Topological Kohonen maps and β turns

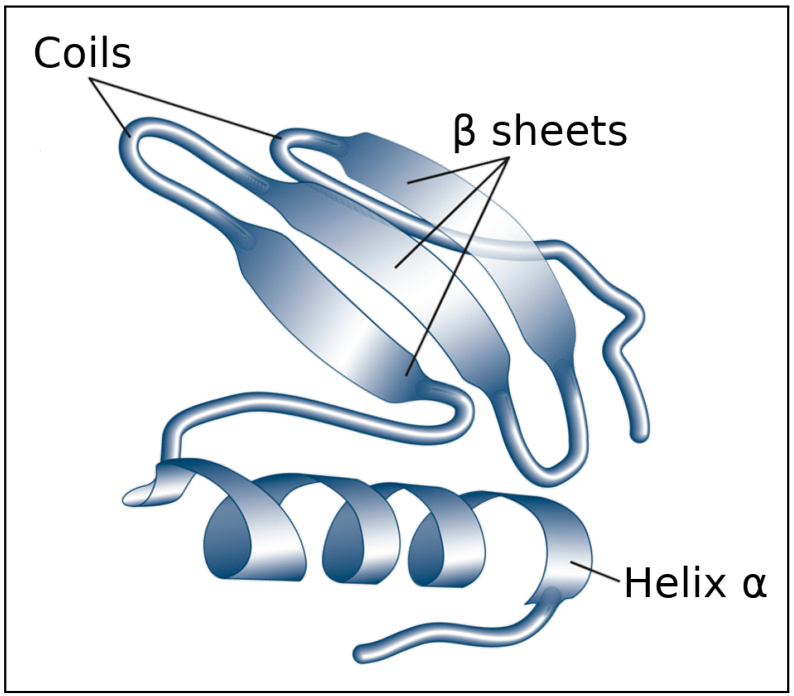
Nicolas Silva

Université de Paris

silva.nicolas.j@gmail.com

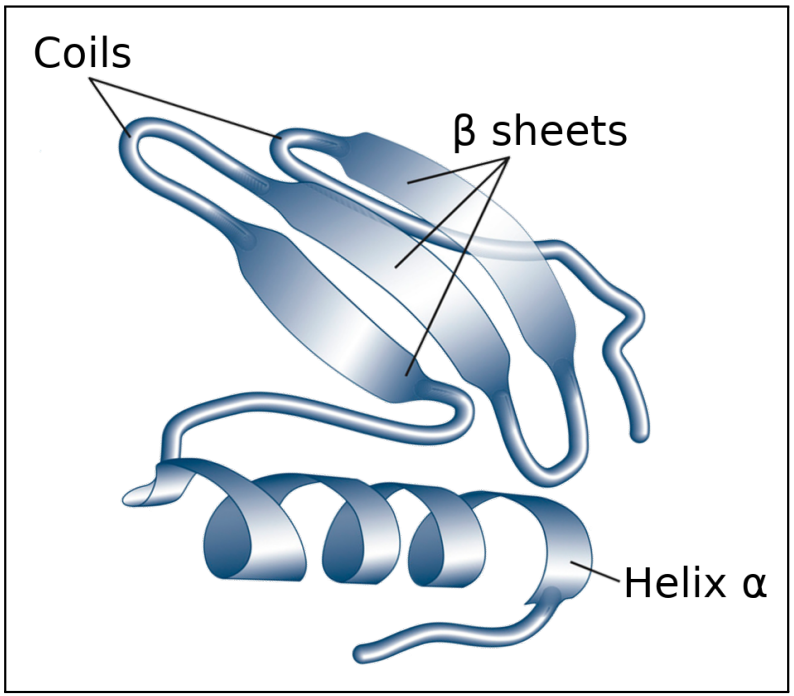
https://github.com/nicolassilva/projetLong_Kohonen-BetaTurn

