

# Structural Bioinformatics

Elodie Laine

Master BIM-BMC Semestre 3, 2014-2015

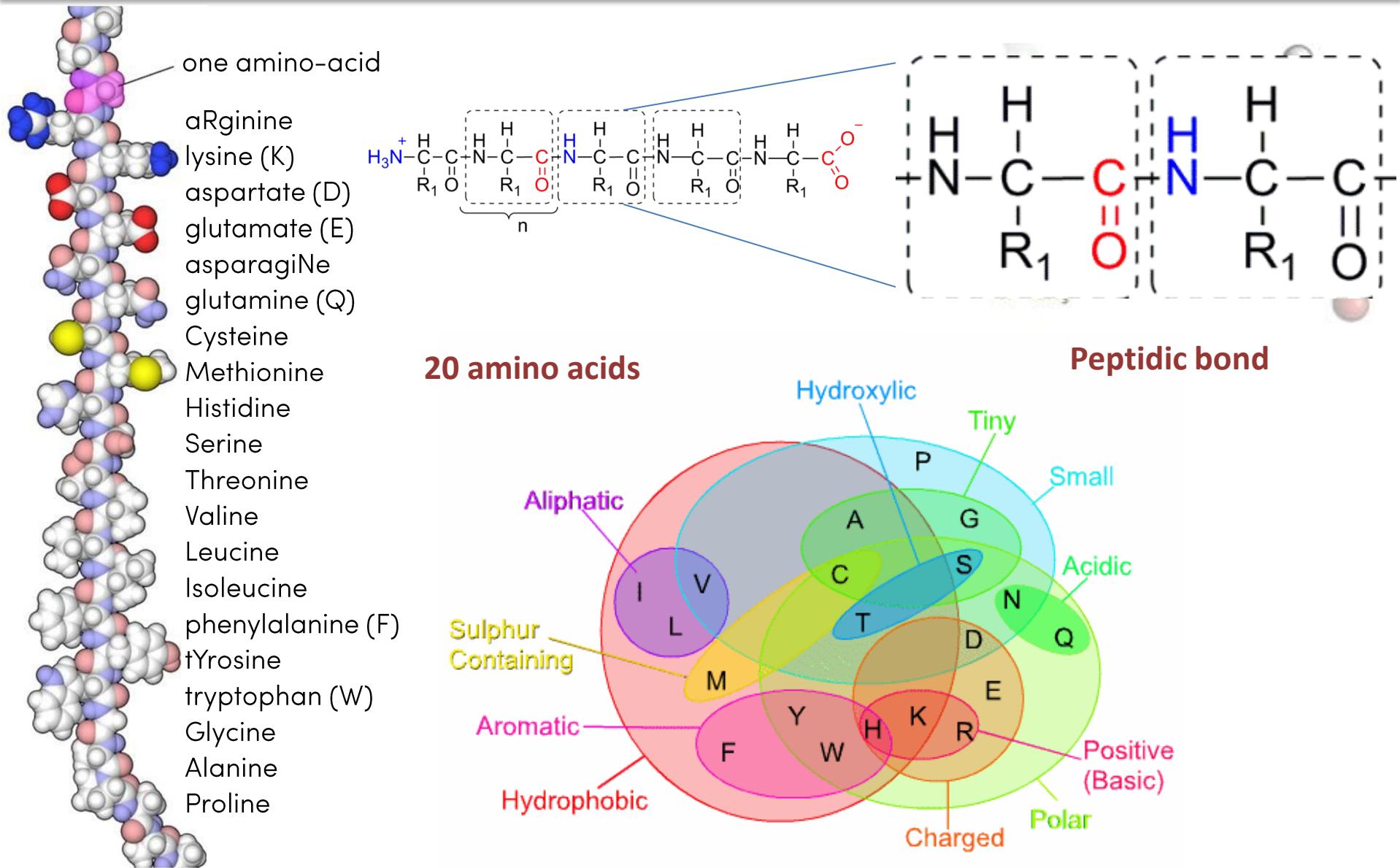
Laboratoire de Biologie Computationnelle et Quantitative (LCQB)

*e-documents:* <http://www.lcqb.upmc.fr/laine/STRUCT>

*e-mail:* [elodie.laine@upmc.fr](mailto:elodie.laine@upmc.fr)

# Introduction

# Protein composition

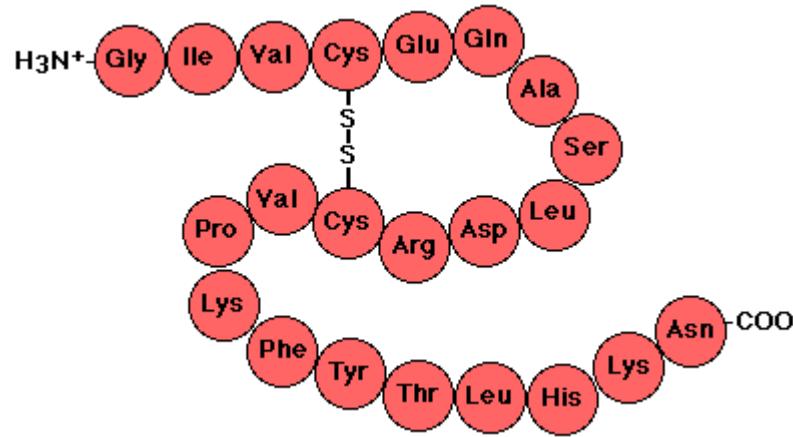


# Protein structure

1<sup>st</sup> level of organisation : primary structure

...QNCQLRPSGWQCRPTRGDCDLPEFCPGDSSQCPDVSLGDG...

