Biophysics and structural bioinformatics I

D. Gilis & M. Rooman

Unité de bioinformatique génomique & structurale

UD3.203/UD3.204 (bâtiment U, campus du Solbosch)

Tél: 02/650.36.15 - 02/650.20.67

e-mail: dgilis@ulb.ac.be / mrooman@ulb.ac.be

Documents at: http://uv.ulb.ac.be and

http://babylone.ulb.ac.be/~dgilis/MA1bioinfo.php

Part 9

1. Structure of membrane proteins: introduction

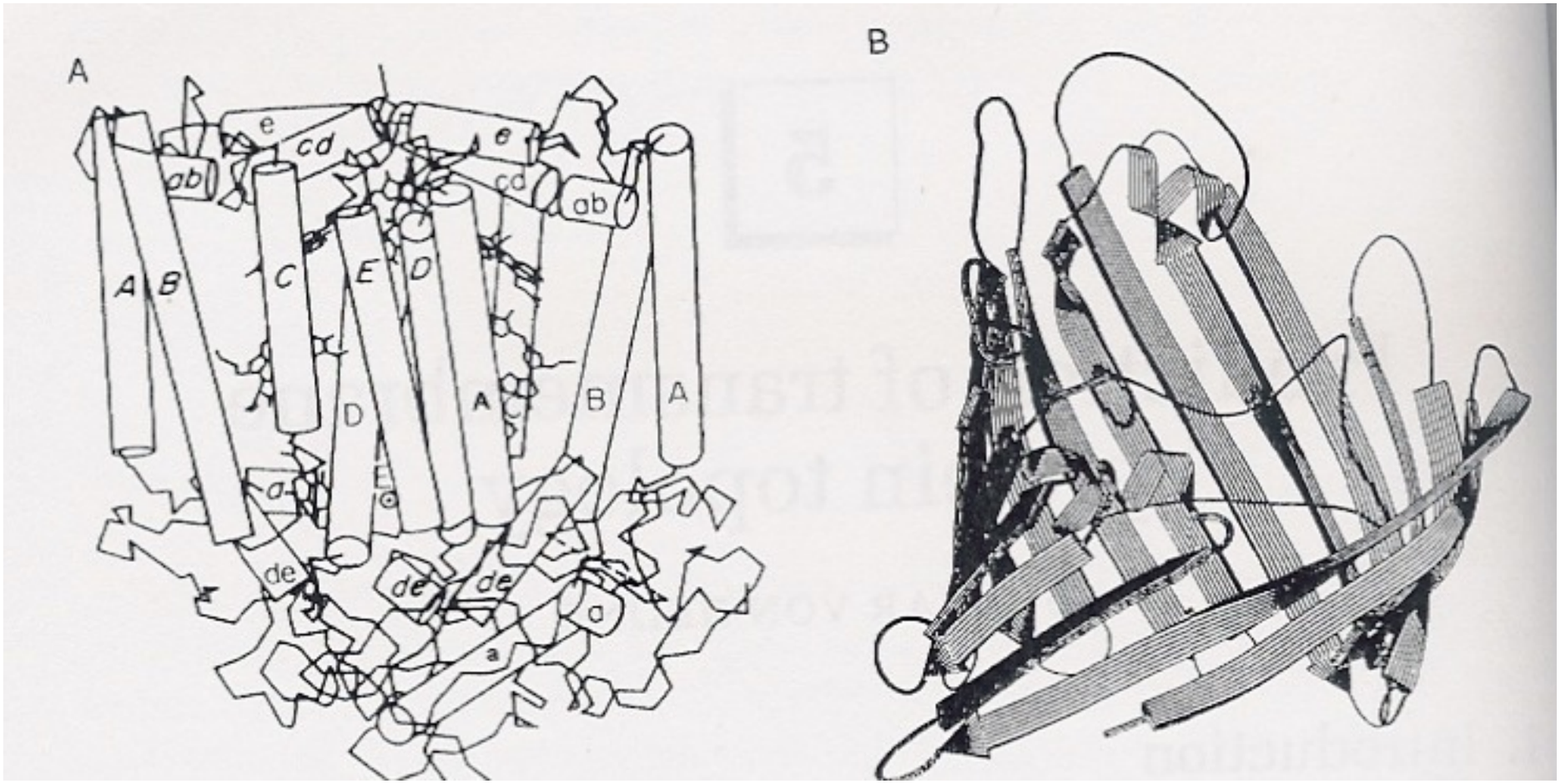
Membrane proteins are associated to a large number of important functions in the cell. Membrane proteins account for ~30% of the genome of living organisms. The number of resolved 3D structure is about 415 (source: http://blanco.biomol.uci.edu/mpstruc/listAll/list; september 2013). Half of current drugs have an effect on transmembrane proteins.

