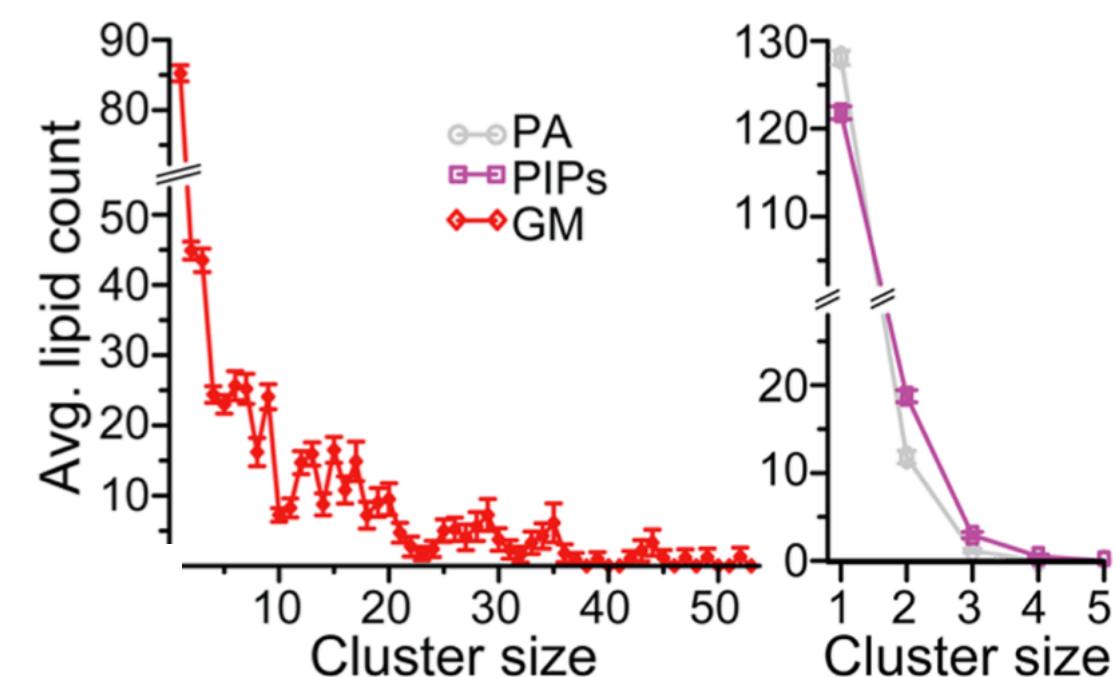


Martini Examples: Visualising a Plasma Membrane

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



bilayer boundary



Observations

1. GlycoLipids (GM) clusterise more than PhosphoLipids (PA/PIP_s).
2. Cholesterol diffuses very quickly between the inner and outer leaflets
3. Cholesterol is more concentrated in the outer leaflet than in the inner one (54:46)



