Session 0,5:

Principles of protein structure and conformational space

From last practical

You can try the exercises to practice for the practical exams

QUESTIONS FROM THE TUTORIAL

Now we can compare all the results and answer the following questions:

1) Why are the e-values different in *target_pdb.out* than in the fifth iteration in *target_pdb_5.out*?

2) Why do we need to run psiblast with uniprot_sprot.fasta before searching in pdb_seq?

3) When obtaining the file *target_pdb_sprot5.out* why we didn't run 5 iterations as before?

4) Search in the SCOP database with the PDB code of the best match of the target sequence. Do all the files *target_pdb_specific.out, target_pdb_sprot5.out, target_pdb_5.out* and *target_pdb.out* produce the same result?

5) Can you use the file target_sprot5.out to obtain the name of the fold in SCOP? Why?

6) What are the folds of the following sequences?

   a. *problem1/serc_myctu.fa*

   b. *problem2/p72_mycmy.fa*

   c. *problem3/lip_staau.fa*

   d. *problem4/orc1_human.fa*

From last practical

How can I know the fold of one protein?

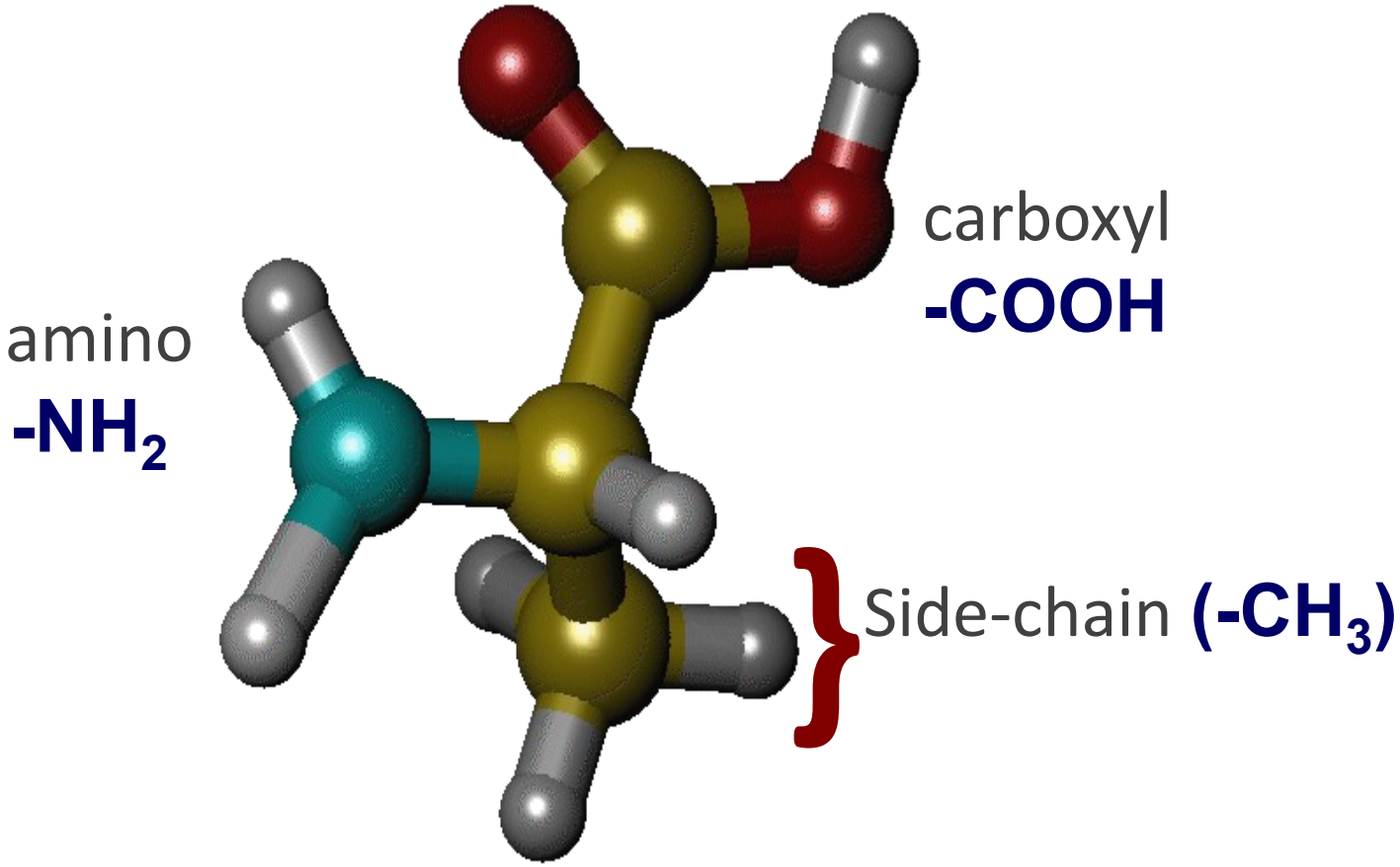
Using the SCOP database

How to use the SCOP database?

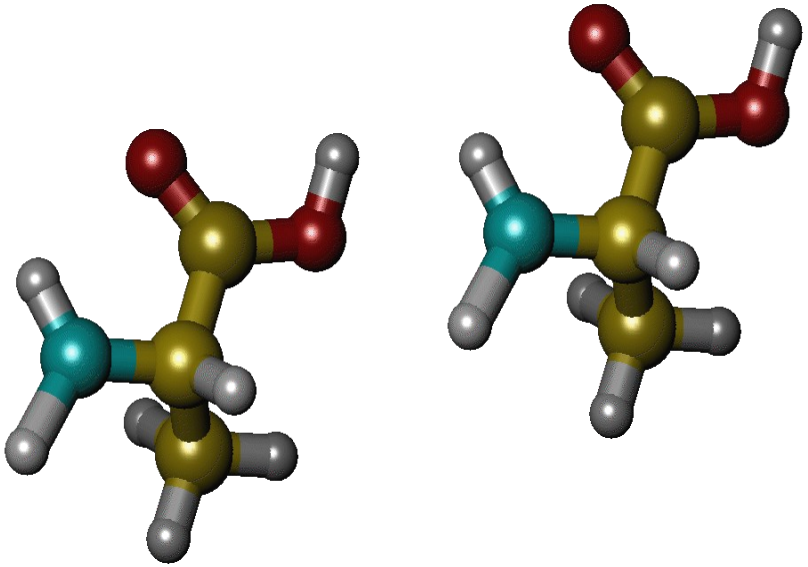
https://scop.mrc-lmb.cam.ac.uk/

Amino acids

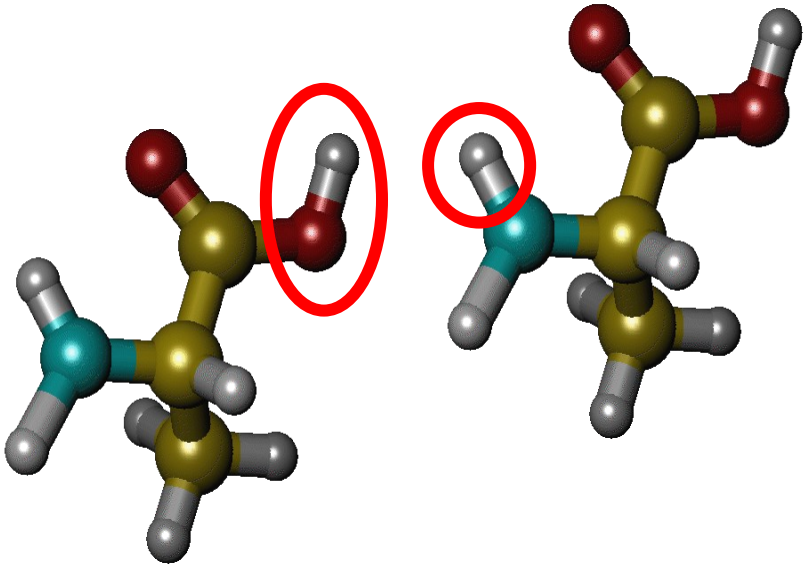
Amino acids are made by a carbon atom bound to a carboxyl group (COOH), an amino group (NH2), a hydrogen atom and a variable side chain



