

Seeing the PDB

Received for publication, April 3, 2021, and in revised form, April 26, 2021 Published, Papers in Press, May 4, 2021,
<https://doi.org/10.1016/j.jbc.2021.100742>

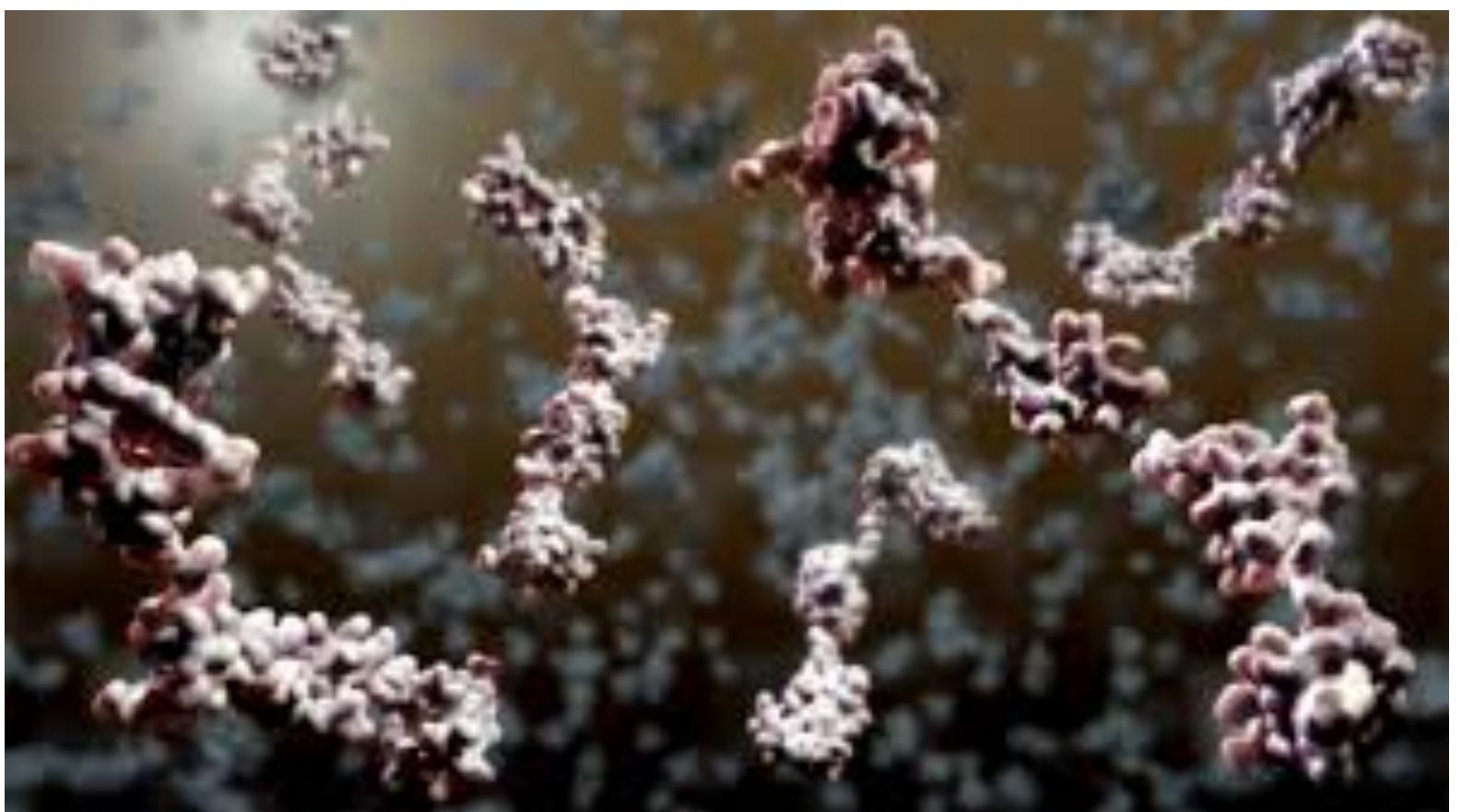
Jane S. Richardson^{1,*}, David C. Richardson¹, and David S. Goodsell^{2,3,*}

From the ¹Department of Biochemistry, Duke University, Durham, North Carolina, USA; ²Department of Integrative and Computational Biology, The Scripps Research Institute, La Jolla, California, USA; ³Research Collaboratory for Structural Bioinformatics Protein Data Bank, Rutgers, the State University of New Jersey, Piscataway, New Jersey, USA

Edited by Karin Musier-Forsyth

Observing the microscopic world is hard:

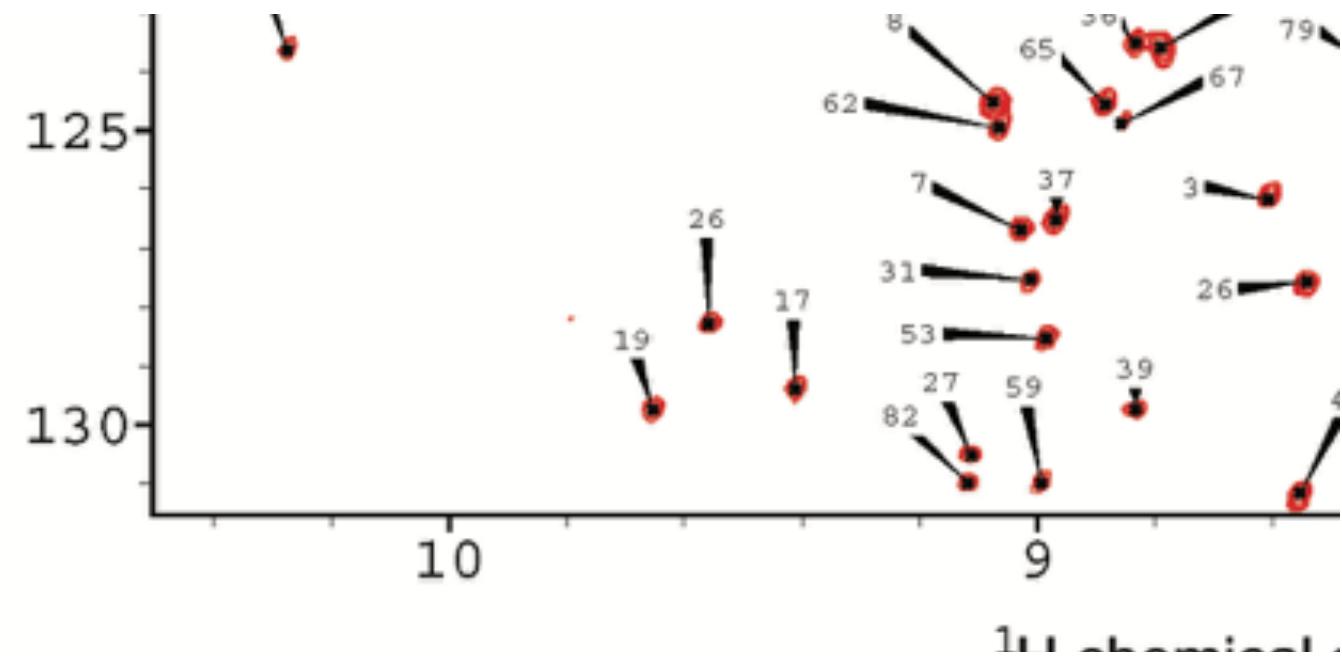
- We do not have a light microscope, so we don't see directly the molecules, we need to interpret other kind of signals (shorter wavelength, electrons, magnetic fields, ...)
- We cannot do movies, so we cannot see molecules in motion, but we take picture with more or less long exposure times
- We do not look at a single molecule at a time but at many (~fraction of Avogadro number)



So how do we “take pictures” of molecules?

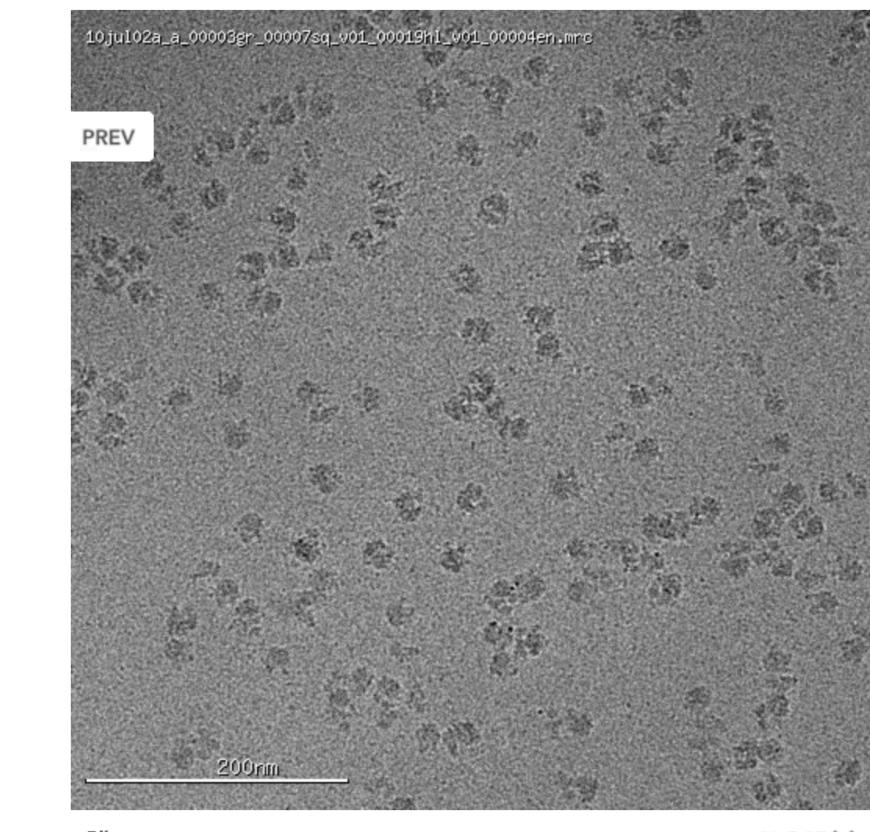
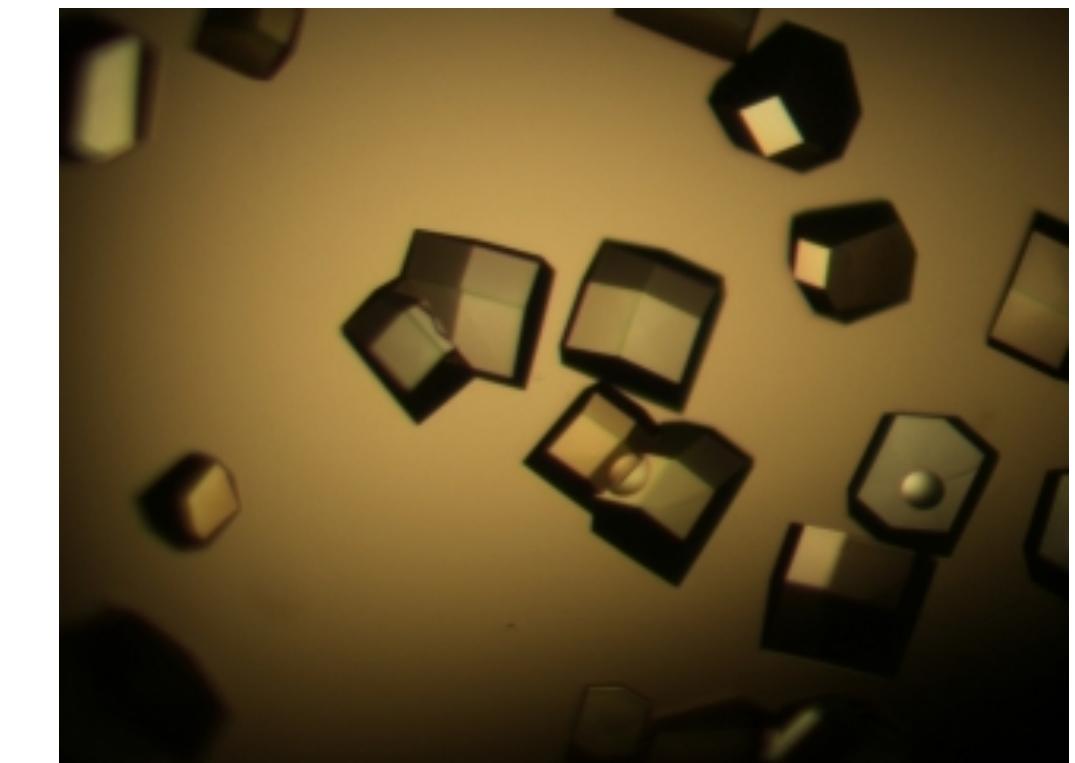
Issues are signal/noise, small size, motion on many time scales

We make all molecules adopt exactly the same structure
building up highly symmetric crystals, to be observed
using X-rays



We employ the radio frequency absorption and emission properties of nuclei in presence of a magnetic field. In this way we can look at the signal of the same atom in all the molecules.

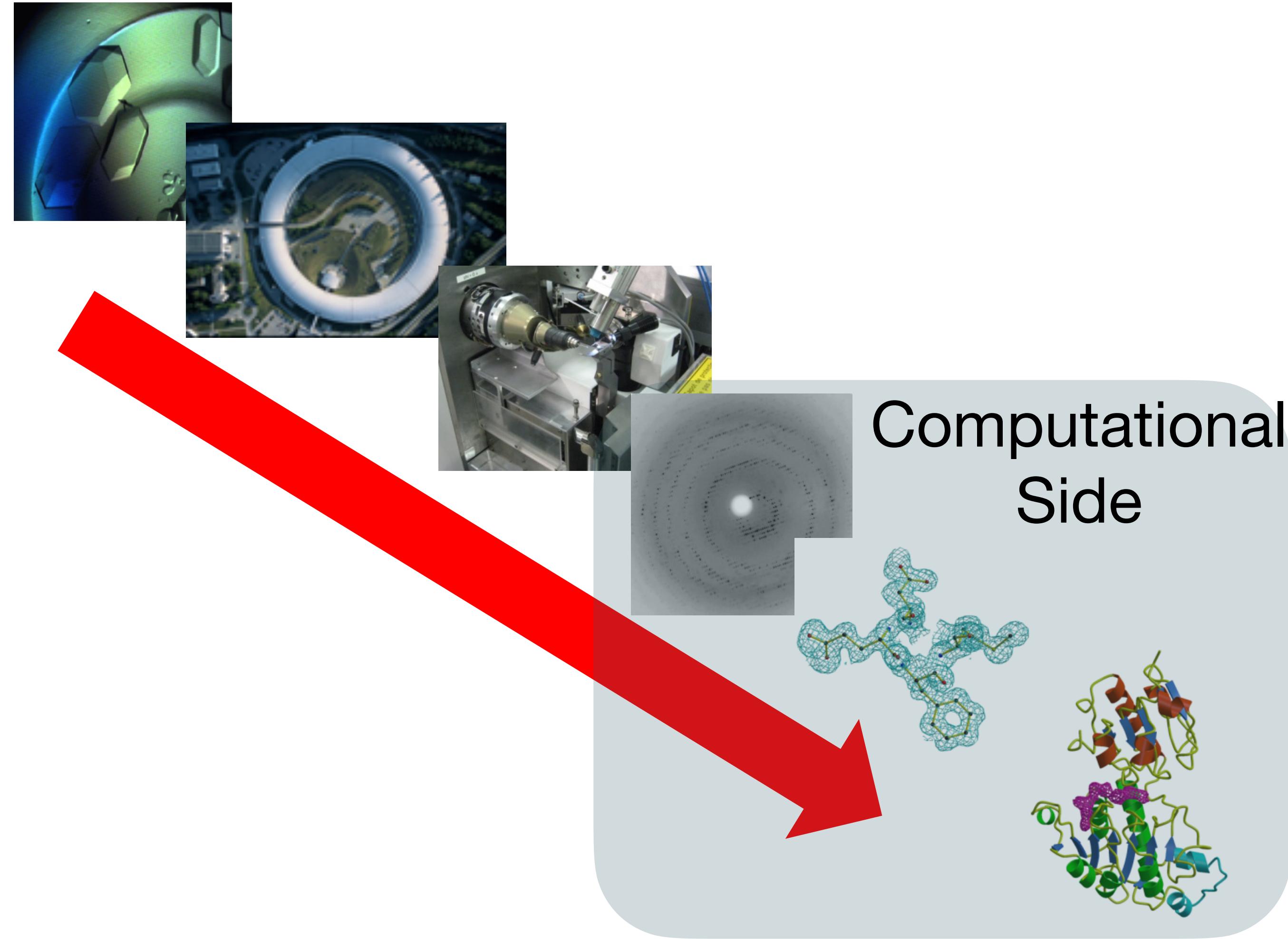
We vitrified the sample on a single layer so to take pictures with an electron microscope using a very low dose of electrons



UNIVERSITÀ
DEGLI STUDI
DI MILANO



X-ray



UNIVERSITÀ
DEGLI STUDI
DI MILANO



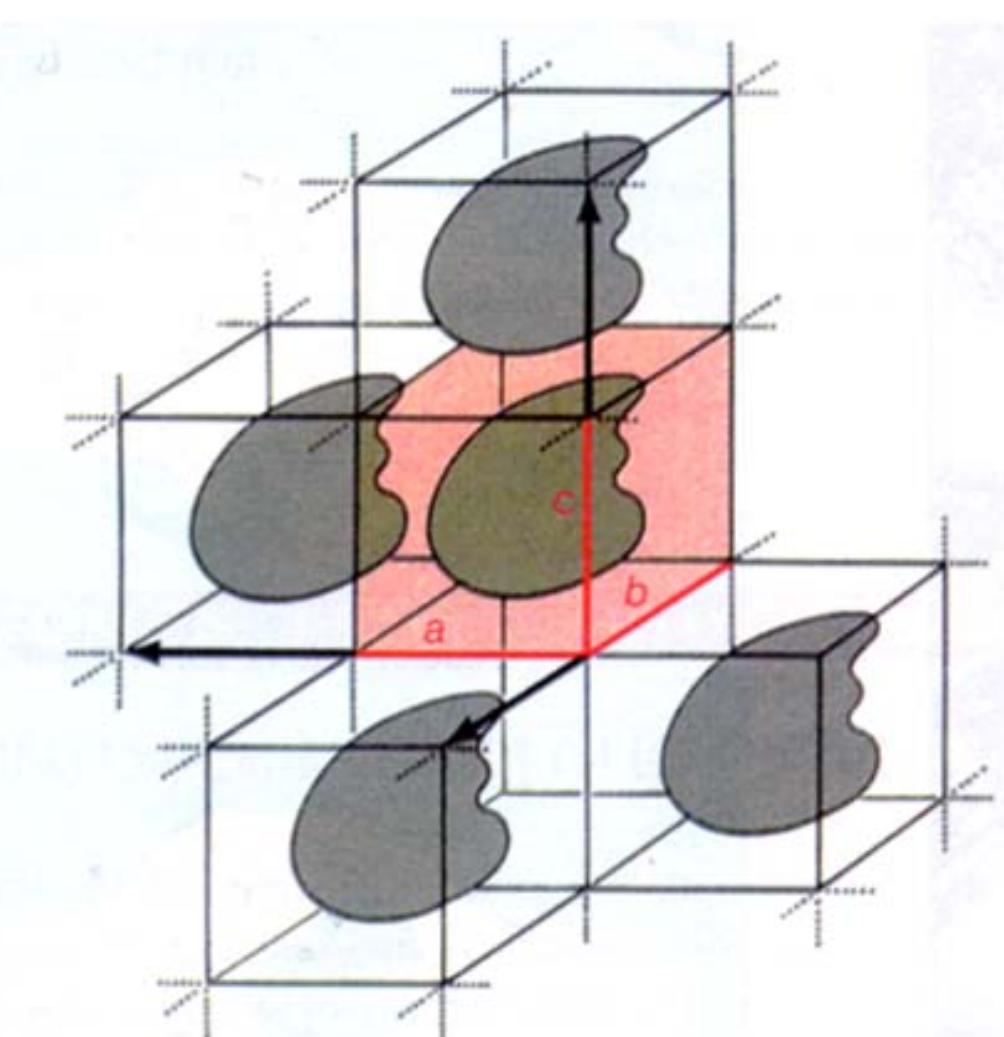
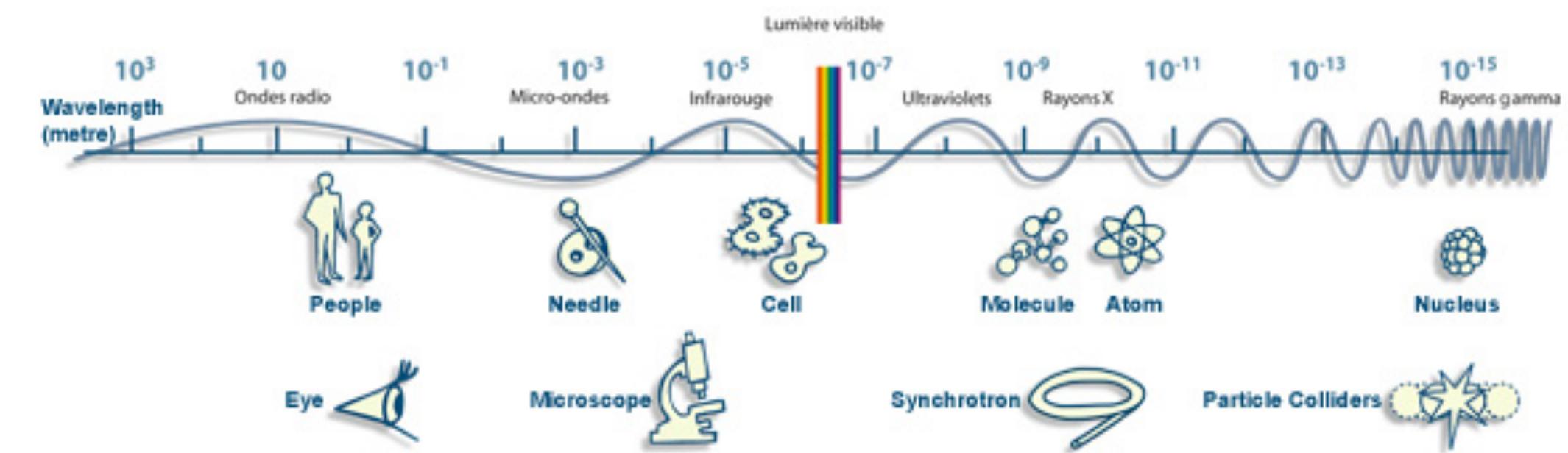
X-ray

Why X-rays ?