Structure-Based models

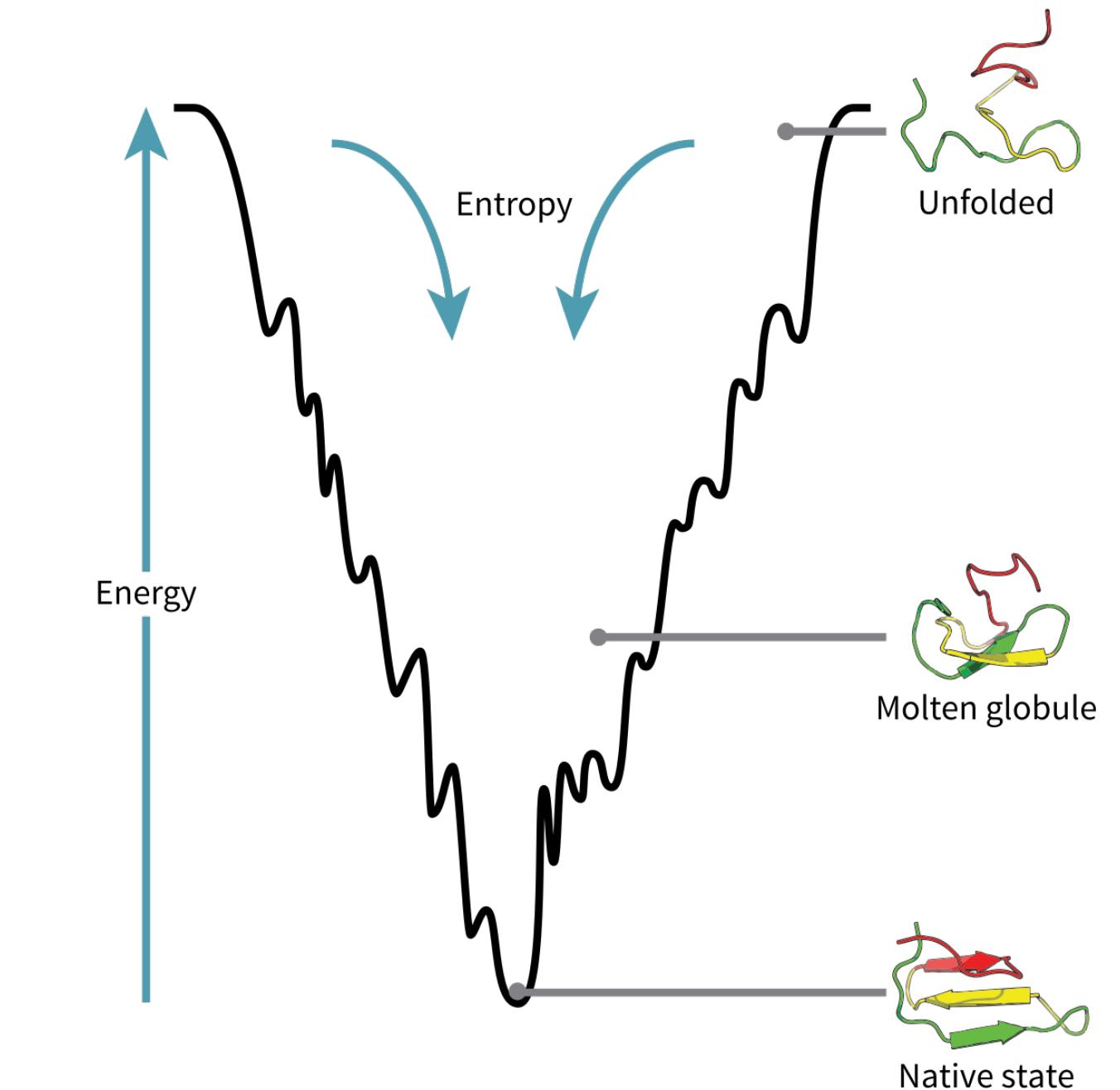
Structure based models for protein originated from the observation that

- folded proteins posses a very well defined 3D structure
- this 3D structure is very stable so it must be a free energy and a potential energy minimum for that sequence of amino acid

As a consequence the spatial organisation of the atoms in 3D should reflect almost optimal interactions among the atoms.

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

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



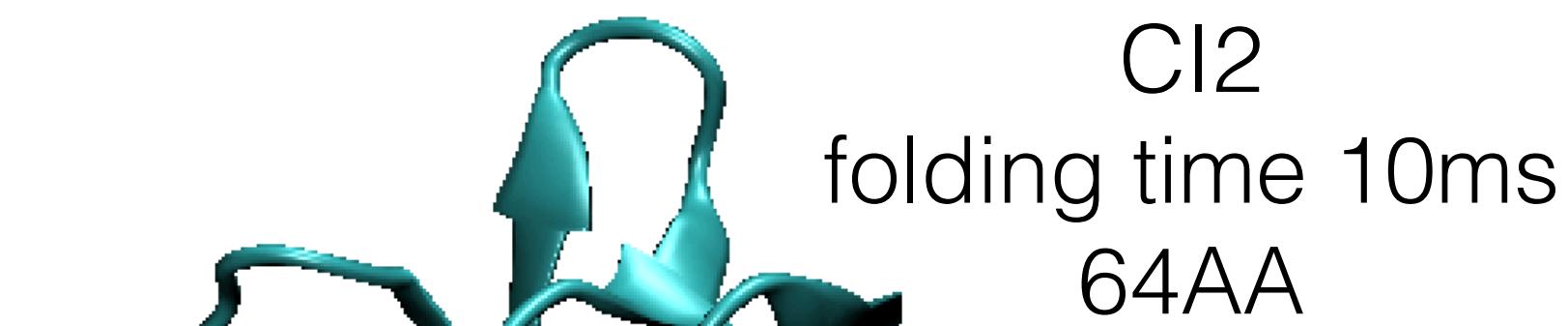
Instead of having parameters valid in general, we have parameters valid for the specific protein structure, in this way the most stable configuration is by definition the protein structure.

Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? an investigation for small globular proteins. *J Mol Biol* **298**, 937–953 (2000).