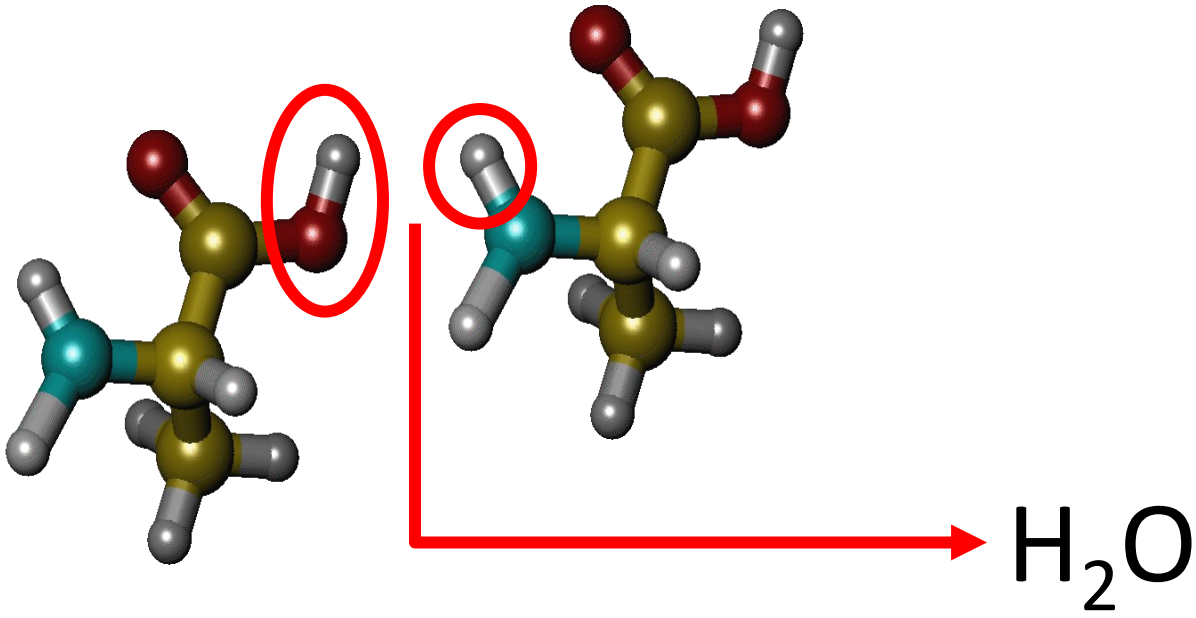
Peptídic bond



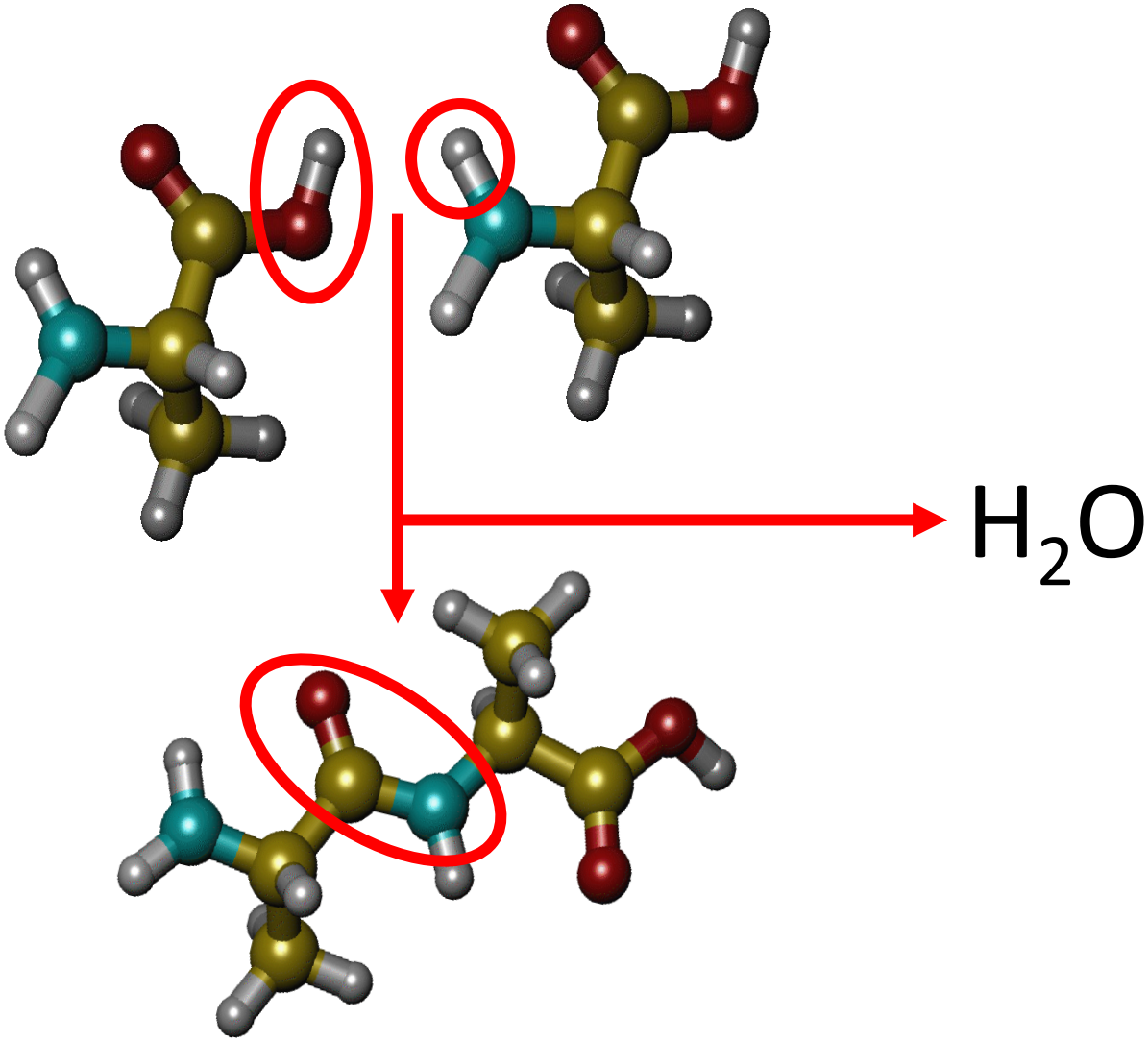
Peptídic bond



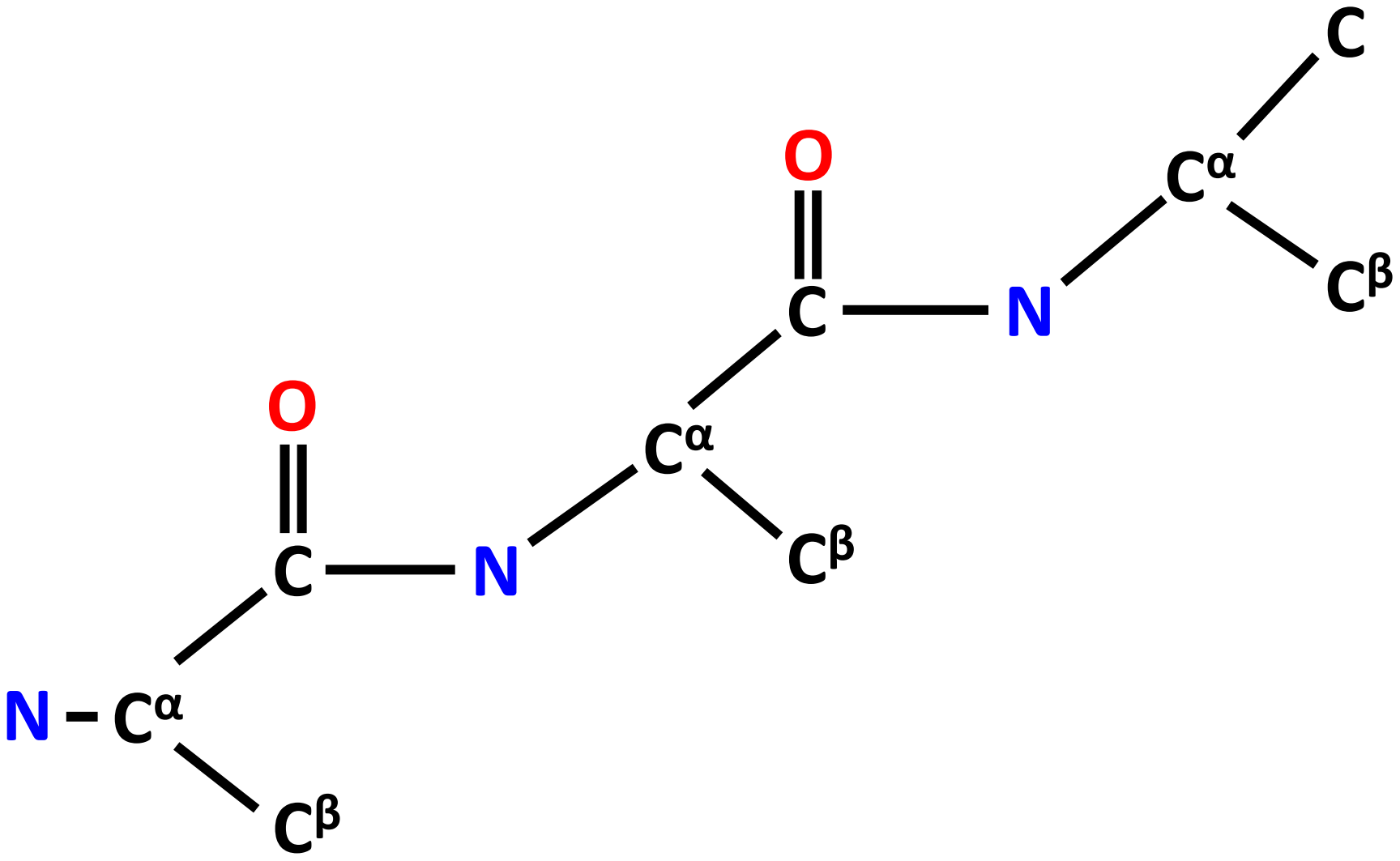
Peptídic bond



Peptídic bond

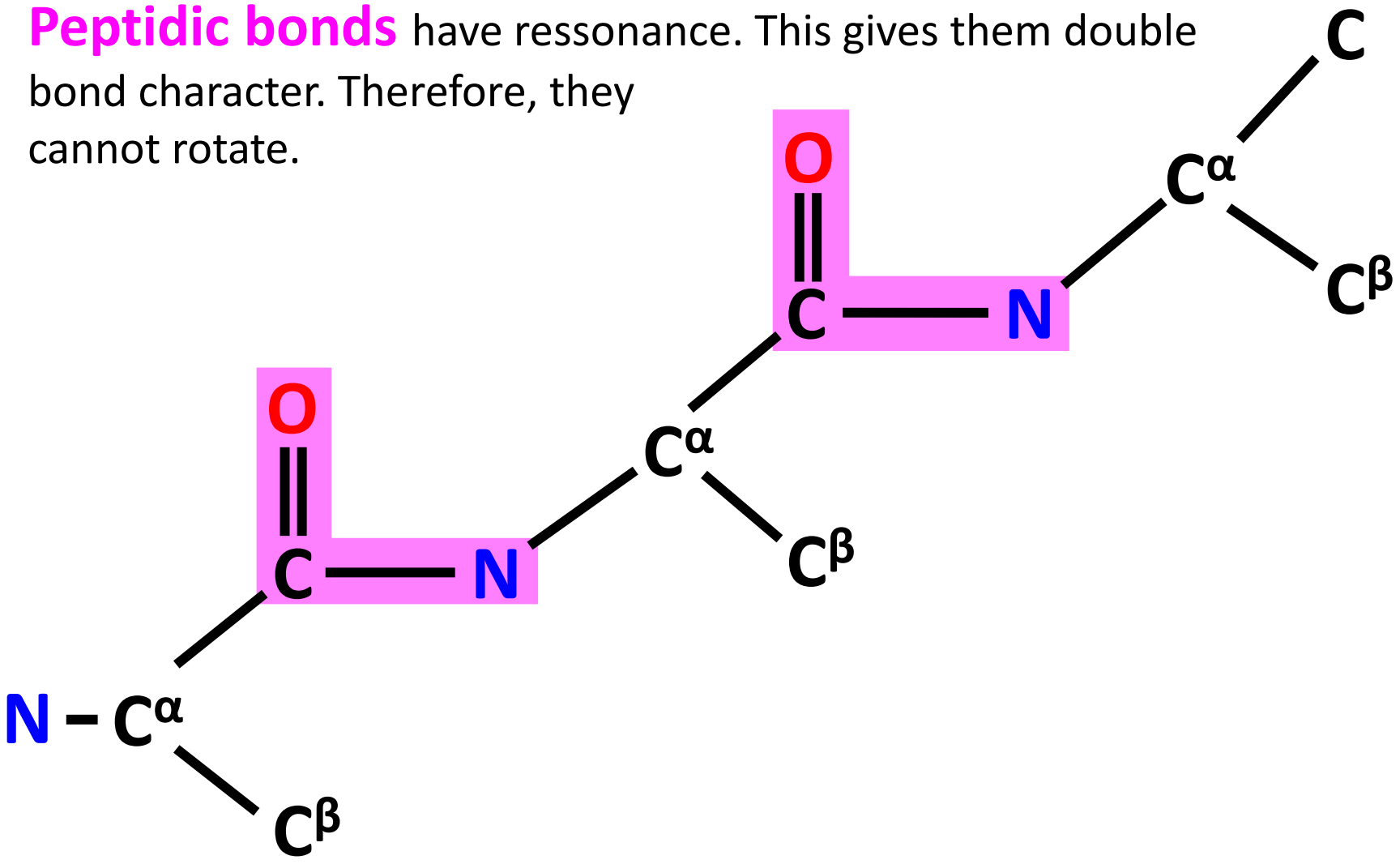


Understanding protein geometry



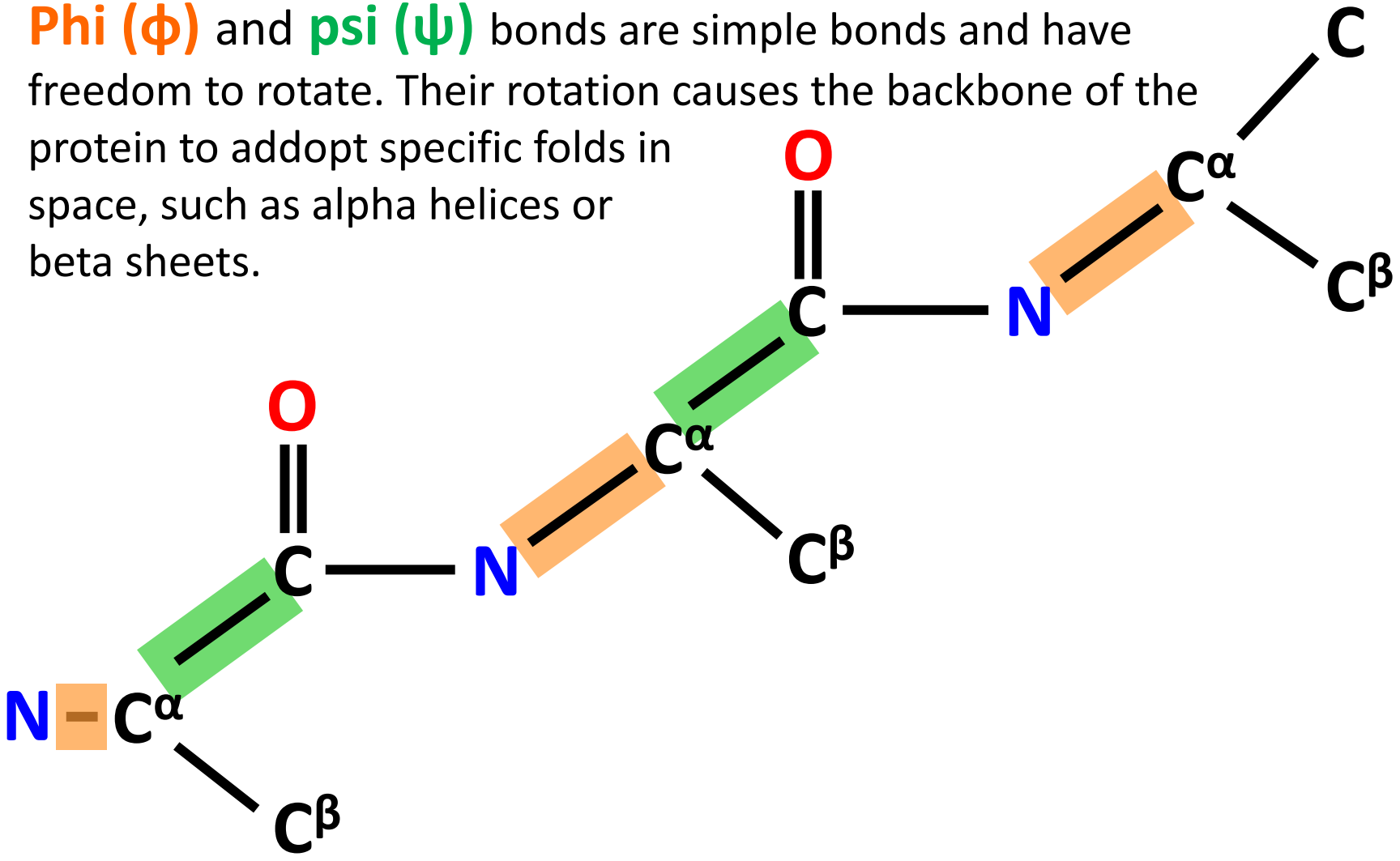
Understanding protein geometry

Peptidic bonds have ressonance. This gives them double bond character. Therefore, they cannot rotate.



Understanding protein geometry

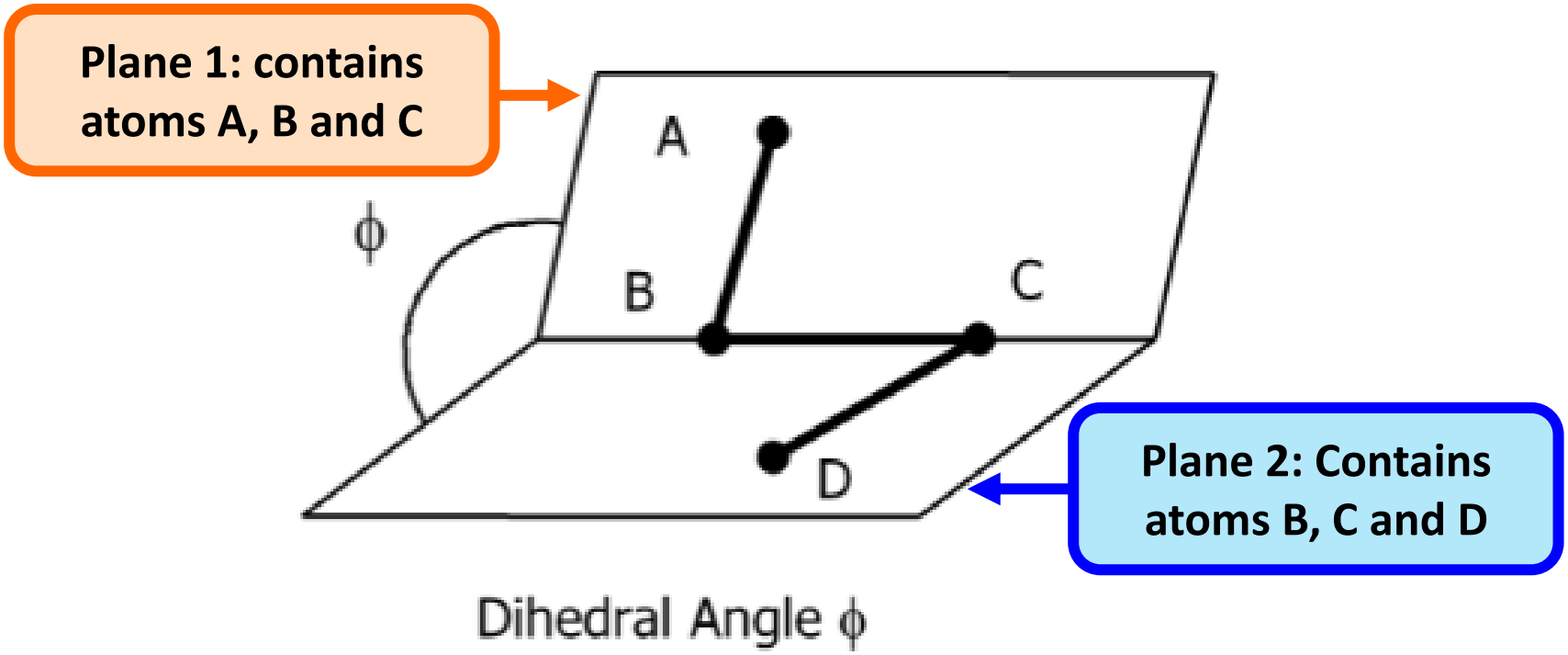
Phi (ϕ) and **psi (ψ)** bonds are simple bonds and have freedom to rotate. Their rotation causes the backbone of the protein to addopt specific folds in space, such as alpha helices or beta sheets.