The main categories are

- integral membrane proteins (transmembrane and permanently attached from one side);
- peripheral membrane proteins (temporarily attached to the membrane or to integral membrane proteins);
- polypeptide toxins (they are water soluble but they can associate to membrane and form transmembrane channels)

The structures of membrane proteins are collected on this website: http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html.

Membrane proteins are either composed of “helix bundles” (A) or β-barrels (B).



Not yet identified a mix of helices and β-strands.

In general helicoidal transmembrane segments are easier to predict than β strands, because the latter are shorter and sometimes less hydrophobic.

The biosynthesis of these folds and the mode of membrane insertion is also different between membrane proteins composed of helices or strands.

Helix bundles

The ribosomes are linked to a “translocon” in a target membrane (internal membrane for bacteria, endoplasmic reticulum for eukaryotic cells).

The protein is then transferred from the translocon to the lipid bilayer. This could be seen as a partition phenomenon between an aqueous medium and the lipidic membrane.

Once the protein is inserted in the membrane, the helices form a compact structure.

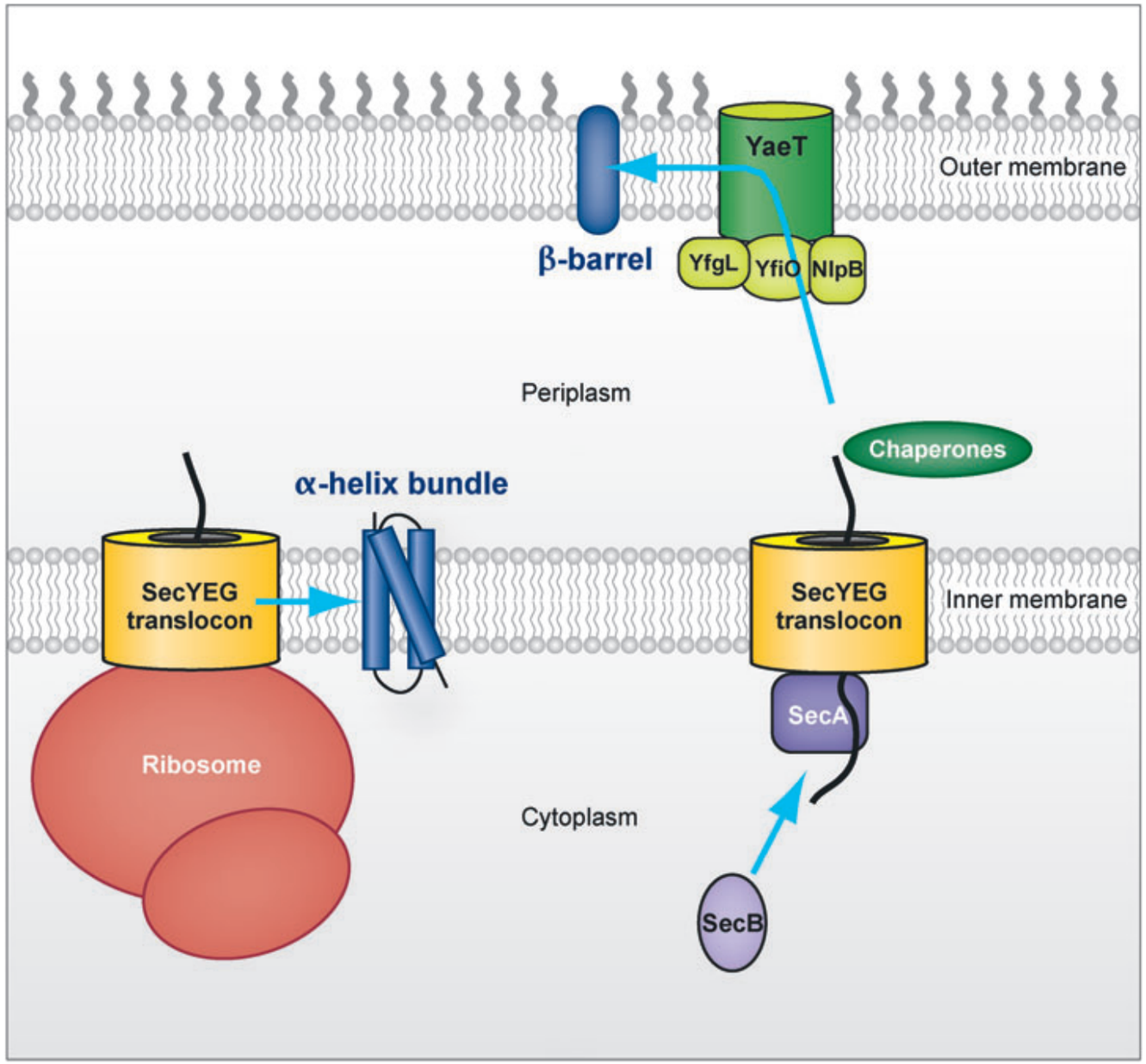
β-barrels

The proteins are transferred from the ribosome to a cytoplasmic chaperone (secB).