~10's to ~1000's of  
amino acid residues

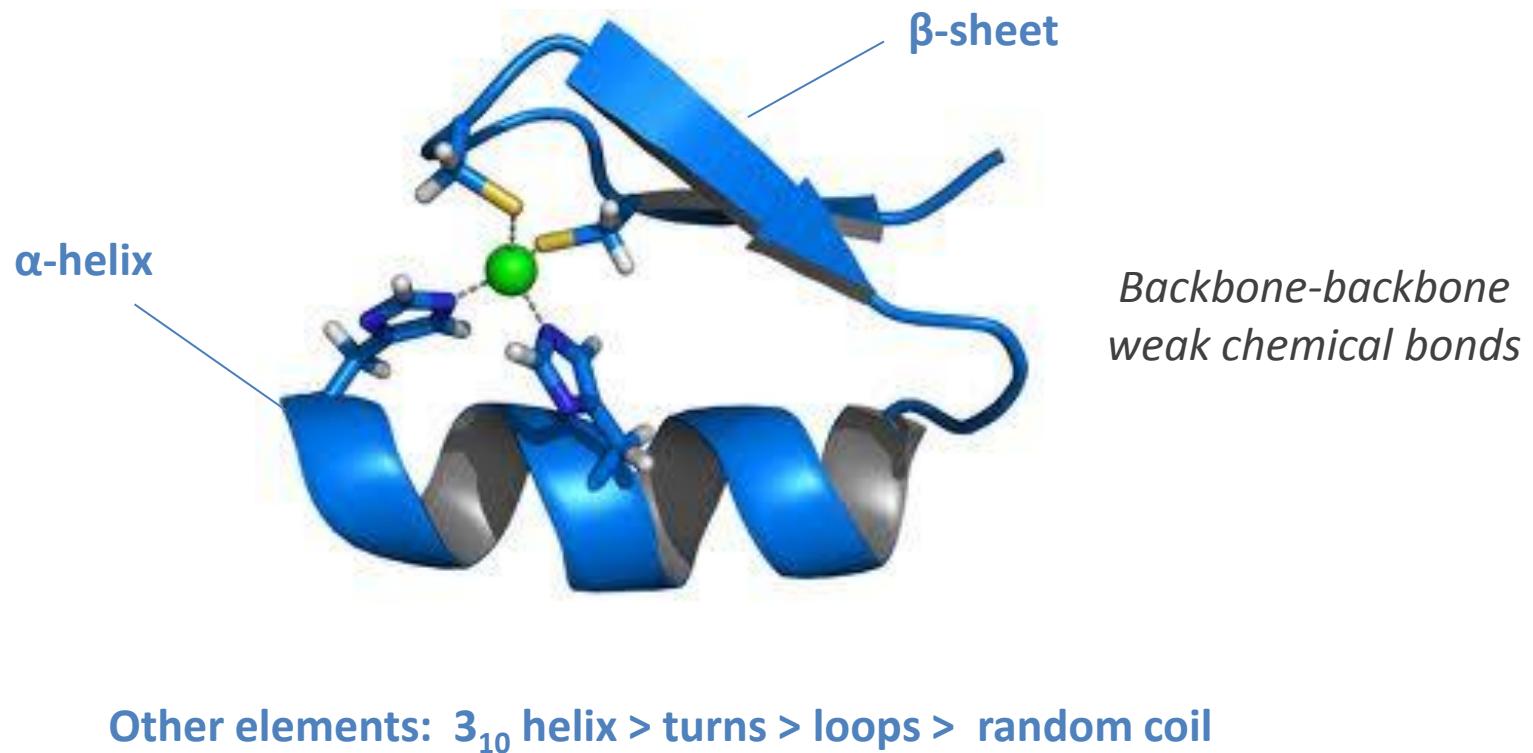


*Covalent bonds*

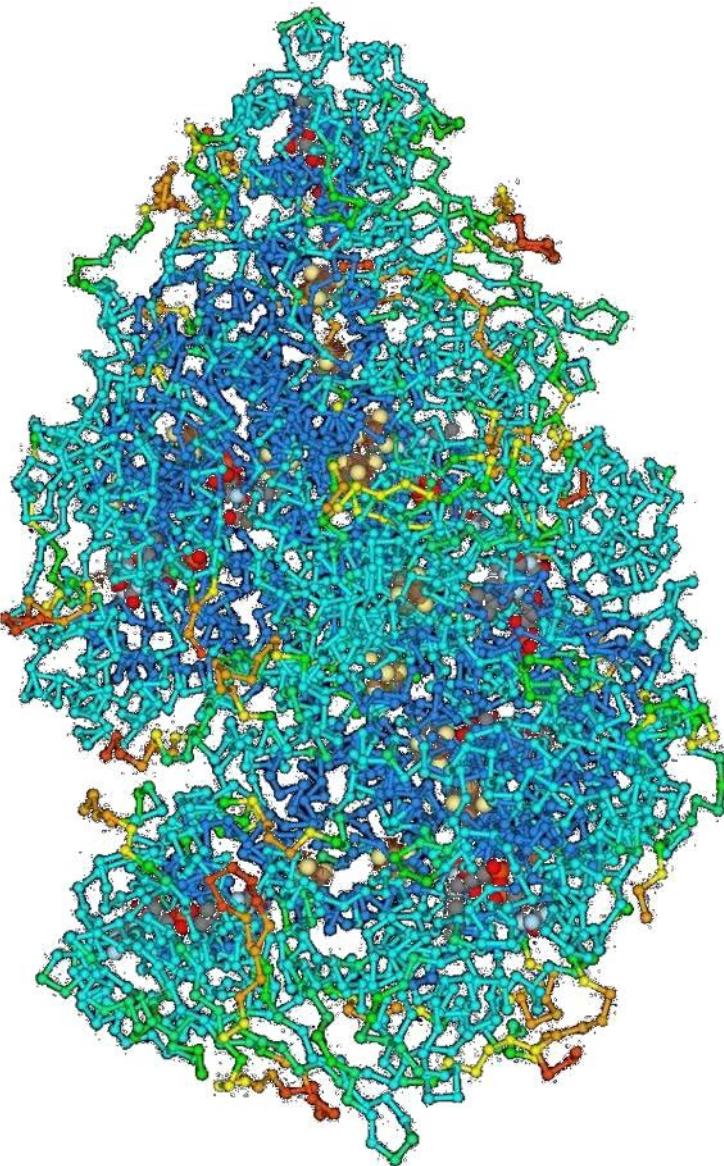
**1 protein = 1 polypeptidic chain**

# Protein structure

2<sup>nd</sup> level of organisation : secondary structure



# Protein structure

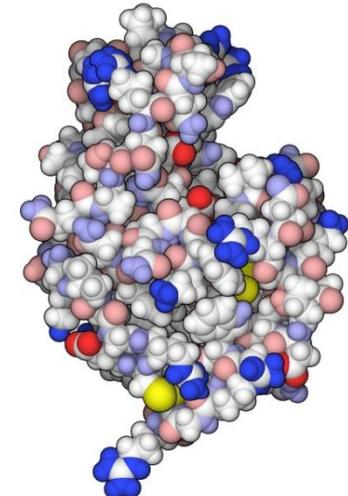


## 3<sup>rd</sup> level of organisation : tertiary structure

A protein sequence adopts a particular fold in solution, which corresponds to a free energy minimum

*Types of non-covalent interactions:*

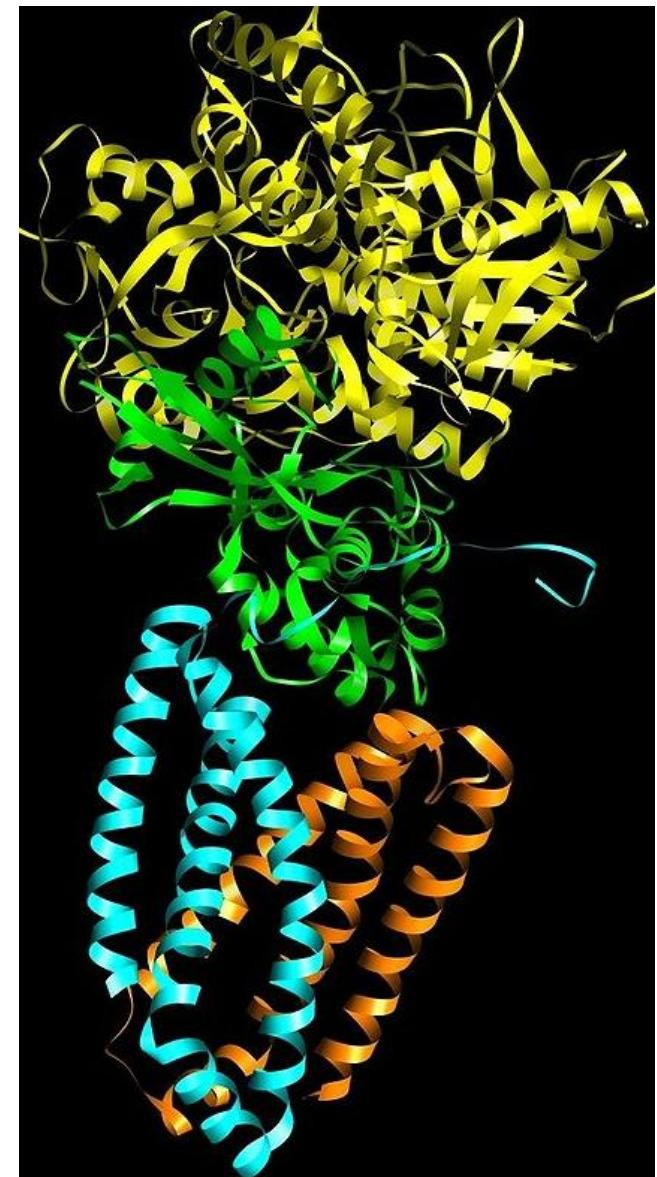
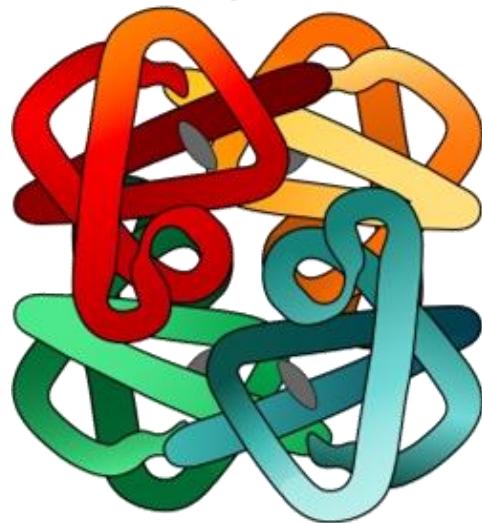
- *salt bridges*
- *hydrogen bonds*
- *hydrophobic contacts*
- *pi-pi stacking...*



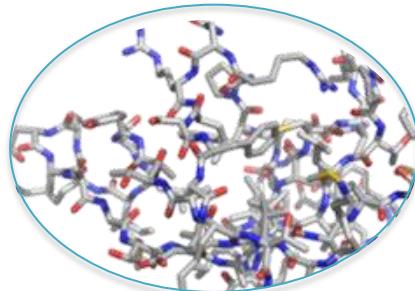
# Protein structure

## 4<sup>th</sup> level of organisation : quaternary structure

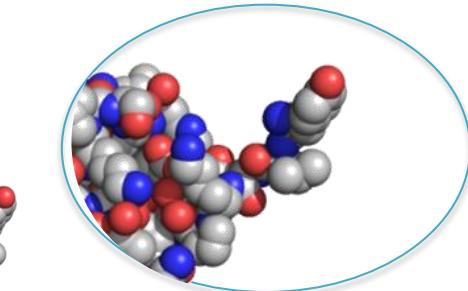
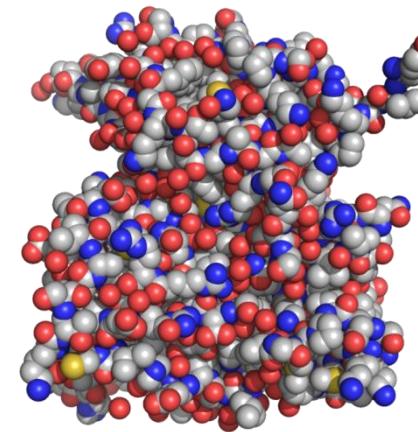
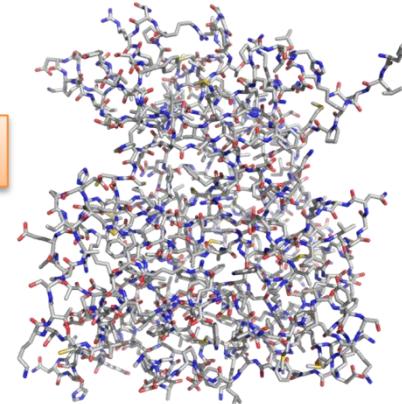
Arrangements of domains within a protein or of proteins within a macromolecular assembly