Secondary structures and turns



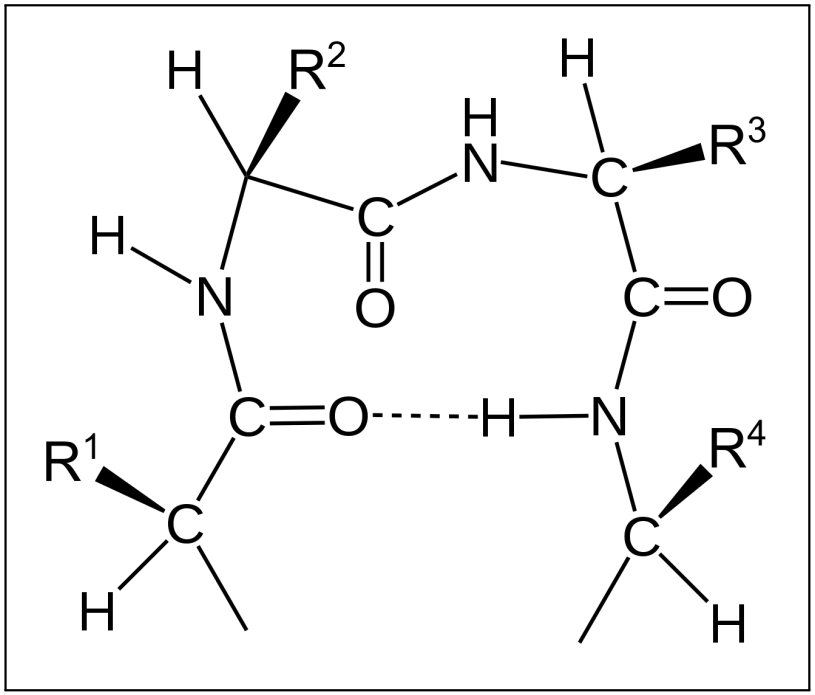
Expected: 50% of coils ; 1/5 of β-sheets and 1/3 of helix-α.

Secondary structures and turns



Expected: 50% of coils ; 1/5 of β-sheets and 1/3 of helix-α.

Multiple turn types: γ-turn ; β-turn ; α-turn and π-turn.

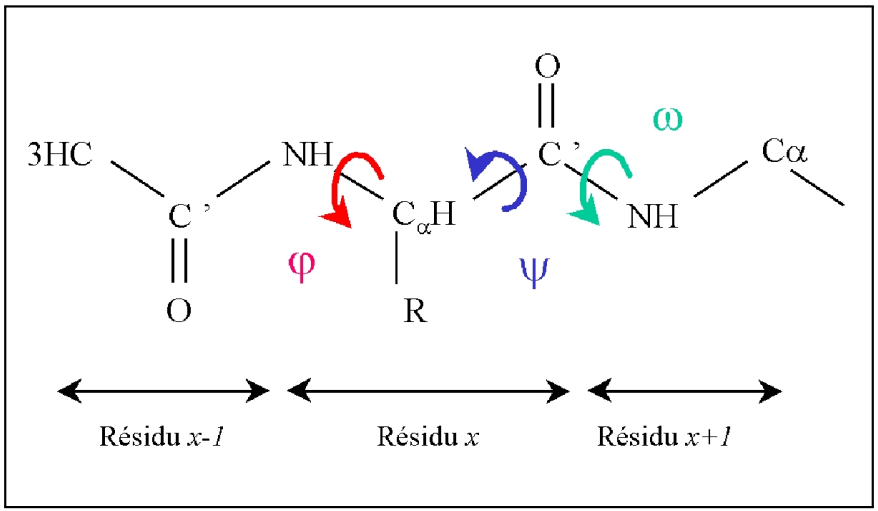


β-turns

- 13 β-turn types: I, I', II, II', IV_1, IV_2, IV_3, IV_4, IV_{misc}, VI_{a1}, VI_{a2}, VI_b and $VIII$