They are then transferred to the translocon (after the translation phase), with the help of an ATPase (secA).

The β-strands being not hydrophobic enough, these proteins are transferred to the lipid bilayer via chaperones.

Finally, these proteins are inserted in the bilayer via an membrane integration complex (YaeT).



Source: Elofsson & von Heijne, Ann. Rev. Biochem. (2007), 76, 125-140.

2. Helix bundles: prediction of membrane segments and topology

Helicoidal transmembrane segments are composed of long segments that are mainly apolar.

Example:

photosynthetic reaction center : 11 transmembrane helices

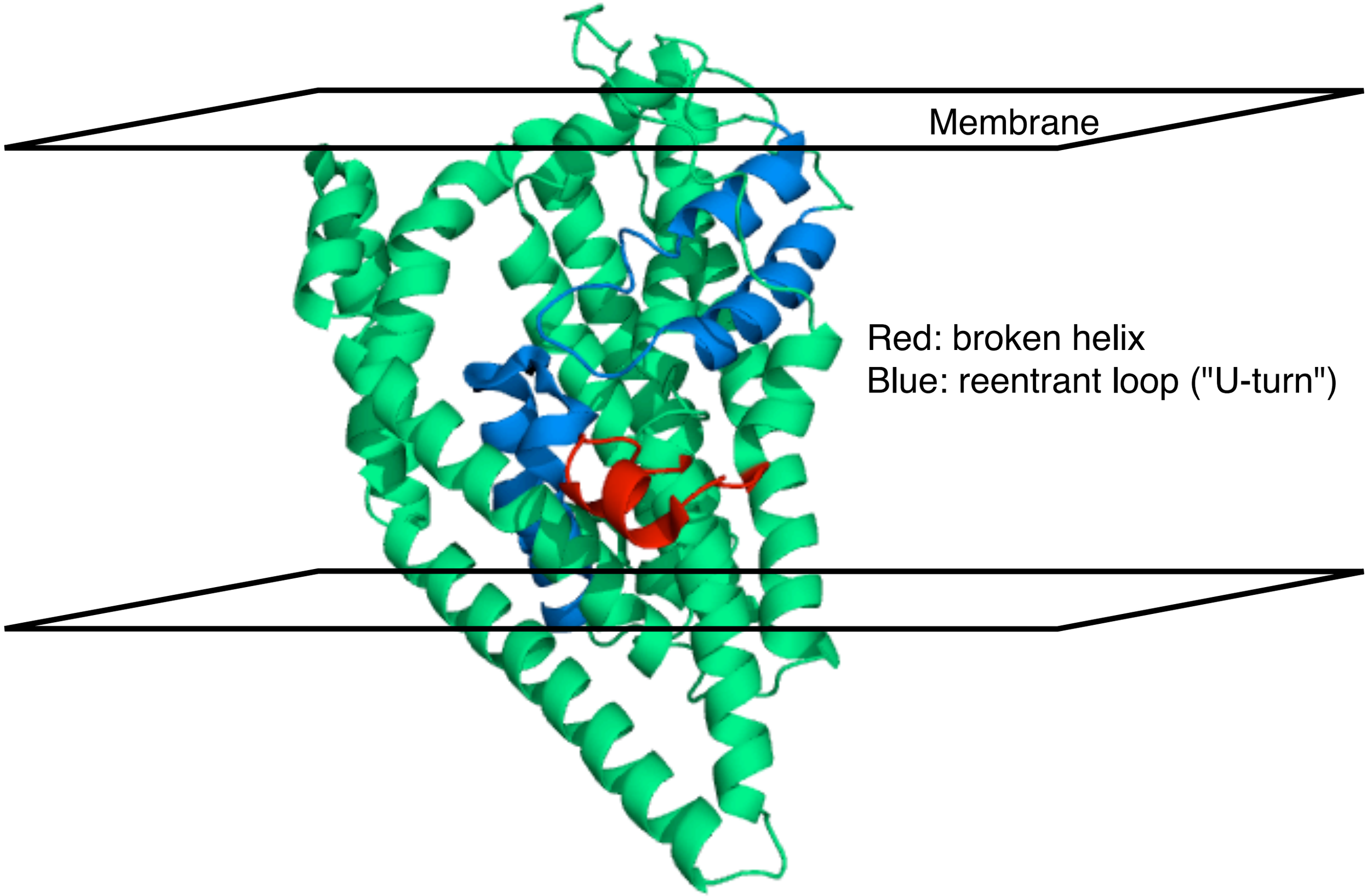
bacteriorhodopsin : 7 helices

Most segment are ~perpendicular to the membrane and are composed of 15-30 residues.