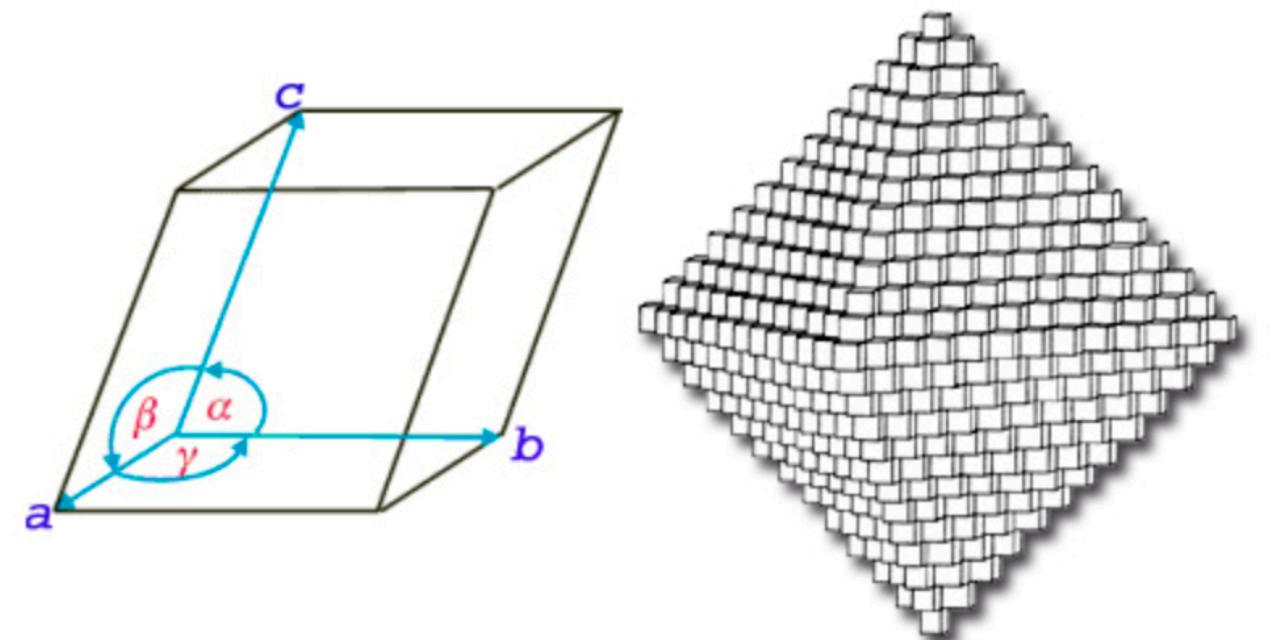
The wavelength of a X-rays is roughly **1 Å** (10^{-10} m), which is on the scale of a single atom, and it allows to have **sufficient resolution to determine the atomic positions**

Why crystals ?

X-ray crystallography requires a crystal to amplify the signal (**10^{15} - 10^{16} identical molecules**); the periodicity of the electron density is used to diffract the X-rays with manageable measurement error



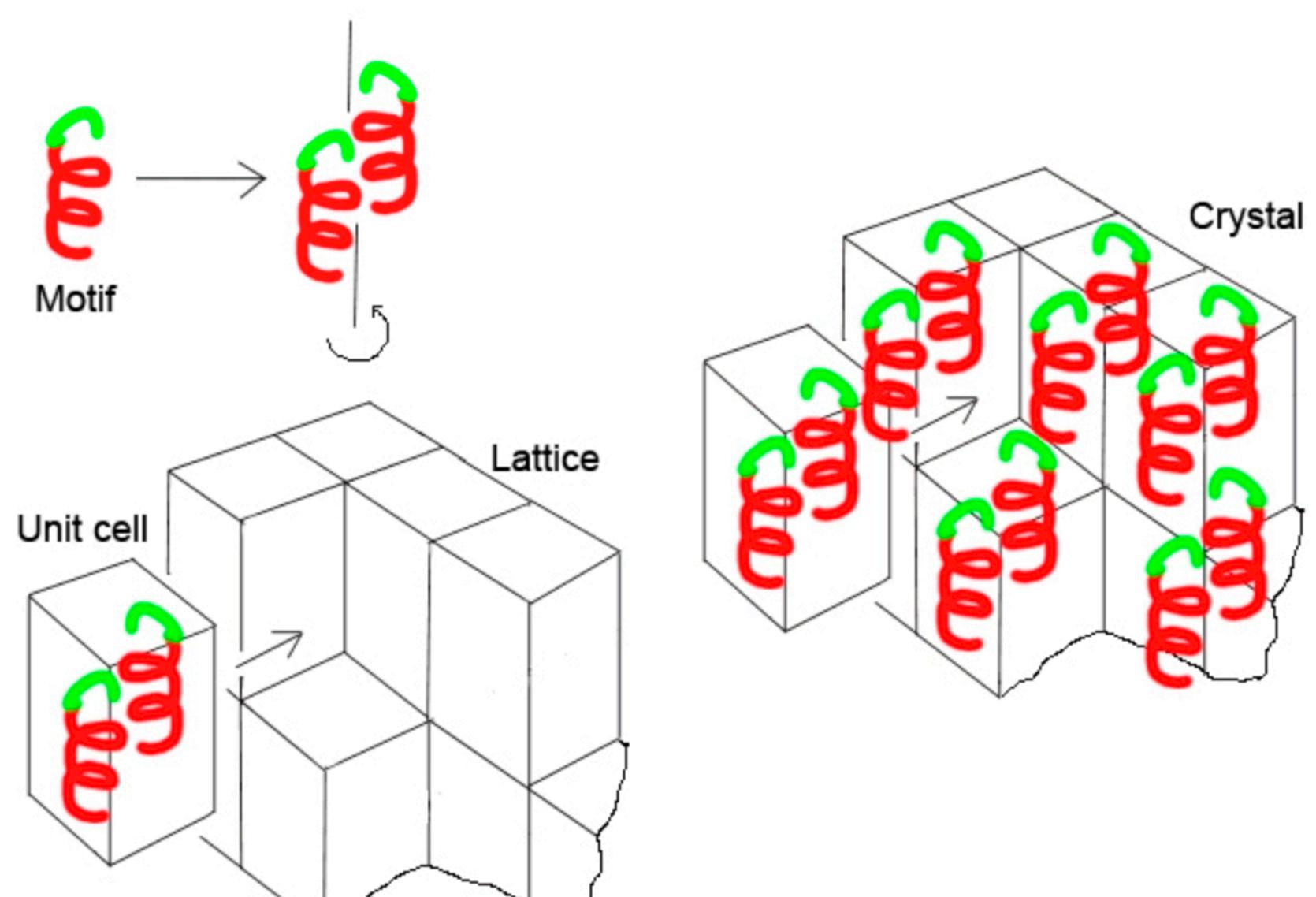
X-ray



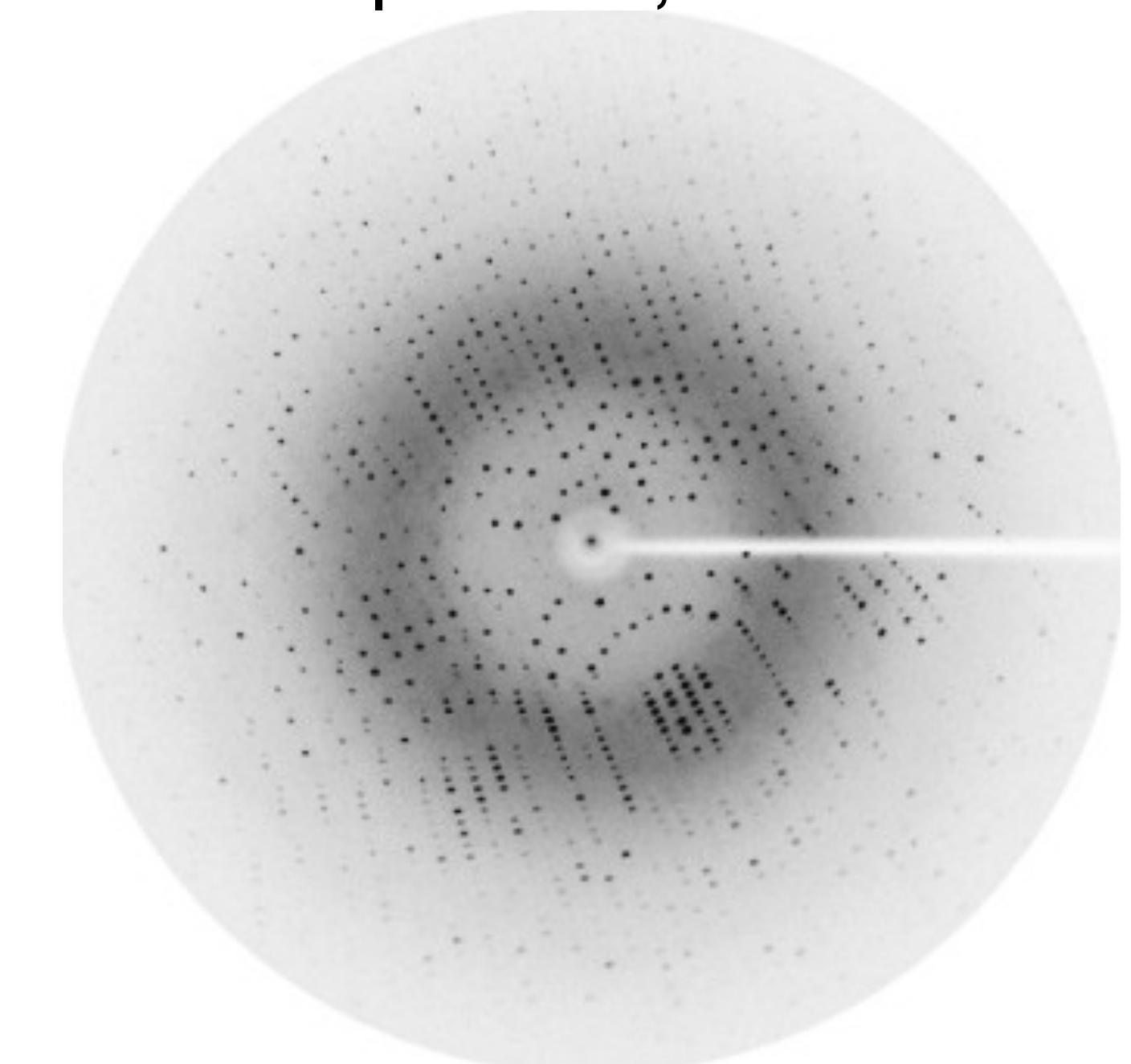
Unit cell (left) whose three-dimensional stacking builds a crystal (right)

The first working hypothesis is that the crystal is ideal, that is the unit cell is always the same, no ‘relevant’ impurities, etc

The distance from the center is proportional to the resolution



Motifs (molecules, etc.) do repeat themselves by symmetry operators inside the unit cell.
Unit cells are stacked in three dimensions, following the rules of the lattice, building the crystal.

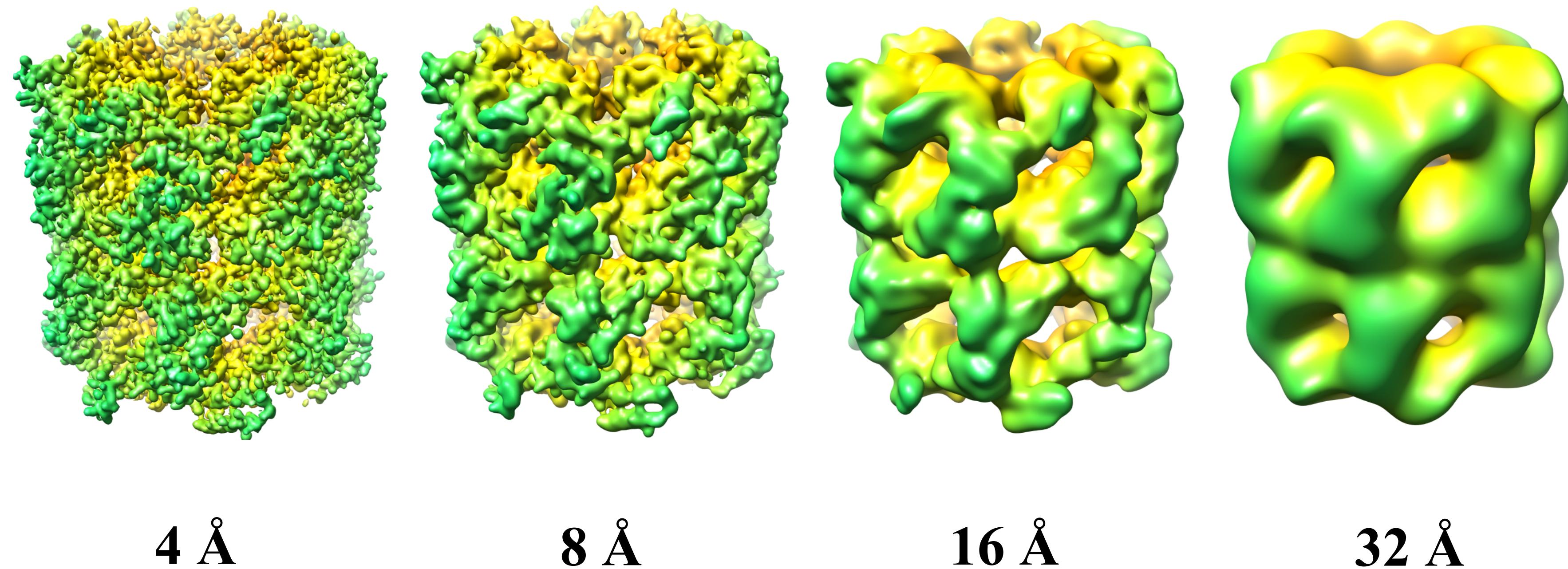


The crystal is rotated and multiple images are collected to find out the symmetry of the crystal, that is to each black spot of each image you can assign 4 numbers: the intensity and three coordinates h , k , l

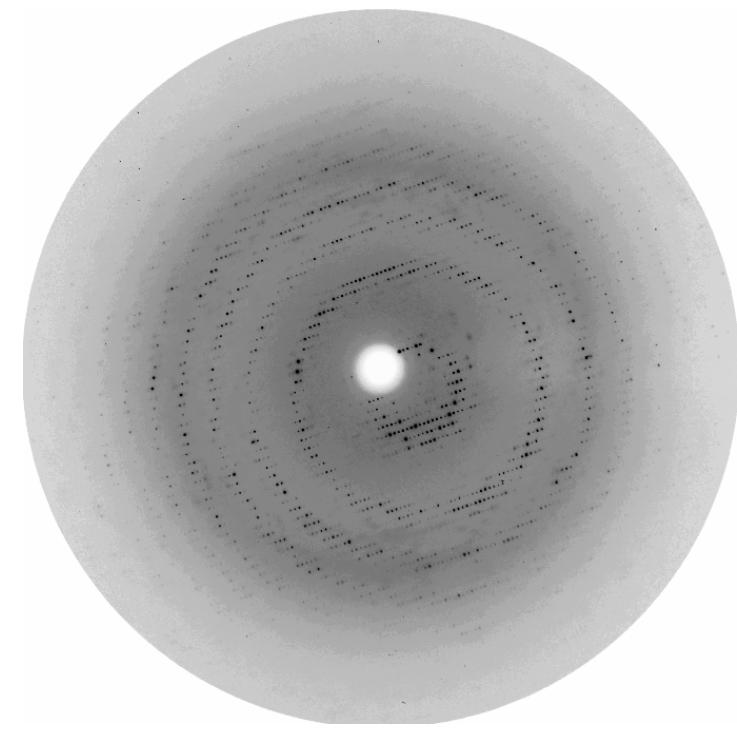
X-ray diffraction (crystallography)

- What does the diffraction experiment provide?

➤ Resolution:



Structure determination



$$I_{hkl} \propto I_0 \frac{V_{xtal}}{V_{Cell}} |F_{hkl}|^2$$
$$\rho(x,y,z) = \frac{1}{V_{hkl}} \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp[-2\pi i(hx+ky+lz)]$$

not measured

We need to know the phase!

Each spot is proportional to the intensity of the source (I_0), the volume of the crystal divided for the volume of the cell and the square of the structure factor

Heavy atoms method:

Make a crystal soaked with heavy atoms that scatter a lot. If the crystal is the same as the former (same symmetry) then it is possible to identify the exact position of the heavy atoms and get an approximate phase. Iteratively it is possible to then start adding atoms and continue to improve the reconstruction.

Molecular replacement:

$$|F_{hkl}| \exp(i\alpha_{hkl}) = \sum_{j=1}^{N \text{ atoms}} f_j \exp[2\pi i (hx_j + ky_j + lz_j)]$$

Atomic form factor

We need to know the structure!