β-turns

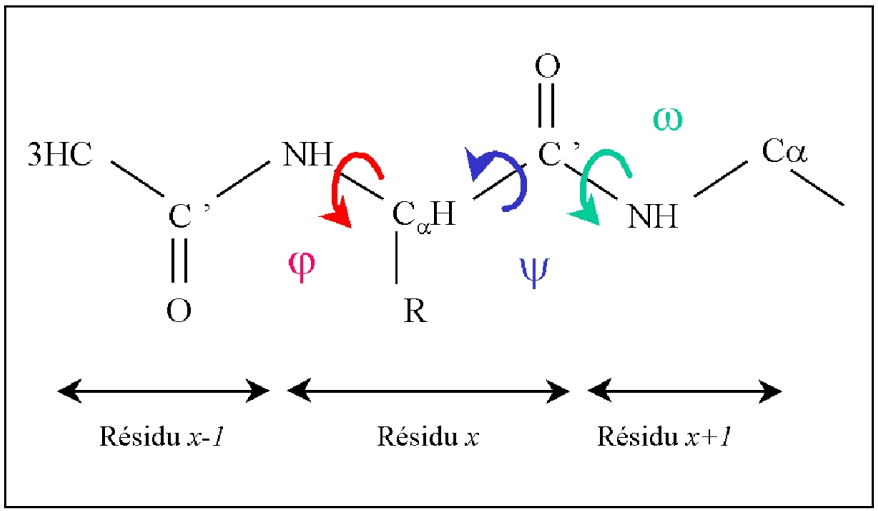
- 13 β-turn types: I, I', II, II', IV_1, IV_2, IV_3, IV_4, IV_{misc}, VI_{a1}, VI_{a2}, VI_b and $VIII$



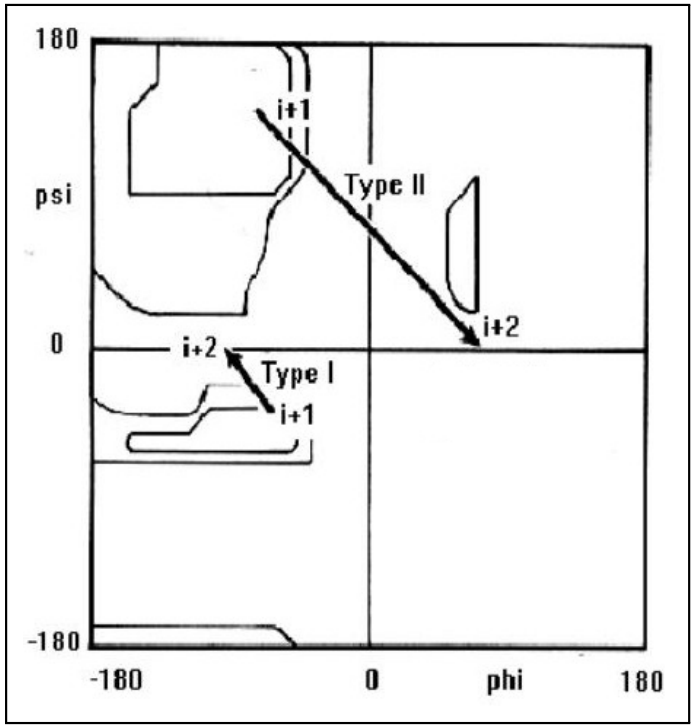
Dihedral angles φ, ψ and ω of the residus X.

β-turns

- 13 β-turn types: *I*, *I′*, *II*, *II′*, IV_1, IV_2, IV_3, IV_4, IV_{misc}, VI_{a1}, VI_{a2}, VI_b and *VIII*