Most of prediction tools rely on rules derived from known experimental structures.

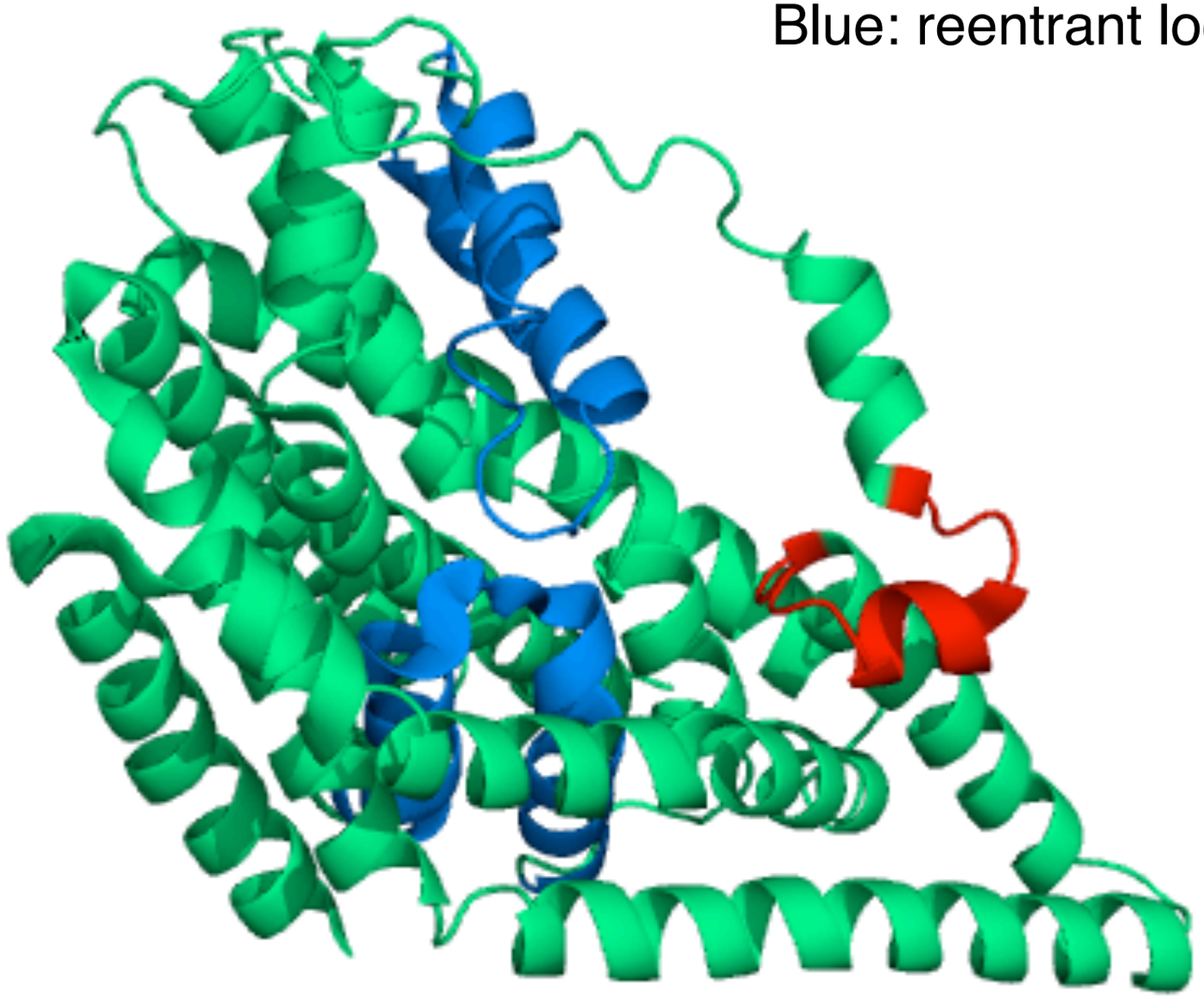
But the number of 3D structures of membrane proteins is quite low => some 3D structural features are probably still unknown.