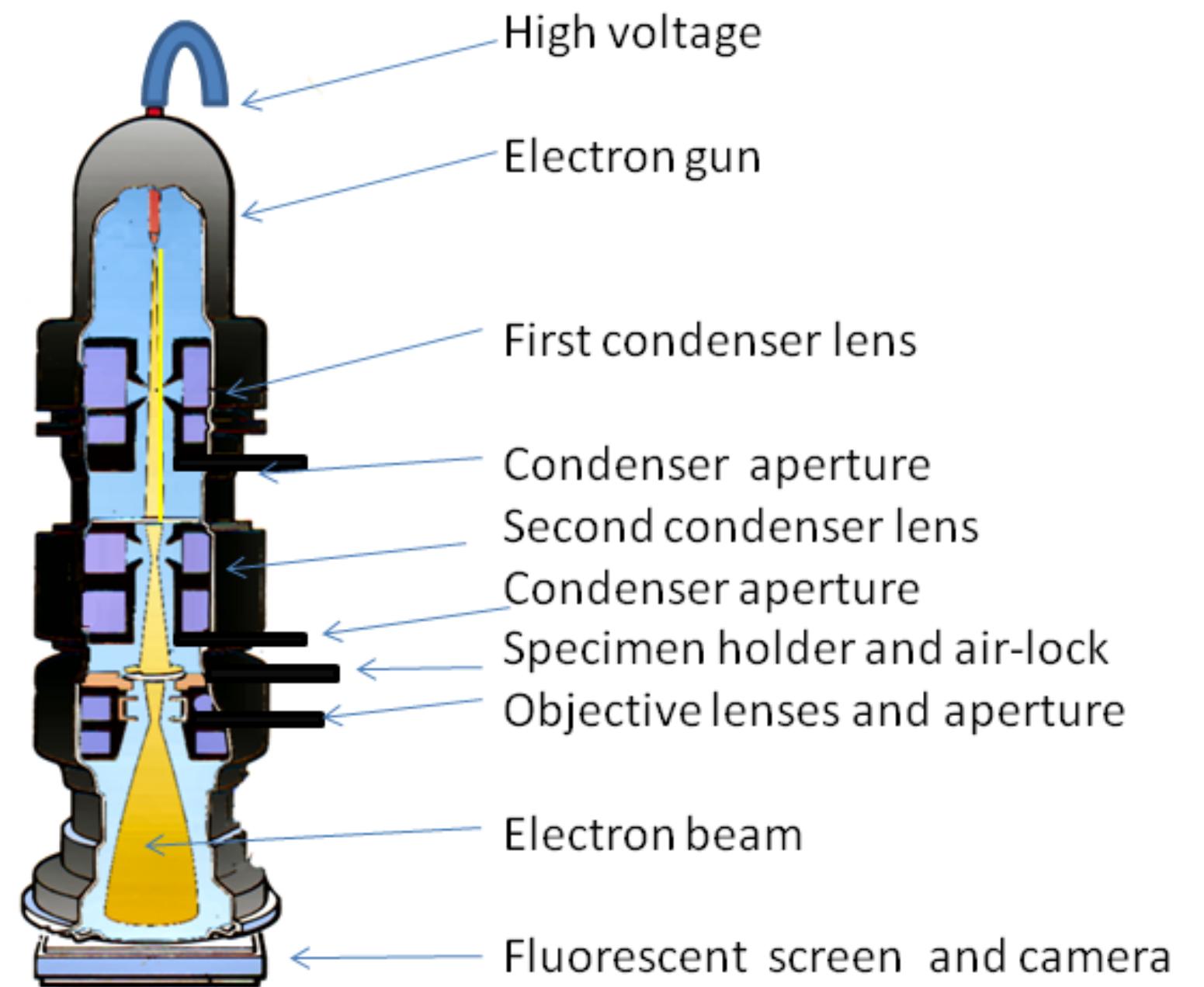
One begins with a model, generates the density and hopefully iteratively converge



Electron microscopy

- **Electrons:**

- Wavelength ($< 0.04 \text{ \AA}$) - on the atomic separation scale
- Negatively charged → can be **focused** by electromagnetic lenses → can be viewed **directly (no need for 3D crystals)**



Transmission Electron Microscope

Introduction to Proteins, Kessel and Ben-Tal, 2018

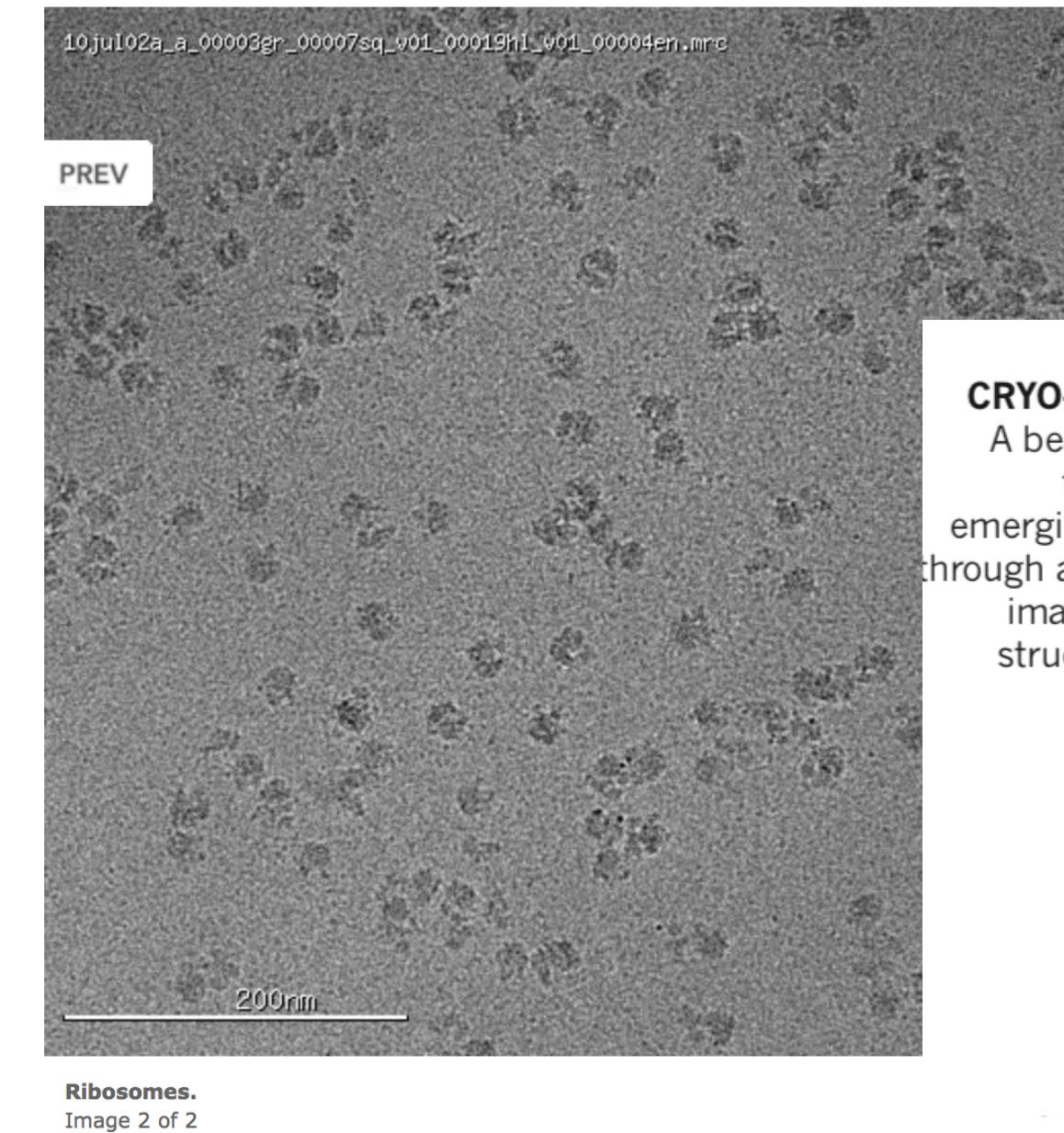


UNIVER
DEGLI S
DI MILA

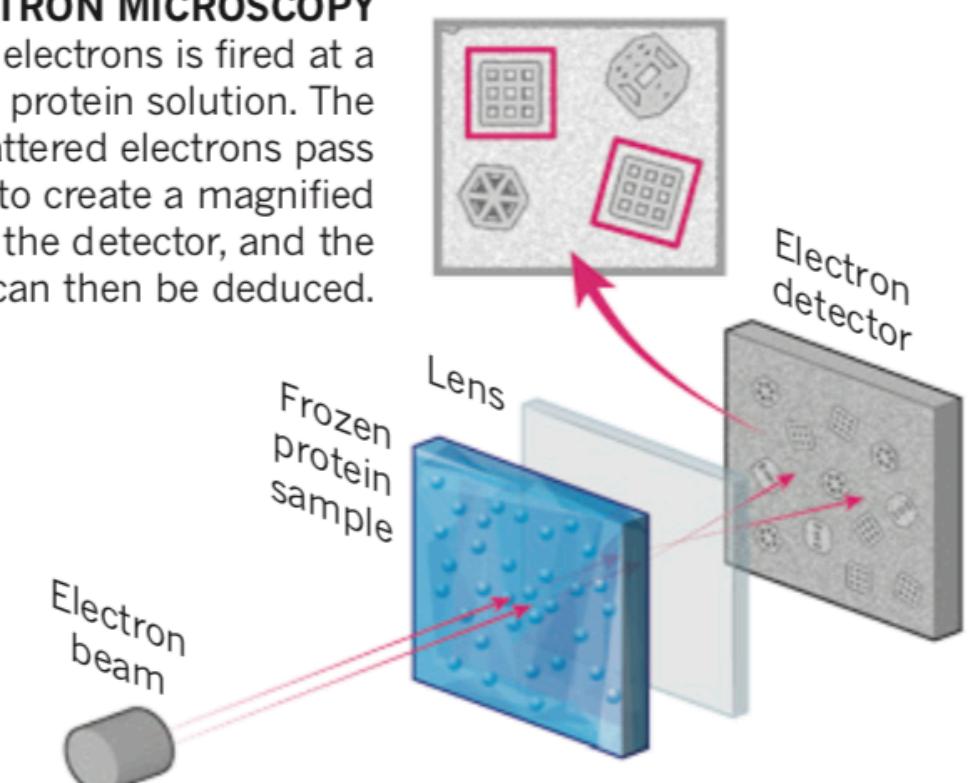
Cryo-EM

Solution is vitrified by liquid nitrogen on a plate, essentially all the molecules are on a plane, and electron (very few) are sent through the sample to avoid molecular motions so we get clear 2D shadows

1. Shadows with the same shape are clustered and averaged to increase the signal
2. Averaged 2D images are used to generate a 3D envelope and this is reprojected in 2D and iterate to improve the 2D classification and so the 3D reconstruction
3. The result is directly a density map, atomic models can be built inside the density



CRYO-ELECTRON MICROSCOPY
A beam of electrons is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, and the structure can then be deduced.



X-RAY IMAGE: SPL

Few electrons on the other hand means low signal-to-noise ratio

See also:

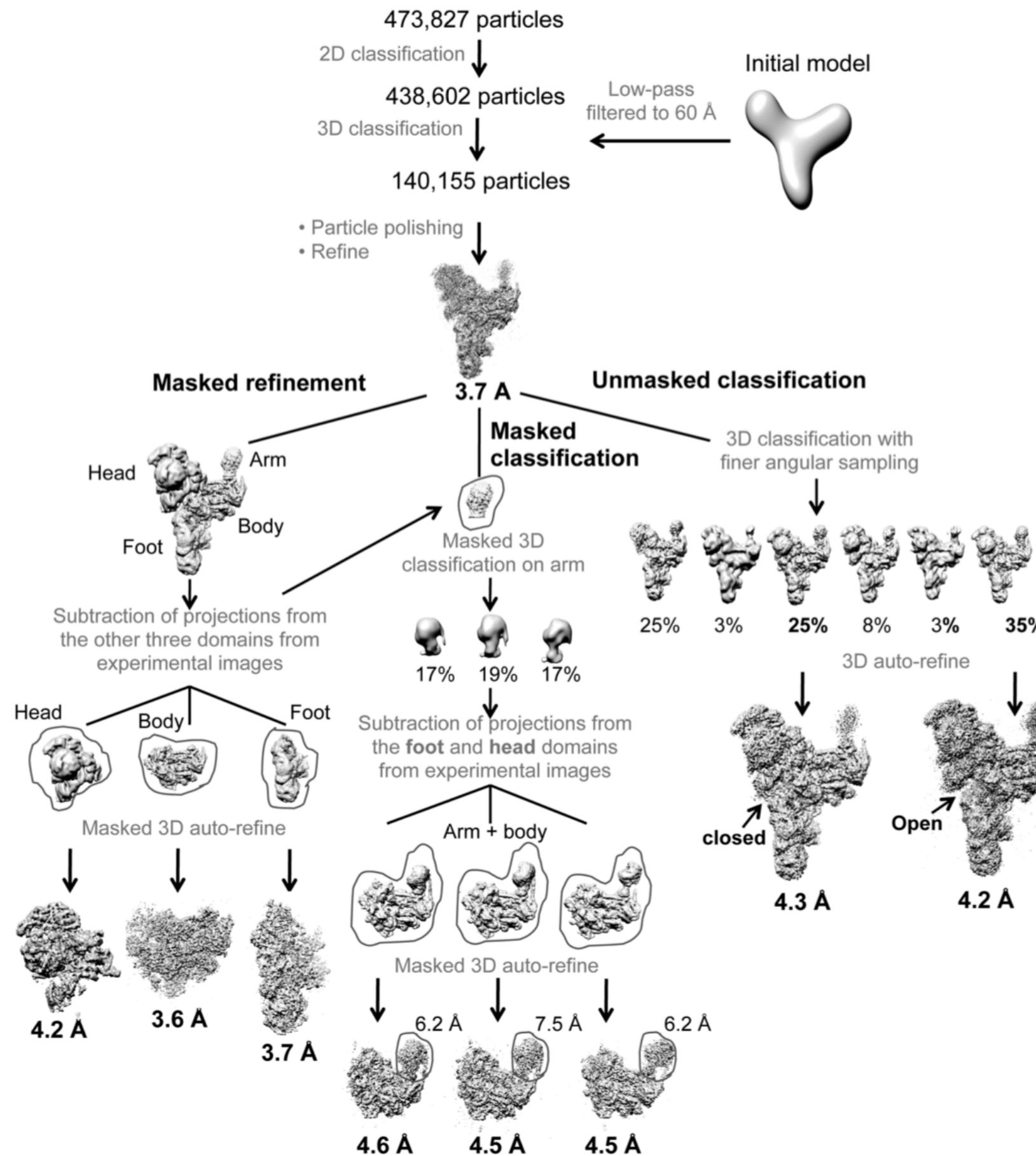
<https://www.youtube.com/watch?v=BJKkC0W-6Qk>



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Cryo-EM



Cryo-EM structure determination is facilitated if one can identify subdomains, and if for those structures are already available.



NMR spectroscopy

- **Principles:**

- A strong **external electric field** applied by the NMR machine to the sample results in **spatial alignment** of protein atoms with nuclear spin (e.g. ^1H , ^{13}C , ^{15}N)
- An **energy pulse** transiently disturbs this alignment
- The nuclei revert to their original state while **emitting radio waves** that are sensed by the NMR machine
- The waves are different for each atoms, and affected by their atom neighbors (**chemical shift**)
- By using the known chemical shifts of nuclei at different chemical environments, **the structure of the protein can be deciphered** from its NMR spectrum



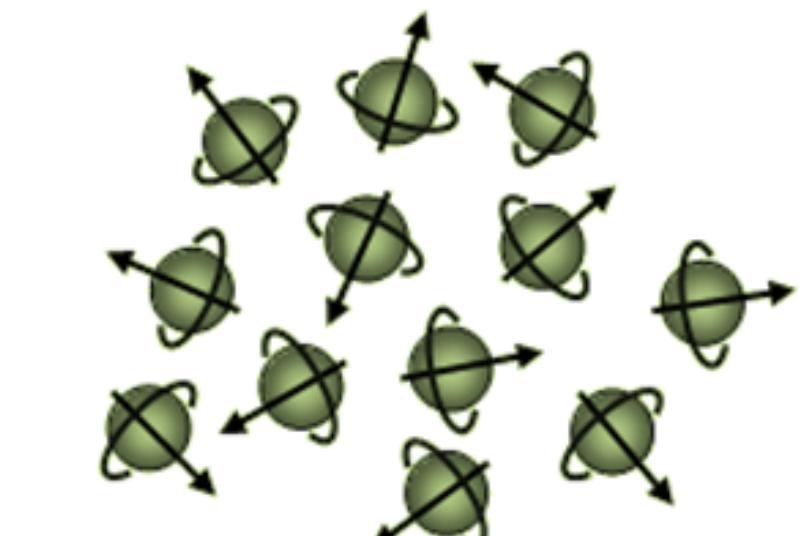
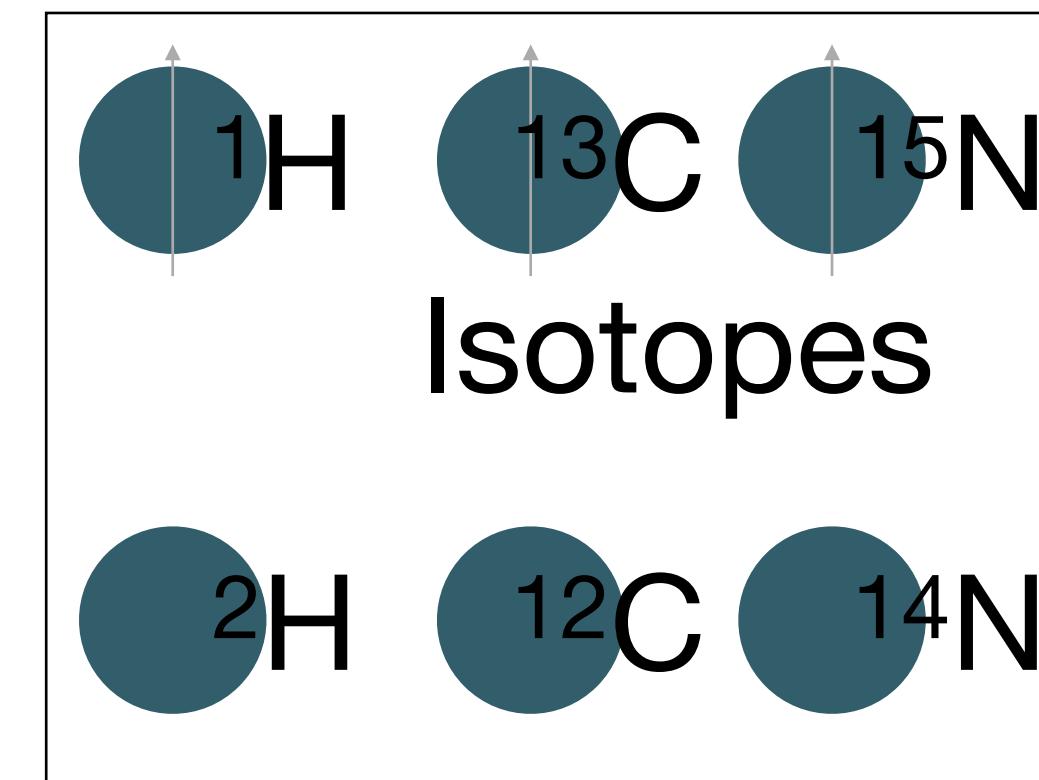
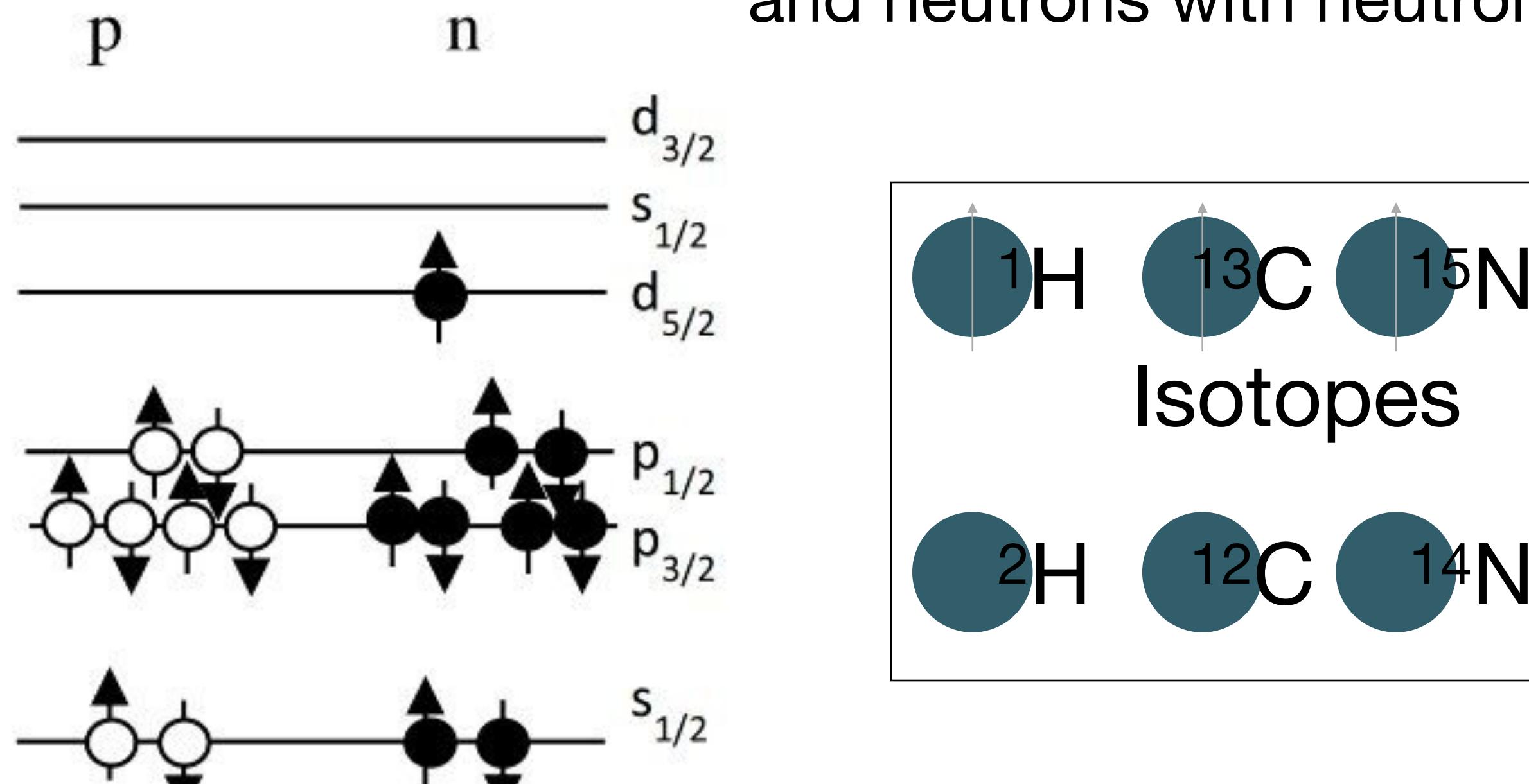
NMR

Nuclei have an intrinsic magnetic moment (spin) that is the result of the sum of the intrinsic magnetic moments (spin) of their protons and neutrons

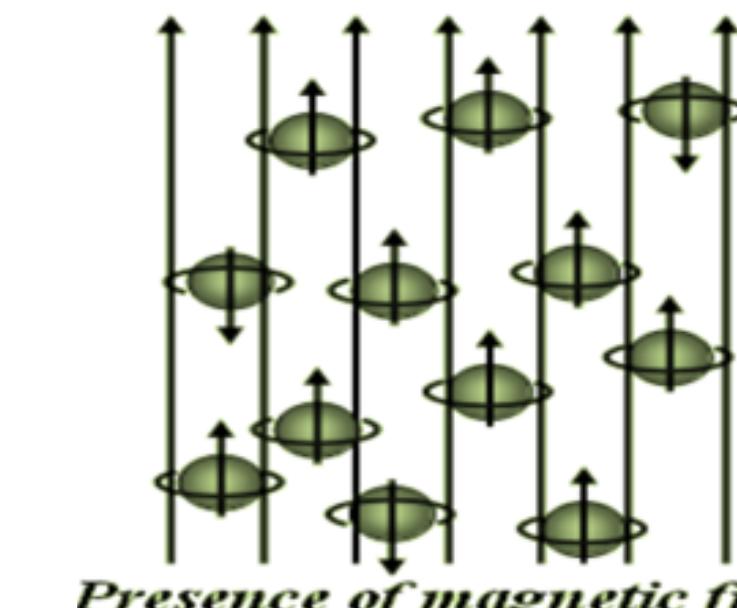
Nuclei with an odd number of protons+neutrons have a fractional spin $1/2, 3/2, \dots$

Fractional spins align to a static magnetic field and generates an absorption spectrum that then we can excite by a variable magnetic field.