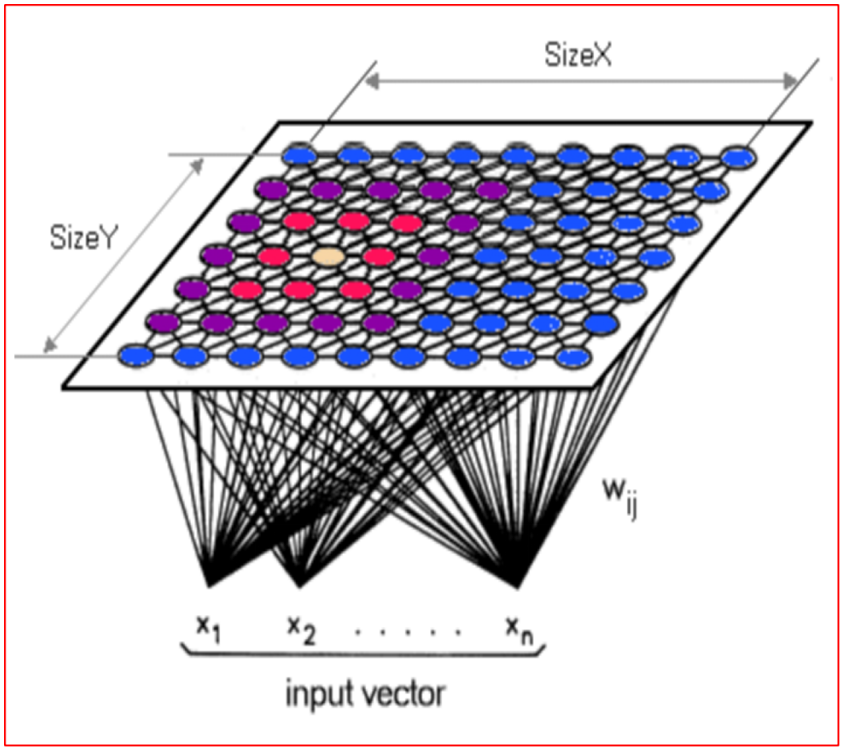


Dihedral angles φ, ψ and ω of the residus *X*.

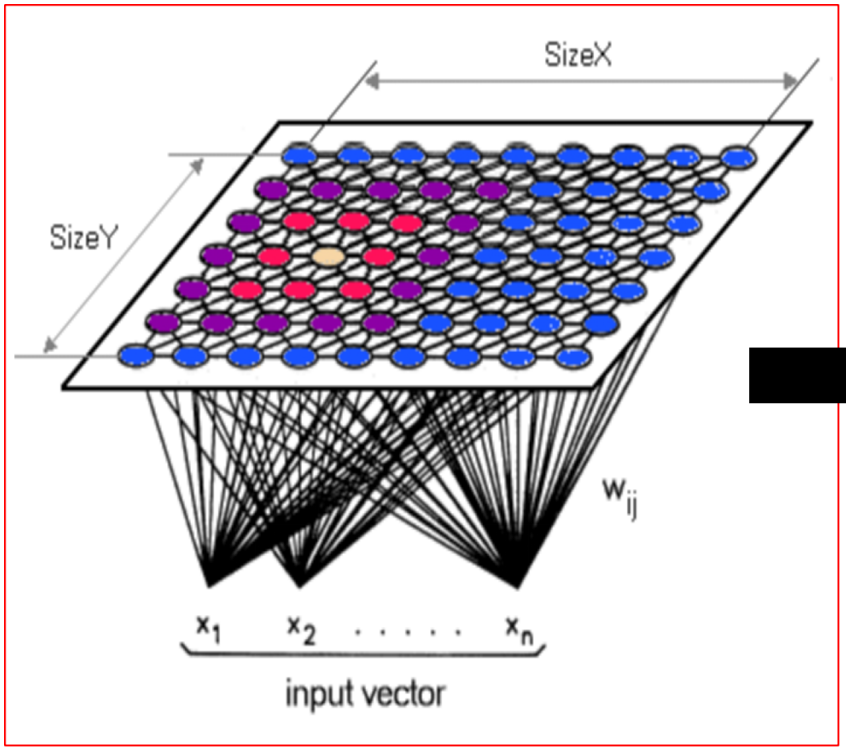


Example of Ramachandran plot [Ramachandran, 1963] for β-turn type *I* and *II*.