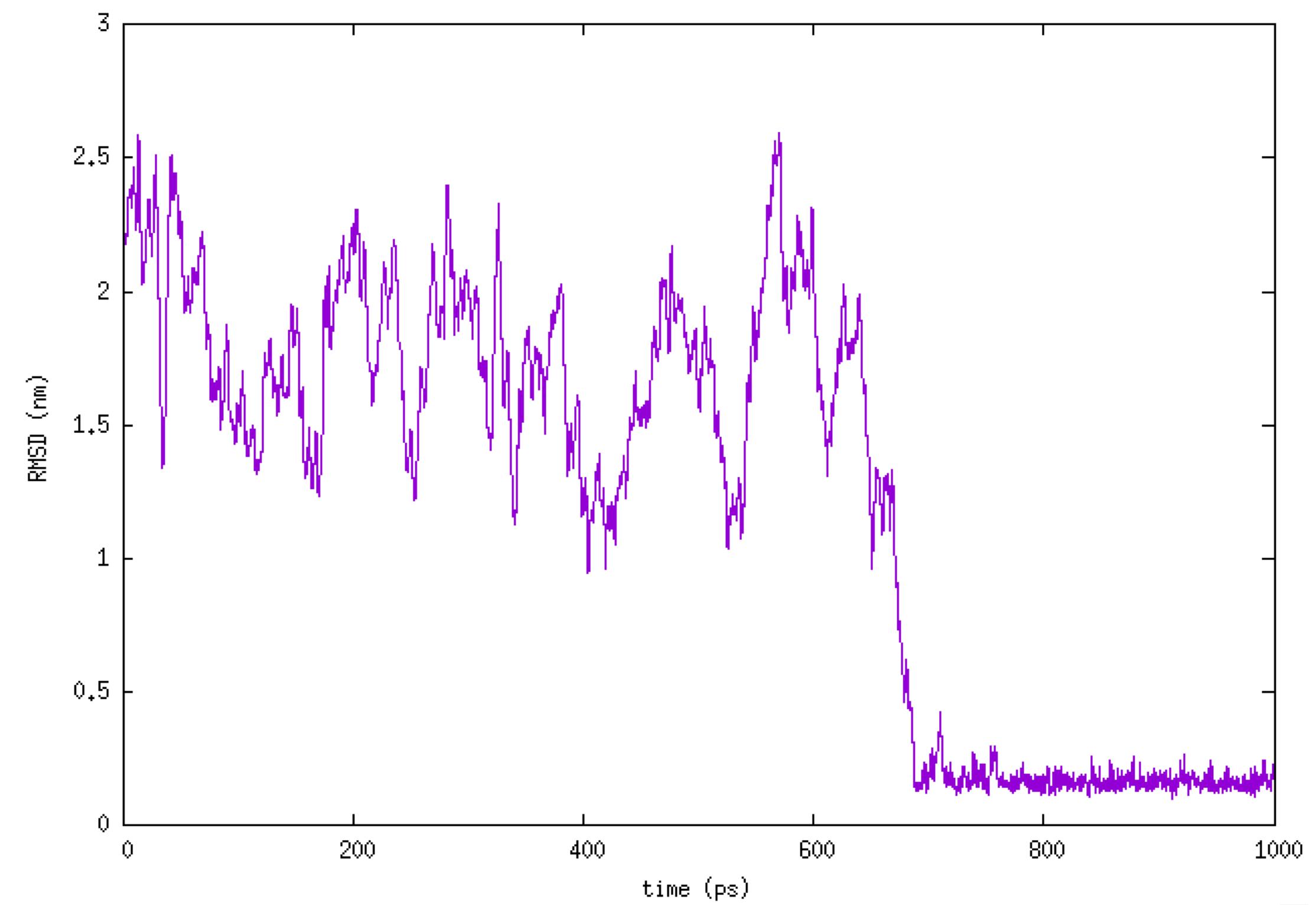
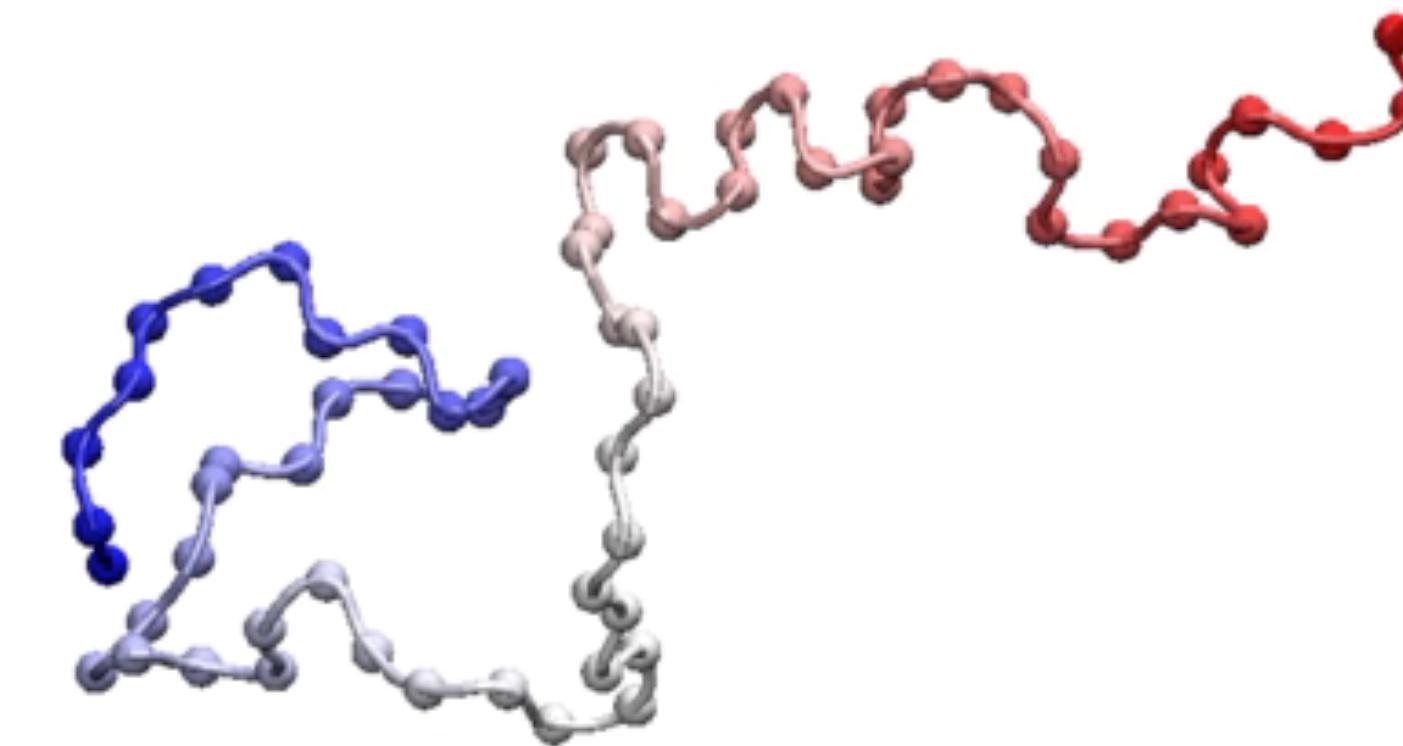