Example: glutamate transporter homolog de Pyrococcus horikoshii (code PDB: 1XFH)



Red: broken helix
Blue: reentrant loop ("U-turn")

Red: broken helix
Blue: reentrant loop ("U-turn")

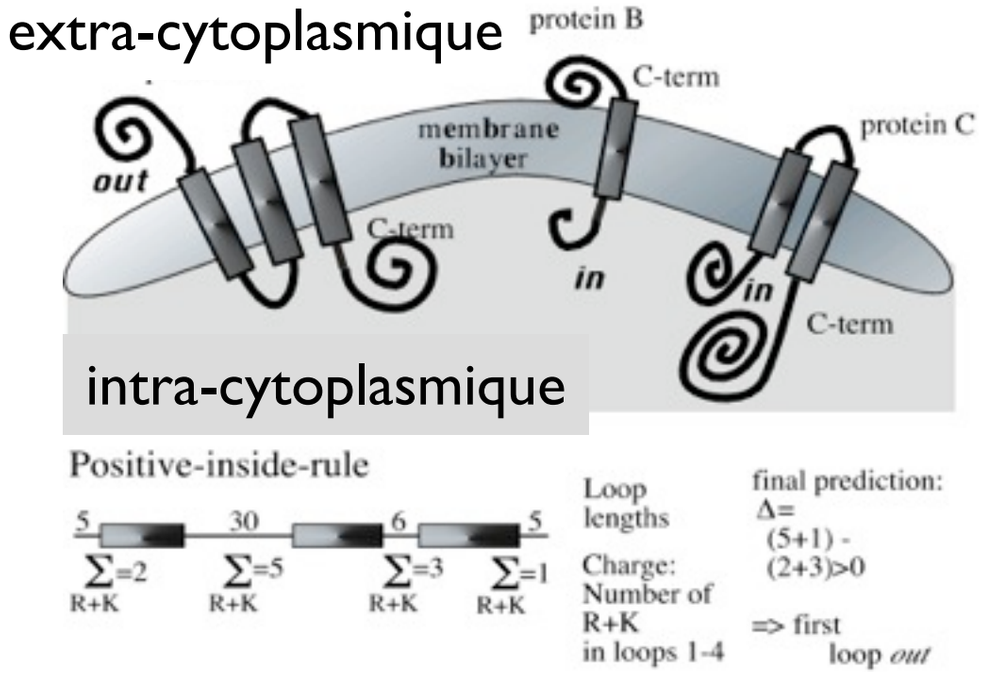


There are also examples with an helix that is parallel to the membrane.

2.1. Parameters that are used in prediction tools

2.1.1. Empirical rules

✓ In general, positive amino acids (Arg, Lys) are more frequent in non-translocated regions (cytoplasmic regions). Positively charged amino acids that are at the border of apolar regions can be used to identify the topology of the membrane insertion (extracellular or intracellular N-terminal region).



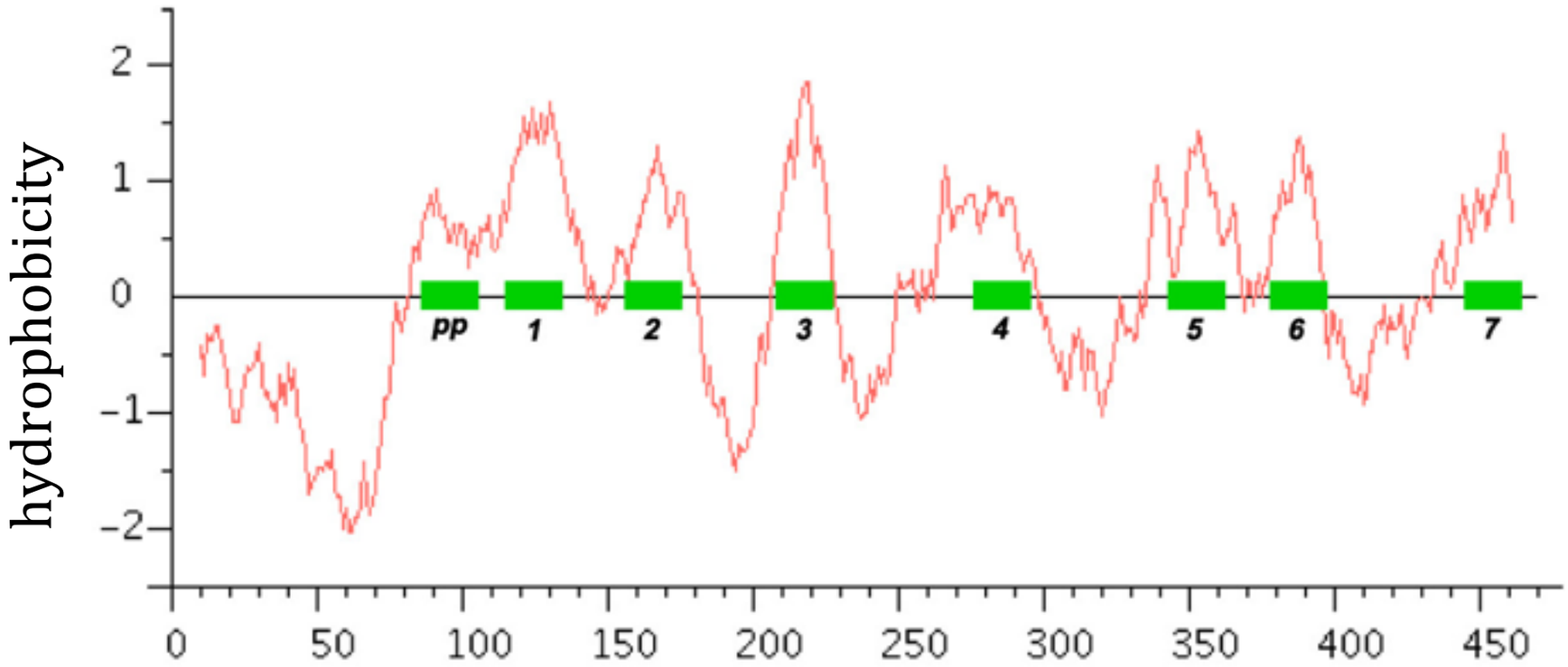
This rule can be used to predict the topology of the insertion of the protein into the membrane.