Self-Organized Maps

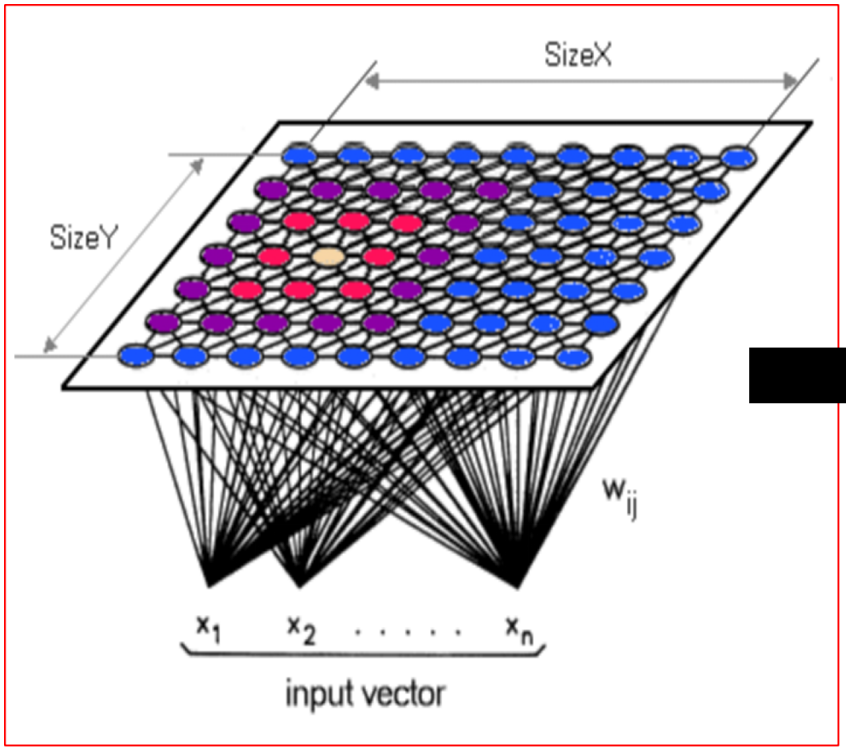


Self-Organized Maps



Neurons' creation for specialised classification of β-turns.

Self-Organized Maps



Neurons' creation for specialised classification of β-turns.

The main objectif is to study a classification method based on angles and to bring a different learning throughout Kohonen maps.

Data set and frequencies

Variables of interest from the data set:

- Amino acid type and position
- Secondary structure
- φ angle
- ψ angle
- (X, Y, Z) coordinates of the carbone-α

Data set and frequencies

Variables of interest from the data set:

- Amino acid type and position
- Secondary structure
- φ angle
- ψ angle
- (X, Y, Z) coordinates of the carbone-α

Data frequencies in the sequences:

- 20 Amino acids frequencies [King & Jukes, 1969]
- 3 Secondary structures frequencies [de Brevern, 2016]
- 13 β-turn types frequencies [de Brevern, 2016]

Turn assignment

Conditions:

- Distance i and $i+3 < 7$Å
- $i+1$ and $i+2$ not Helix
- All residus not β-sheets

Turn assignment

Conditions:

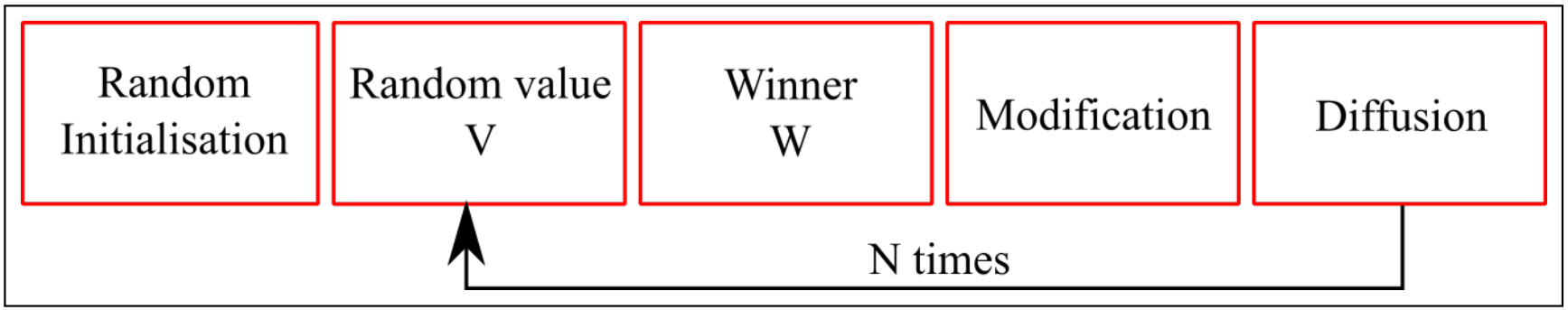
- Distance i and $i+3 < 7$Å
- $i+1$ and $i+2$ not Helix
- All residus not β-sheets

Degree of angle liberty:

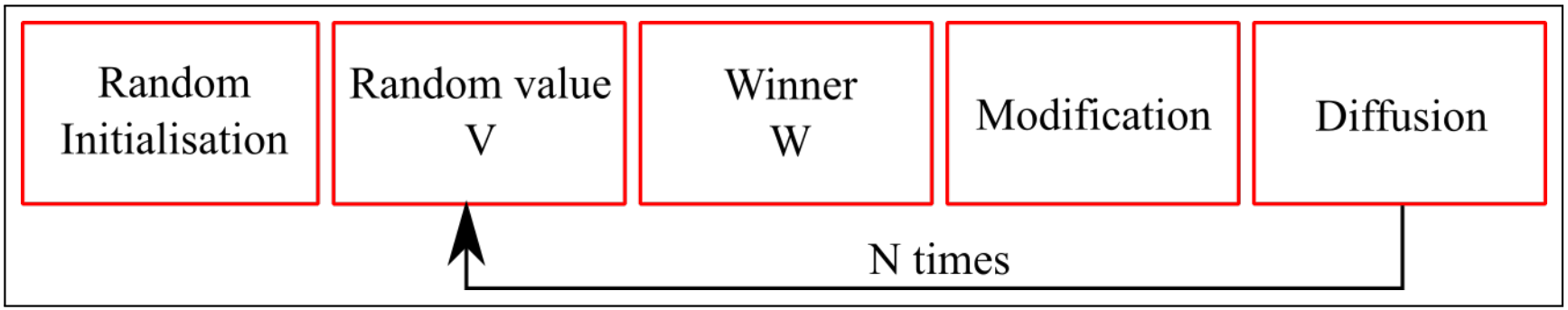
- $\pm 30°$ for 3 of the angles
- $\pm 45°$ for 1 of the angles

Type	**Résidu i+1**		**Résidu i+2**	
	φ	ψ	φ	ψ
Type I	-60	-30	-90	0
Type I'	60	30	90	0
Type II	-60	120	80	0
Type II'	60	-120	-80	0
Type IV_1	-120	130	55	41
Type IV_2	-85	-15	-125	55
Type IV_3	-71	-30	-72	-47
Type IV_4	-97	-2	-117	-11
Type VI_{a1}	-60	120	90	0
Type VI_{a2}	-120	-120	-60	0
Type VI_b	-135	135	-75	160
Type VIII	-60	-30	-120	120
Type IV_{misc}	*Aucunes des autres catégories*			

Learning


Random Initialisation
Random value V
Winner W
Modification
Diffusion
N times

Learning

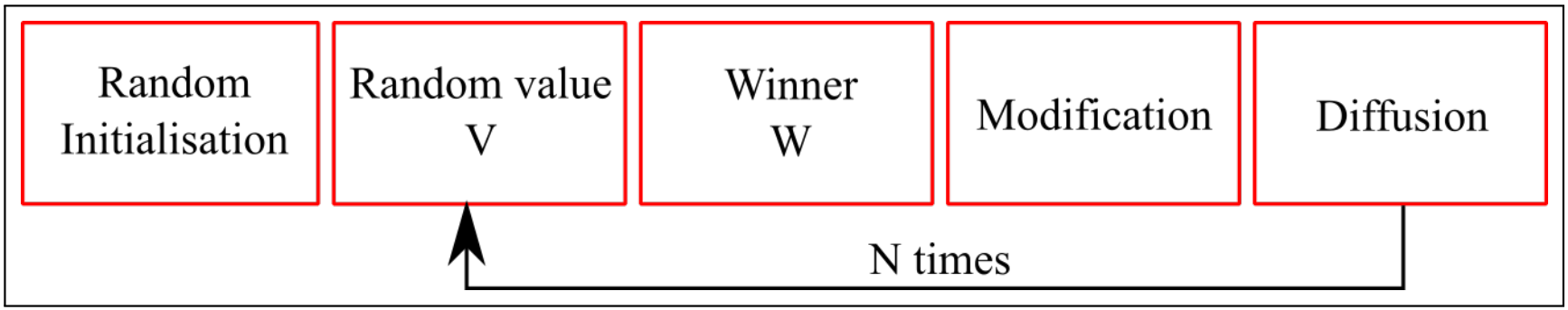


Winner:

$$D = |V_i - W_i|$$

$$\textit{If} \quad D < 180, \quad \textit{then} \quad D = D$$
$$\textit{If} \quad D > 180, \quad \textit{then} \quad D = 360 - D$$