Structure-Based models

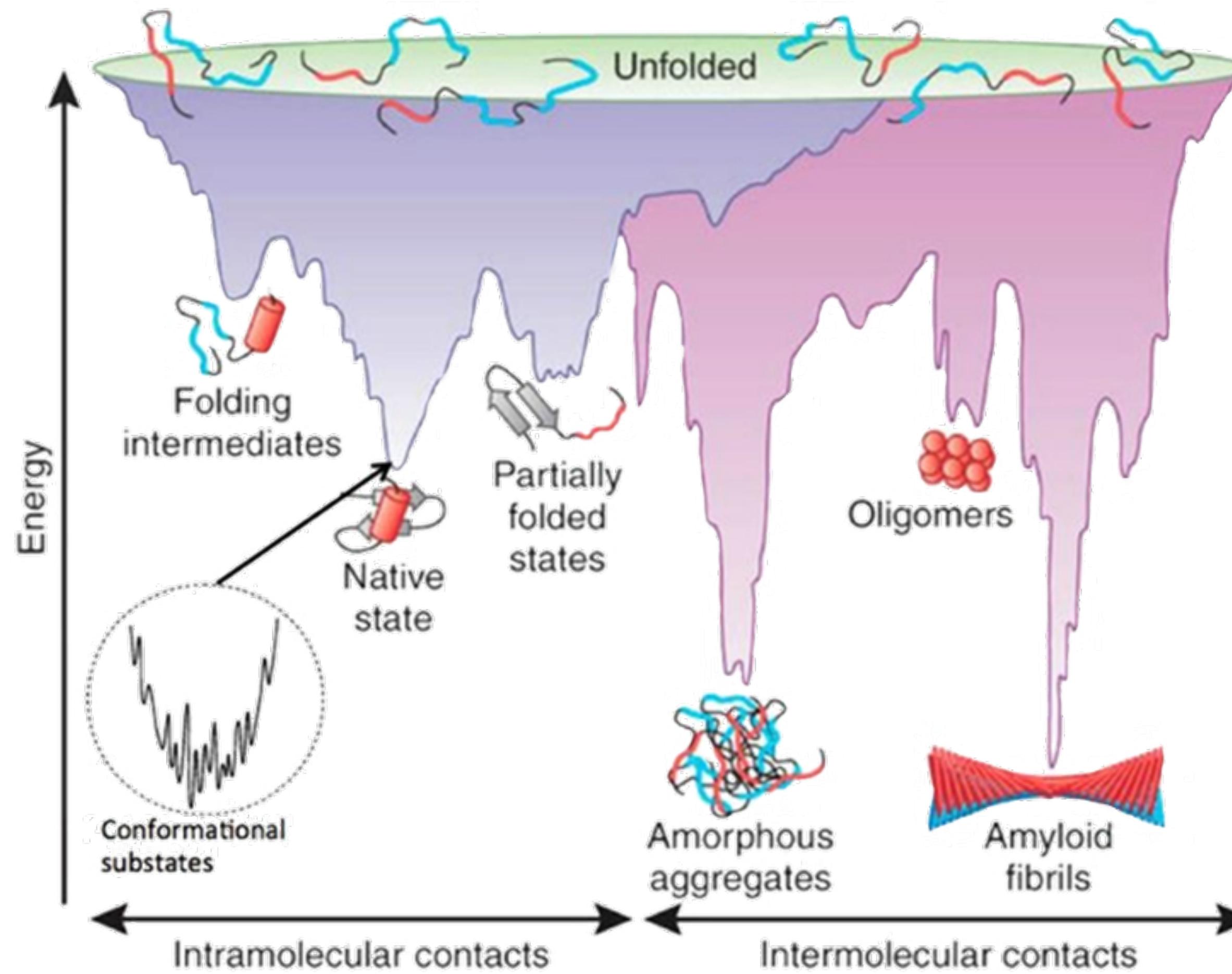
In the '90-early 2000 this approach was used to study the mechanism of protein folding, nowadays it is used (not very often) to study the motion of very large proteins



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Structure-Based