# Protein structure representations

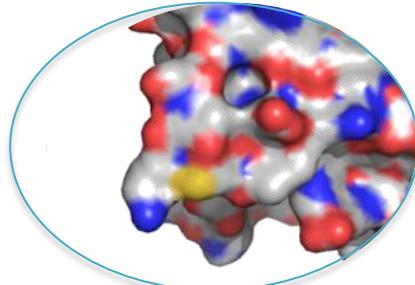
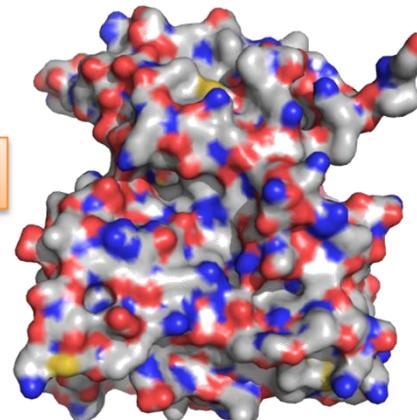


sticks

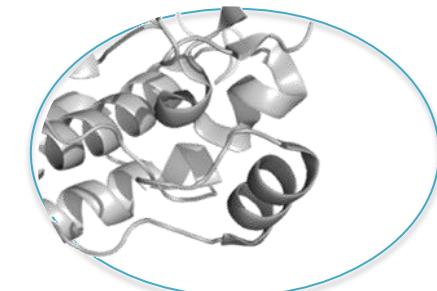
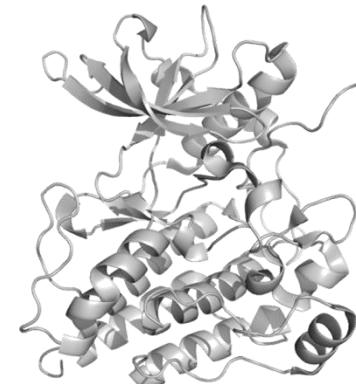