Understanding protein geometry

Remember biochemistry, dihedrals are the angle created by two planes

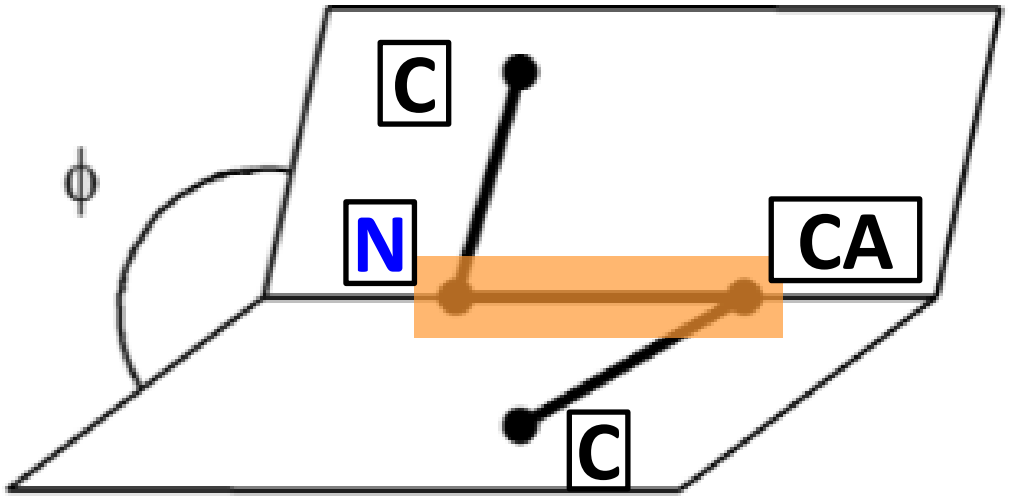
To calculate the dihedral of a bond you select the atoms involved in the bond (B and C) and one atom before and another atom after (A and D)



Understanding protein geometry

Remember biochemistry, dihedrals are the angle created by two planes

For the **Phi (ϕ)** angle, the involved atoms are the carbonil carbon and the nitrogen of one amino acid; and the alpha carbon and carbonil carbon of the next amino acid

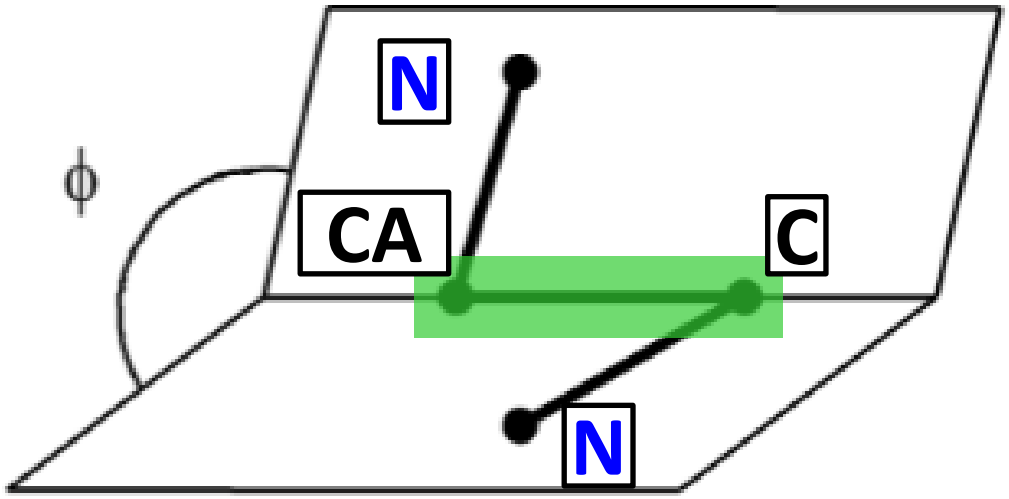


Dihedral Angle ϕ

Understanding protein geometry

Remember biochemistry, dihedrals are the angle created by two planes

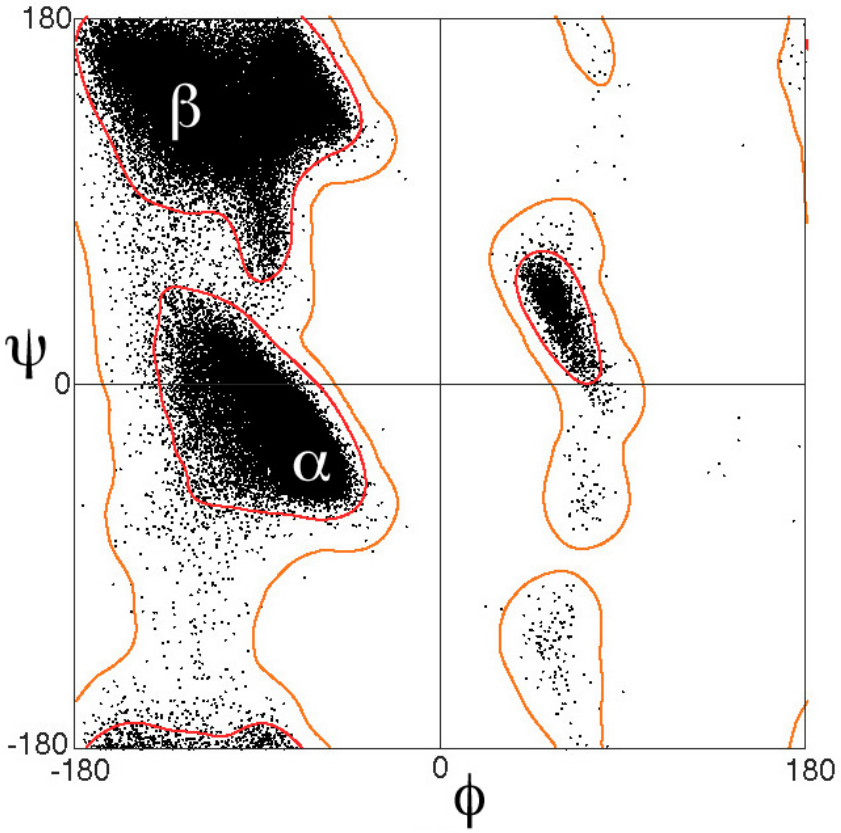
For the **Psi (ψ)** angle, the involved atoms are the the nitrogen of one amino acid; and the alpha carbon, carbonil carbon and nitrogen of the next amino acid



Dihedral Angle ϕ

The Ramachandran plot

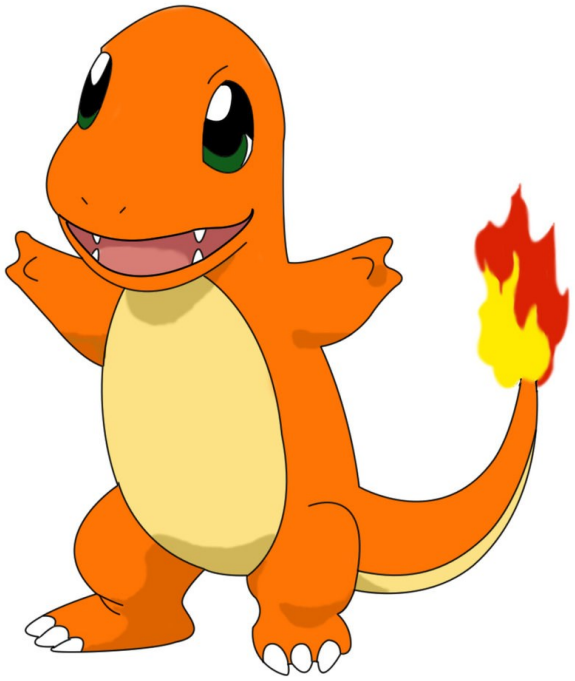
The Ramachandran plot is a representation of the phi (ϕ) and psi (ψ) dihedrals in a protein.

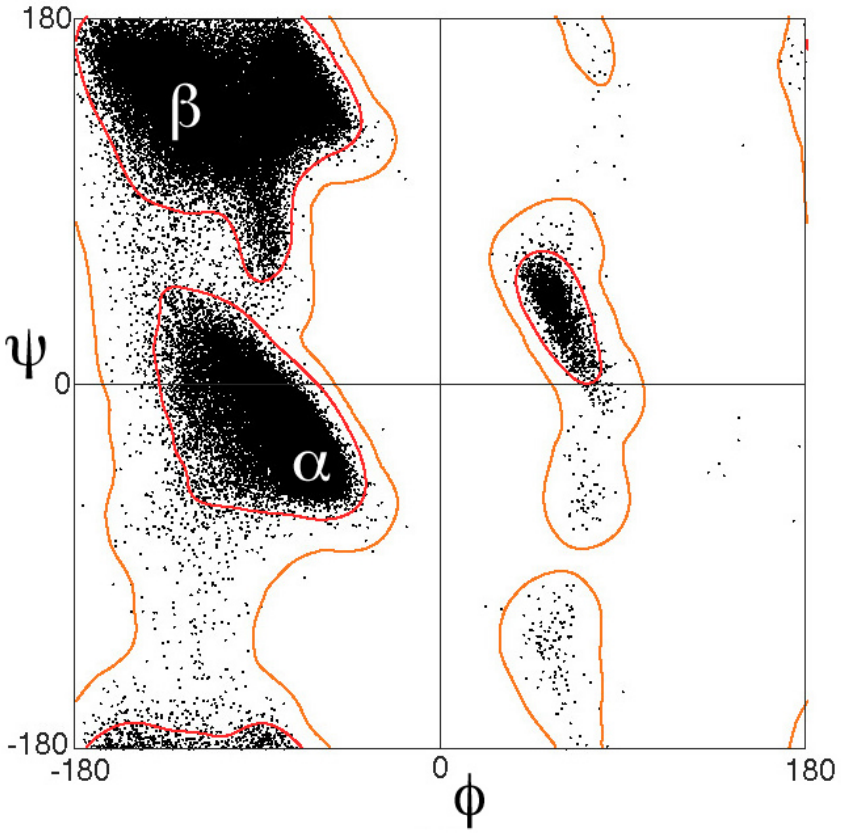


The Ramachandran plot

The Ramachandran plot is a representation of the phi (ϕ) and psi (ψ) dihedrals in a protein.

Do not confound Ramachandran with charmander (it happened to a friend!)

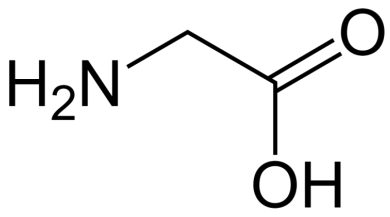


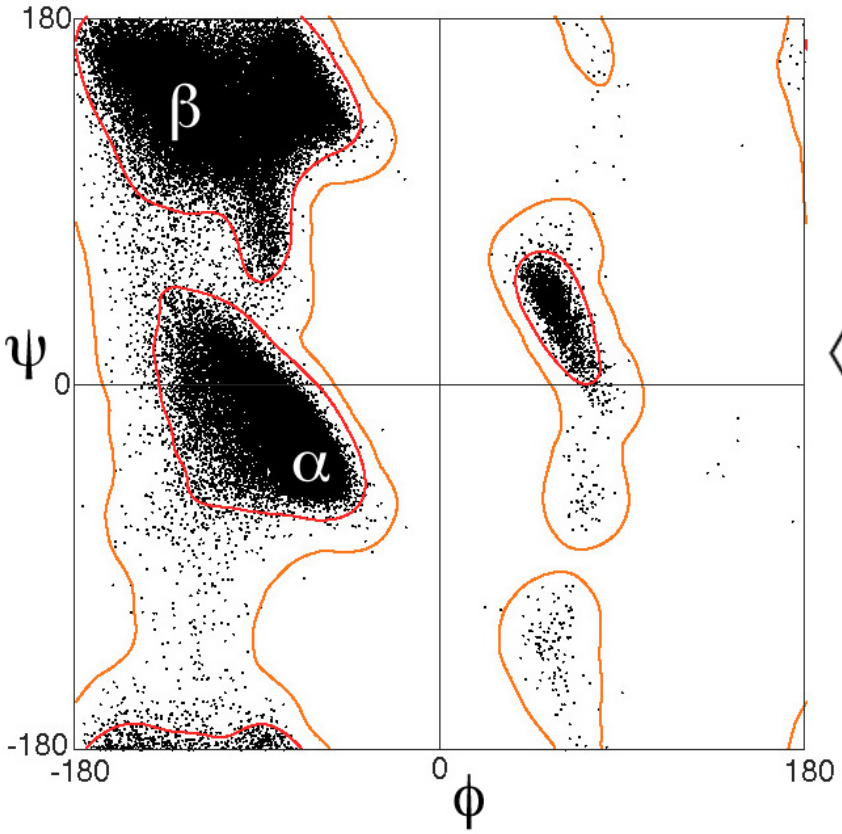


The Ramachandran plot

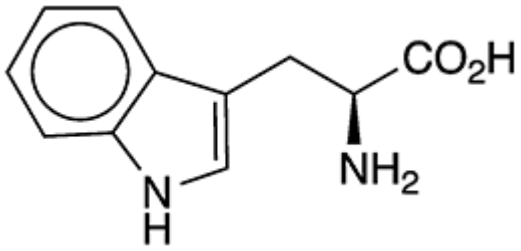
How would the Ramachandran plot be for a protein that only has glycines? And for a protein that only has tryptophans?