models for protein aggregation and more



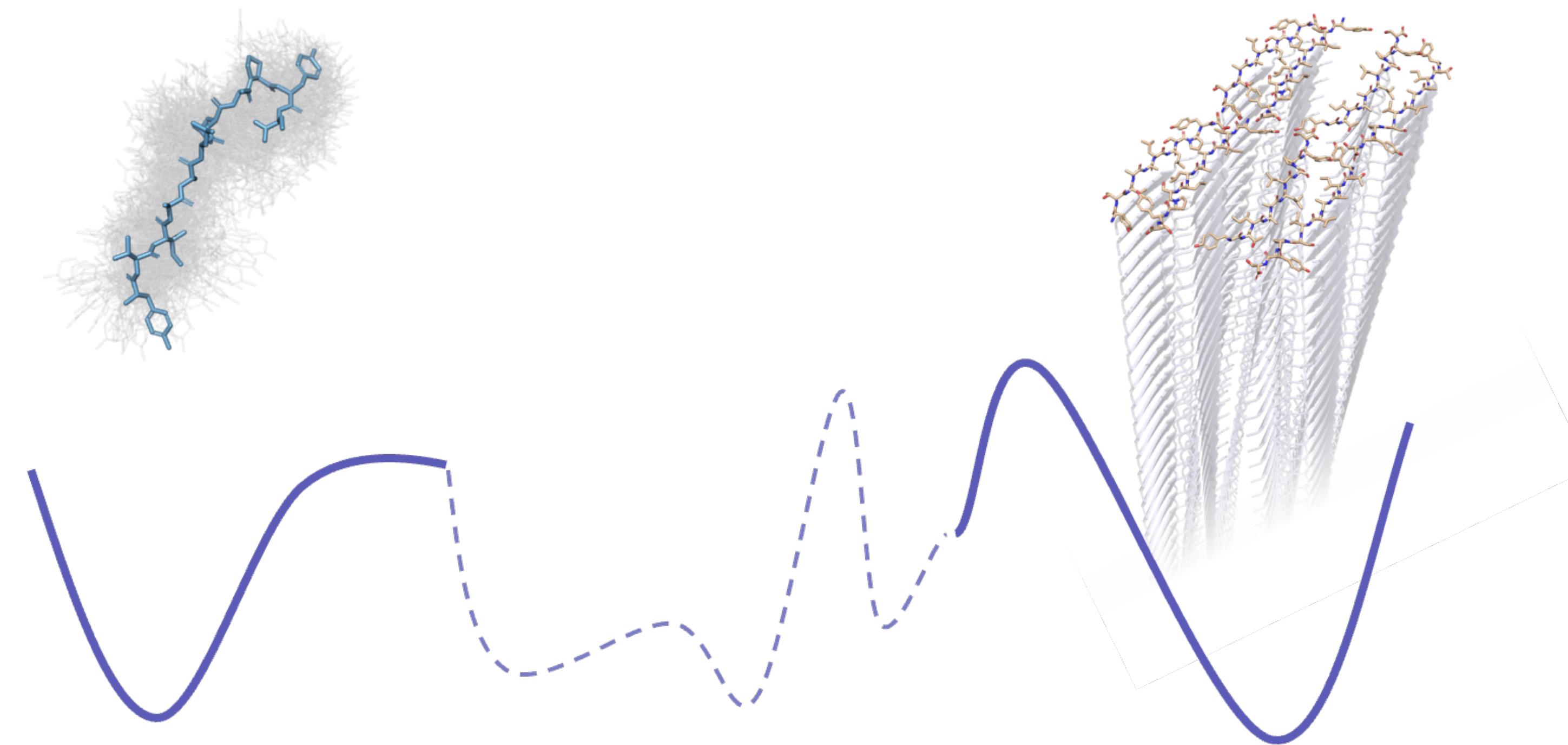
The general understanding of the behaviour of a protein in solution is that its native and amyloid states are the only free energy minima characterised by a well defined structure.