spheres

surface



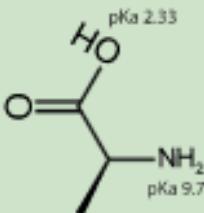
cartoon



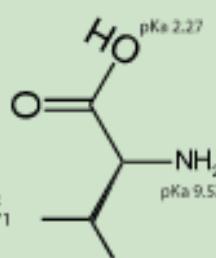
# Biochemical units

## Hydrophobic side chain

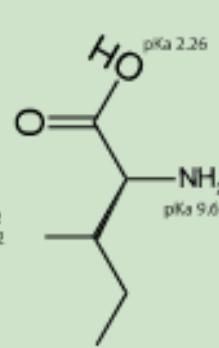
Alanine  
(Ala) **A**



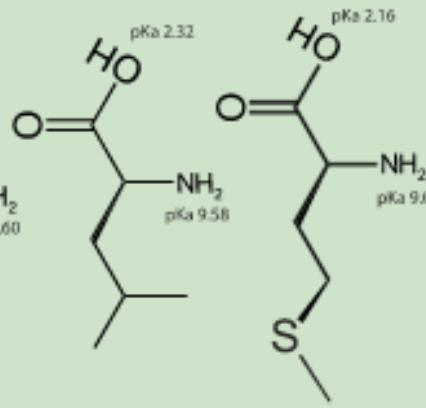
Valine  
(Val) **V**



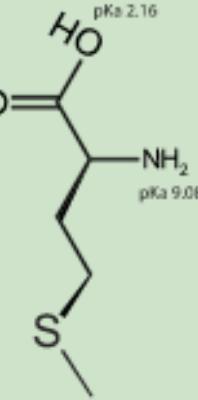
Isoleucine  
(Ile) **I**



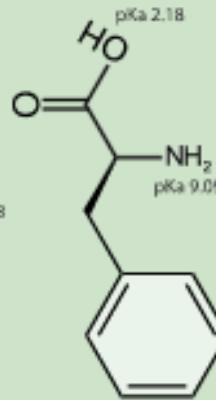
Leucine  
(Leu) **L**



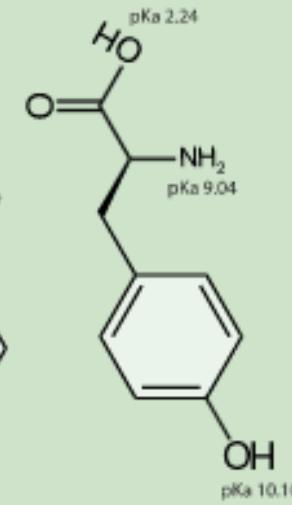
Methionine  
(Met) **M**



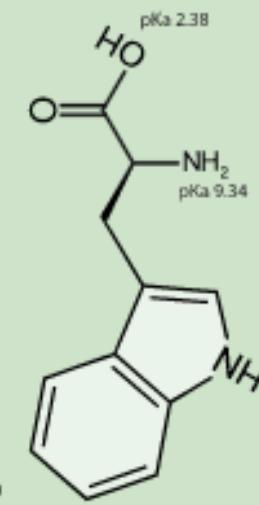
Phenylalanine  
(Phe) **F**



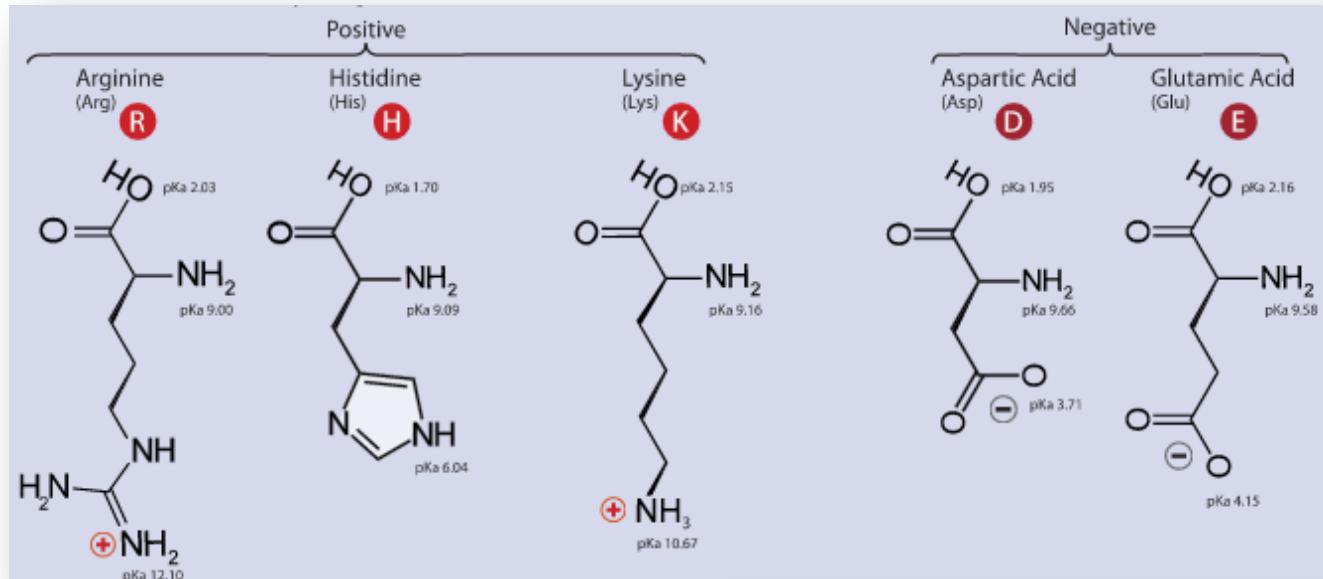
Tyrosine  
(Tyr) **Y**



Tryptophan  
(Trp) **W**

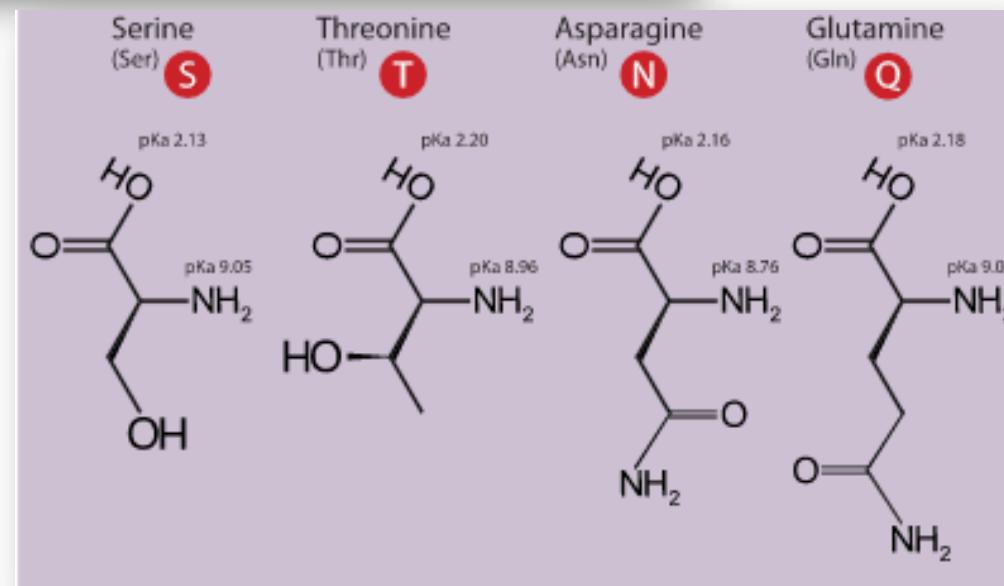


# Biochemical units



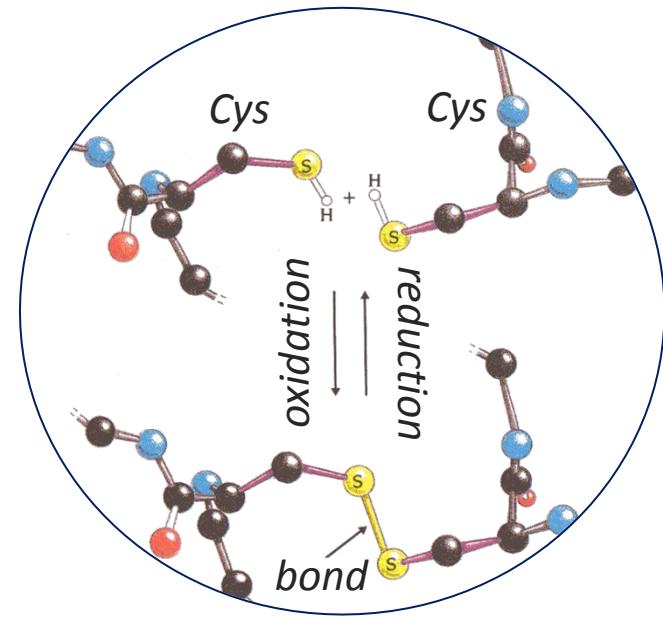
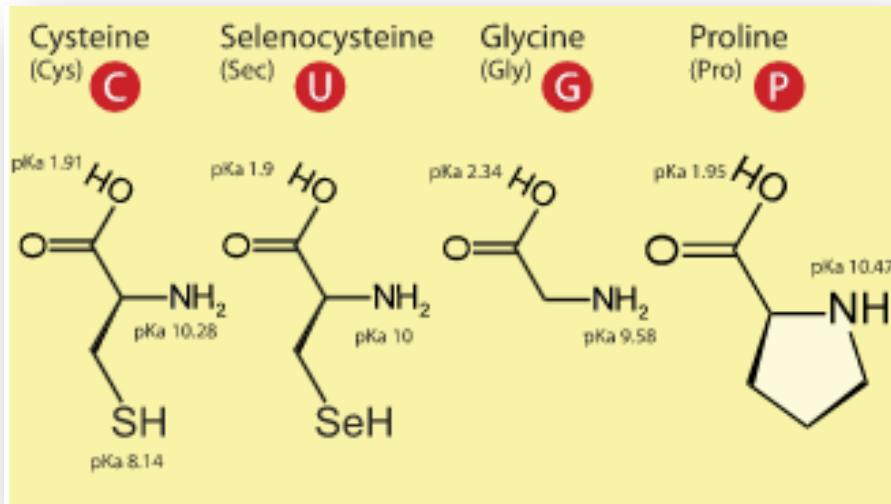
Charged side chain

Polar side chain

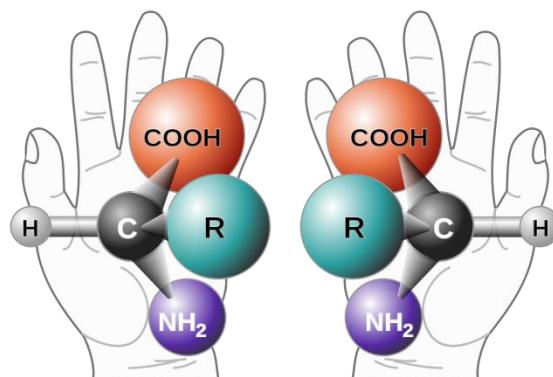


# Biochemical units

## Special



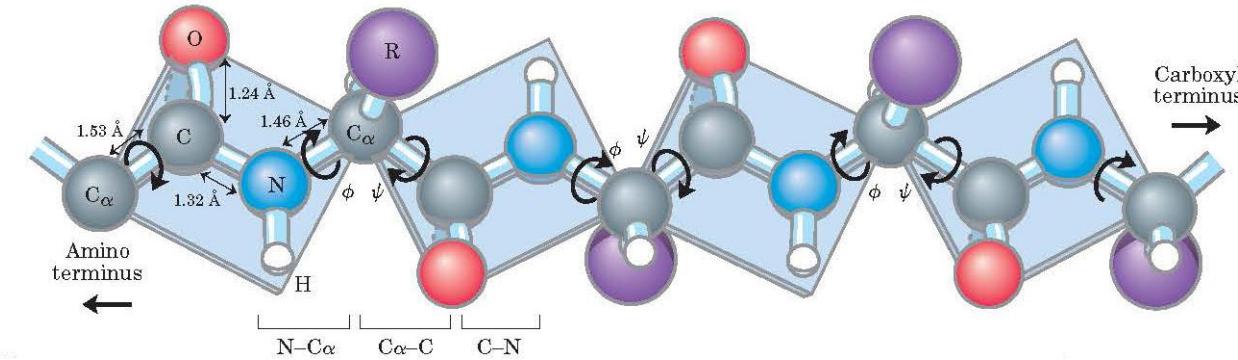
## Chirality



The translation machinery of protein has evolved to utilize only one of the chiral forms of amino acids : the **L-form**

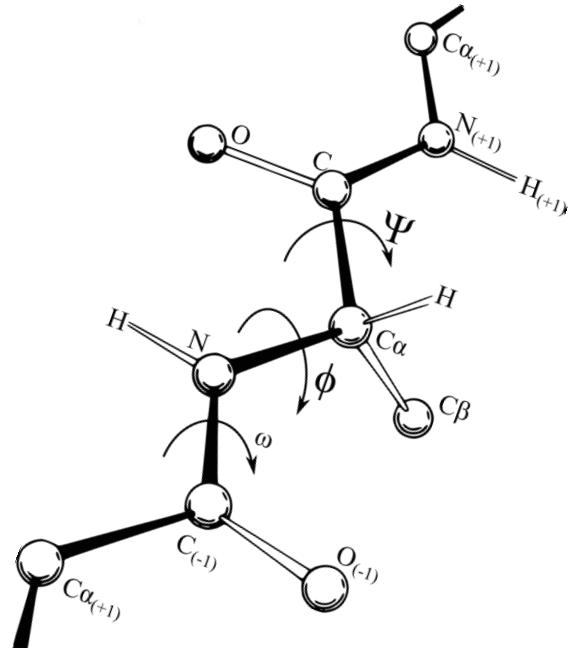
# Structural units

## Planar peptide units



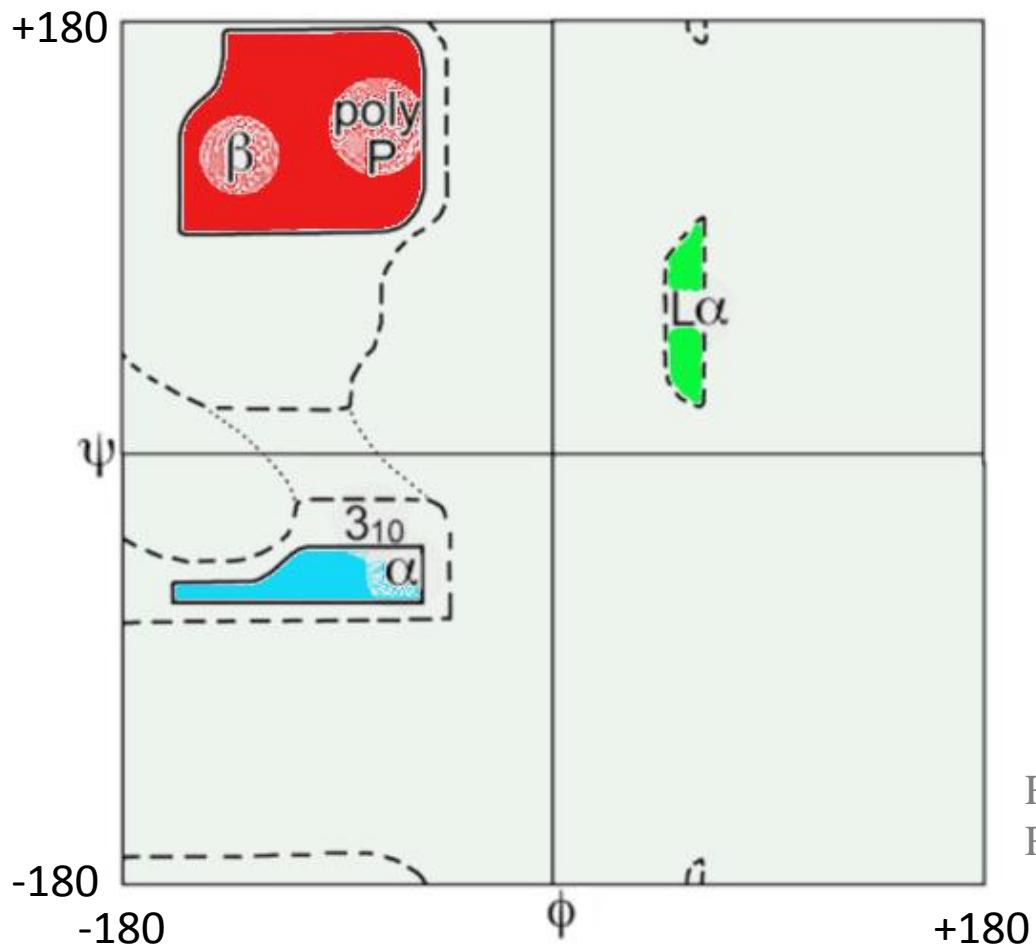
## Backbone torsion angles

The conformation of the whole main chain of the polypeptide is completely determined when the **rotation angles  $\varphi$  ( $N-C\alpha$ ) and  $\psi = (C\alpha-C')$**  are defined with high accuracy.