Protons and neutrons have spin $1/2$ but they are organised in the nucleus to cancel it (protons with protons and neutrons with neutrons)



Absence of magnetic field



Presence of magnetic field



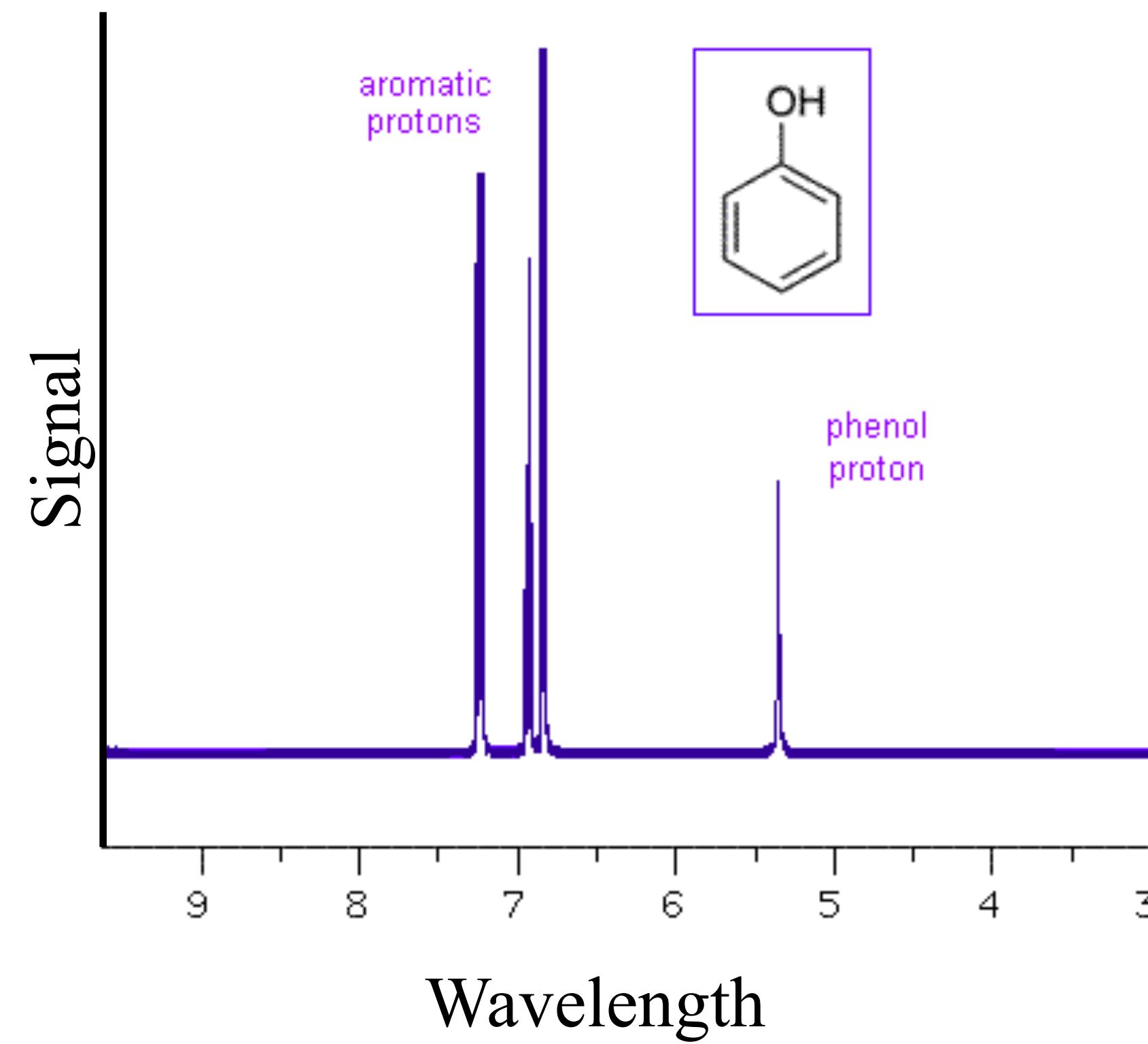
UNIVERSITÀ
DEGLI STUDI
DI MILANO

So ^{17}O has spin $1/2$ while ^{16}O has spin 0

NMR spectroscopy

- **Principles:**

- 1D NMR spectrum of protons from different locations in a phenol molecule:



NMR spectroscopy

- **Principles:**

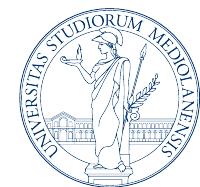
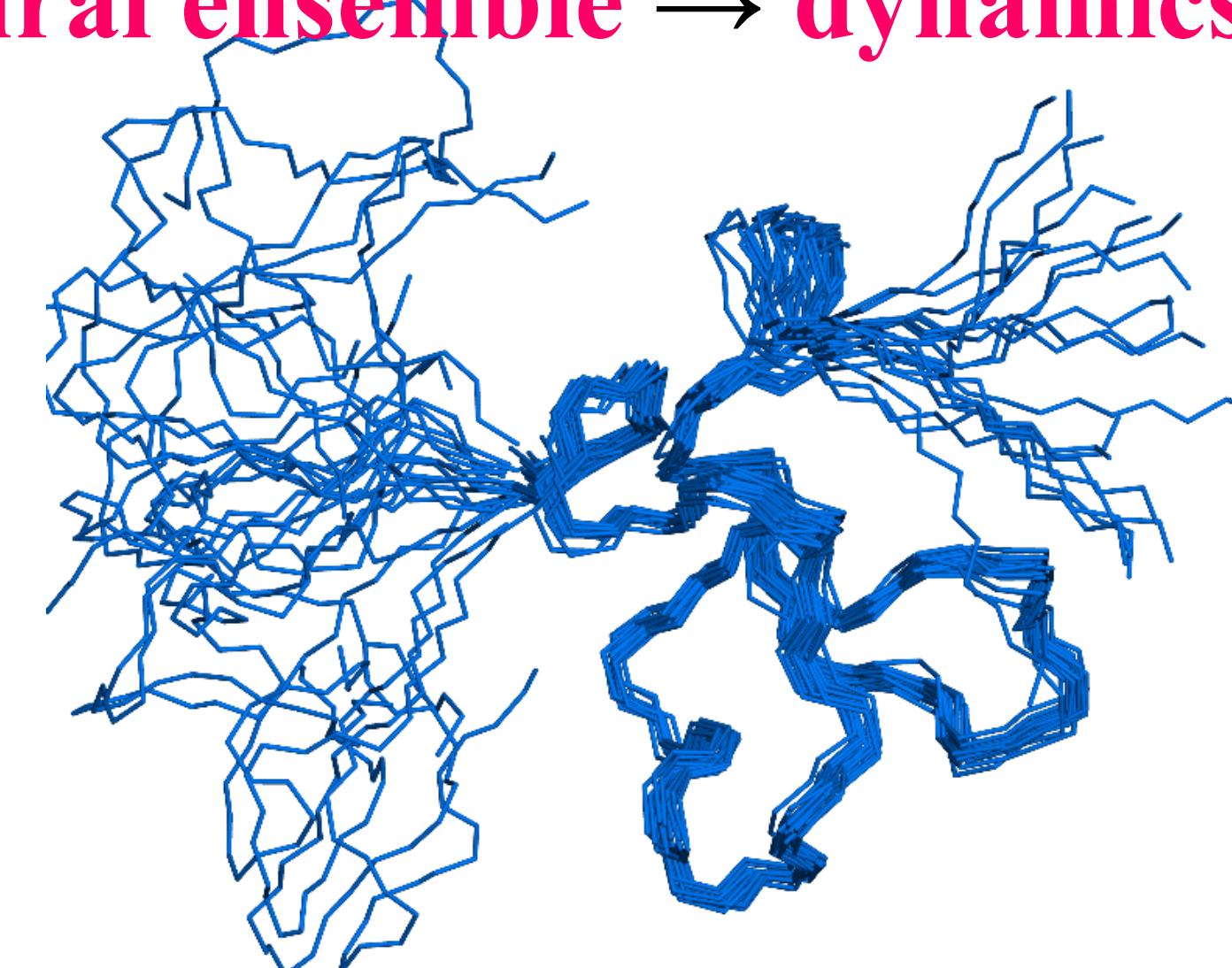
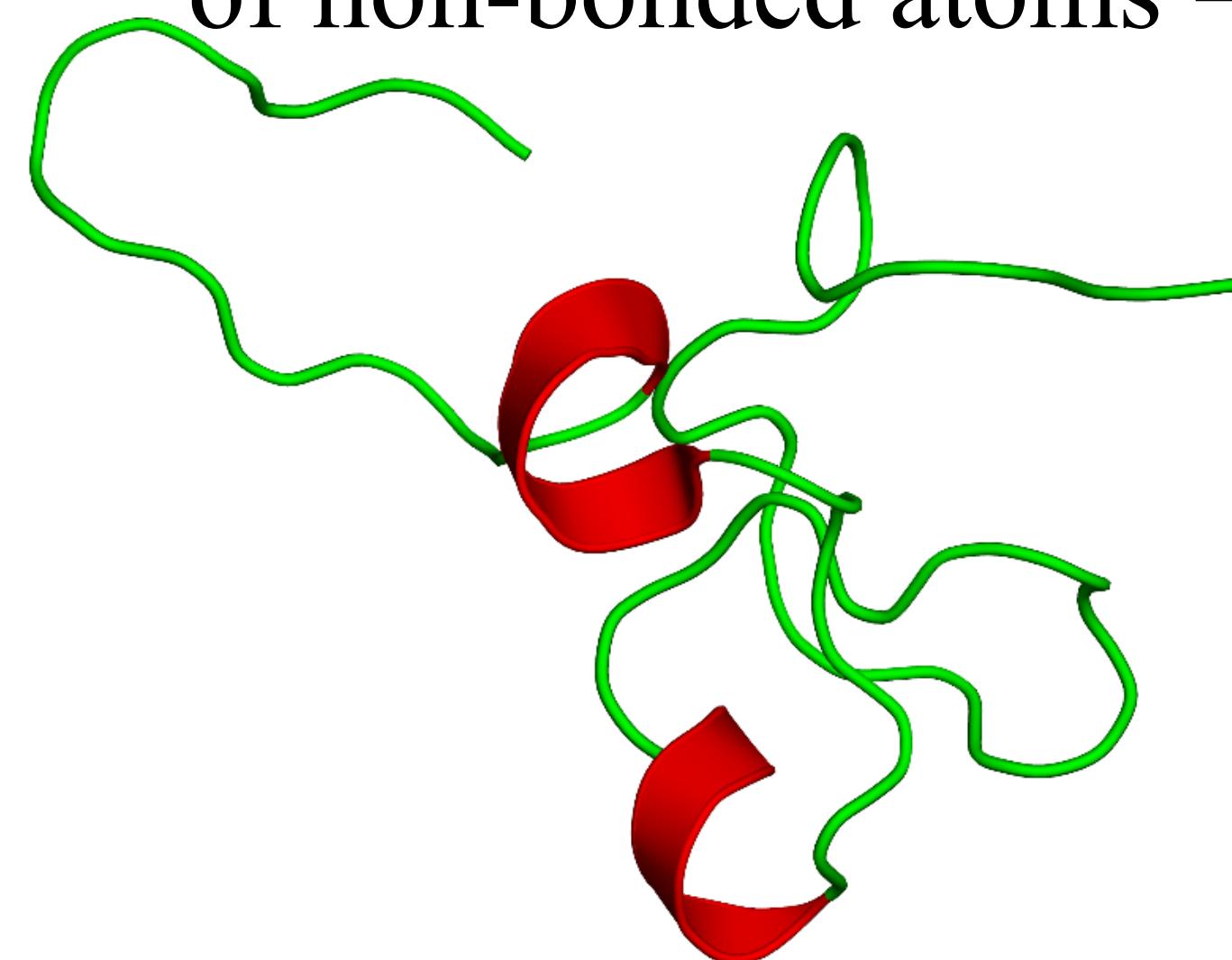
- In macromolecules, deciphering the structure from the signal requires analyzing the **couplings** between different atoms:
 - **J-coupling:** exists between covalently bonded nuclei; its measurement (**COSY**) reveals atom **connectivity** in the molecule



NMR spectroscopy

- **Principles:**

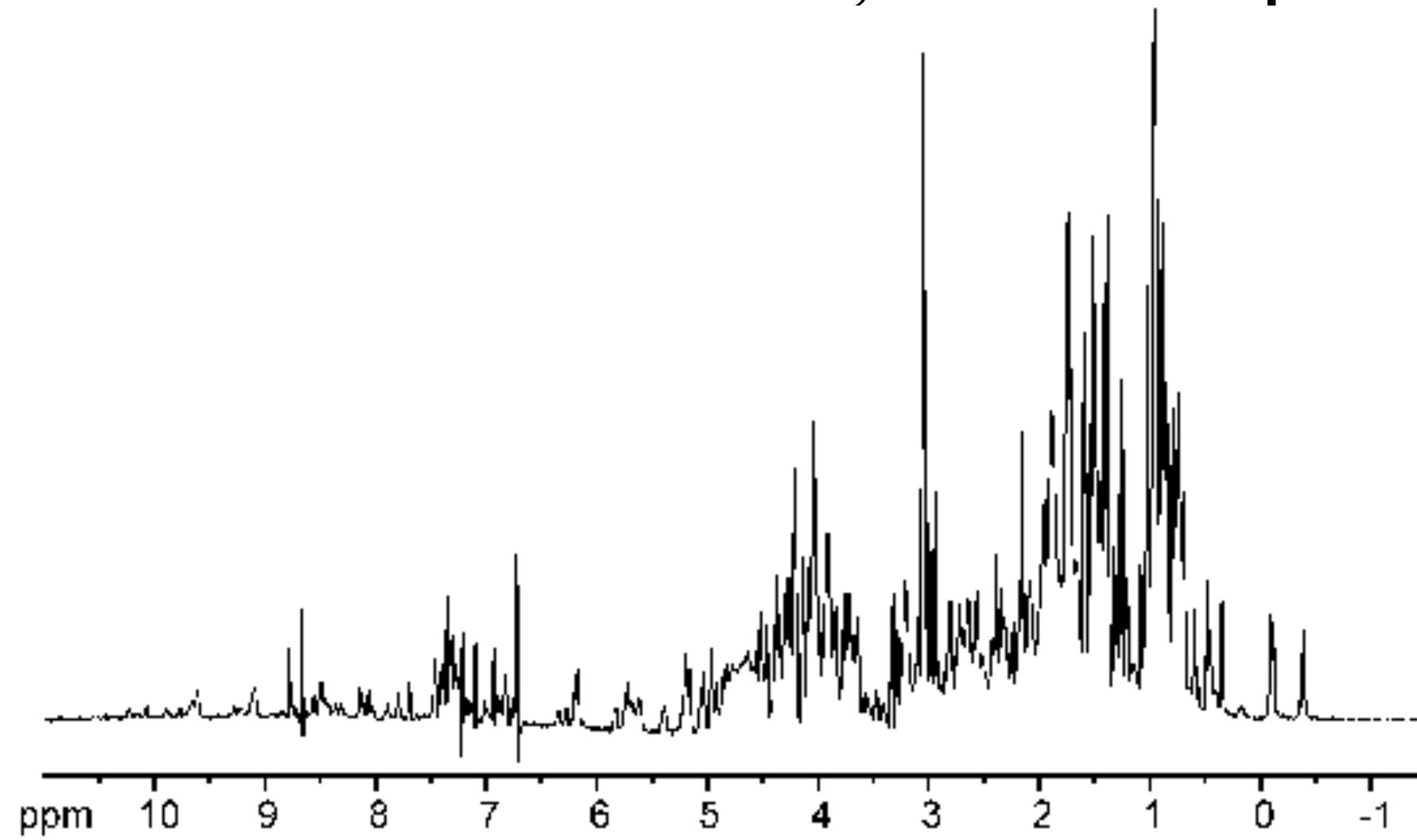
- In macromolecules, deciphering the structure from the signal requires analyzing the **couplings** between different atoms:
 - **NOE coupling:** occurs between proximal ($< 5 \text{ \AA}$) non-bonded nuclei; its measurement (**NOESY**) reveal all possible locations of non-bonded atoms → **structural ensemble** → **dynamics**