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

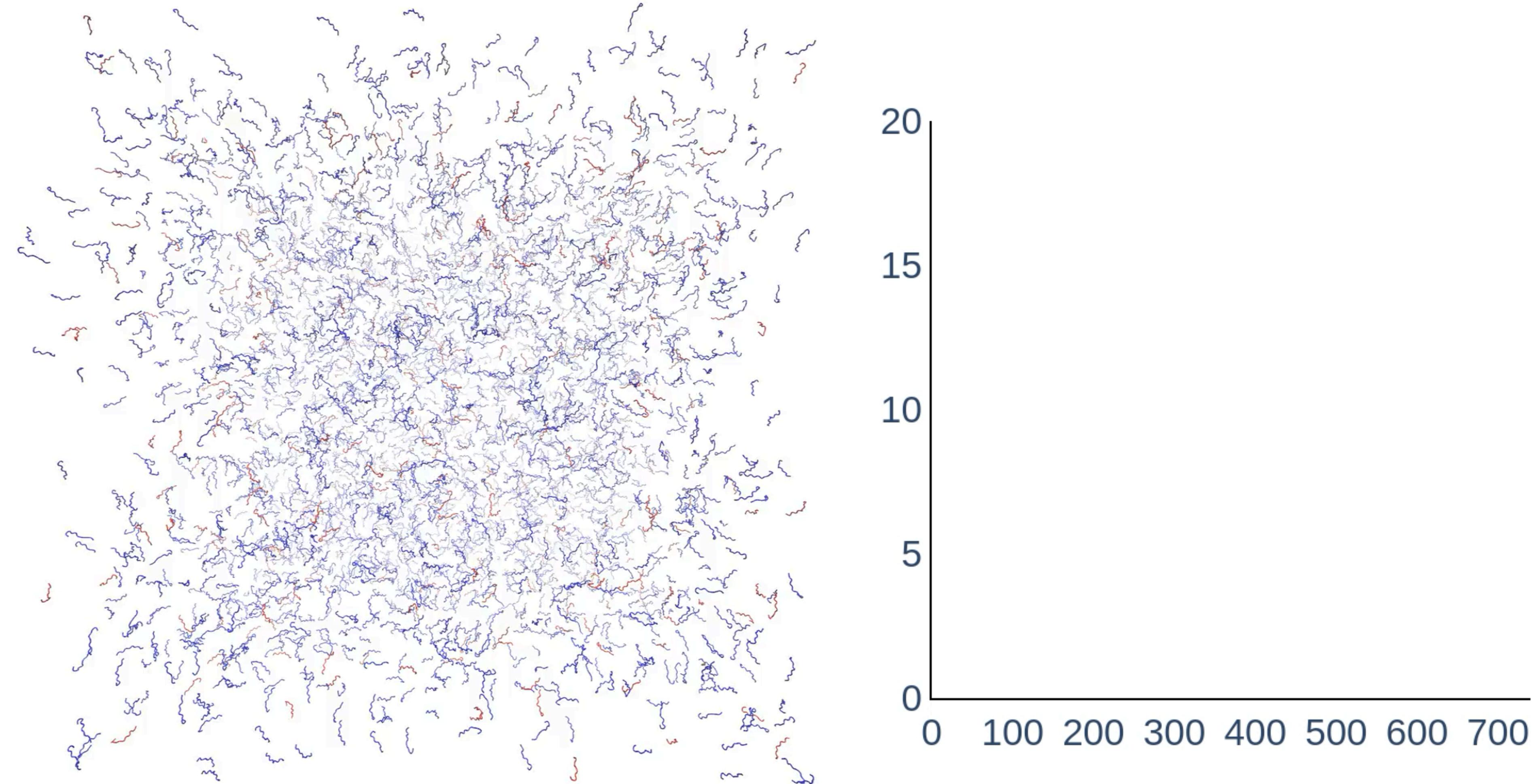


Structure-Based models for protein aggregation and more

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