# Structural units

## Ramachandran diagram

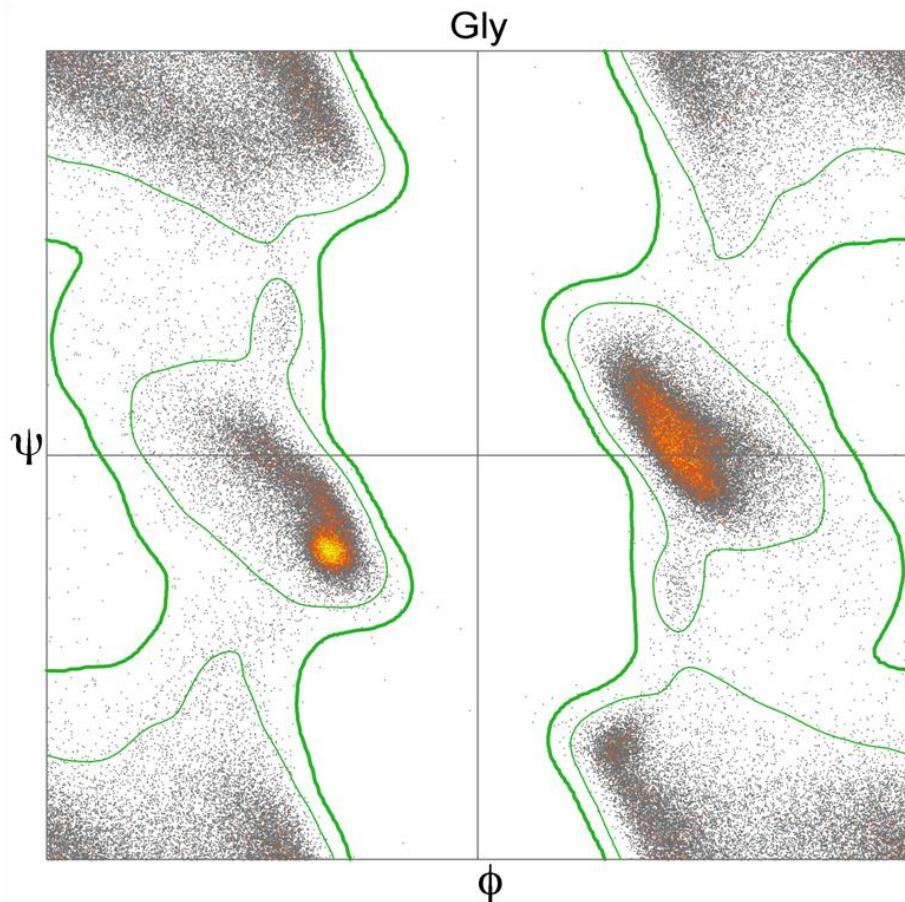


Most combinations of  $\phi$  and  $\psi$  for an amino acid are not allowed because of steric collisions between the side chain and main chain

Ramachandran *et al.* (1963) *J. Mol. Biol.*  
Ramachandran *et al.* (1968) *Adv. Protein. Chem.*

# Structural units

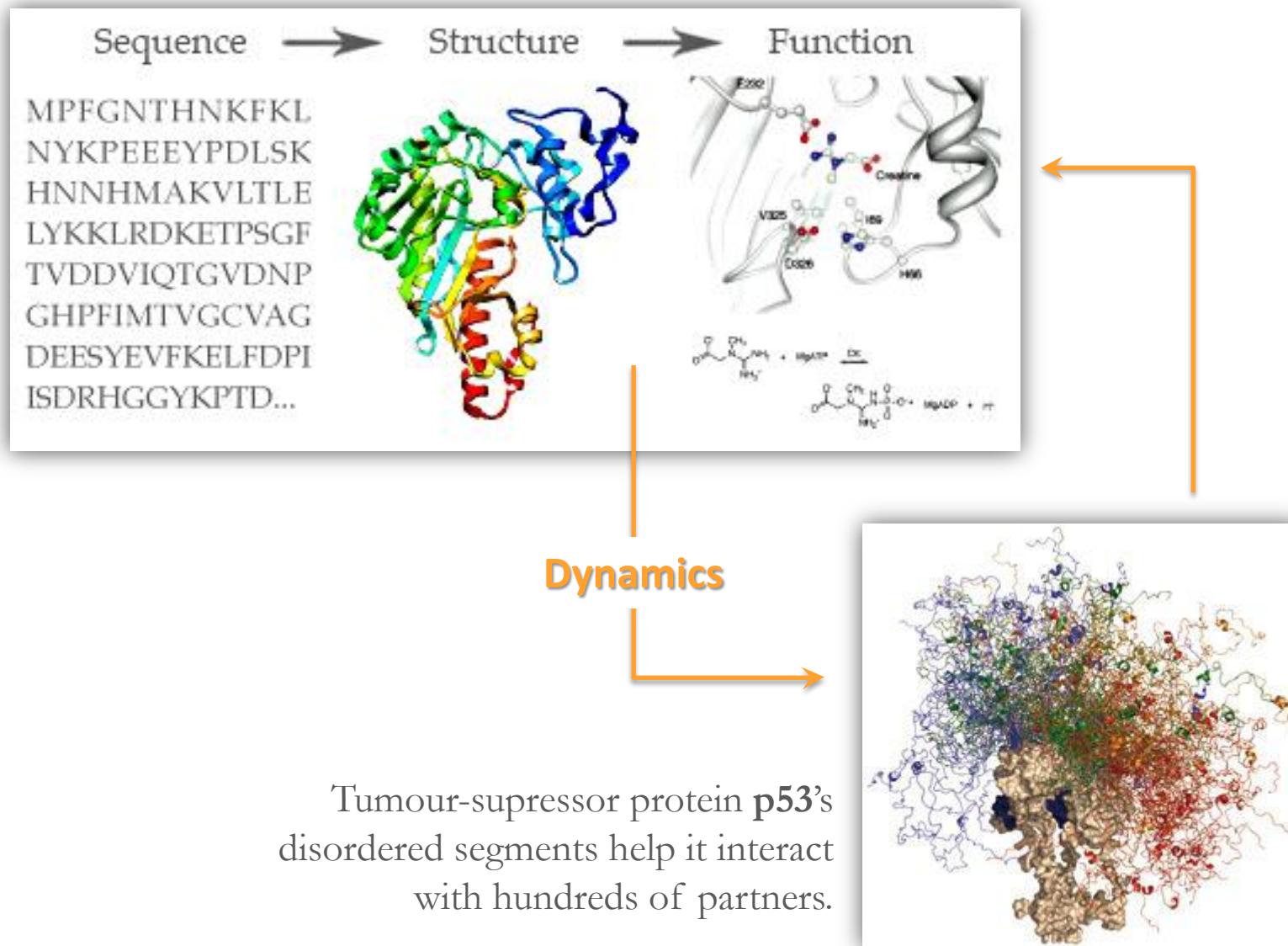
## Ramachandran diagram



**Glycine** can adopt a much wider range of conformations than the other residues and thus plays a structurally very important role

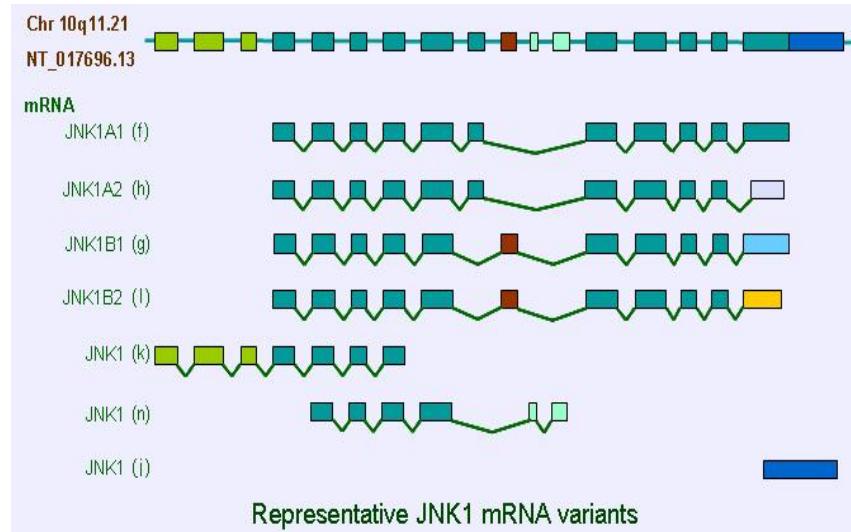
Ramachandran *et al.* (1963) *J. Mol. Biol.*  
Ramachandran *et al.* (1968) *Adv. Protein. Chem.*