Remark: sometimes, long loops (60-70 residues) are translocated outside the membrane and are composed of positively charged amino acids => not in agreement with the "rule".

✓ Transmembrane helices are mainly apolar and are composed of ~15-30 amino acids.

2.1.2. Hydrophobicity plots

Transmembrane helices can be identified with a plot of the hydrophobicity of amino acids the versus the position along the sequence.



=> necessary to use an hydrophobicity scale (large number of scales).

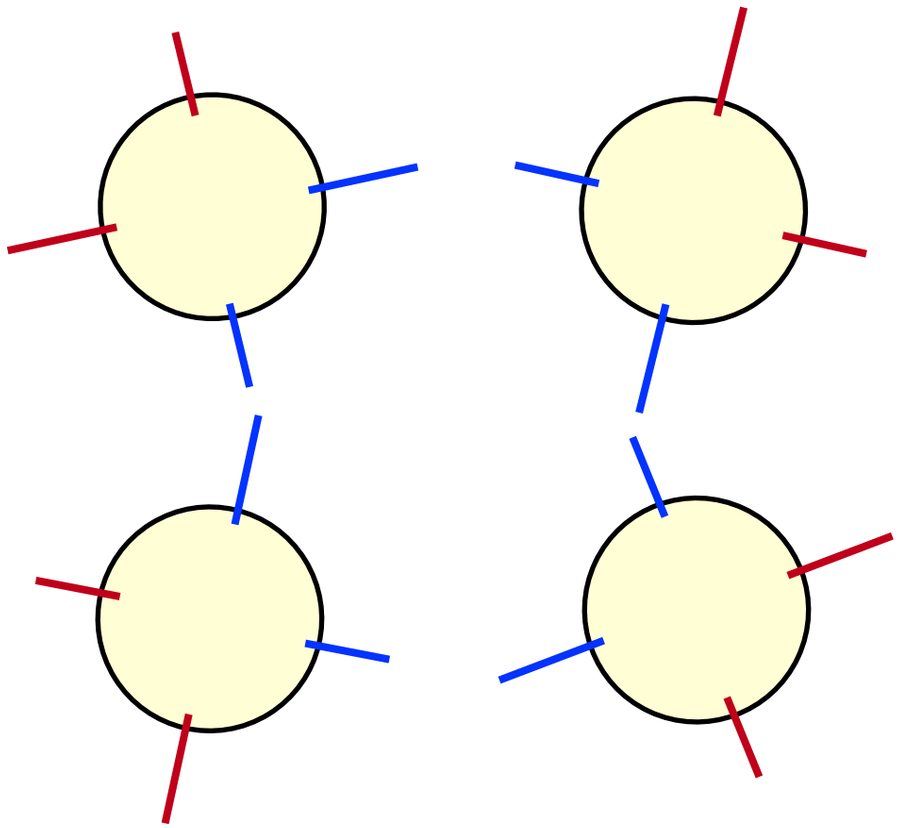
Methods that are only based on an hydrophobicity plot are not reliable enough to identify the number of tansmembrane segments.