Glycine





Tryptophan



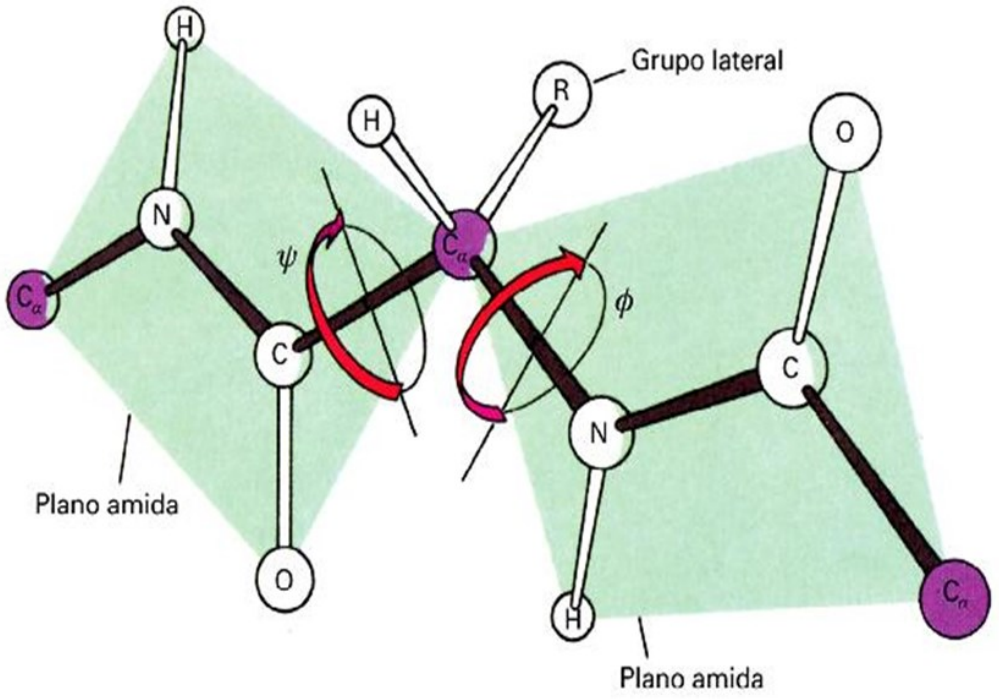
Conformational space

The conformational space are all the possible conformations that a peptide can undergo

Conformational space is limited by the following aspects:

Peptidic bonds must always remain in the same plain

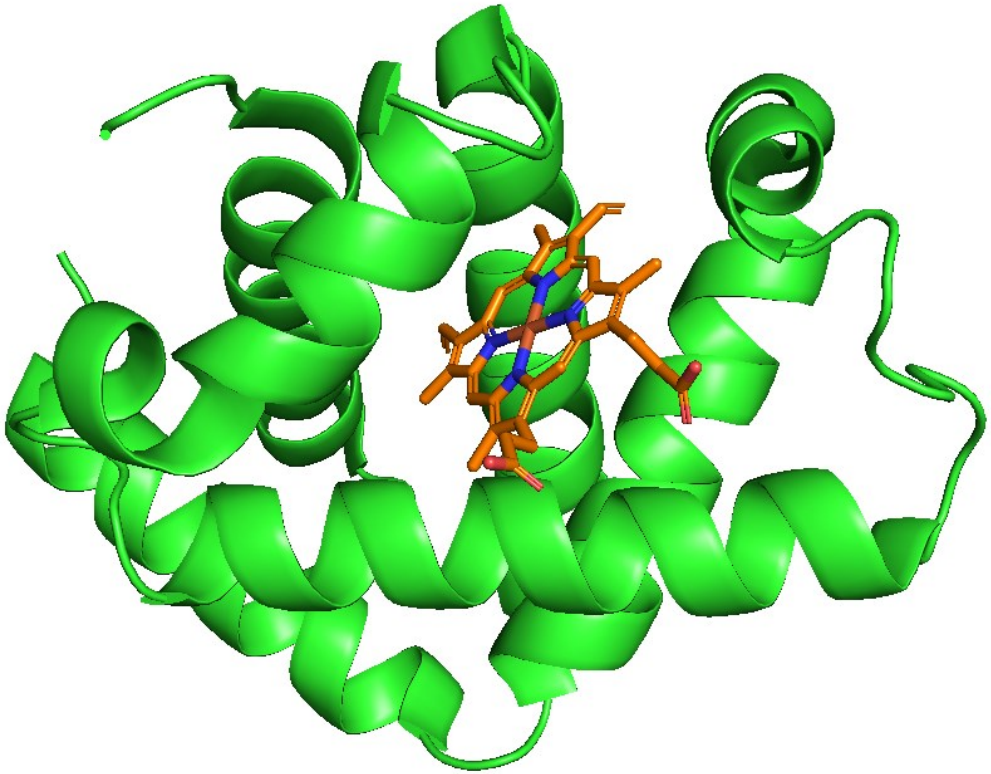
Side chains impair the full rotation of phi (ϕ) and psi (ψ) angles



Native conformation

The native conformation is the conformation at which one protein is stable and can carry out its function

This is an hemoglobin



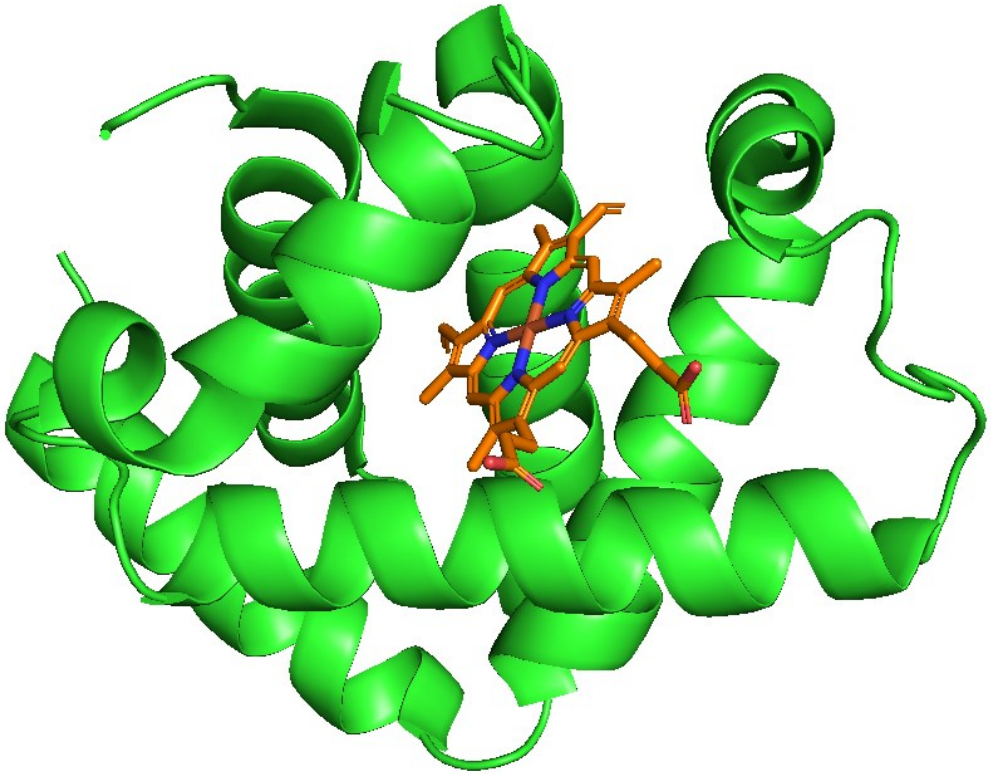
Native conformation

The native conformation is the conformation at which one protein is stable and can carry out its function

This is an hemoglobin

↓

Its function is to transport oxygen



Native conformation

The native conformation is the conformation at which one protein is stable and can carry out its function

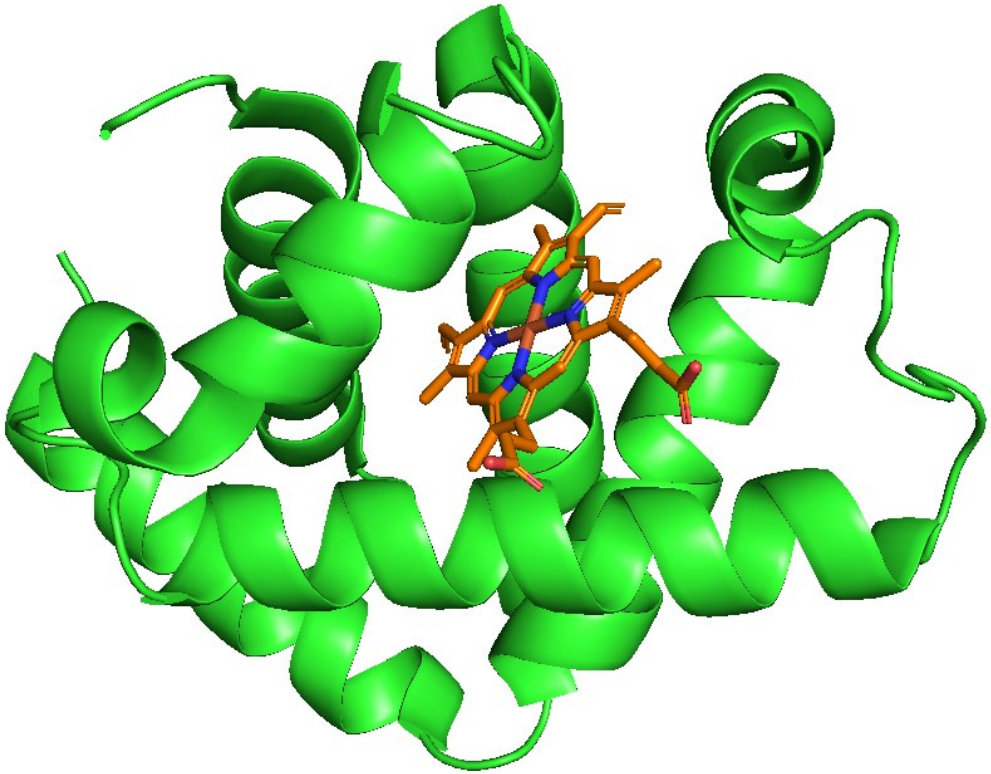
This is an hemoglobin

↓

Its function is to transport oxygen

↓

Oxygen is bound to the hemo group (orange)



Native conformation

The native conformation is the conformation at which one protein is stable and can carry out its function

This is an hemoglobin

↓

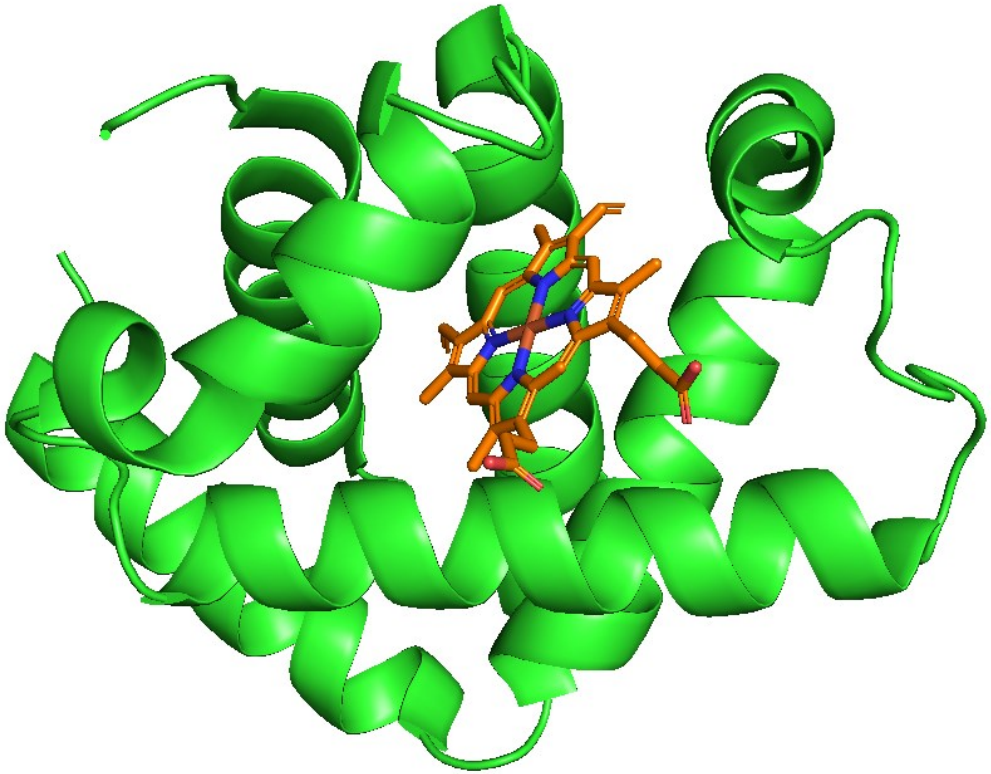
Its function is to transport oxygen

↓

Oxygen is bound to the hemo group (orange)

↓

The native conformation is the one that allows the proper interaction with the hemo group

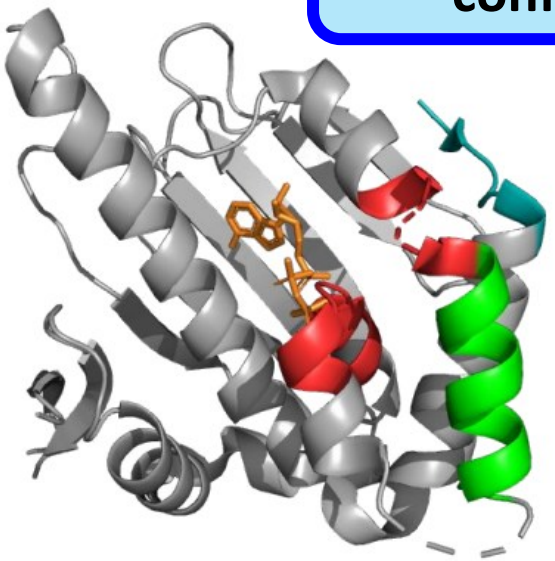


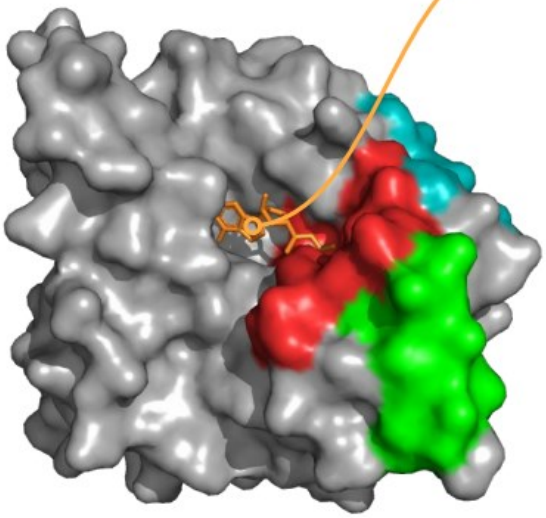
Native conformation

The native conformation is the conformation at which one protein is stable and can carry out its function

Some proteins regulate their function by having active and inactive conformations. In these cases, both conformations are native conformations.