# Sequence-structure-function paradigm

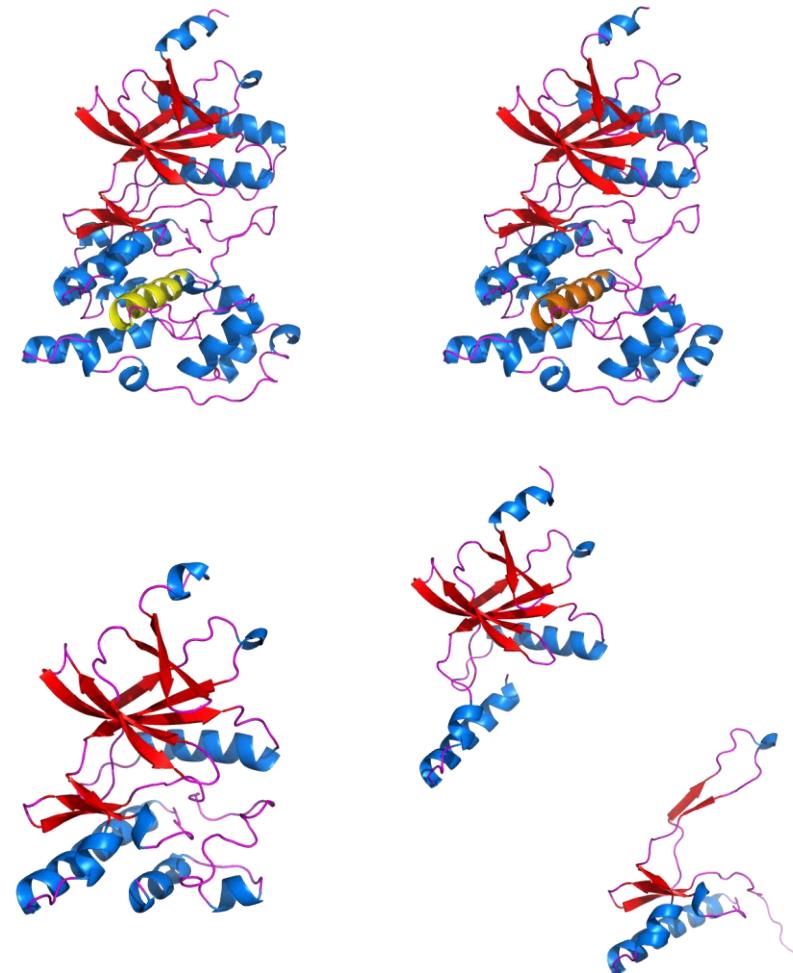


# Multiple isoforms from a single gene

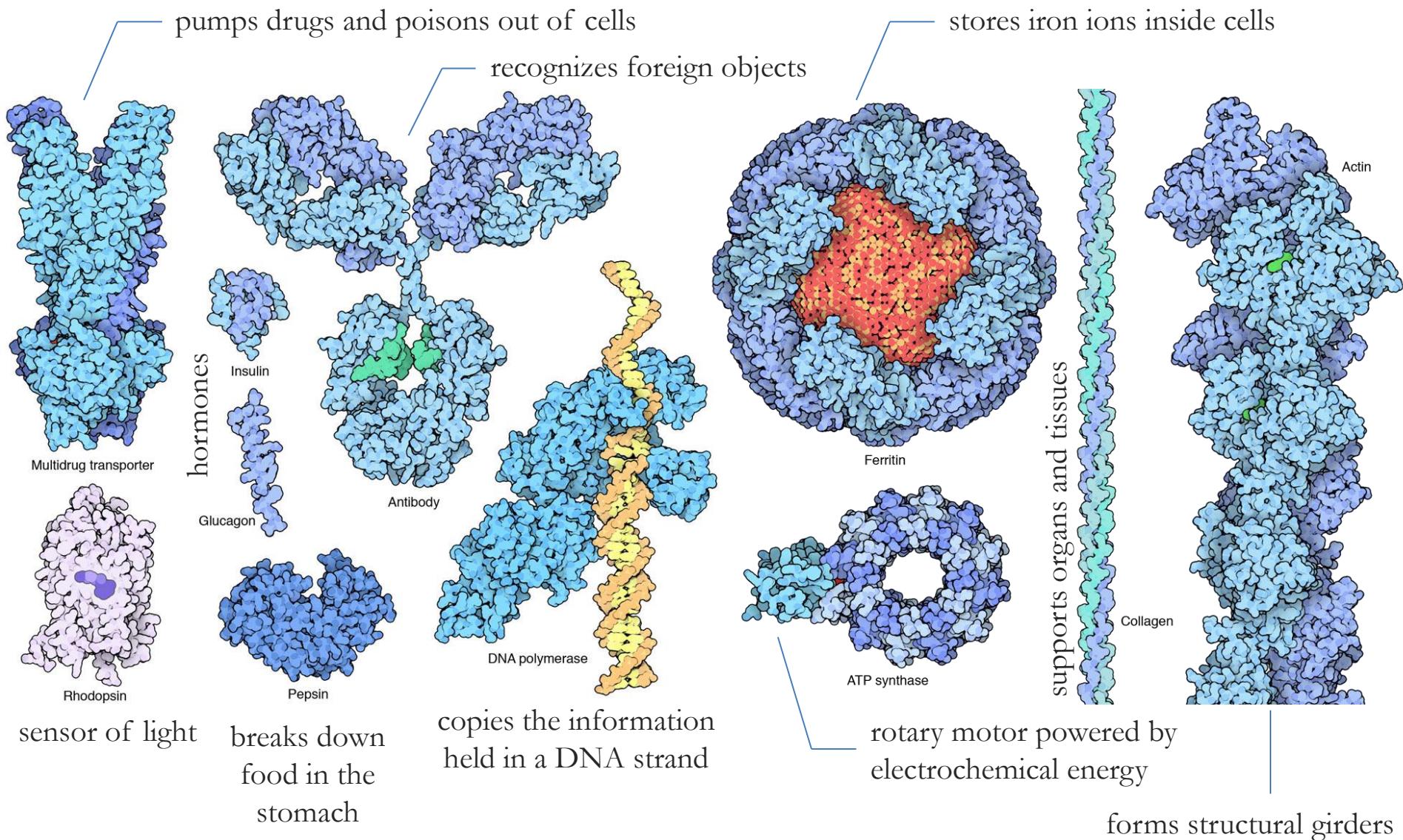
Does one sequence code for one protein?



Alternative splicing produces several isoforms from the same gene, by combining subsets of exons in different ways

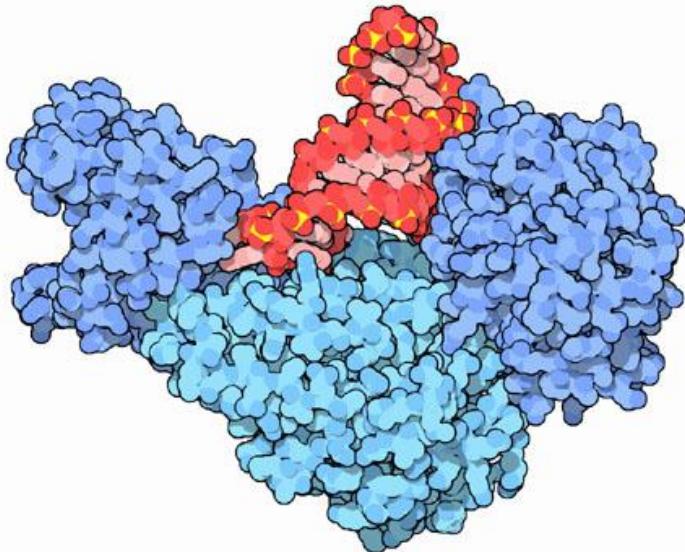
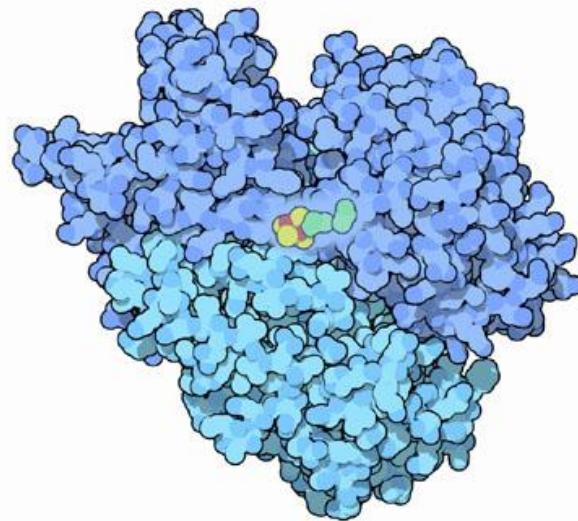


# Protein functions



# Protein functions

The enzyme **aconitase** is a key player in the central pathway of energy production. It converts citrate into isocitrate.



The **iron regulatory protein 1** interacts with messenger RNA to control the levels of iron inside cells.