2.1.3. Analysis of the hydrophobic moment

It corresponds to the amphiphilicity of an helicoidal segment. It measures the distribution of the hydrophobic side chains. The amphiphilic helices tend to group according to the following scheme:

— hydrophilic amino acids

— hydrophobic amino acids



2.2. Example of programs

2.2.1. TopPred

✓ computes an hydrophobicity plot;

✓ identifies the "undoubted" transmembrane segments : hydrophobicity larger than a threshold C1;

✓ identifies putative transmembrane segments: hydrophobicity larger than a threshold C2 (C2<C1);

✓ builds all the possible topologies, including the "undoubted" regions and including or excluding the putative segments;

✓ computes the difference between the number of extracellular and intracellular Arg+Lys for each topologies; does not take into account the long loops;

✓ chooses the topology with the largest difference;

✓ if there is a large number of long loops, their amino acid composition could indicate their intra- or extracellular localization.

2.2.2. TMHMM and HMMTOP

These programs rely on hidden markov models (HMM)

It is sometimes difficult to make a distinction between transmembrane segments and signal peptide. The Phobius program predicts both types.

HMM's can include evolutionary information and some experimental results in view of improving the predictions. It can also include hydrophobicity, charge bias, helix length and other constraints.

✓ A set of states is defined in the HMM, each residue can be in one of these states.

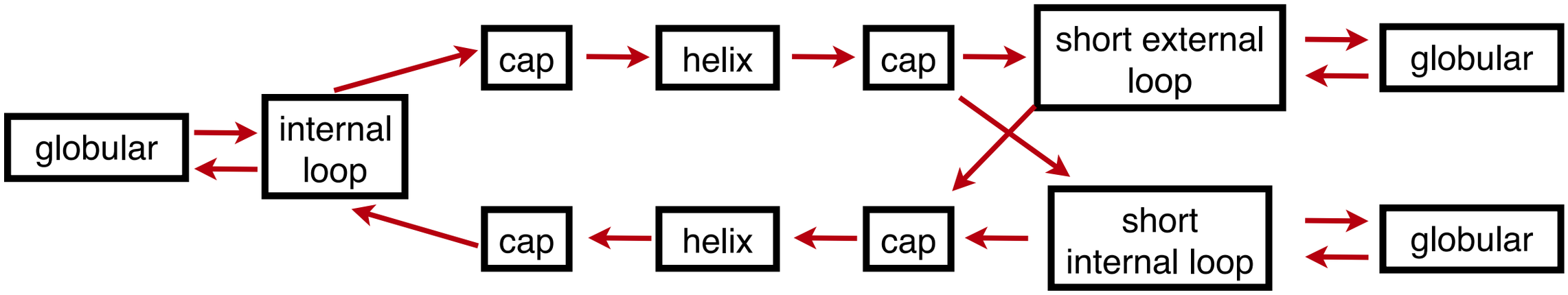
✓ Different states can be defined, according to the model. Example: state 1= intracellular loop, state 2=extracellular loop, state 3=transmembrane helix.

✓ The distribution probability in each state is computed for the 20 amino acids.

✓ A connectivity between the states is defined, takin into account the "biology of the system".

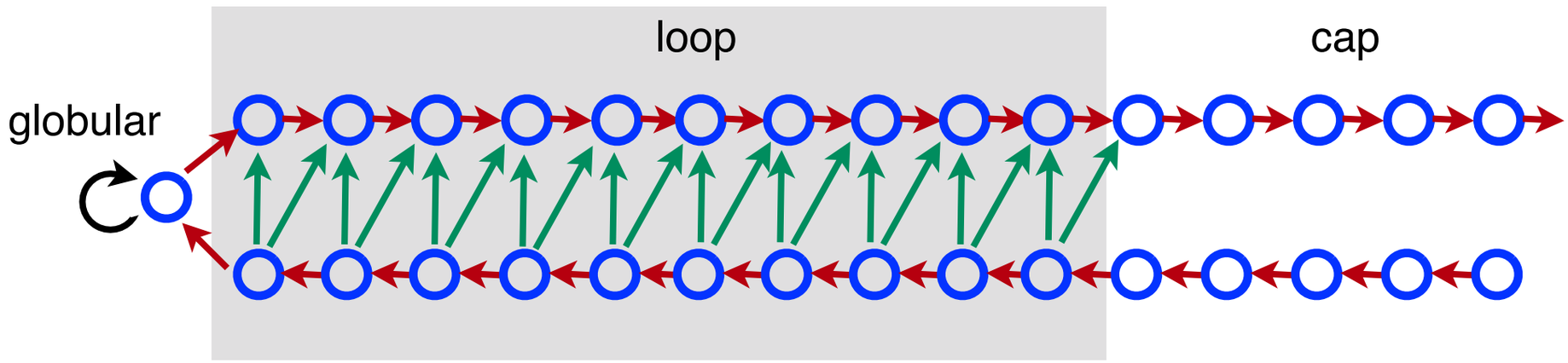
✓ The transition probability between the states is computed.