Chaperone in inactive conformation





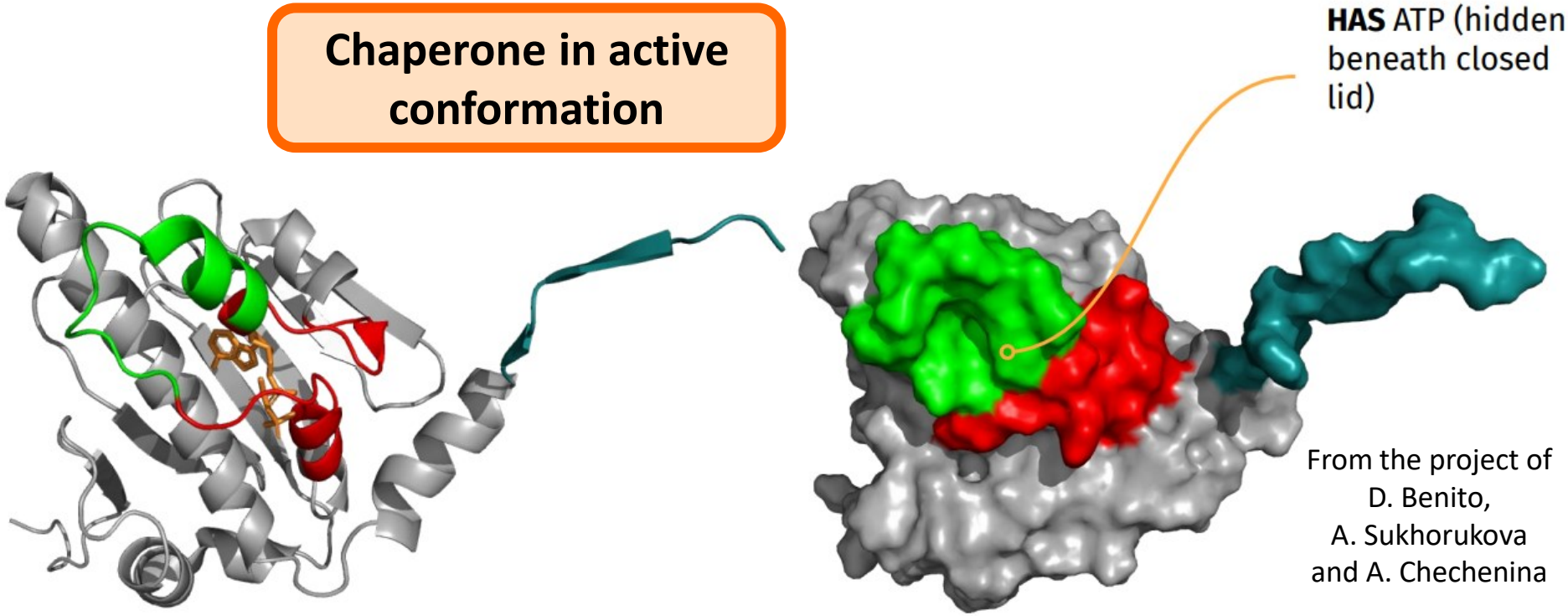
From the project of D. Benito, A. Sukhorukova and A. Chechenina

Native conformation

The native conformation is the conformation at which one protein is stable and can carry out its function

Some proteins regulate their function by having active and inactive conformations. In these cases, both conformations are native conformations.

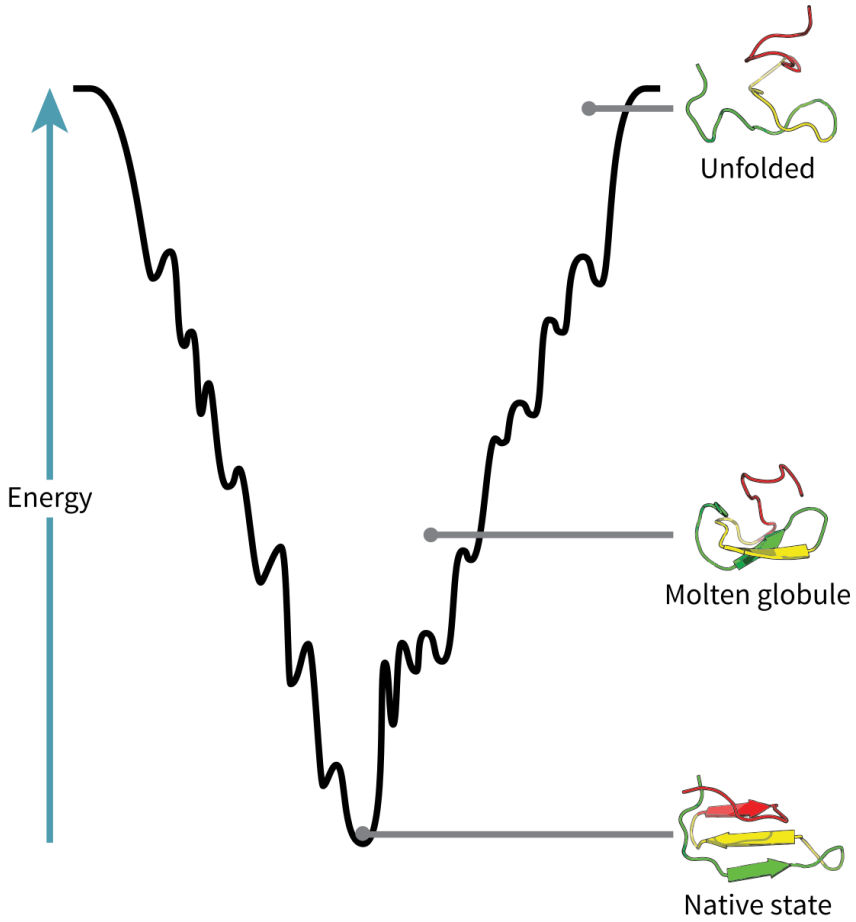
Chaperone in active conformation



From the project of D. Benito, A. Sukhorukova and A. Chechenina

Native conformation

The native conformation usually correspond with minimums within the energy landscape

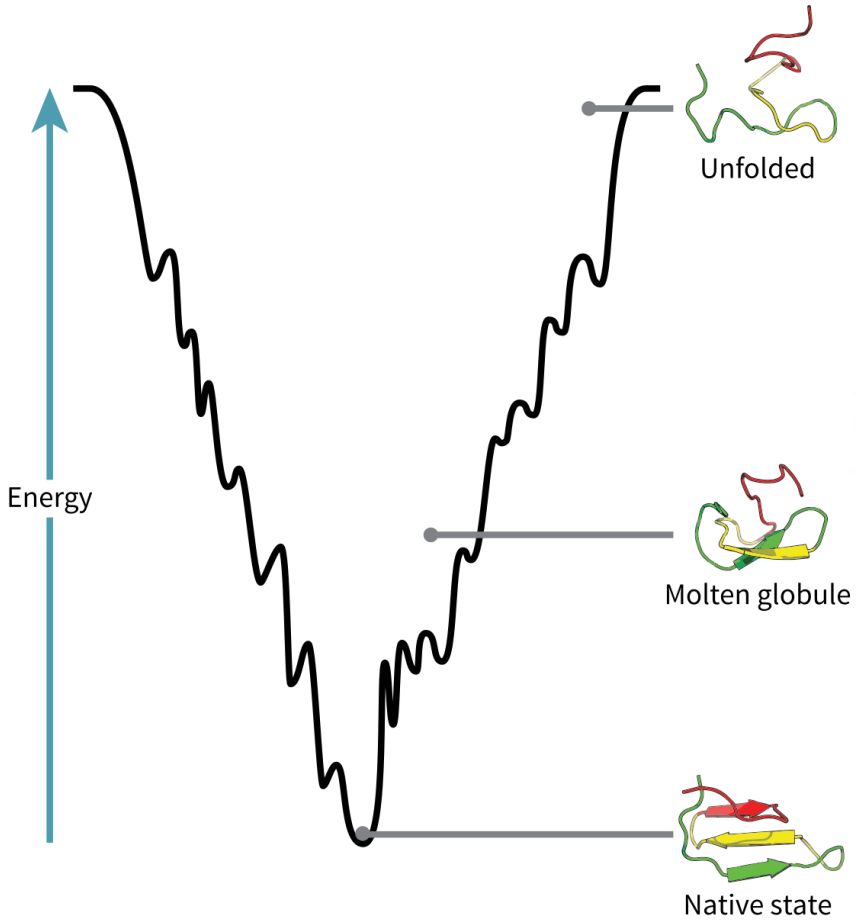


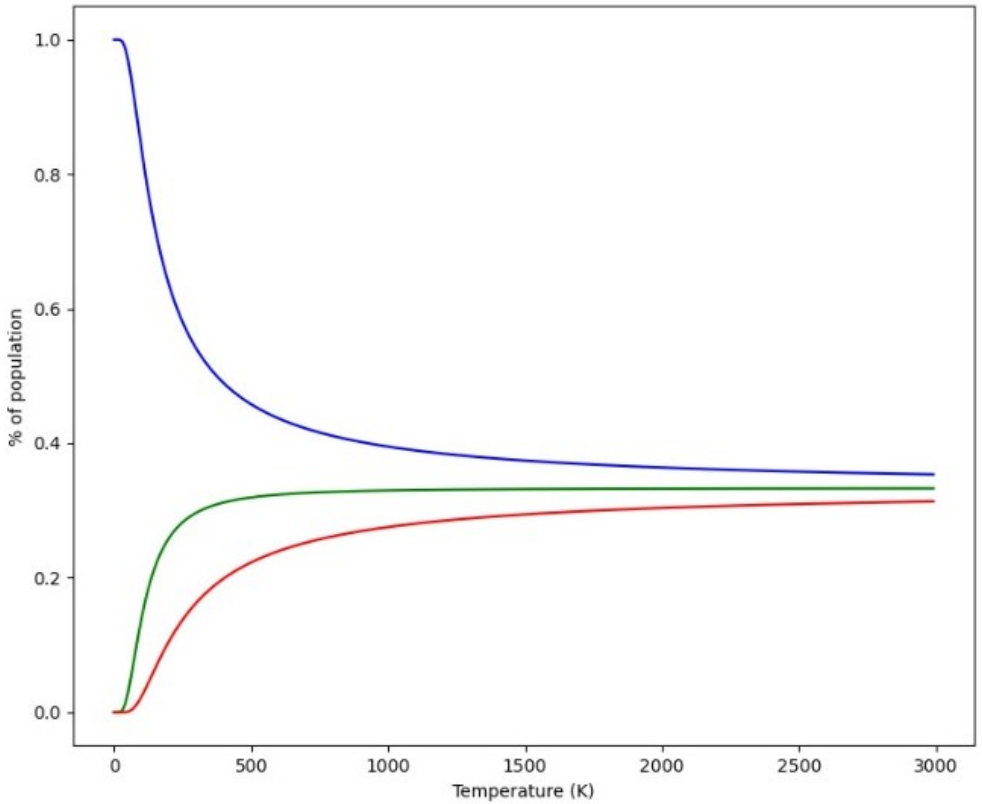
The X-axis represents the conformational space of the protein

You can relate this with the Boltmann distribution we saw in biophysics

Native conformation

The native conformation usually correspond with minimums within the energy landscape



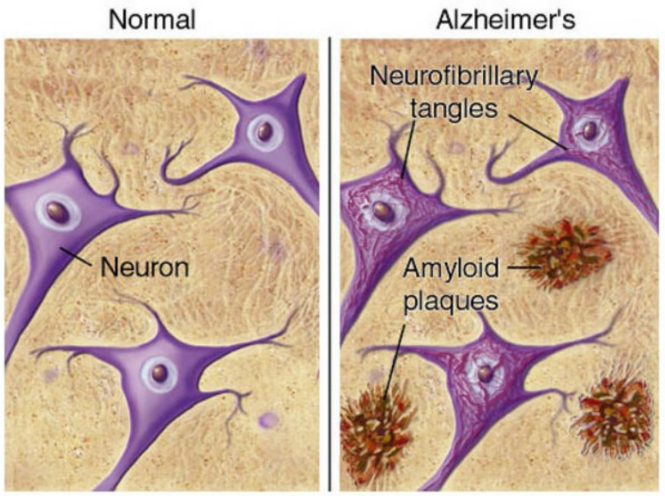


Native conformation

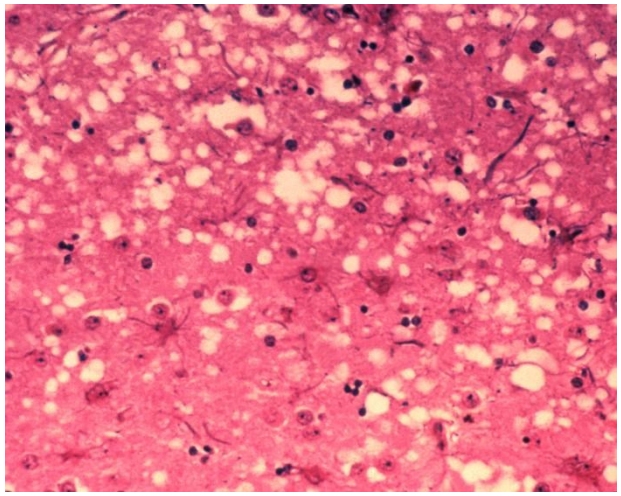
Do not confound native conformations with pathologically superstable conformations

These are proteins that are at minimums in the energy landscape, but they are not functional and can be the couse of several diseases such as:

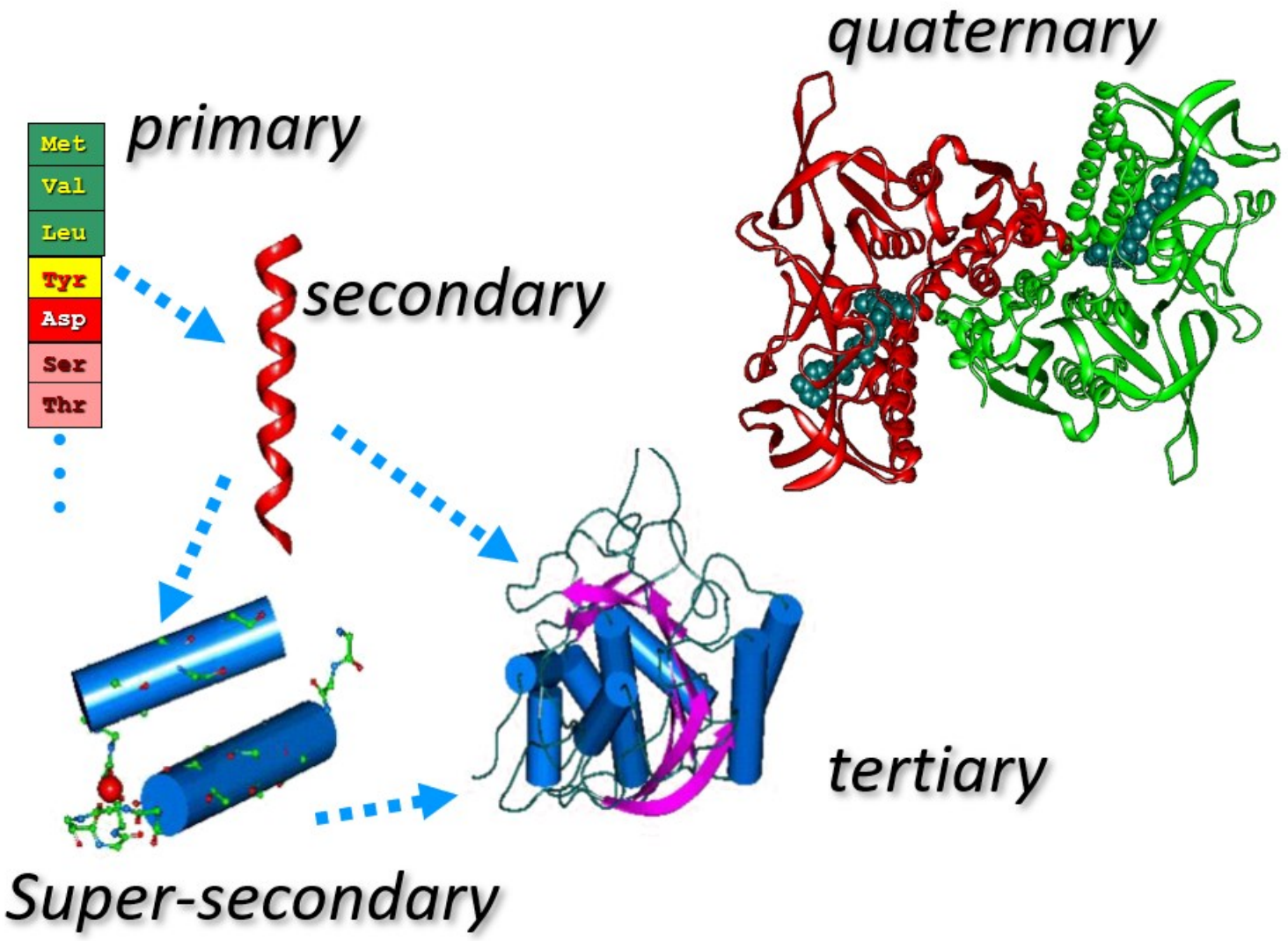
Alzheimer's disease
Beta amyloid accumulates creating extracelular plaques