**Moonlighting proteins** lead double lives, performing two entirely different functions

# Protein folding prediction problem

The geometric configuration of a protein's *native state* determines its macroscopic properties, behaviour and function.

The number of possible conformations for a given protein is astronomical.

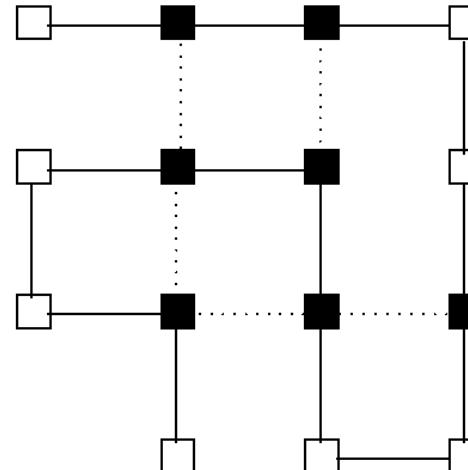
ex: 100aa, 3 conf/aa =>  $5 \cdot 10^{47}$  conf

1 fold/ $10^{-13}$  sec =>  $10^{27}$  years (age of Universe:  $10^{10}$  years)

And yet protein do fold spontaneously in a matter of milliseconds.

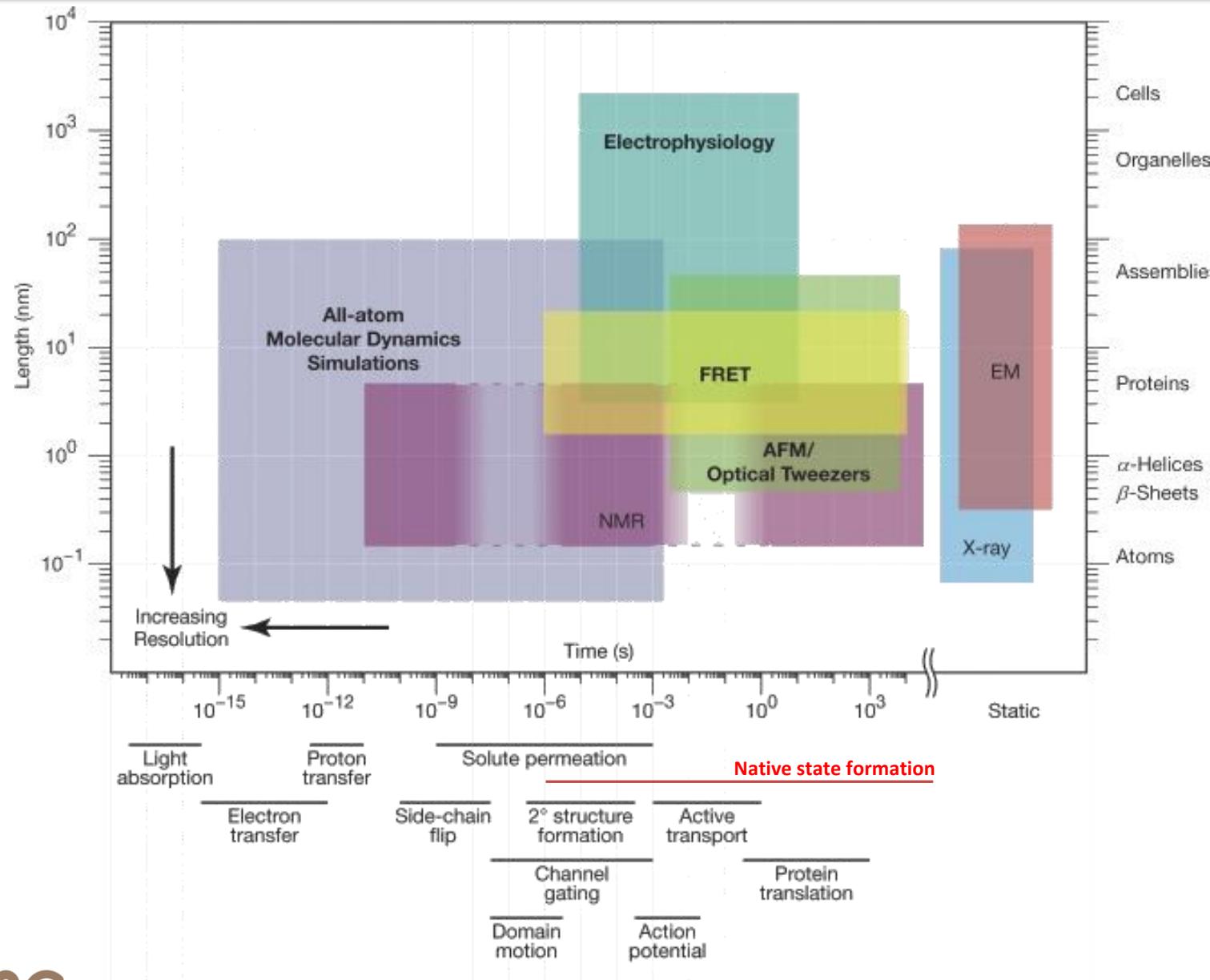
How can it be ?

Levinthal's paradox



Hydrophobic-polar (HP) 2D-lattice model

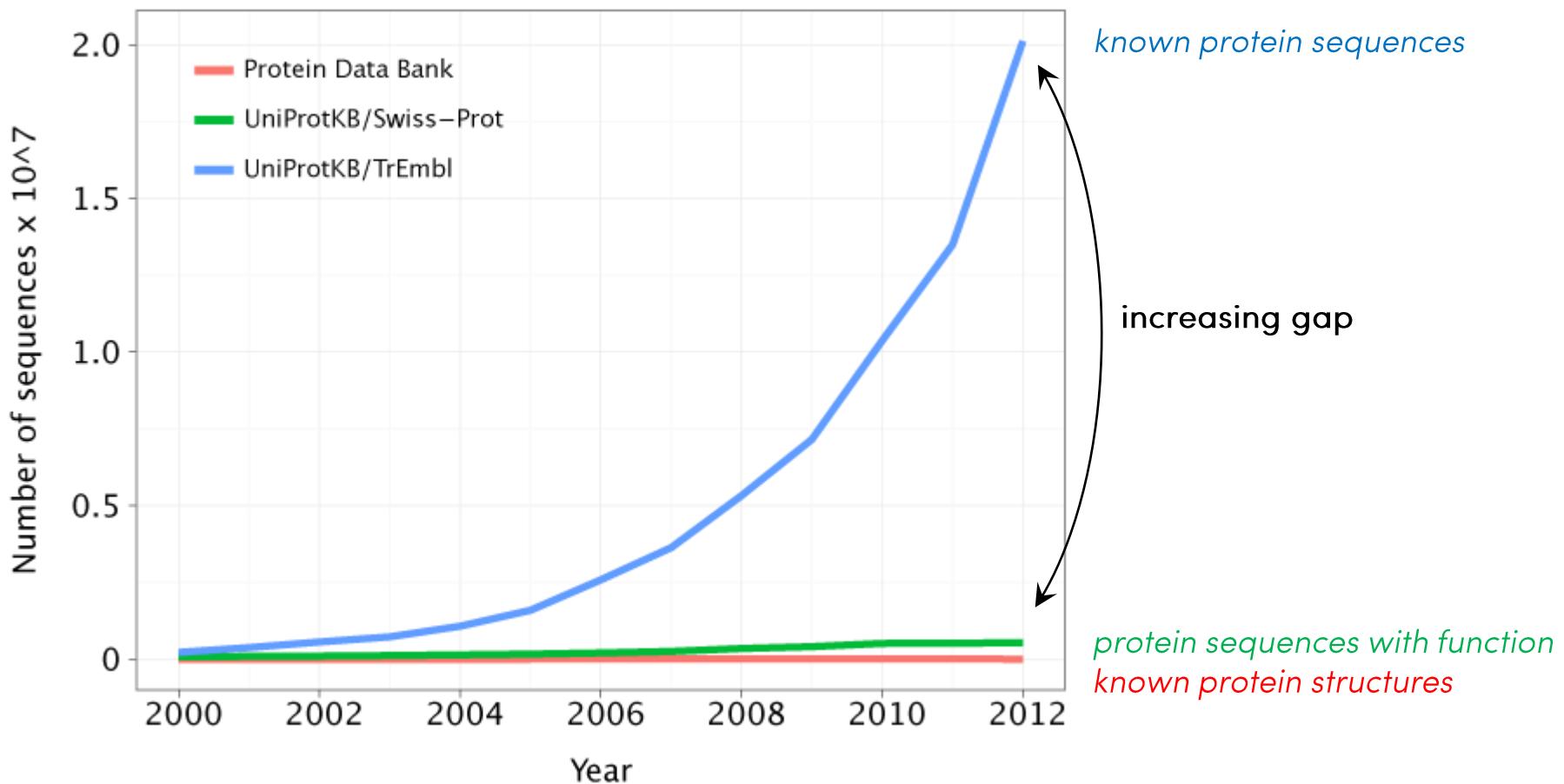
# Protein dynamics spatio-temporal scales



# Algorithms in structural bioinformatics: what for ?

## ➤ To predict protein structures

- experimental data analysis and 3-dimensional model building
- secondary or tertiary structure prediction based on the sequence