NMR spectroscopy

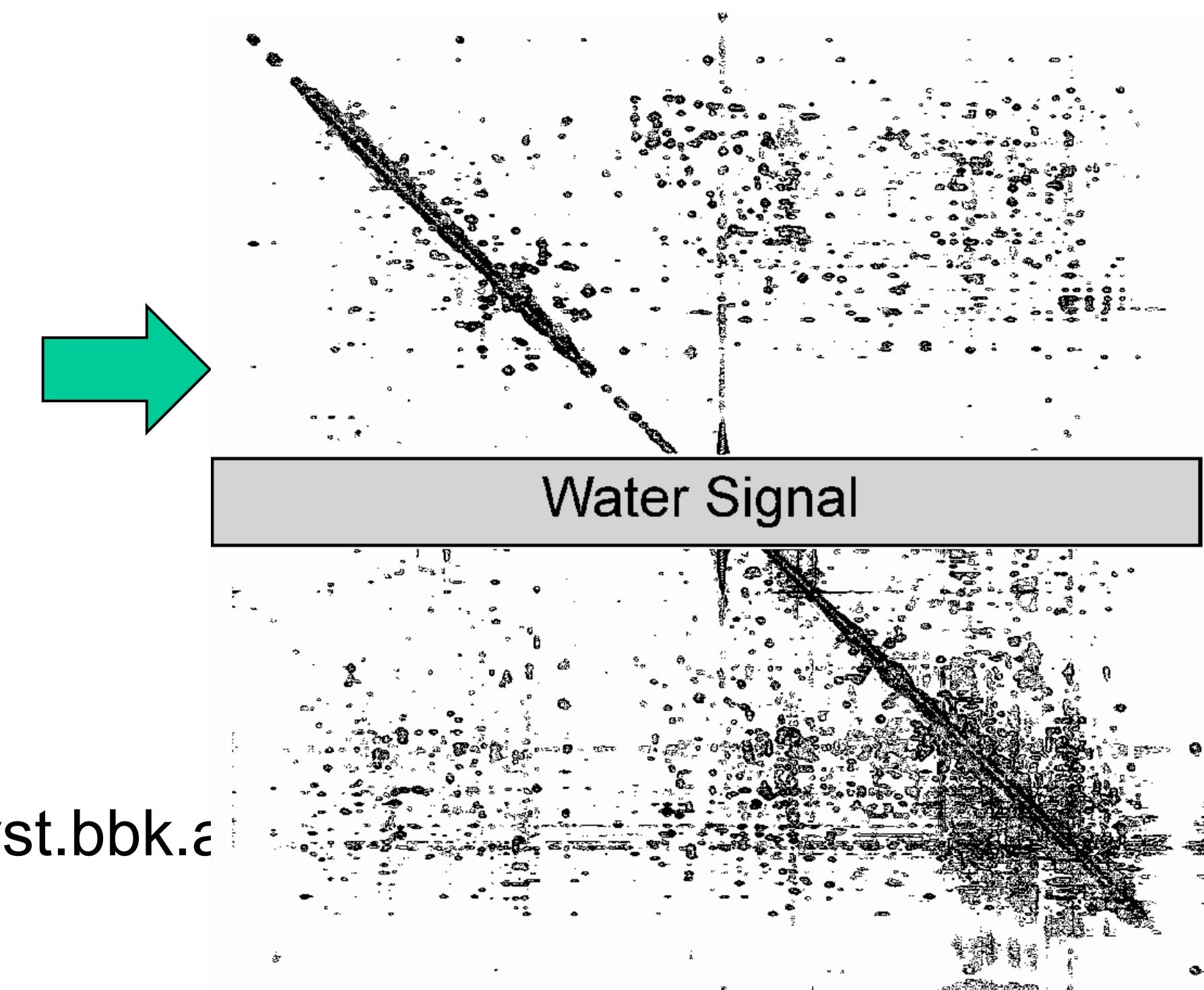
- **Principles:**

- The multiple peaks can also be resolved by conducting 2 or more measurements (2D/3D/4D NMR)
- Further resolution can be achieved by labeling the sample with ^{13}C or ^{15}N , and/or replace ^1H with ^2H



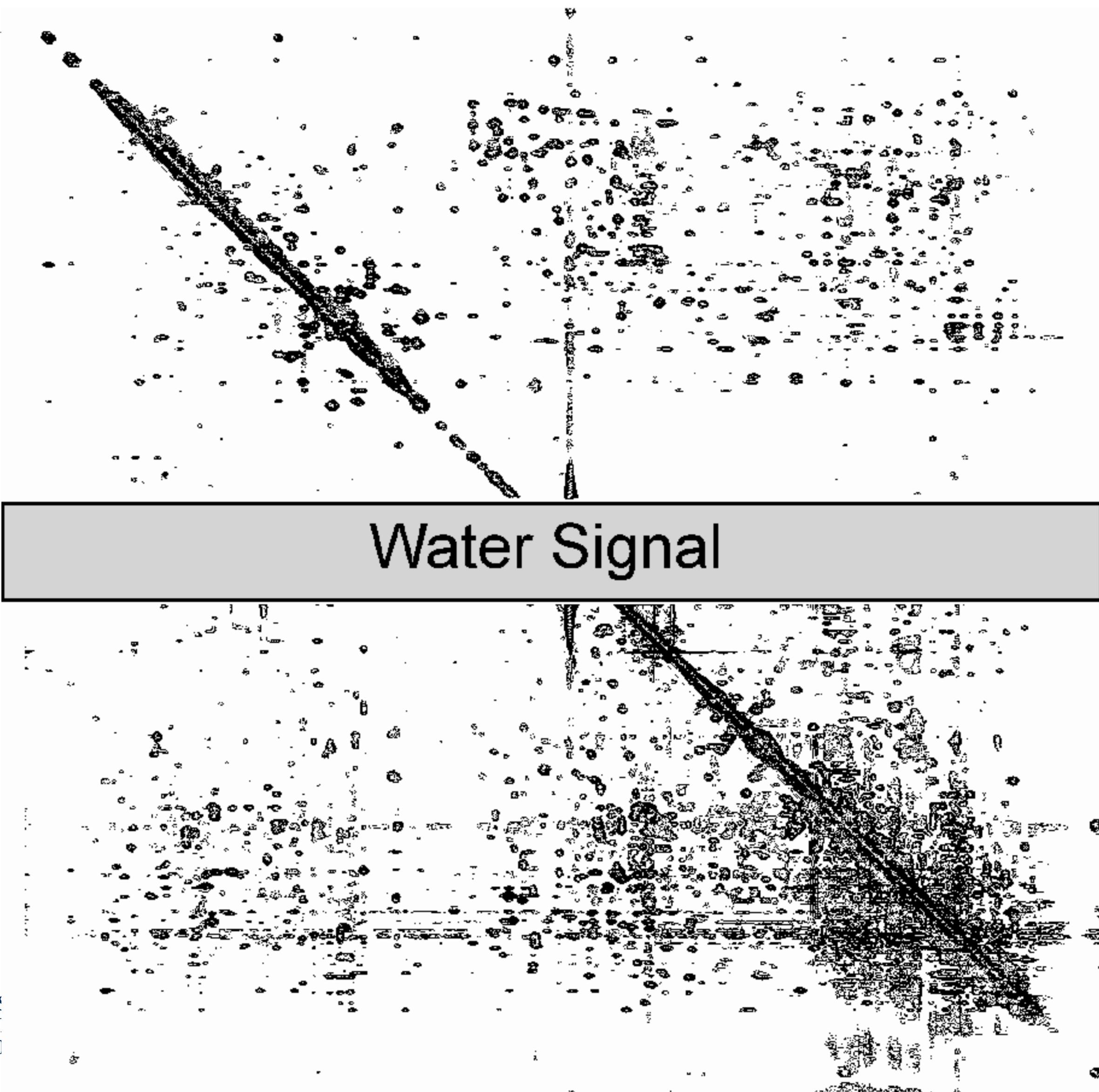
1D NMR

(www.cryst.bbk.ac.uk)



UNIVE
DEGLI
DI MILI

NMR



The intensity of each single out-of-diagonal peak is proportional to $1/d^6$ of a couple of atoms

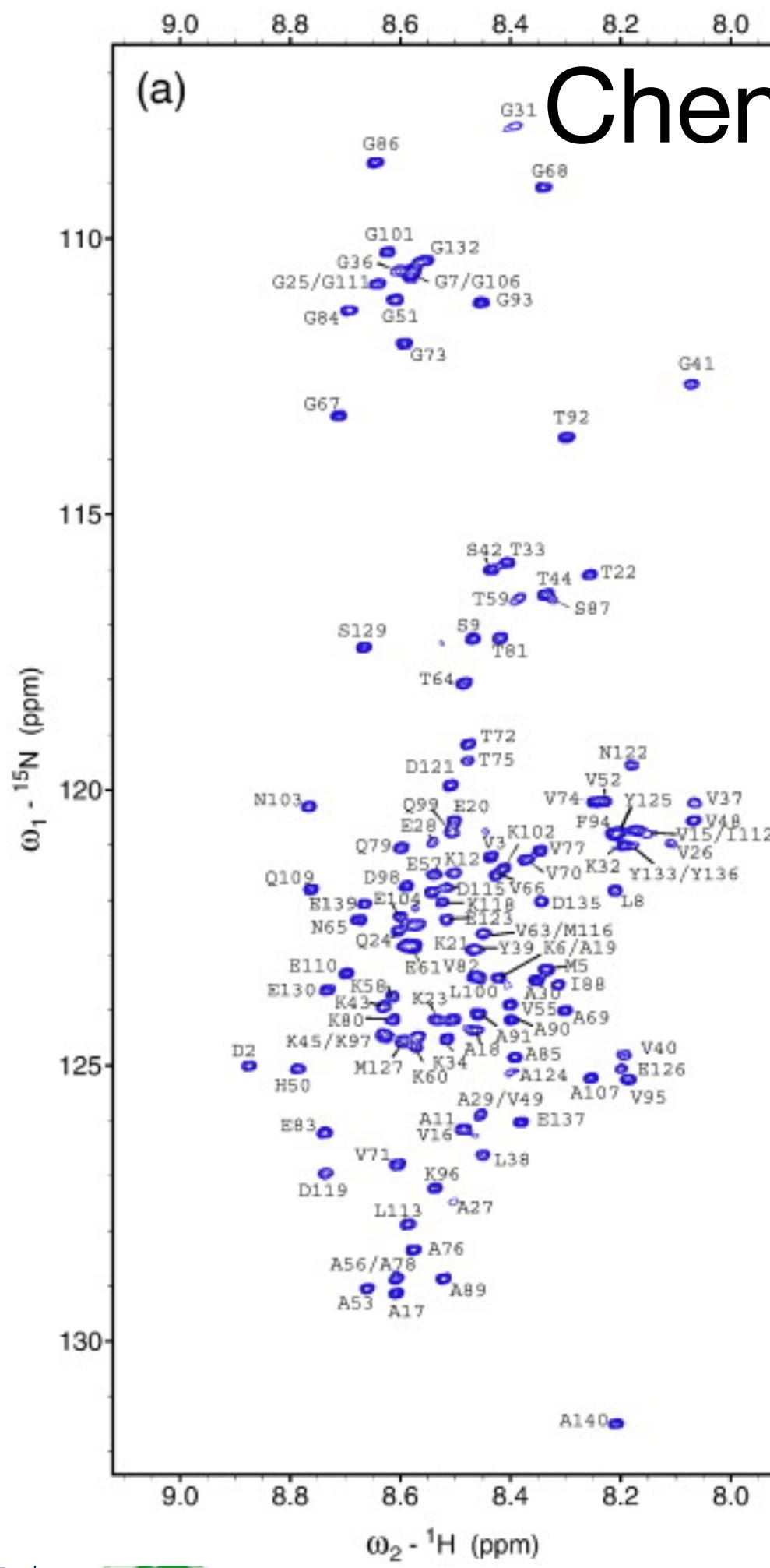
A reference distance is obtained for a rigid couple and all other distances are then obtained as:

$$d_{ij} = d_{ref} \left(\frac{I_{ref}}{I_{ij}} \right)^{1/6}$$

These are then used only as upper limits for the distance

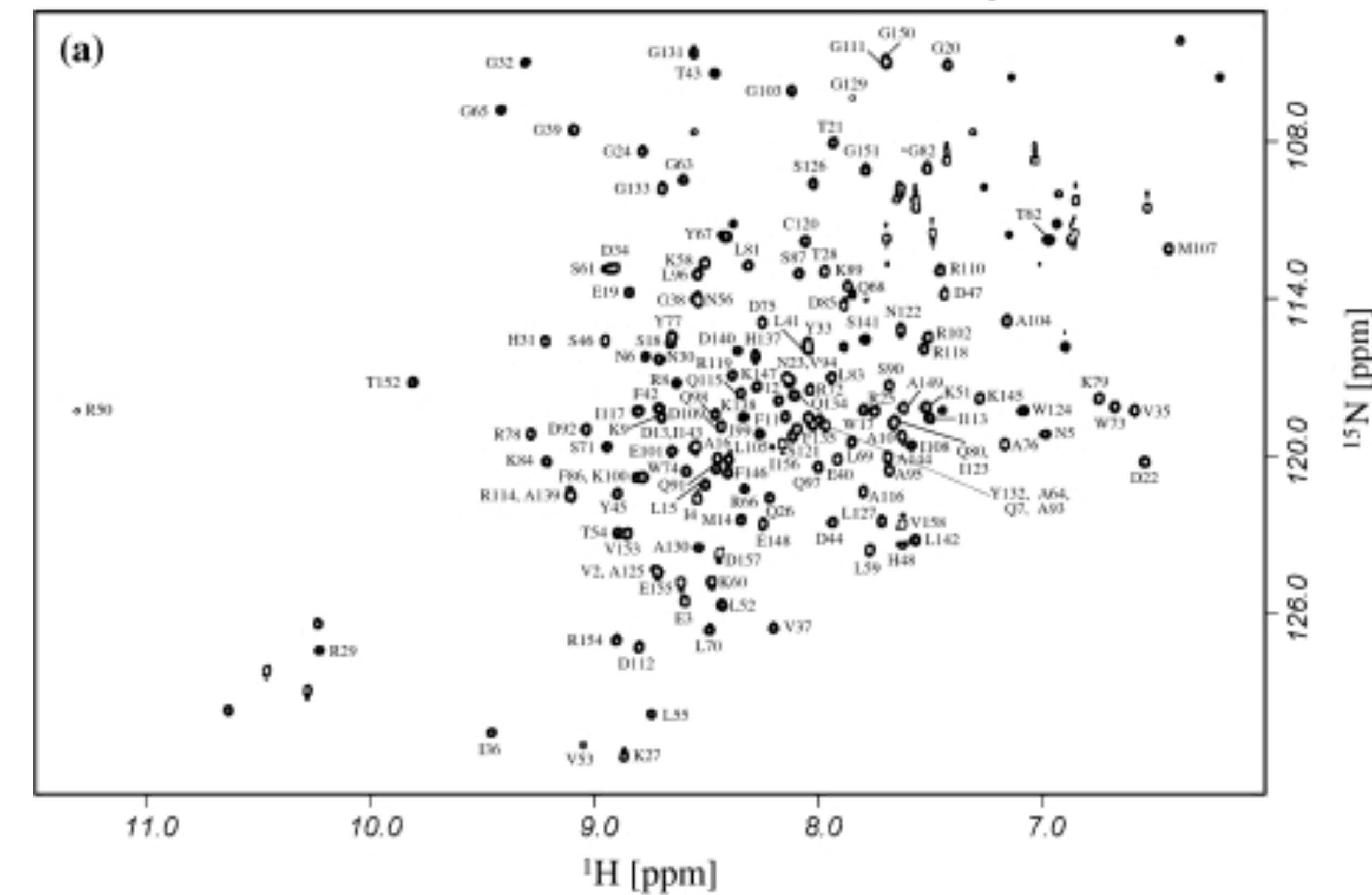


NMR

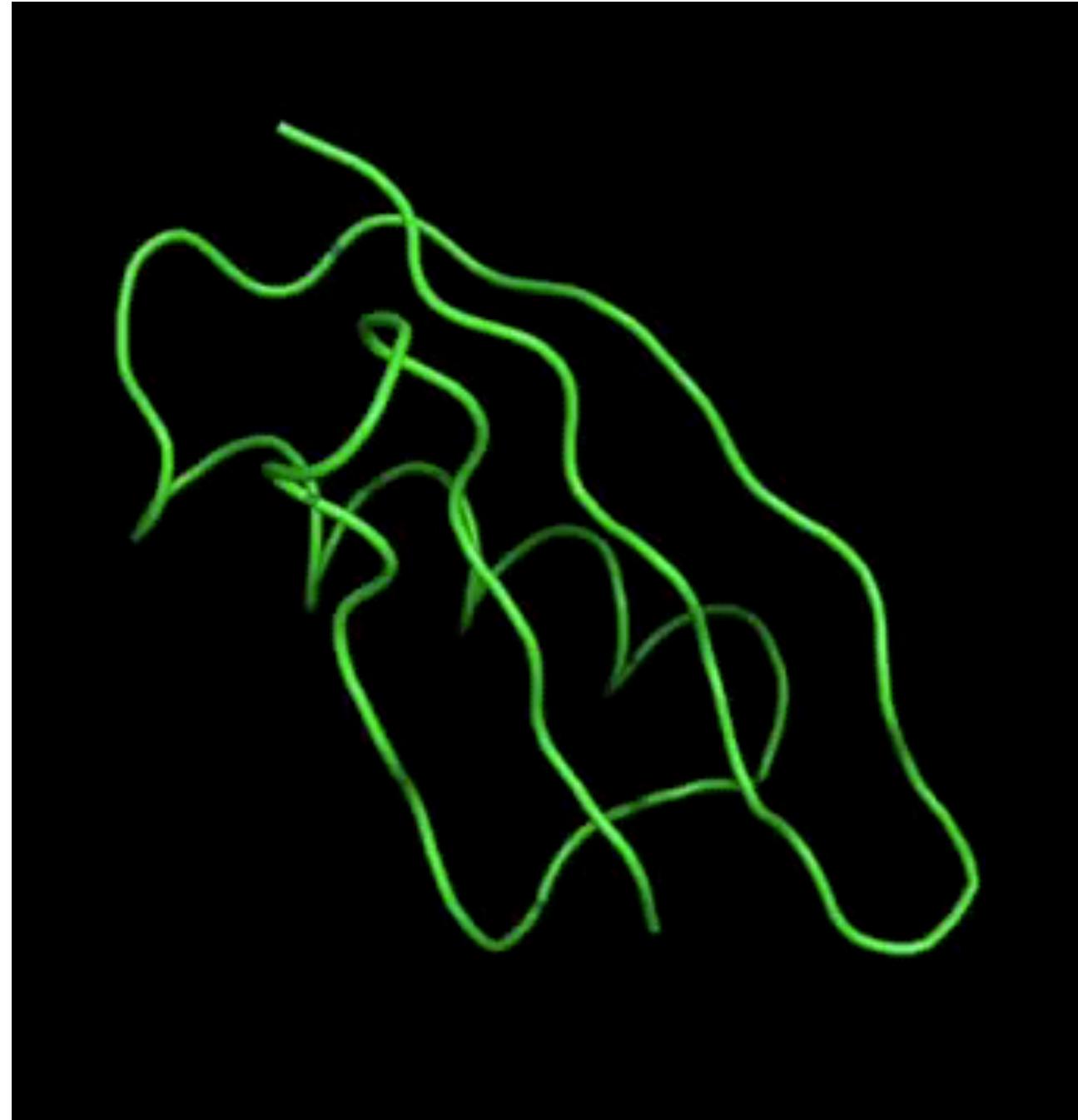


Chemical shifts for an intrinsically disordered protein

Chemical shifts for a folded protein



From the experiment to the Structure



By using distances (from NOE intensity), dihedral angles from J coupling it is possible to drive a sequence of amino acids towards an average structure of the protein in solution (if a solution exists).