Bovine spongiform encephalopathy (AKA mad cow disease)

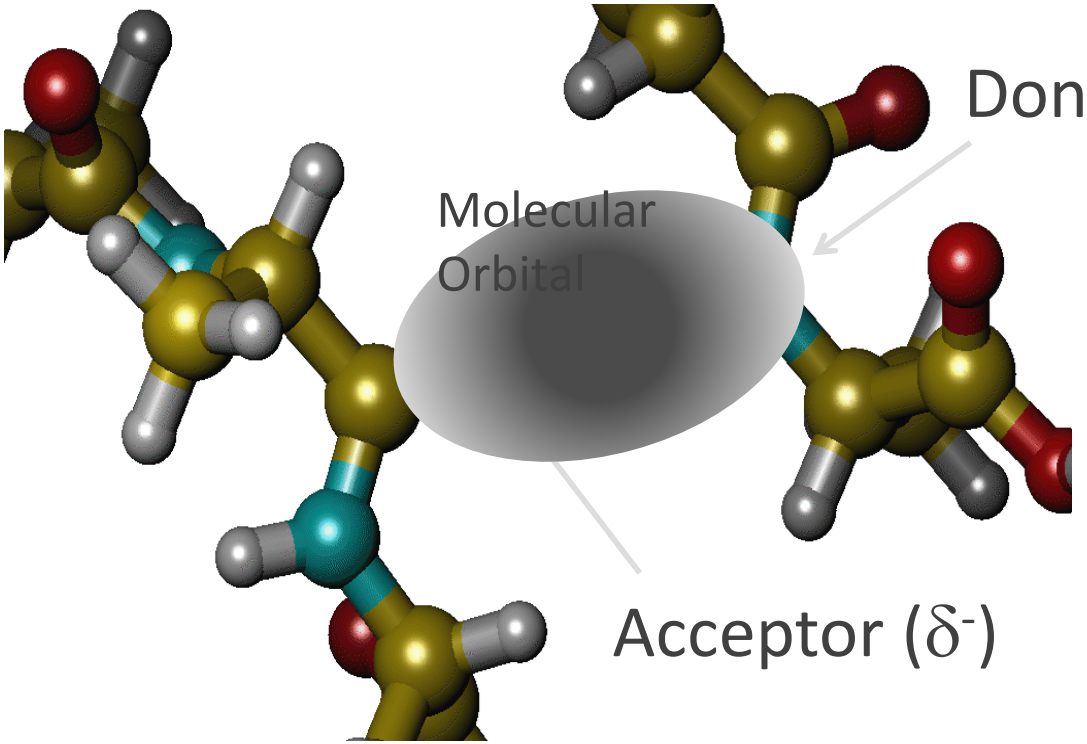


Levels of structure



Hydrogen bond

Hydrogen bonds are the main force contributing to the creation of secondary structures

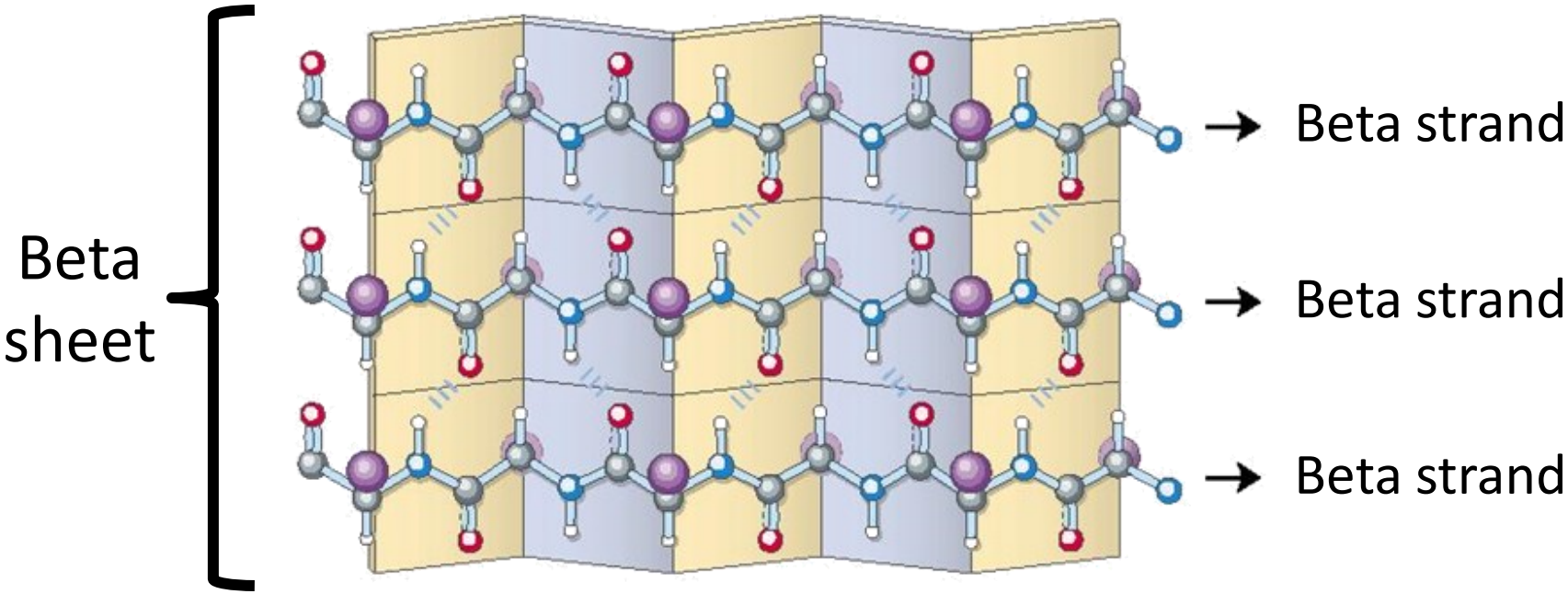


85% ionic bond
15% covalent
Approx. 4 kcal/mol

Restriction:
Angle OHN = 180º+/-35º
Distance ON = 2.8A

Secondary structures: Beta sheets

Beta sheets involve that the hydrogen bonds are made between distant amino acids within the sequence

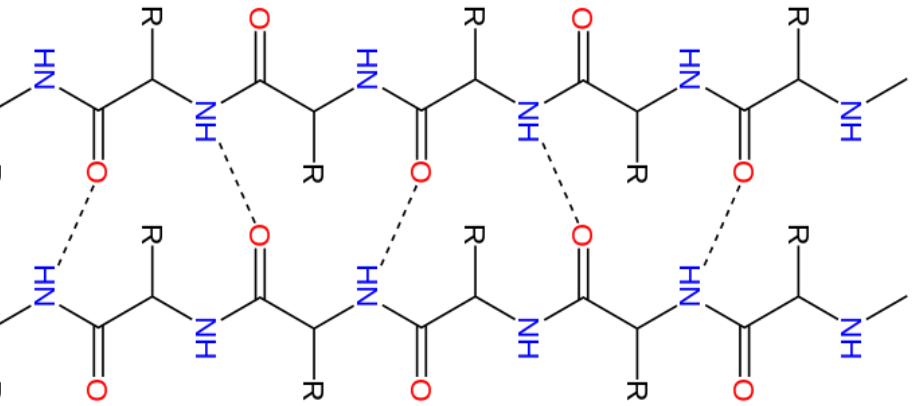


Secondary structures: Beta sheets

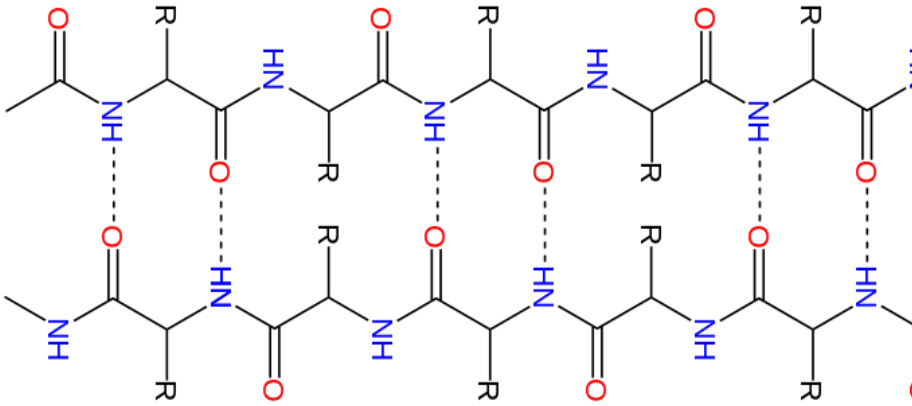
Beta sheets involve that the hydrogen bonds are made between distant amino acids within the sequence

Depending on the direction of the beta strands, they can be parallel or antiparallel, which involves a different arangement of hydrogen bonds

Parallel beta sheet



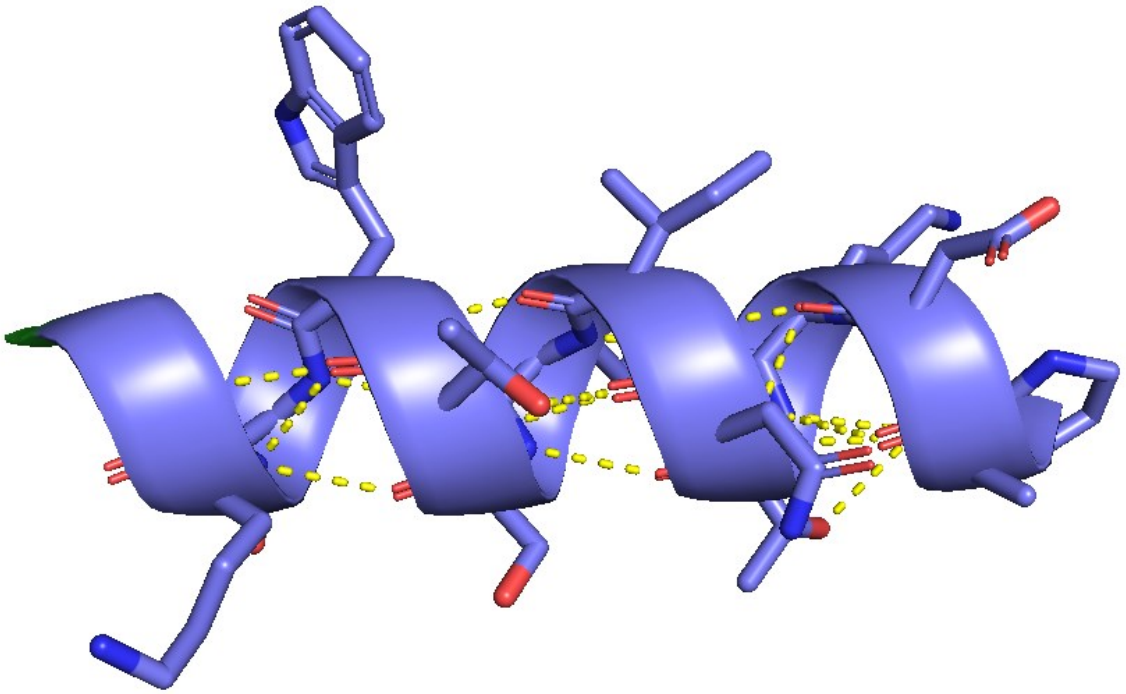
Anti-parallel beta sheet



Secondary structures: Helices

In helices, hydrogen bonds are made between amino acids that are close in the sequence

Helices are cilindric structures, with the side chains of the amino acids pointing outside the cilinder



Secondary structures: Helices

You can have different types of helices depending on the distance between amino acids that make hydrogen bonds

Helix 3_{10}

- Hydrogen bonds are made every 3 Aa
- One turn is made every 3 Aa
- Thinner helix the alpha helix
- Not very common

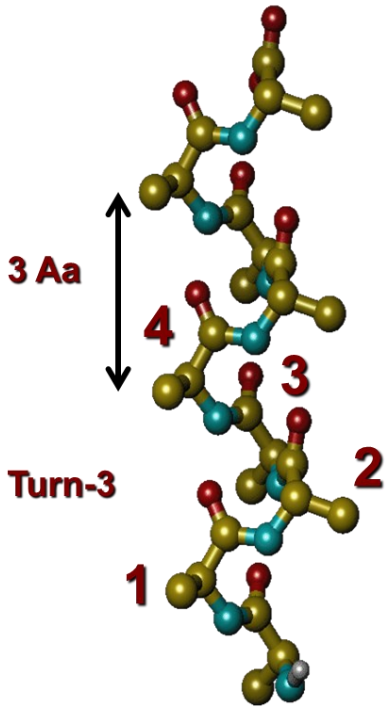
Alpha helix

- Hydrogen bonds are made every 4 Aa
- One turn is made every 3.6 Aa
- Thicker helix than the Helix 3_{10}
- Very common

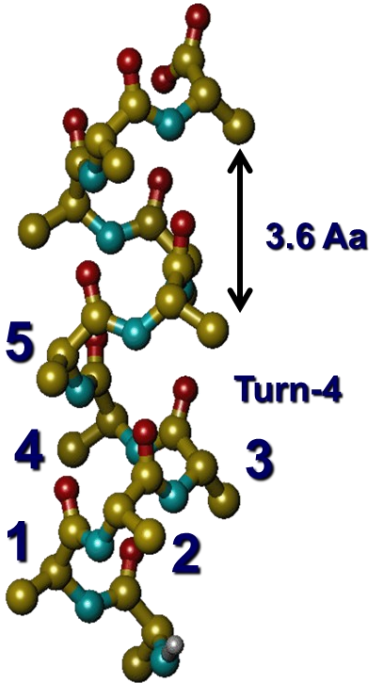
Secondary structures: Helices

You can have different types of helices depending on the distance between amino acids that make hydrogen bonds