Example: considered states in *TMHMM*

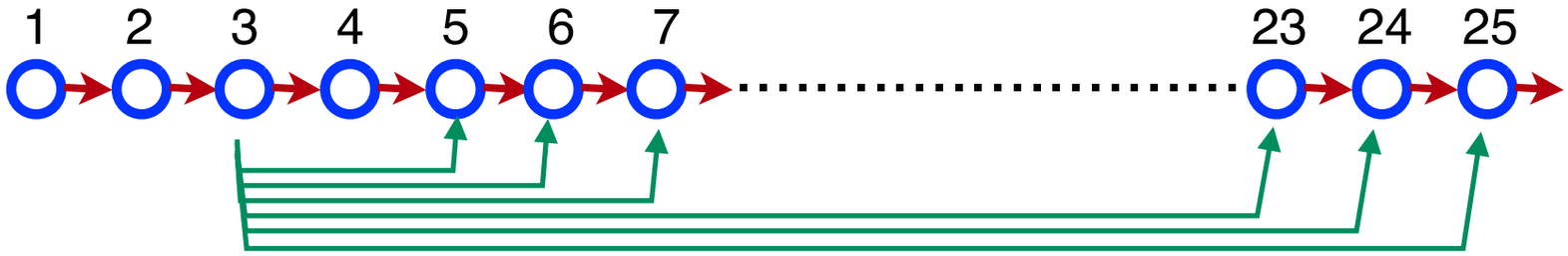


Each box corresponds to one or several states.

Loop: maximum 20 residues; if it is larger, it is considered as a globular state:



Helix core: between 5 et 25 residues (including cap's, it corresponds to helix of 15 to 35 residues)



The transition probabilities between the states and of the 20 amino acids for each state are computed from a set of 160 proteins.

3. β-barrels: prediction and topology

Several 3D structures of transmembrane β-barrel proteins have been resolved (porins, channels, enzymes, transporters).

Most of proteins contain an even number of strands. In general the angle between the β-strands and membrane is about 45°.

In general, one face of the strand is composed of hydrophobic amino acids and faces the lipids. One amino acid out of two is hydrophobic, the other hydrophilic.
Aromatic amino acids are frequently found at the beginning and at the end of the β-strands.

A face hydrophobicity profile can be computed:

$$H_s(i)=1/4\ [h(i-2) + h(i) + h(i+2) + h(i+4)]$$

where h is the hydrophobicity ; when residues at positions i-2 or i+4 are aromatic, the hydrophobicity is increased by 1,6, which biases the prediction towards strands that contains aromatic residues at their borders.

Some HMM models have also been developed to predict β-barrel transmembrane segments.

4. Performances of the methods

The most performing methods are those based on HMM. They predict correctly the topology (number of helices and orientation in the membrane) of ±70% of membrane proteins. This score can be improved by adding experimental constraints.

In the case of β-barrels, HMM's are also the most performing. l

5. Approaches to predict the 3D structure

Some recent methods have been developed to predict the structure of reentrant loops.

✓ They appear more frequently in ionic and water channels. They are less frequent in receptors.

✓ They contain, on average, smaller residues, they present peculiar amino acid composition in comparison with other transmembrane regions and they contain peculiar sequence motifs.

- *Ab initio* and fold recognition methods

The main difficulty is to take into account the hydrophobic medium => difficult to use methods developed for soluble globular proteins

Tasser (fold recognition method) has been used to predict all the human GPCR structures (Zhang et al. PLoS Comput. Biol. (2006) 2, e13). Accuracy ???

Rosetta (*ab initio* method) has been used to predict the structure of a voltage dependent potassium channel that undergoes a conformational change.

- Comparative modeling

The main problems:

✓ low number of known structures (and of possible templates);

✓ sometimes, low sequence indentity between the template and the target. Below 30%, the quality of the model decreases, due to errors in the sequence alignment or to differences between the structure of the template and the "real" structure of the target.

How to improve the sequence alignment ? Identification of functional sequence positions and align these positions without taking into account the amino acid types at these positions.

Another problem: loop modeling. If the loop contains more that ~12 residues, the prediction is less reliable, several loops could also interact.

Swiss Model 7TM: comparative modeling approach that focus on membrane receptors containing 7 transmembrane helices.

When a template is available and when the sequence identity is large enough, the quality of the model is similar to that of soluble proteins.