NMR spectroscopy

- **Advantages:**

- The structure is determined in **native environment** (solution, membrane powder) → dynamics, no crystals required
- The protein's **hydration layer** can be studied
- NMR is sensitive to weak interactions → studying binding events
- The radio waves do not harm the sample

- **Shortcomings:**

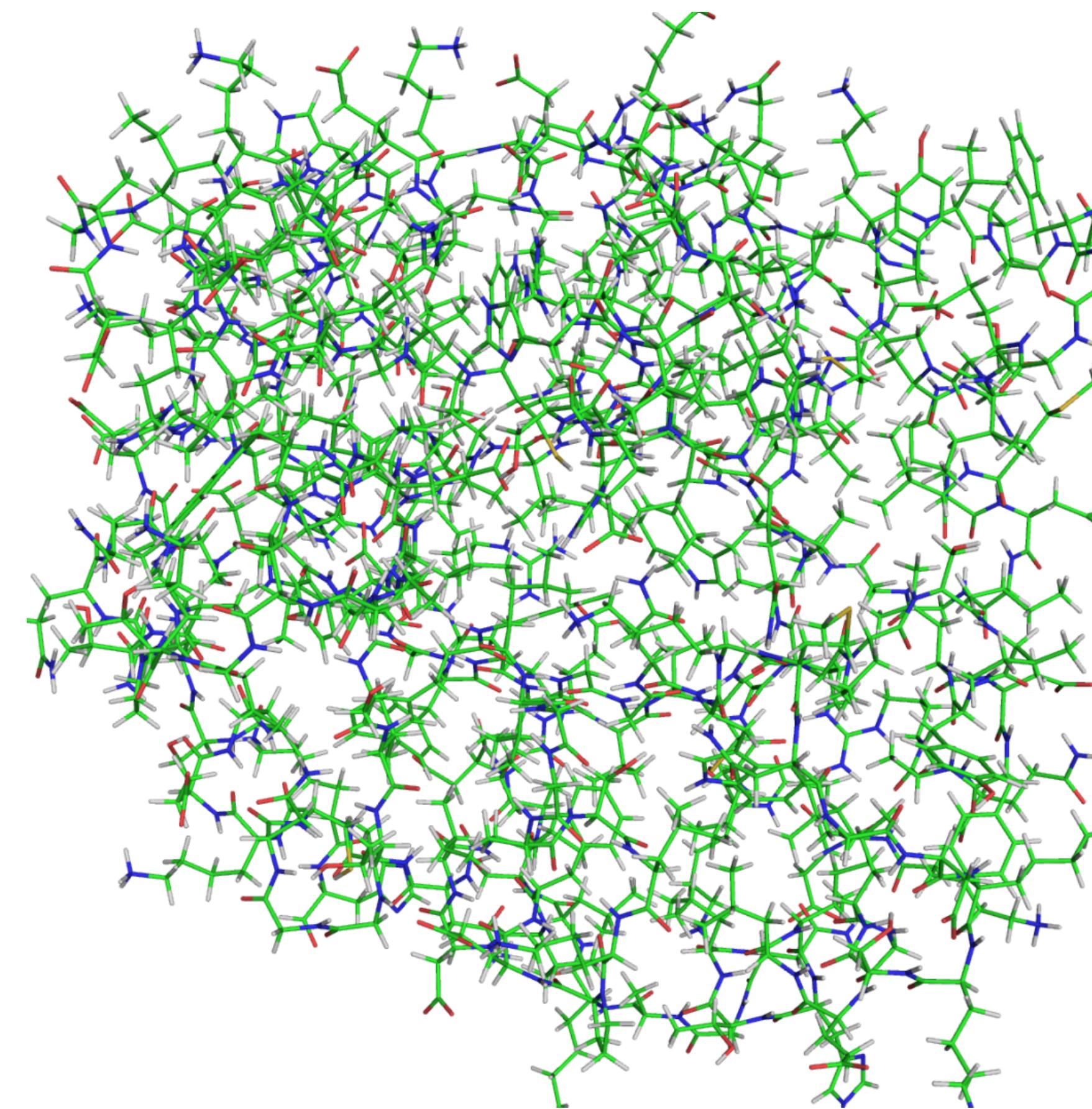
- Limited to small proteins ($\leq \sim 40$ kDa)



Visualisation

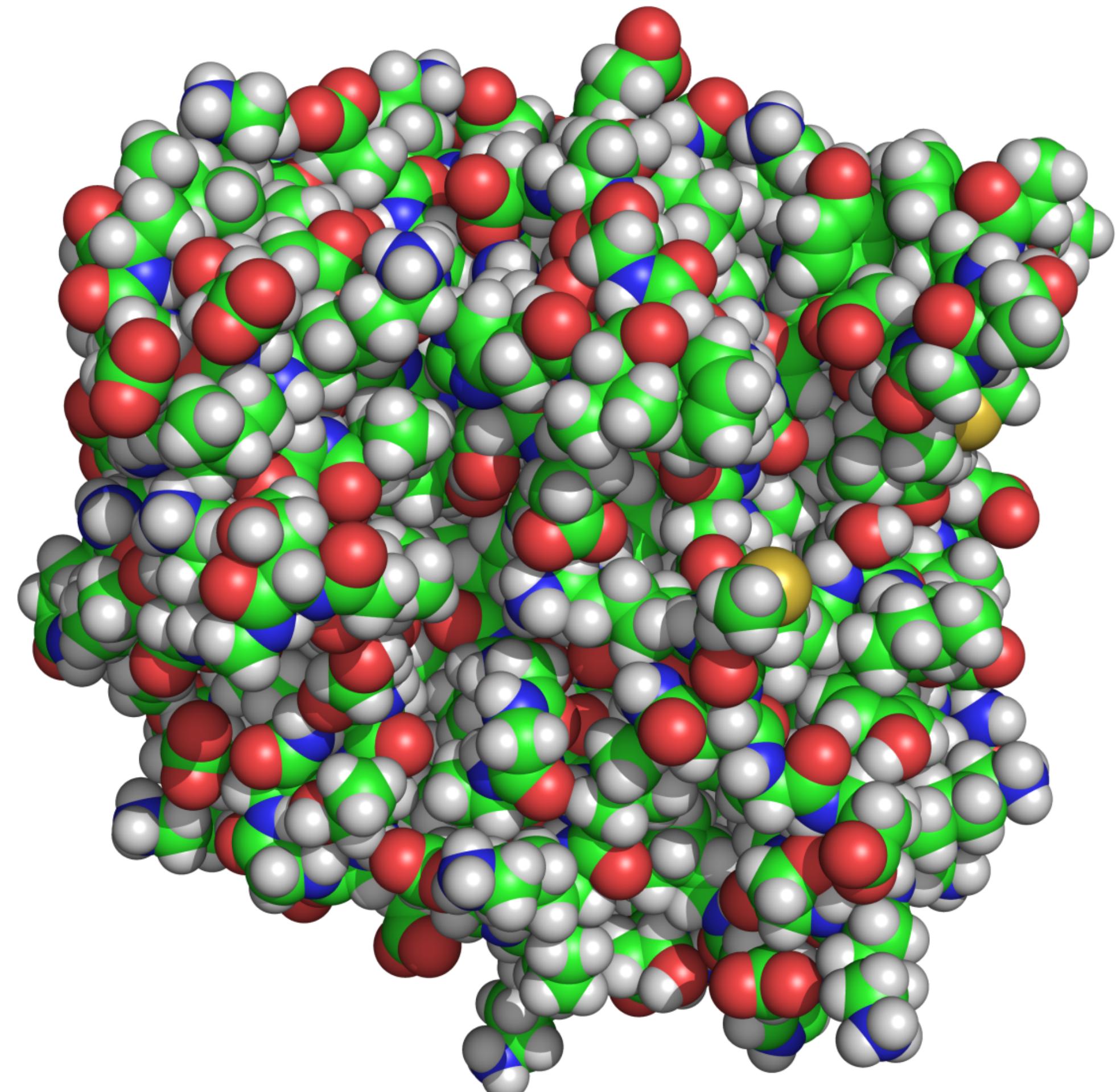
- **Wireframe (bonds connectivity)**

The covalent bonds in the protein are shown as wires, coloured by the type of atoms they connect. The atoms are not shown.



Different graphic representations of proteins reveal different properties

- Space-fill (general shape, size)



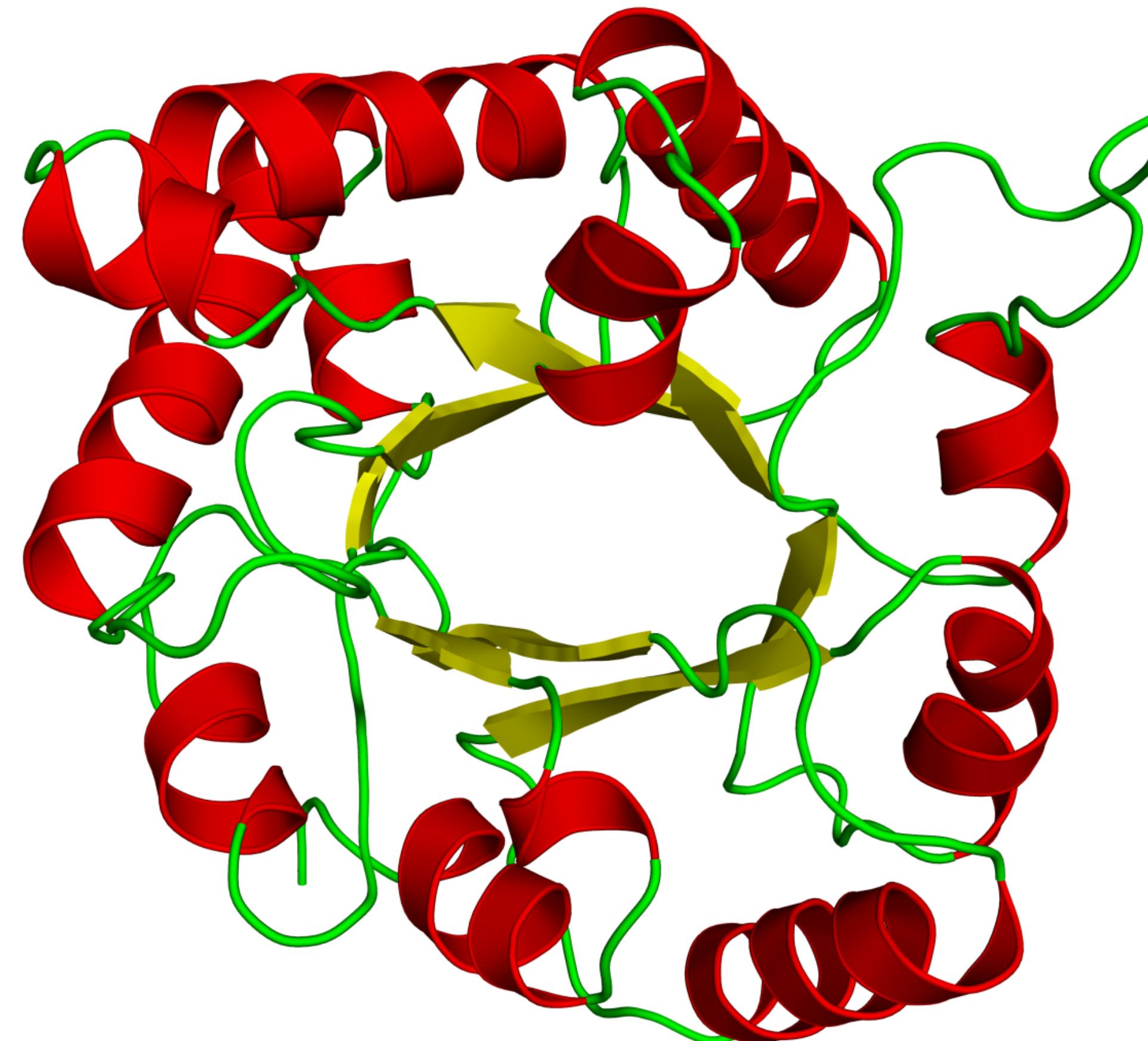
The atoms are represented as spheres with a van der Waals radius, and coloured according to the CPK convention



Different graphic representations of proteins reveal different properties

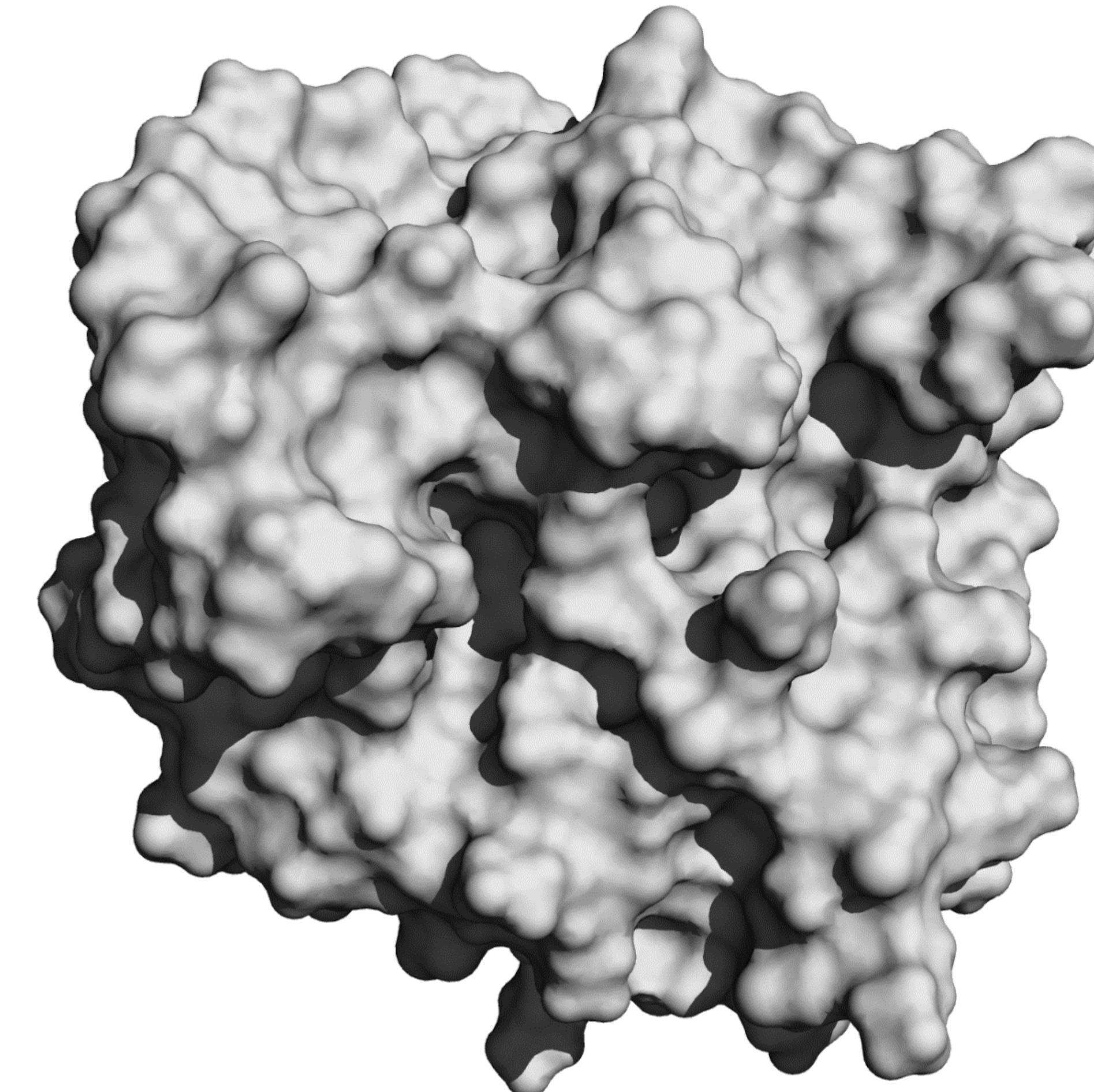
- **Ribbon (topology, secondary structures)**

The atoms and bonds of the protein are not shown. Instead, the backbone is represented as a ribbon with different colors, according to the secondary structure of the chain in that region: α -helices are colored in red, β -strands in yellow, and loops in green, as well as the disordered parts of the chain. The shape of the ribbon was calculated as the line going through all the C_α atoms of the protein.



Different graphic representations of proteins reveal different properties

- Surface (potential binding sites)

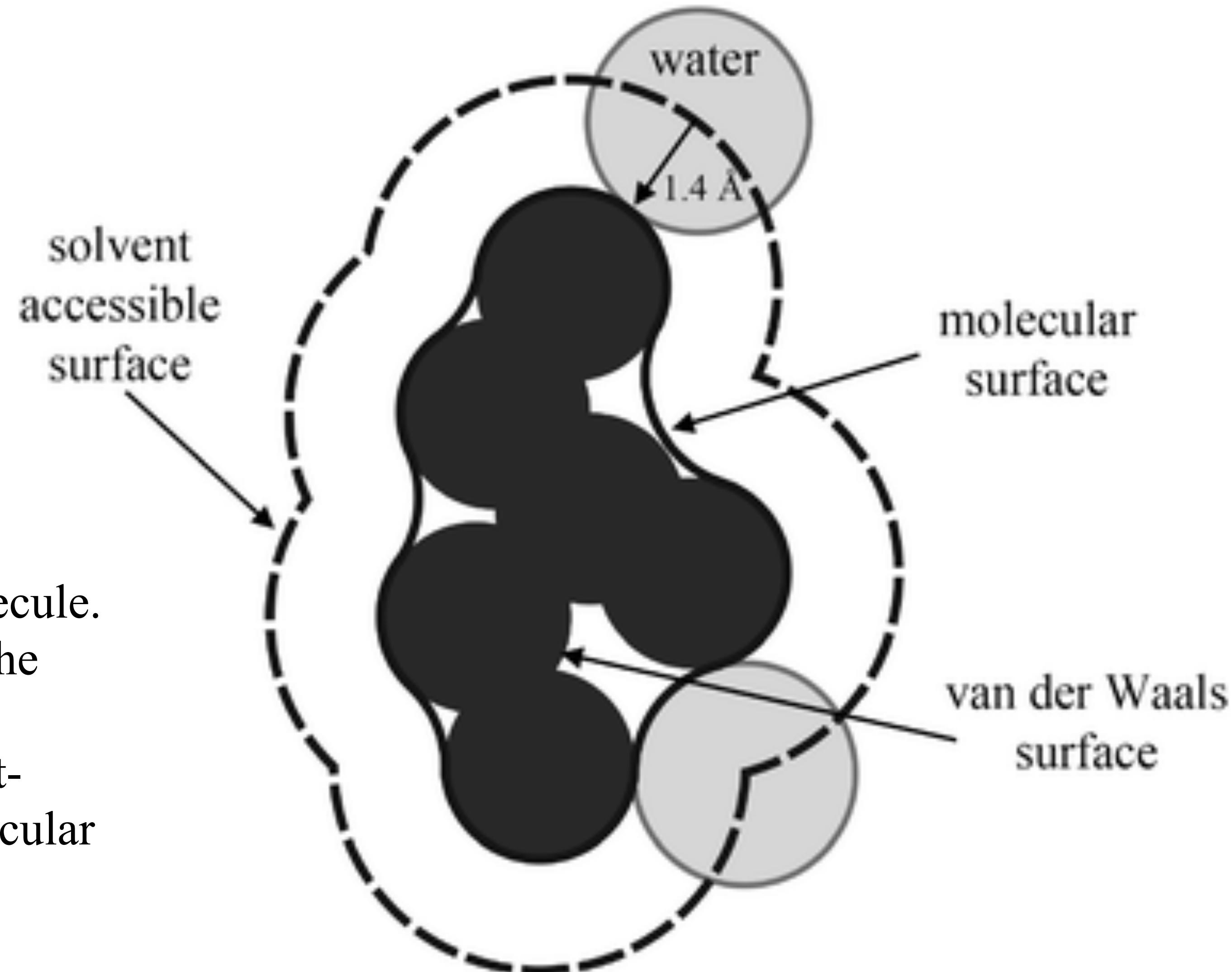


The water-accessible surface is shown, illustrating the indentations and crevices, which may function as binding sites.



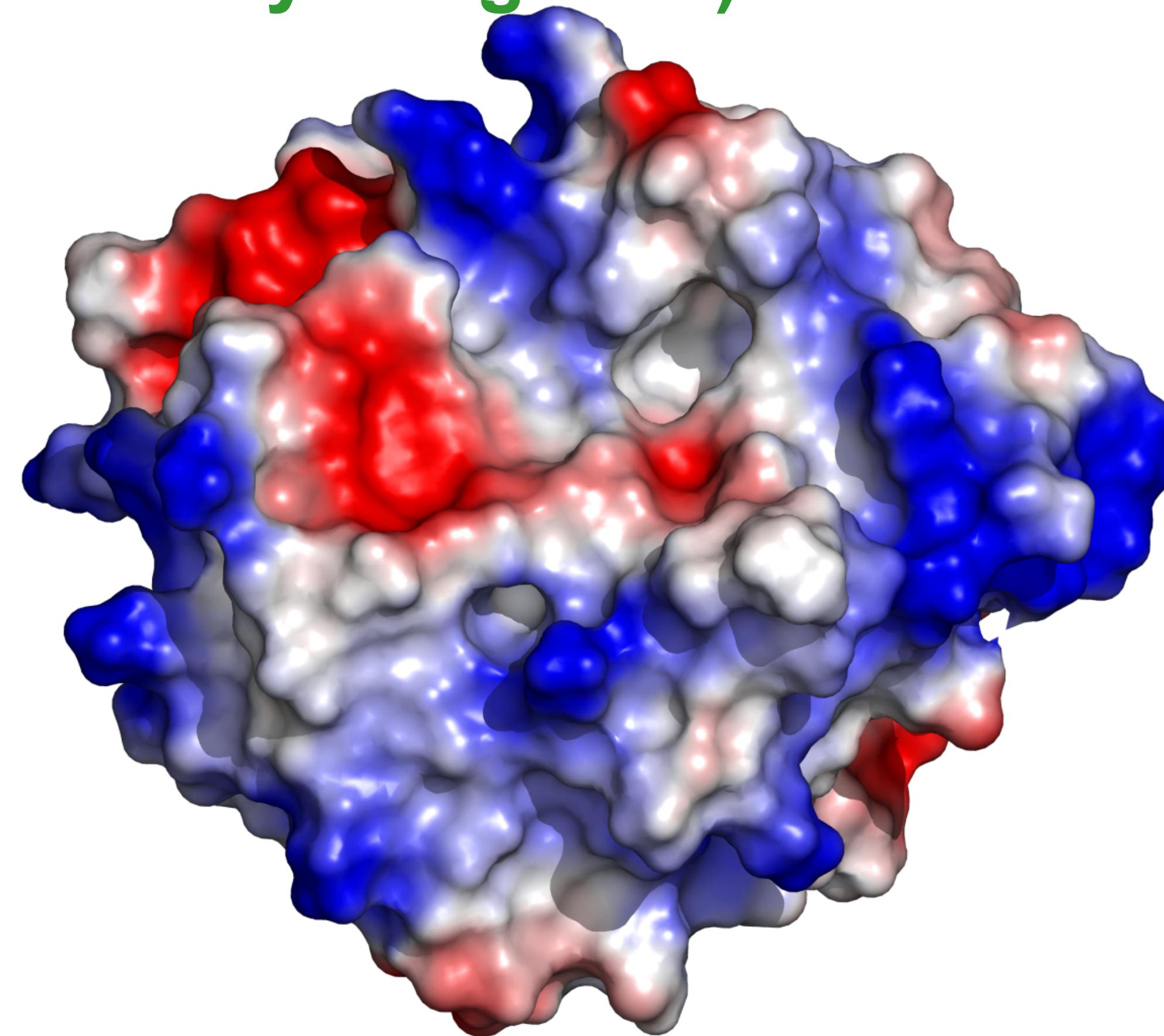
Types of surfaces

The black spheres represent the atoms comprising the molecule. The gray sphere is the water probe used to calculate the solvent-accessible and molecular surfaces.