Helix 3_{10}



Alpha helix

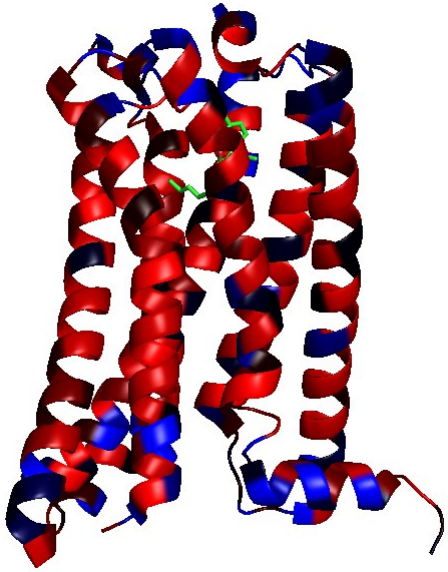


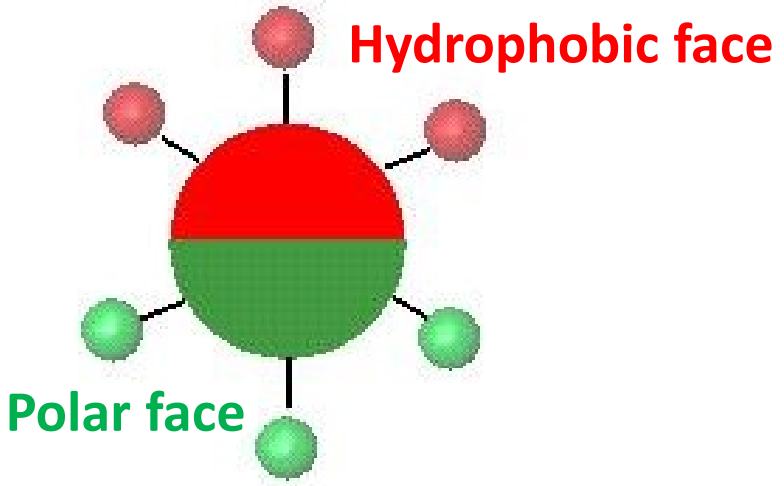
Secondary structures: Helices

By placing hydrophobic and polar amino acids we can obtain helices with clear different functions

Fully hydrophobic
For standard transmembrane domains

Amphypatic helices
For membrane transporters or hydrophobic cores





Amino acid propensities

Table 1
Frequency of amino acids and their propensities for the three secondary structure states in the PDBselect dataset

Amino acid	Frequency	P_α	P_β	P_c
A—Ala	7.73	1.39	0.75	0.8
C—Cys	1.84	0.74	1.31	1.05
D—Asp	5.82	0.89	0.55	1.33
E—Glu	6.61	1.35	0.72	0.86
F—Phe	4.05	1.01	1.43	0.76
G—Gly	7.11	0.47	0.65	1.62
H—His	2.35	0.92	0.99	1.07
I—Ile	5.66	1.04	1.71	0.59
K—Lys	6.27	1.11	0.83	1
L—Leu	8.83	1.32	1.1	0.68
M—Met	2.08	1.21	0.99	0.83
N—Asn	4.5	0.77	0.62	1.39
P—Pro	4.52	0.5	0.44	1.72
Q—Gln	3.94	1.29	0.76	0.89
R—Arg	5.03	1.17	0.91	0.91
S—Ser	6.13	0.82	0.85	1.24
T—Thr	5.53	0.76	1.23	1.07
V—Val	6.91	0.89	1.86	0.64
W—Trp	1.51	1.06	1.3	0.79
Y—Tyr	3.54	0.95	1.5	0.78

S. Costantini et al, 2006

Amino acid propensities

Table 1

Can you think of two amino acids that will be overrepresented in loops?

	1.5	0.79
0.95	1.5	0.78

S. Costantini et al, 2006

Amino acid propensities

Table 1
Frequency of amino acids and their propensities for the three secondary structure states in the PDBselect dataset

Amino acid	Frequency	P_α	P_β	P_c
A—Ala	7.73	1.39	0.75	0.8
C—Cys	1.84	0.74	1.31	1.05
D—Asp	5.82	0.89	0.55	1.33
E—Glu	6.61	1.35	0.72	0.86
F—Phe	4.05	1.01	1.43	0.76
G—Gly	7.11	0.47	0.65	1.62
H—His	2.35	0.92	0.99	1.07
I—Ile	5.66	1.04	1.71	0.59
K—Lys	6.27	1.11	0.83	1
L—Leu	8.83	1.32	1.1	0.68
M—Met	2.08	1.21	0.99	0.83
N—Asn	4.5	0.77	0.62	1.39
P—Pro	4.52	0.5	0.44	1.72
Q—Gln	3.94	1.29	0.76	0.89
R—Arg	5.03	1.17	0.91	0.91
S—Ser	6.13	0.82	0.85	1.24
T—Thr	5.53	0.76	1.23	1.07
V—Val	6.91	0.89	1.86	0.64
W—Trp	1.51	1.06	1.3	0.79
Y—Tyr	3.54	0.95	1.5	0.78

S. Costantini et al, 2006

Amino acid propensities

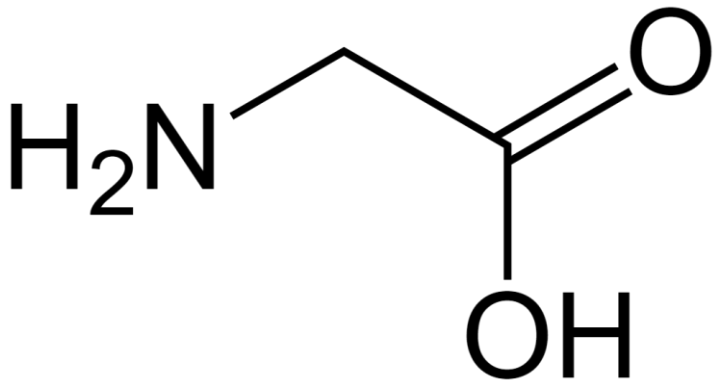
Table 1
Frequency of amino acids and their propensities for the three secondary structure states in the PDBselect dataset

Amino acid	Frequency	P_α	P_β	P_c
A—Ala	7.73	1.39	0.75	0.8
C—Cys	1.84	0.74	1.31	1.05
D—Asp	5.82	0.89	0.55	1.33
E—Glu	6.61	1.35	0.72	0.86
F—Phe	4.05	1.01	1.43	0.76
G—Gly	7.11	0.47	0.65	1.62
H—His	2.35	0.92	0.99	1.07
I—Ile	5.66	1.04	1.71	0.59
K—Lys	6.27	1.11	0.83	1
L—Leu	8.83	1.32	1.1	0.68
M—Met	2.08	1.21	0.99	0.83
N—Asn	4.5	0.77	0.62	1.39
P—Pro	4.52	0.5	0.44	1.72
Q—Gln	3.94	1.29	0.76	0.89
R—Arg	5.03	1.17	0.91	0.91
S—Ser	6.13	0.82	0.85	1.24
T—Thr	5.53	0.76	1.23	1.07
V—Val	6.91	0.89	1.86	0.64
W—Trp	1.51	1.06	1.3	0.79
Y—Tyr	3.54	0.95	1.5	0.78

S. Costantini et al, 2006

Glycine

Is the smallest amino acid. Its small side chain allows for a lot of conformational flexibility, which is the main characteristic of loops.



Amino acid propensities

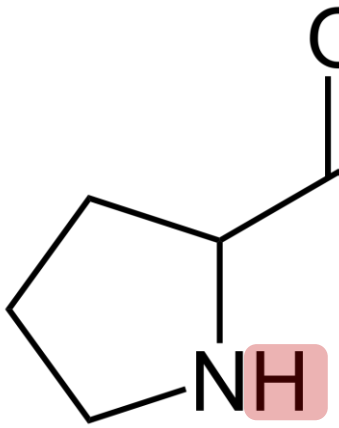
Table 1
Frequency of amino acids and their propensities for the three secondary structure states in the PDBselect dataset

Amino acid	Frequency	P_α	P_β	P_c
A—Ala	7.73	1.39	0.75	0.8
C—Cys	1.84	0.74	1.31	1.05
D—Asp	5.82	0.89	0.55	1.33
E—Glu	6.61	1.35	0.72	0.86
F—Phe	4.05	1.01	1.43	0.76
G—Gly	7.11	0.47	0.65	1.62
H—His	2.35	0.92	0.99	1.07
I—Ile	5.66	1.04	1.71	0.59
K—Lys	6.27	1.11	0.83	1
L—Leu	8.83	1.32	1.1	0.68
M—Met	2.08	1.21	0.99	0.83
N—Asn	4.5	0.77	0.62	1.39
P—Pro	4.52	0.5	0.44	1.72
Q—Gln	3.94	1.29	0.76	0.89
R—Arg	5.03	1.17	0.91	0.91
S—Ser	6.13	0.82	0.85	1.24
T—Thr	5.53	0.76	1.23	1.07
V—Val	6.91	0.89	1.86	0.64
W—Trp	1.51	1.06	1.3	0.79
Y—Tyr	3.54	0.95	1.5	0.78

S. Costantini et al, 2006

Proline

Is the only amino acid with the side chain bound to the amino group. This means that the amino group cannot be a hydrogen bond donnor. This breaks the pattern of hydrogen bonding of secondary structures.

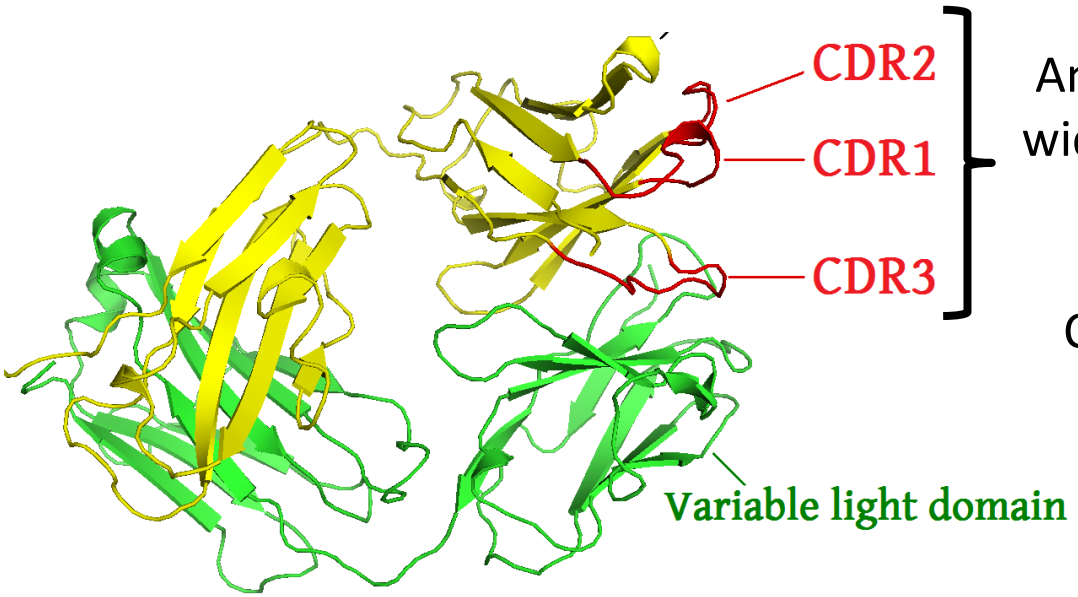


H lost in the peptidic bond

Loops

Loops are highly flexible structures without secondary structure. They connect secondary structures and allow their change in orientation.

Loops usually allow more mutations than other parts of the protein. This allows these regions to be more diverse in terms of sequence and structure.



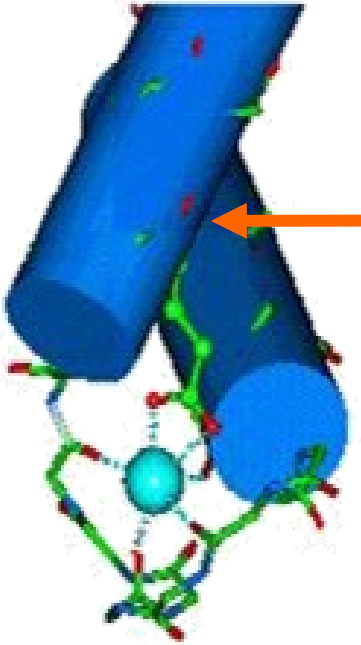
Antibodies use loops to recognize a wide range of antigens thanks to the variability that they provide.

CDR stands for Complementarity Determining Region

Supersecondary structures

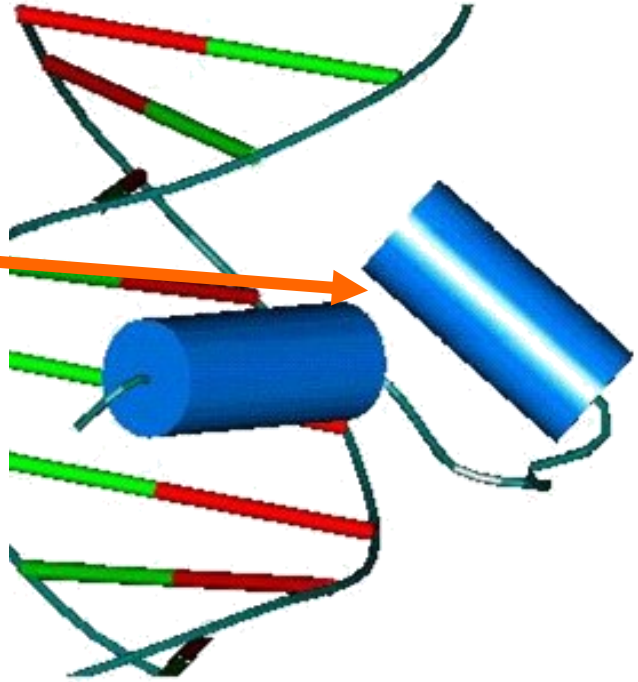
Cluster of 2 to 4 secondary structures united by loops usually involved in a particular function

Some supersecondary structures are stabilized by a hydrophobic core. Here we see examples of alpha-alpha supersecondary structures.



Hydrophobic core
That is the force keeping the two helices together.

The helices must be amphypatic!!