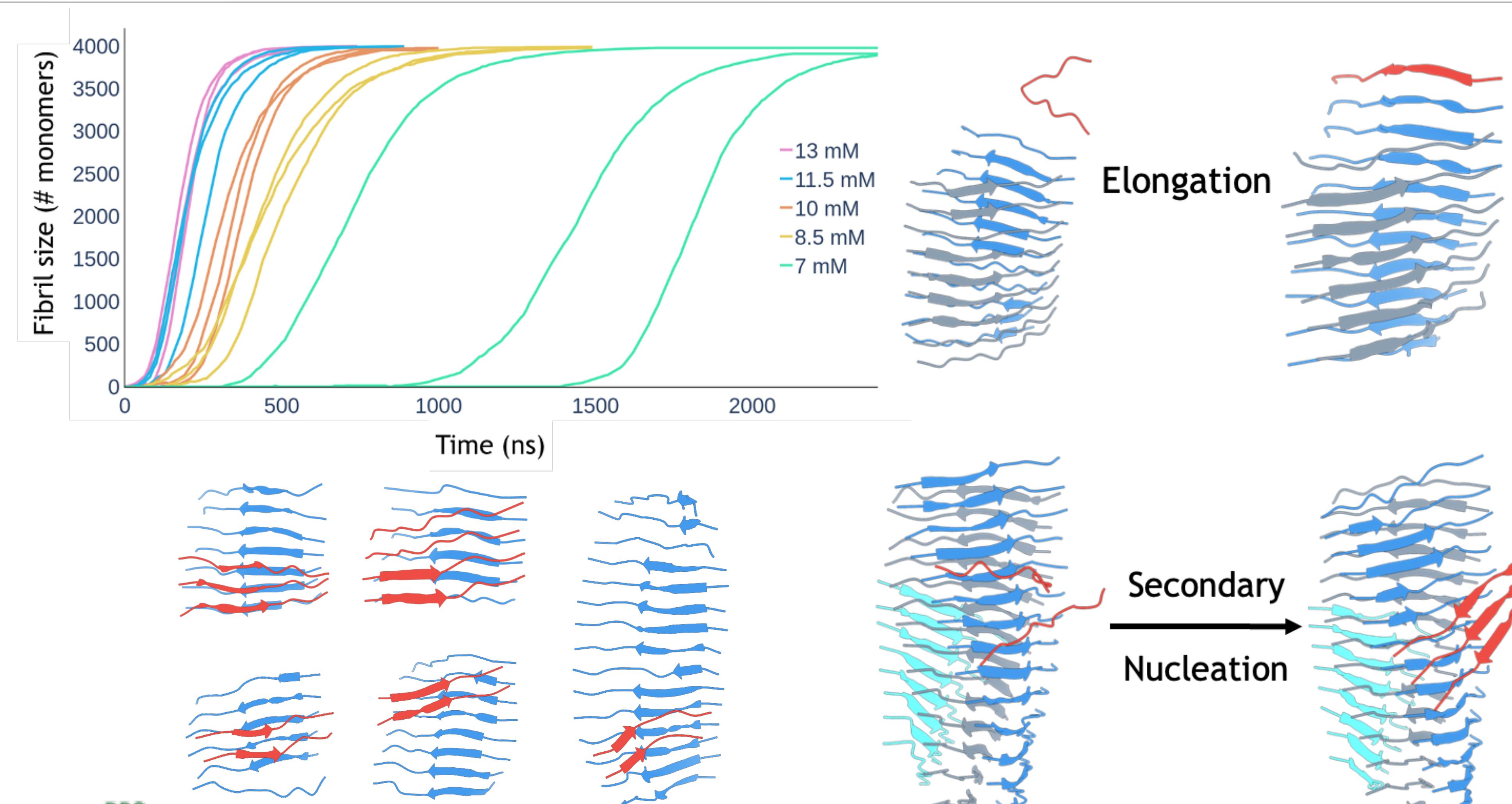


UNIVERSITÀ
DEGLI STUDI
DI MILANO

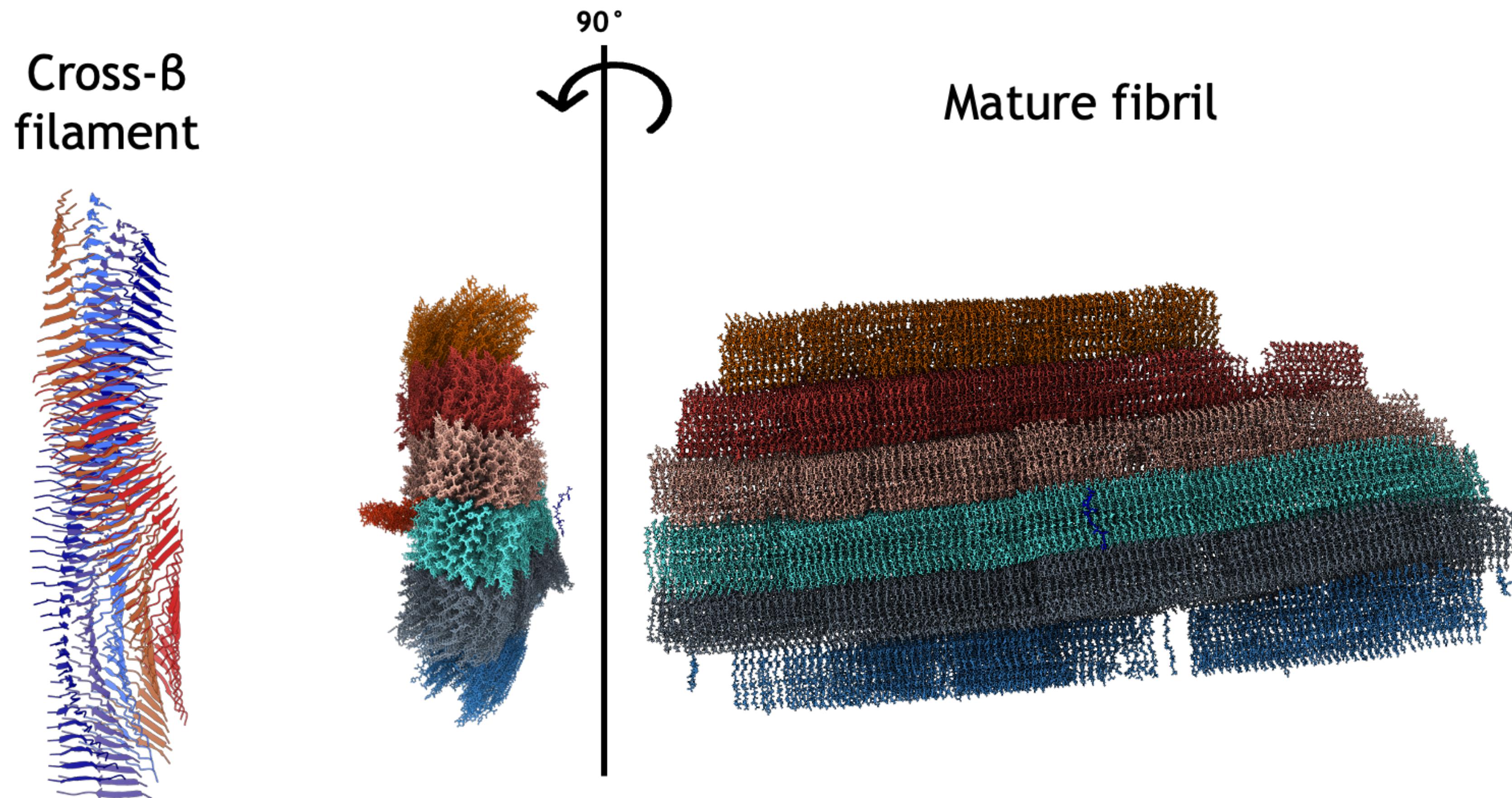


Scalone, E. *et al.* Multi-eGO: An in silico lens to look into protein aggregation kinetics at atomic resolution. *Proc National Acad Sci* **119**, e2203181119 (2022).

Structure-Based models for protein aggregation and more



Structure-Based models for protein aggregation and more



Choose the most appropriate simulation technique(s)