Different graphic representations of proteins reveal different properties

- Surface, colored by electrostatic potential (Φ)
(complementarity to ligands)



The figure was produced using PyMol. Negative potentials ($0k_B T/e > \Phi > -60k_B T/e$) are red, positive potentials ($0k_B T/e < \Phi < 60k_B T/e$) are blue, and neutral potentials are white. The electrostatic potential was calculated using APBS.

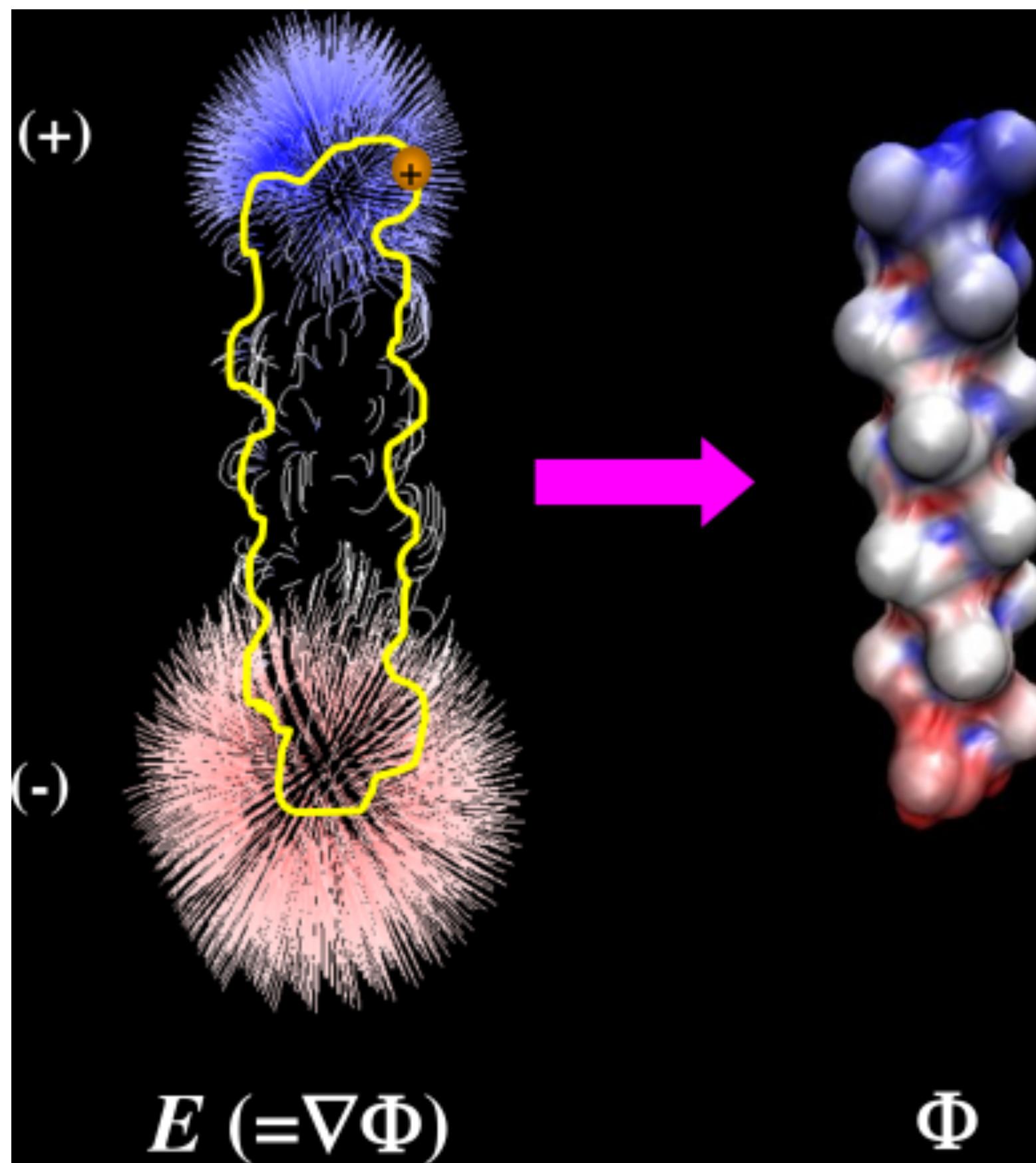


Different graphic representations of proteins reveal different properties

- **Electrostatic potential (Φ)**

- The energy of a probe charge placed within the protein's electric field (E)

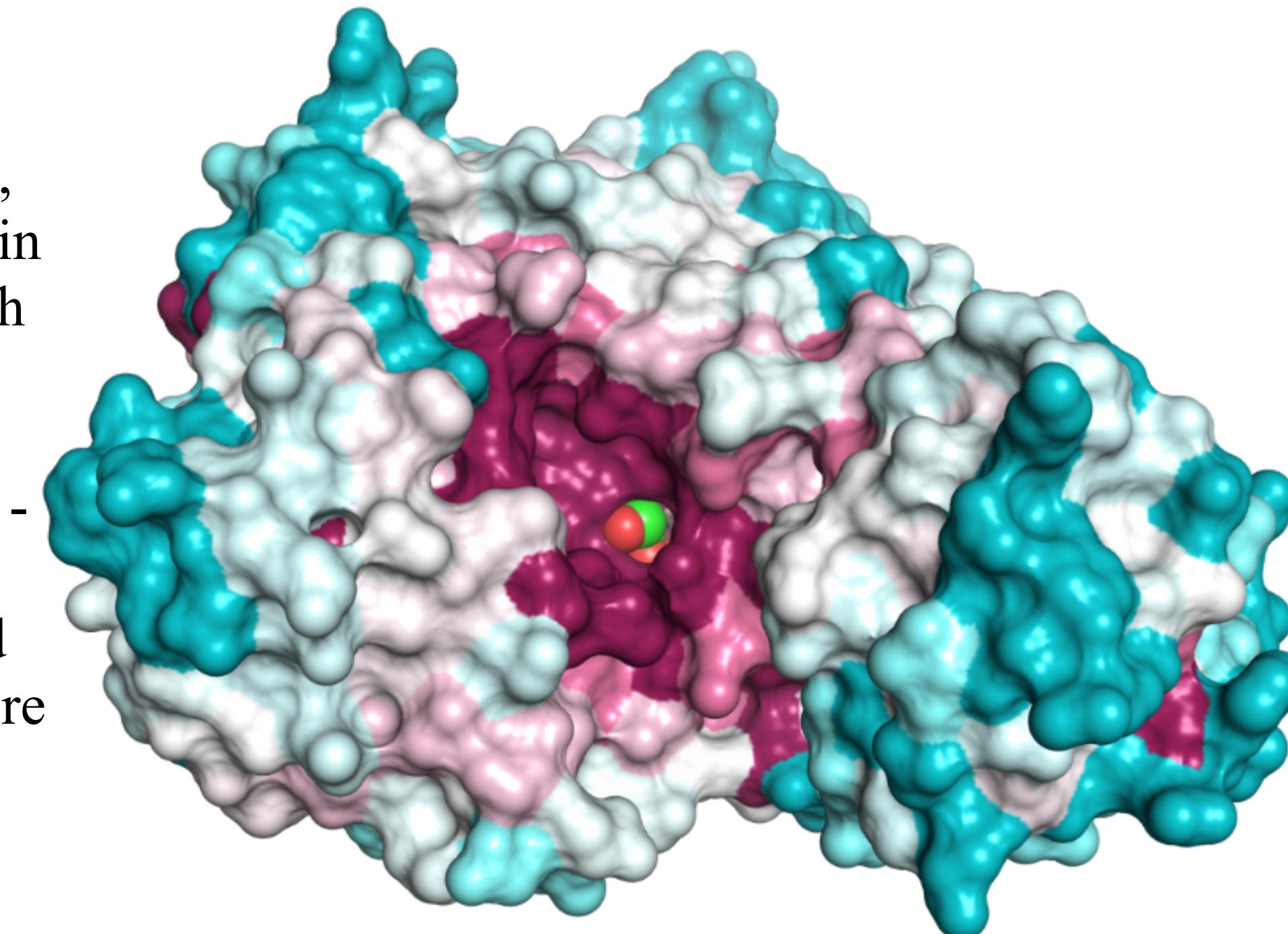
The electrostatic field of a spiral-shaped peptide (α -helix, see main text below) is shown of the left, and represented as lines emanating from the peptide (field lines are colored as in Figure 6). The peptide is represented implicitly by its contour, with the + and - signs representing the charges at its N- and C-termini, respectively. The charges result from the electric dipole, which characterize peptides having this shape (see Figure 2-13 below). The orange sphere positioned onto the contours of the peptide is a charged probe used to calculate the electrostatic potential at that point (shown on the right). As the bottom of the figure specifies, the electrostatic field is the derivative of the potential in three-dimensional space. The figure was produced using VMD.



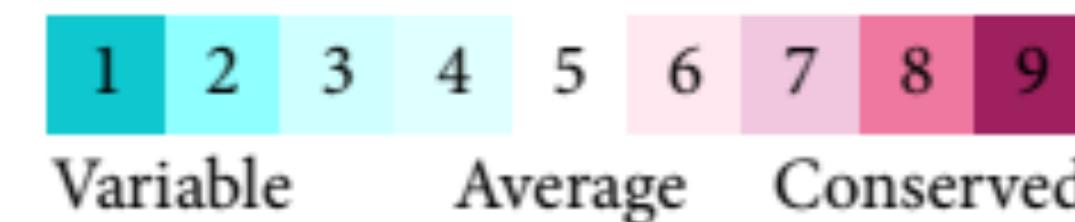
Different graphic representations of proteins reveal different properties

- Space-fill, colored by **evolutionary conservation (biologically important regions)**

The protein in the figure (triose phosphate isomerase, PDB entry 1amk) is shown in spacefill representation, with each residue colored by evolutionary conservation (turquoise - lowest, maroon - highest. See color-code in figure). The most conserved region is in the middle, where the natural substrate of the enzyme is bound.



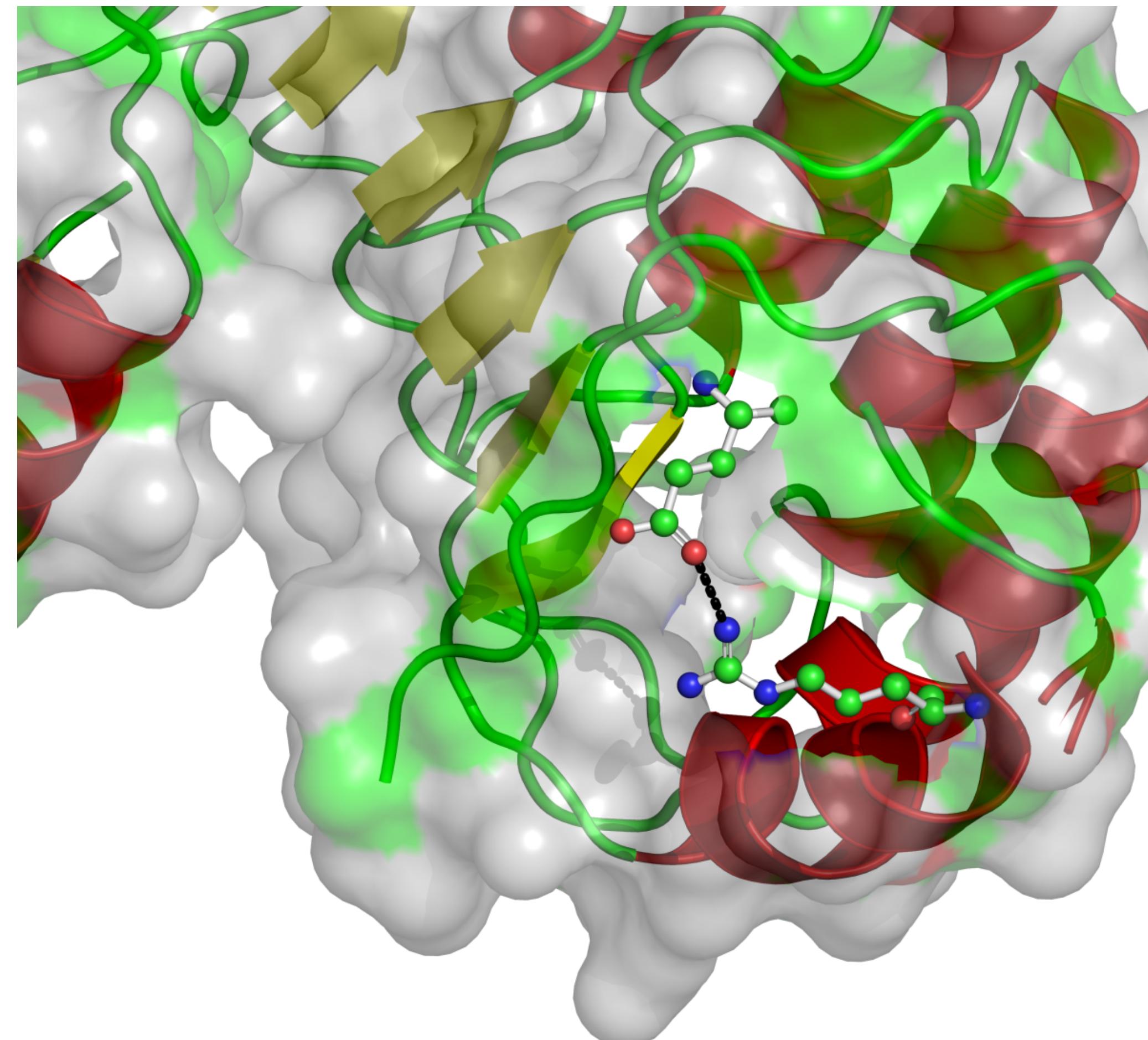
(<http://consurf.tau.ac.il>)



UNIVEI
DEGLI
DI MIL

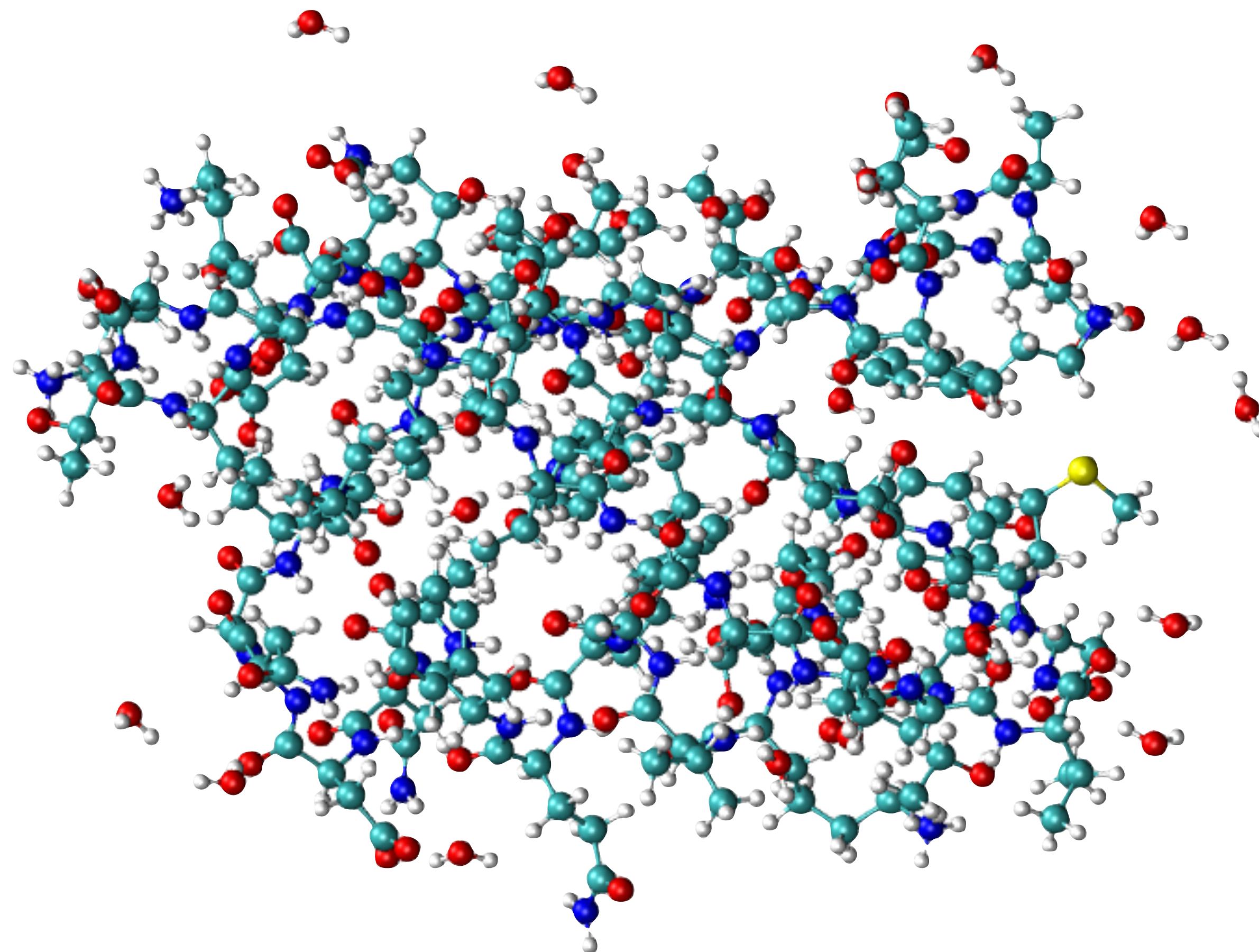
Different graphic representations of proteins reveal different properties

- Integrated (multiple properties)