Supersecondary structures

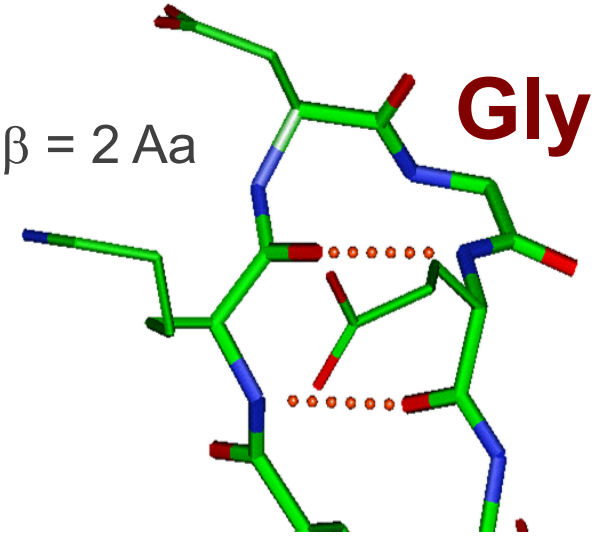
Here we see examples of beta hairpins. In this case, hydrogen bonds are more important for the stability of the hairpin.

Beta hairpin type 1

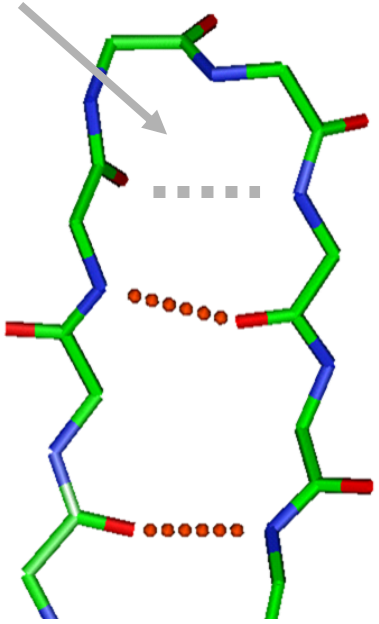
Beta hairpin type 2

Turn β = 2 Aa

Gly



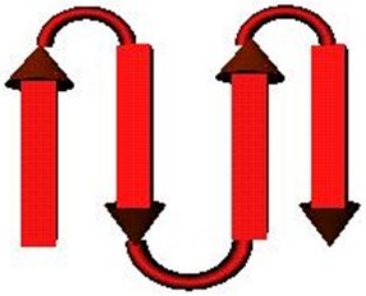
H.B. lost



Supersecondary structures

Here we see examples of supersecondary structures based on beta strands

β meander



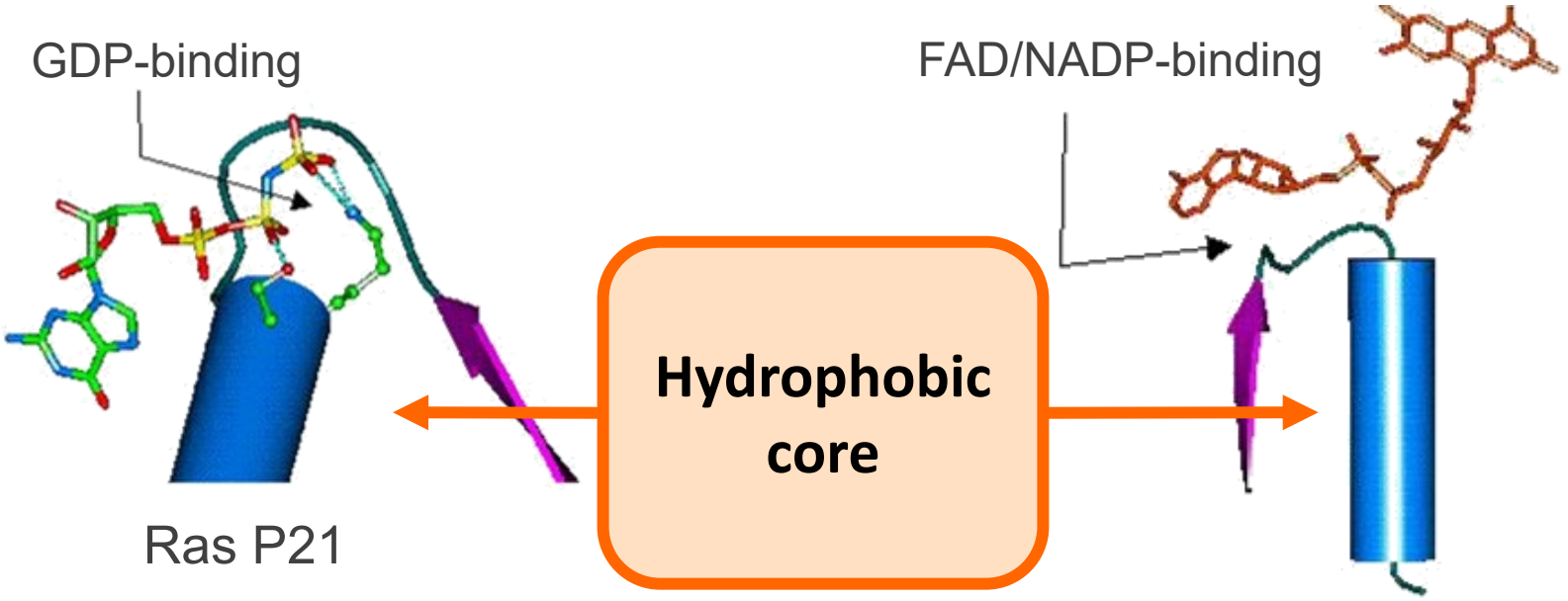
Greek Key

Supersecondary structures

Examples of beta-alpha and alpha-beta supersecondary structures. They are stabilized by a hydrophobic core.

P-loop

FAD/NADP binding loop

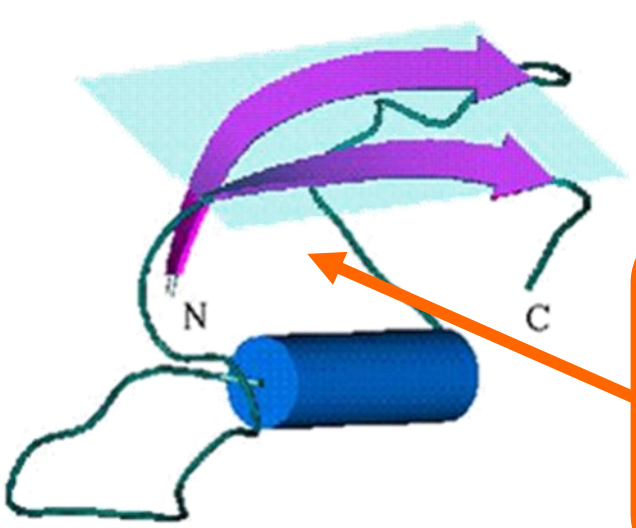


Supersecondary structures

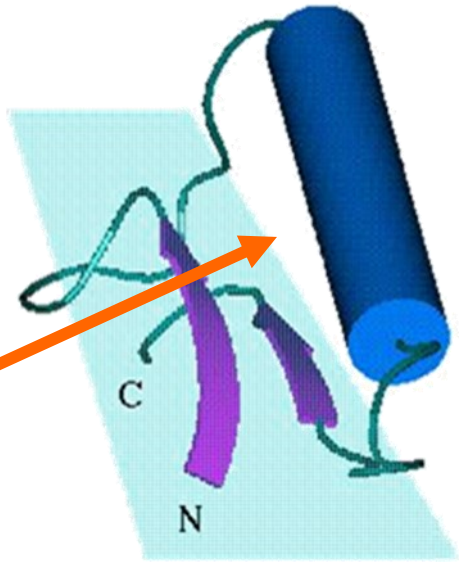
Examples of beta-alpha-beta supersecondary structures. They are stabilized by a hydrophobic core.

Left handed

Right handed



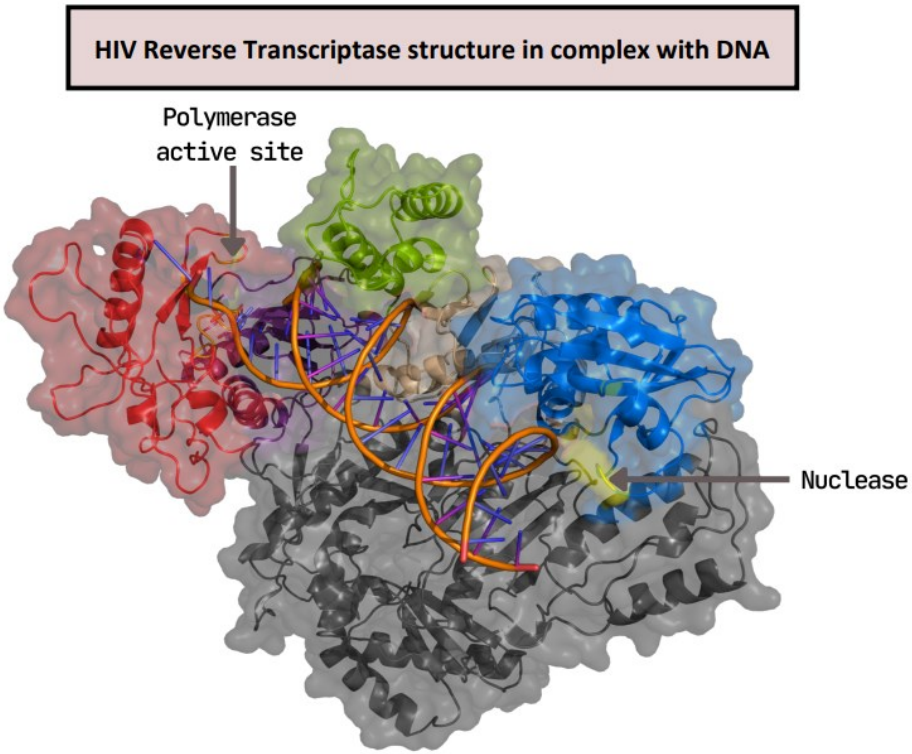
Hydrophobic core



Protein domains

Fundamental unit of 3D structure, able to fold by itself in the right conditions with independence of the rest of the protein.

It often corresponds to functional, local and compact units of a protein.



ACTIVE SITES

Polymerase active site

Fingers | Palm | Thumb

Located in P66 subunit

Nuclease active site

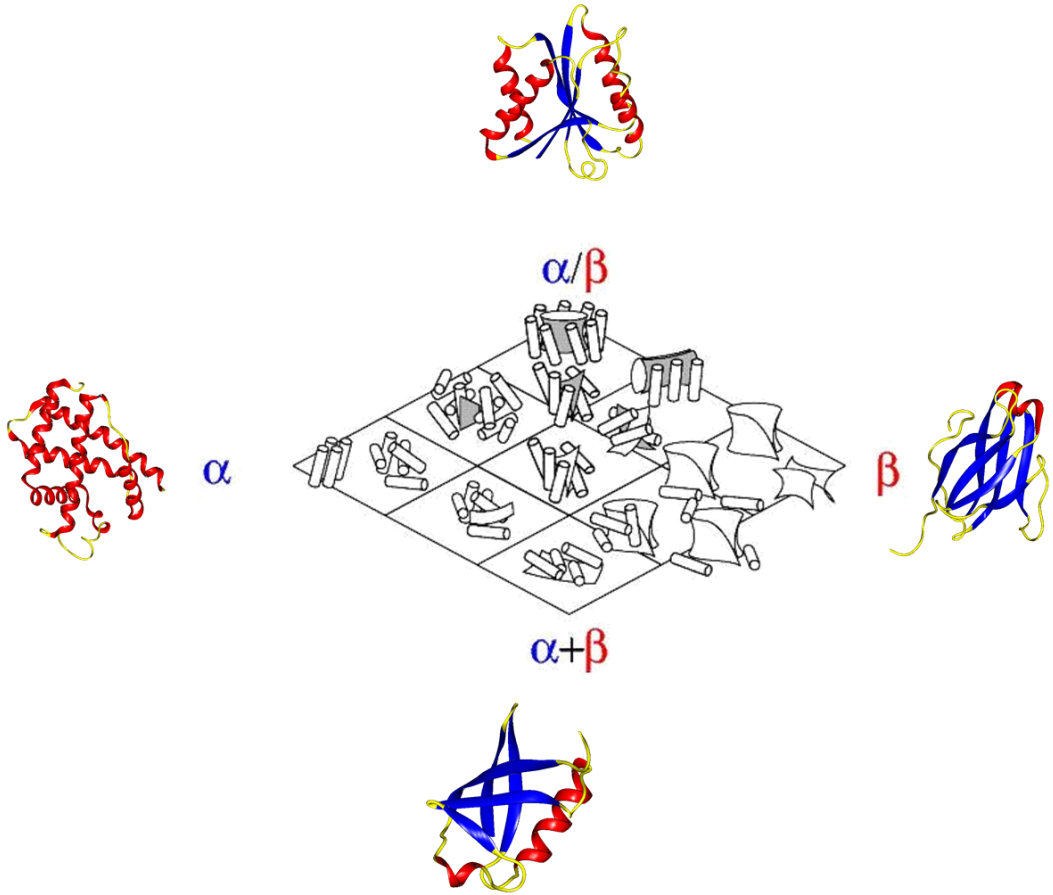
RNase

Helps to degrade the RNA template that the enzyme is copying, allowing the reverse transcription to proceed.

From the project of
X. Crespo,
N. Mitjavila,
M. Torner and
M. Ortigas

Protein domains

We can classify domains accoding with their composition in terms of secondary structure



Types of proteins

Globular proteins

Proteins that fold by compactin themselves arround a hydrophobic core. They are soluble in water. This are the proteins in which we will focus in this subject. One example of this is hemoglobin.

Types of proteins

Globular proteins

Proteins that fold by compactin themselves arround a hydrophobic core. They are soluble in water. This are the proteins in which we will focus in this subject. One example of this is hemoglobin.

Transmembrane proteins

Contain transmembrane domains. The interaction with a membrane is necessary for their proper folding. Examples of this are membrane transporters or receptors.

Types of proteins

Globular proteins

Proteins that fold by compactin themselves arround a hydrophobic core. They are soluble in water. This are the proteins in which we will focus in this subject. One example of this is hemoglobin.

Transmembrane proteins

Contain transmembrane domains. The interaction with a membrane is necessary for their proper folding. Examples of this are membrane transporters or receptors.

Fibrous proteins

They are filament like proteins, with low water solubility and mainly structural and protective functions. Examples of this are collagen, elastin or keratin.

Types of proteins

Globular proteins

Proteins that fold by compactin themselves arround a hydrophobic core. They are soluble in water. This are the proteins in which we will focus in this subject. One example of this is hemoglobin.

Transmembrane proteins

Contain transmembrane domains. The interaction with a membrane is necessary for their proper folding. Examples of this are membrane transporters or receptors.

Fibrous proteins

They are filament like proteins, with low water solubility and mainly structural and protective functions. Examples of this are collagen, elastin or keratin.

Unestructured or disordered proteins

Proteins that don’t have secondary structures. They are extremelly flexible and they can have many different conformations.

The relation between sequence, structure and function