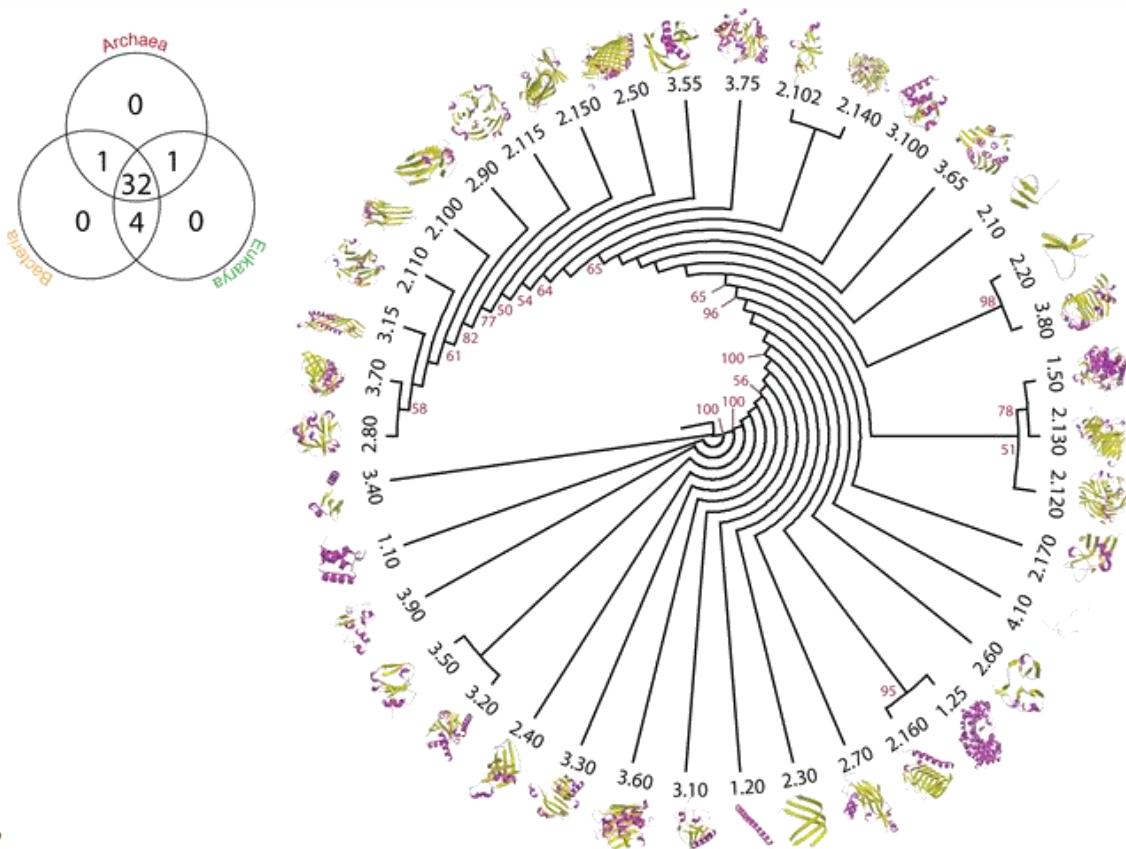


# Algorithms in structural bioinformatics: what for ?

➤ To predict protein structures

➤ To compare protein structures

- classification of proteins (divergent/convergent evolution)
  - identification of active sites, functional motifs or binding sites



# Phylogenetic tree of 38 CATH Architecture domain structures

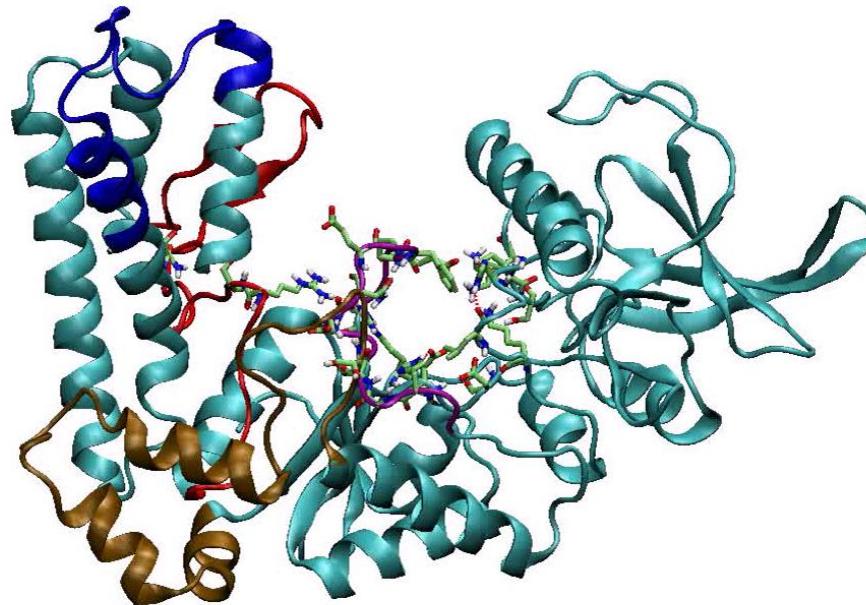
# Algorithms in structural bioinformatics: what for ?

➤ To predict protein structures

➤ To compare protein structures

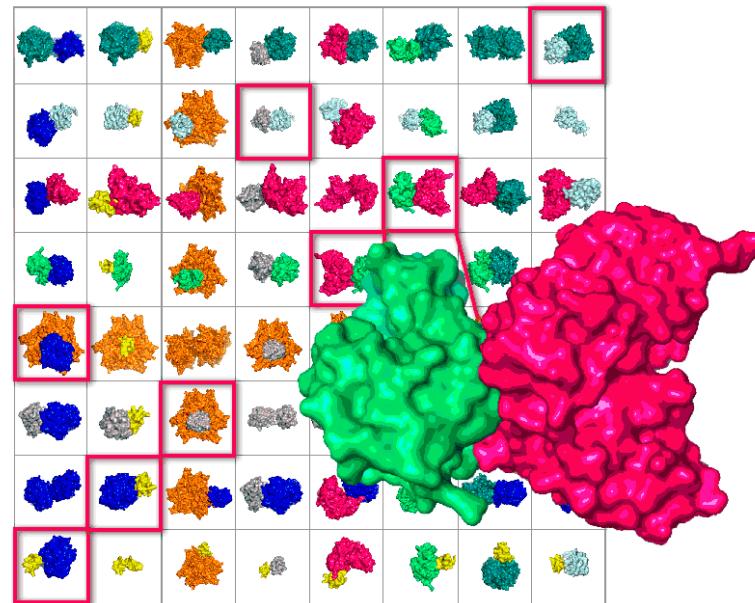
➤ To simulate protein motions

- atomic-level description of the mechanisms underlying protein activity
- characterization of intermediate conformations that can be targeted by drugs



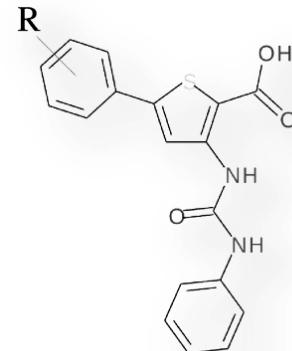
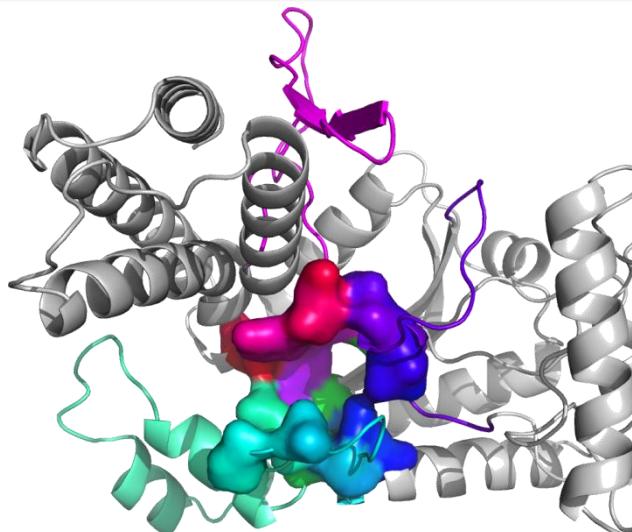
# Algorithms in structural bioinformatics: what for ?

- To predict protein structures
- To compare protein structures
- To simulate protein motions
- To characterize protein interactions
  - protein interaction sites identification and complex structures prediction
  - discrimination between true partners in the cell and non-interactors



# Algorithms in structural bioinformatics: what for ?

- To predict protein structures
- To compare protein structures
- To simulate protein motions
- To characterize protein interactions
- To discover and design drugs
  - putative druggable pockets identification
  - binding mode and relative affinity prediction



# Conclusion

- **Proteins** are composed of amino acid residues linked by a **peptidic bond**. There exist **20 types of amino acids** with different physico-chemical properties.
- A protein is a **polypeptidic chain**. Four **levels of organization** determine the 3D atomic coordinates of a protein structure.
- Proteins fulfill various **biological functions**: structural, enzymatic, of transport... Some proteins can have multiple functions.
- Proteins are **dynamic and flexible objects** which adapt their shape in response to environmental conditions.
- **Algorithm in structural bioinformatics** can help predict and classify protein structures, describe their motions and interactions, design new drugs.