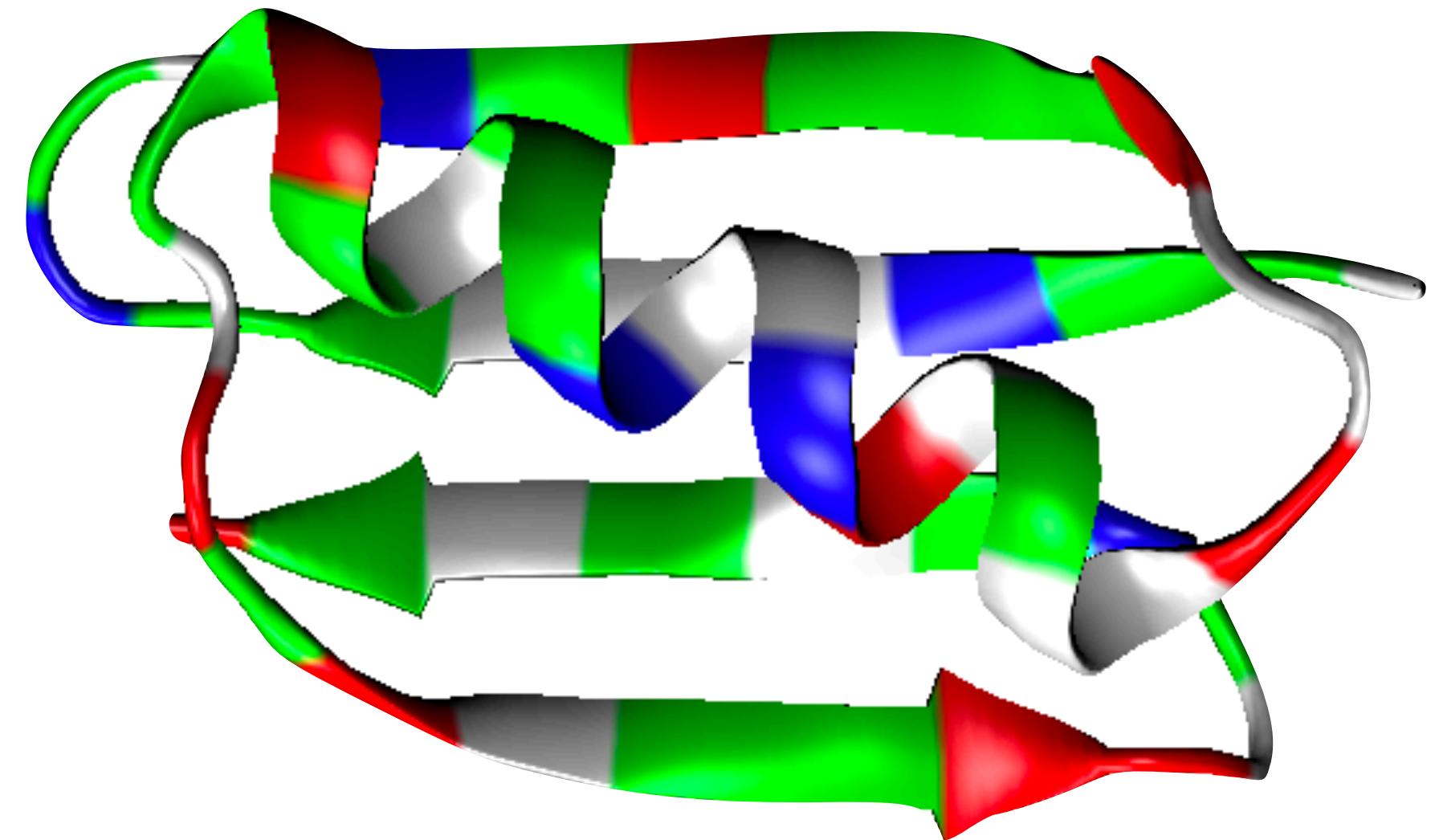
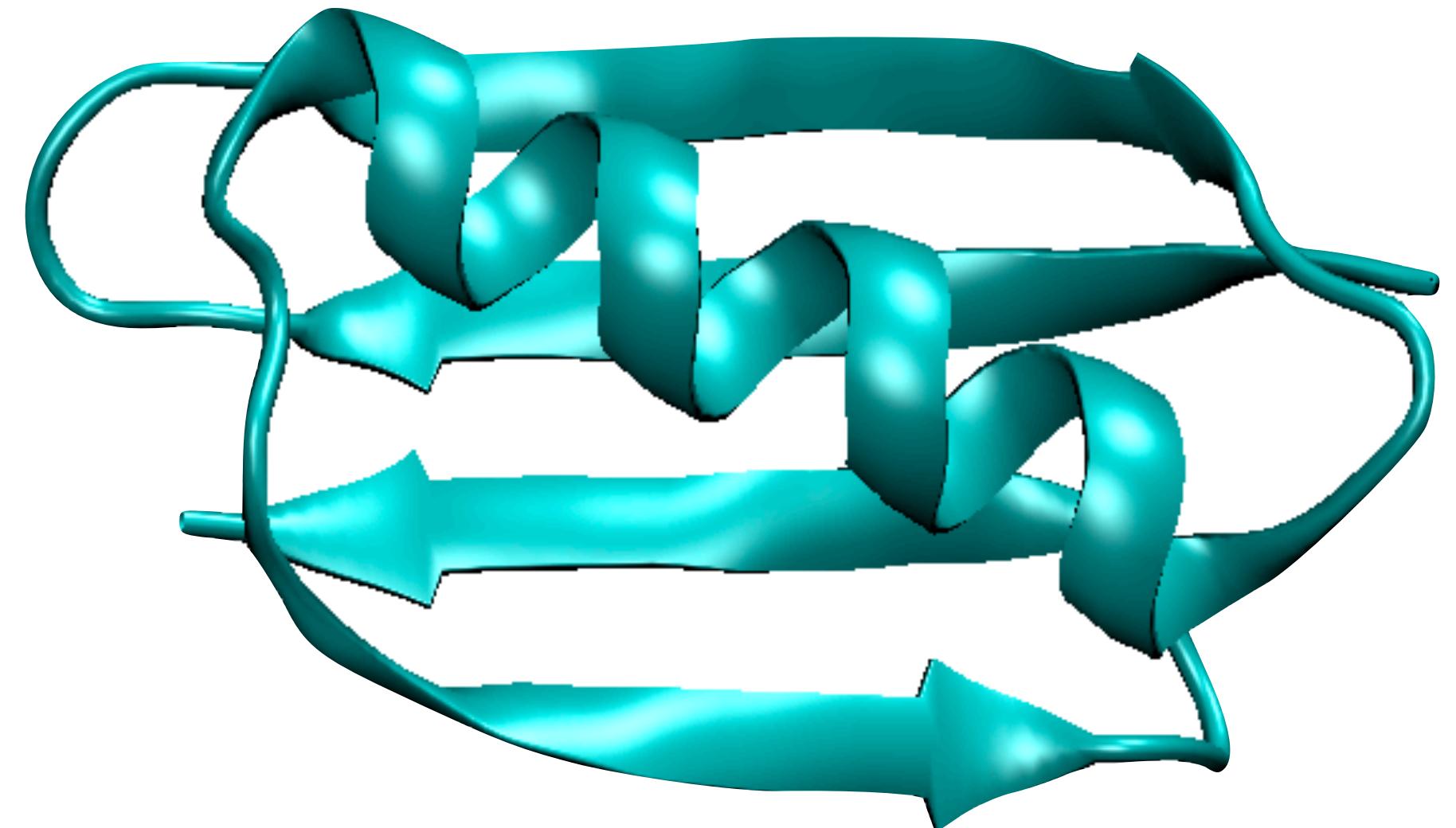
Highlighting interactions:

Ball-and-Stick: too many details

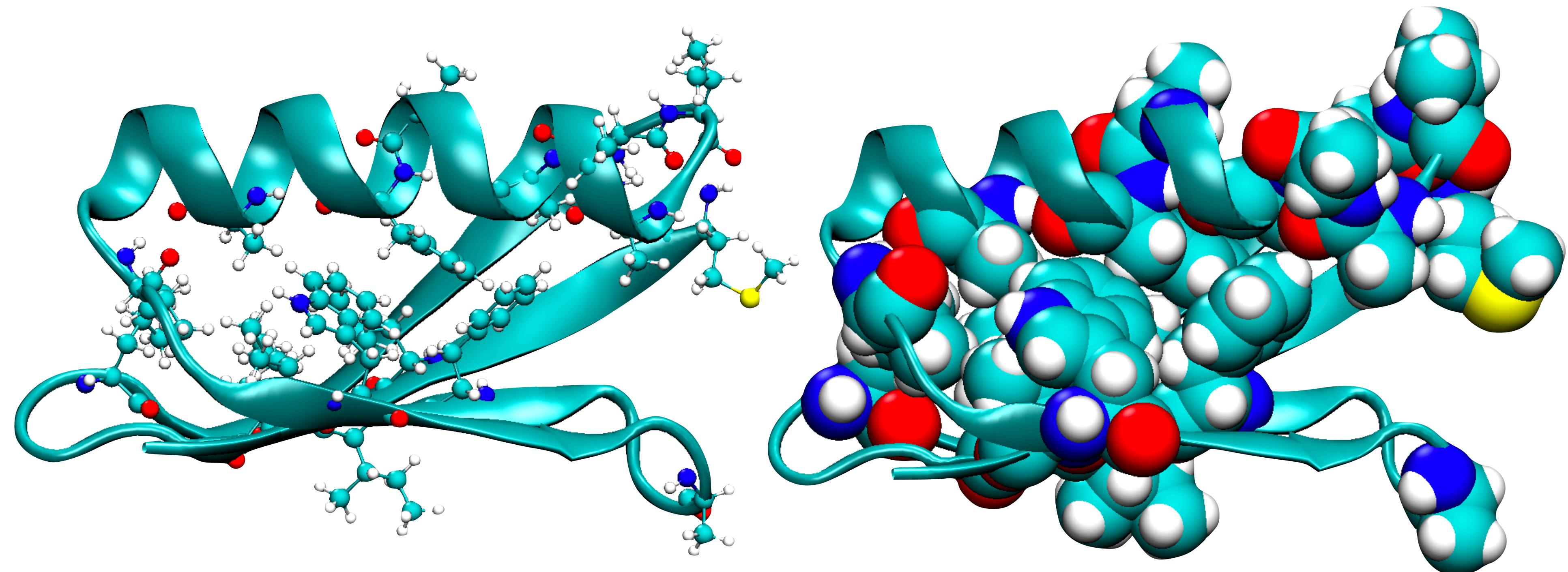


Secondary structures

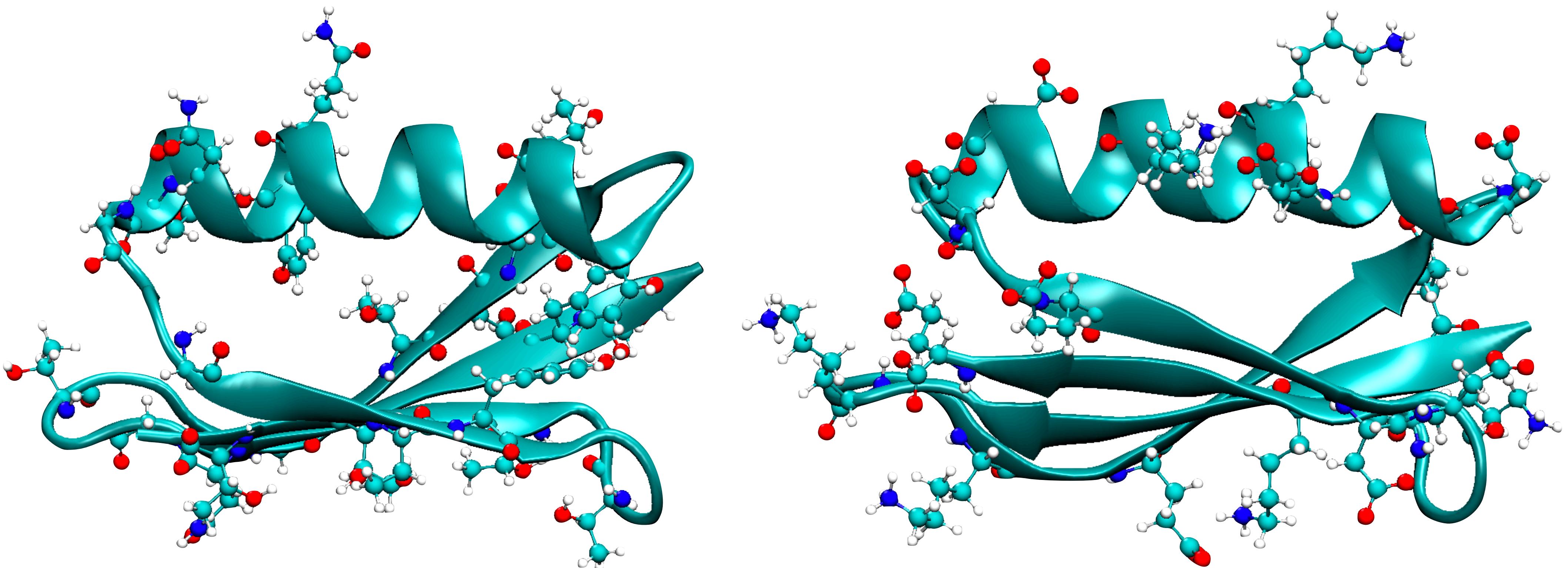
By residue type
(polar/non-polar/
positive/negative)



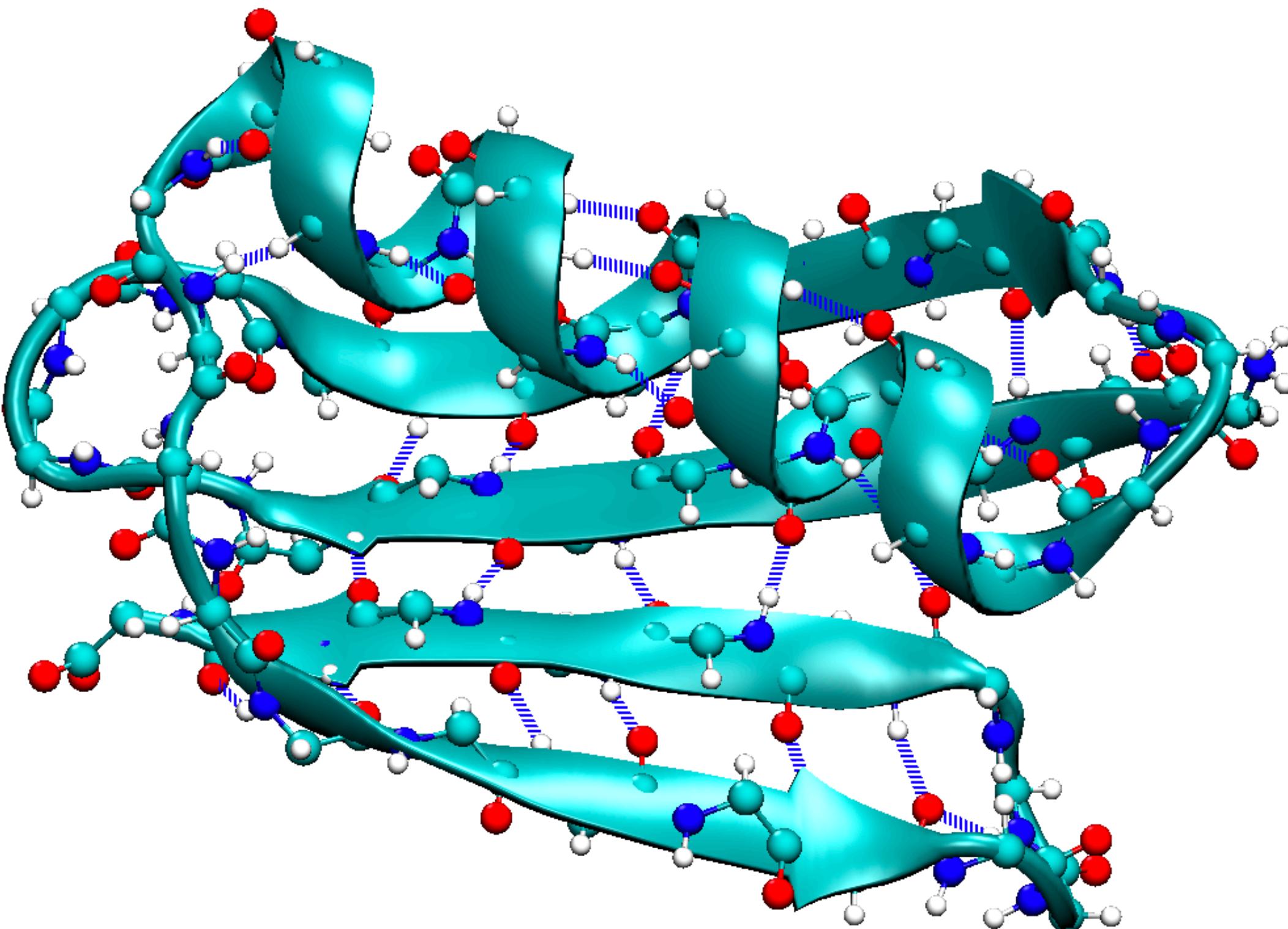
Hydrophobic residues



Polar and Charged



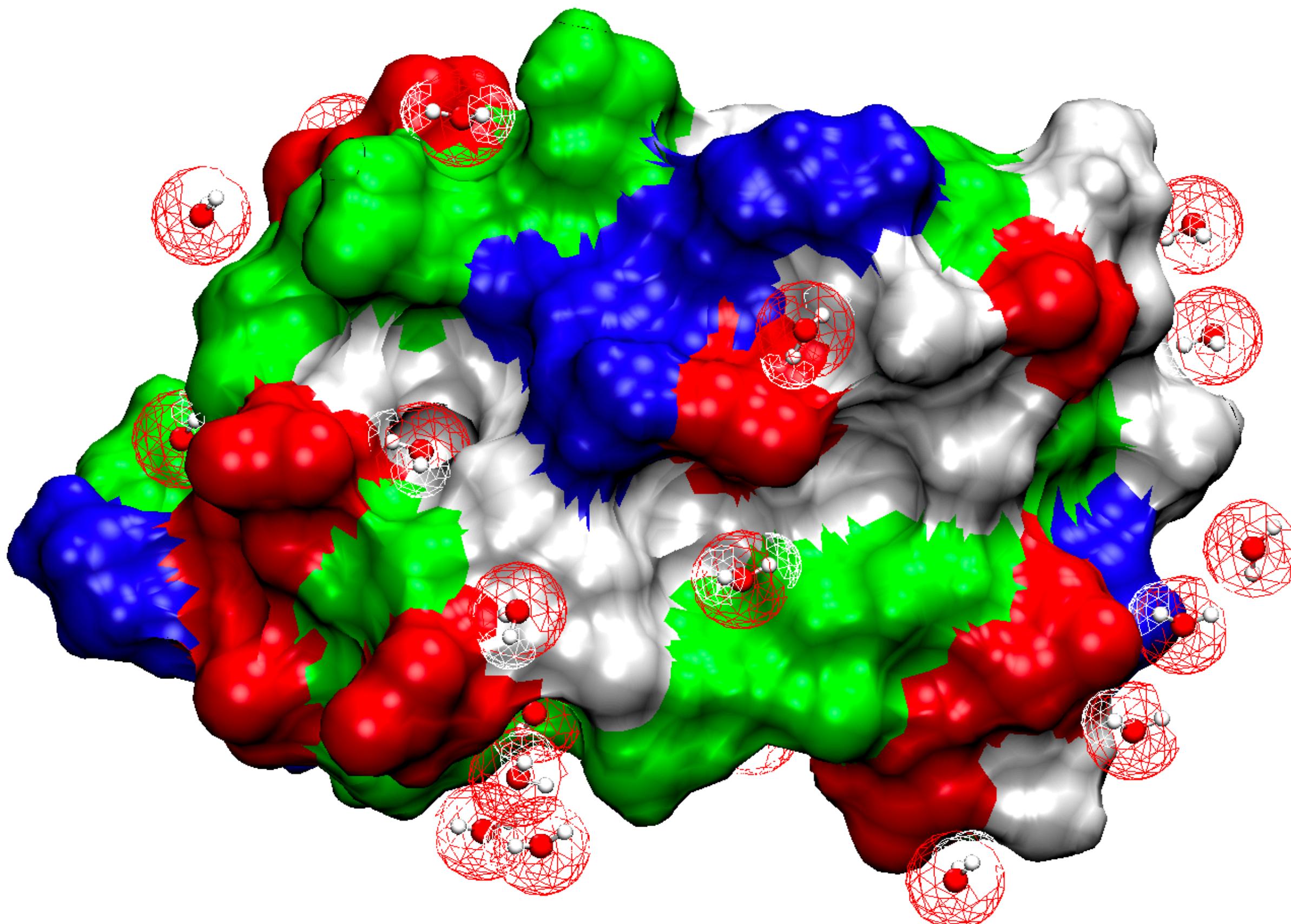
Hydrogen bonds



Hydrogen bonds are usually defined geometrically
distance between donor and acceptor < 3.5
Angle between donor, acceptor and hydrogen < 30 deg



SASA and crystal water



UNIVERSITÀ
DEGLI STUDI
DI MILANO



Visualisation: Software

VMD: Visual Molecular Dynamics, most used for molecular dynamics simulations

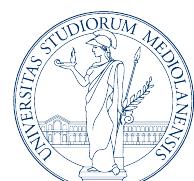
PyMol: high-quality graphics, scriptable in python, very good for protein superposition

ChimeraX: high-quality graphics, very good for cryo-EM

Coot: focused on structure determination for X-Ray and Cryo-EM

UnityMol: based on gaming Unity Engine

NGLView: to be programmed in python



Practicals

Install VMD if you will
use your own laptop

We are going to use
COLAB:

colab.research.google.com

<https://www.ks.uiuc.edu/Research/vmd/>

To use them (essentially to
save the results of your work)

You need a google Drive
account, please do one