Learning



Winner:

$$D = |V_i - W_i|$$

$$\begin{aligned} &\textit{If} \quad D < 180, \quad \textit{then} \quad D = D \\ &\textit{If} \quad D > 180, \quad \textit{then} \quad D = 360 - D \end{aligned}$$

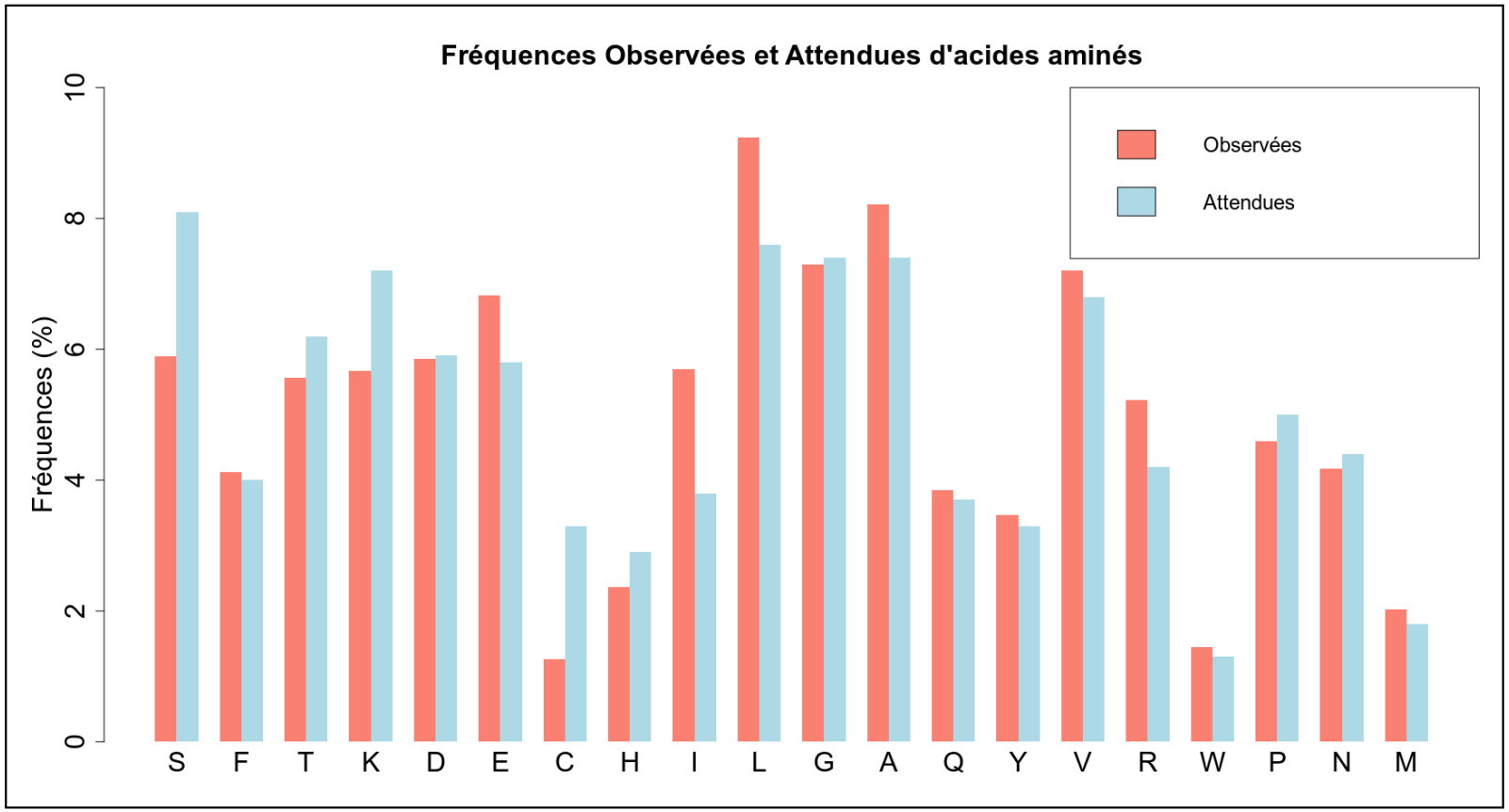
Modification & diffusion:

$$W_{new} = W_i + (V_i - W_i)\alpha(t)\beta(t)$$

with

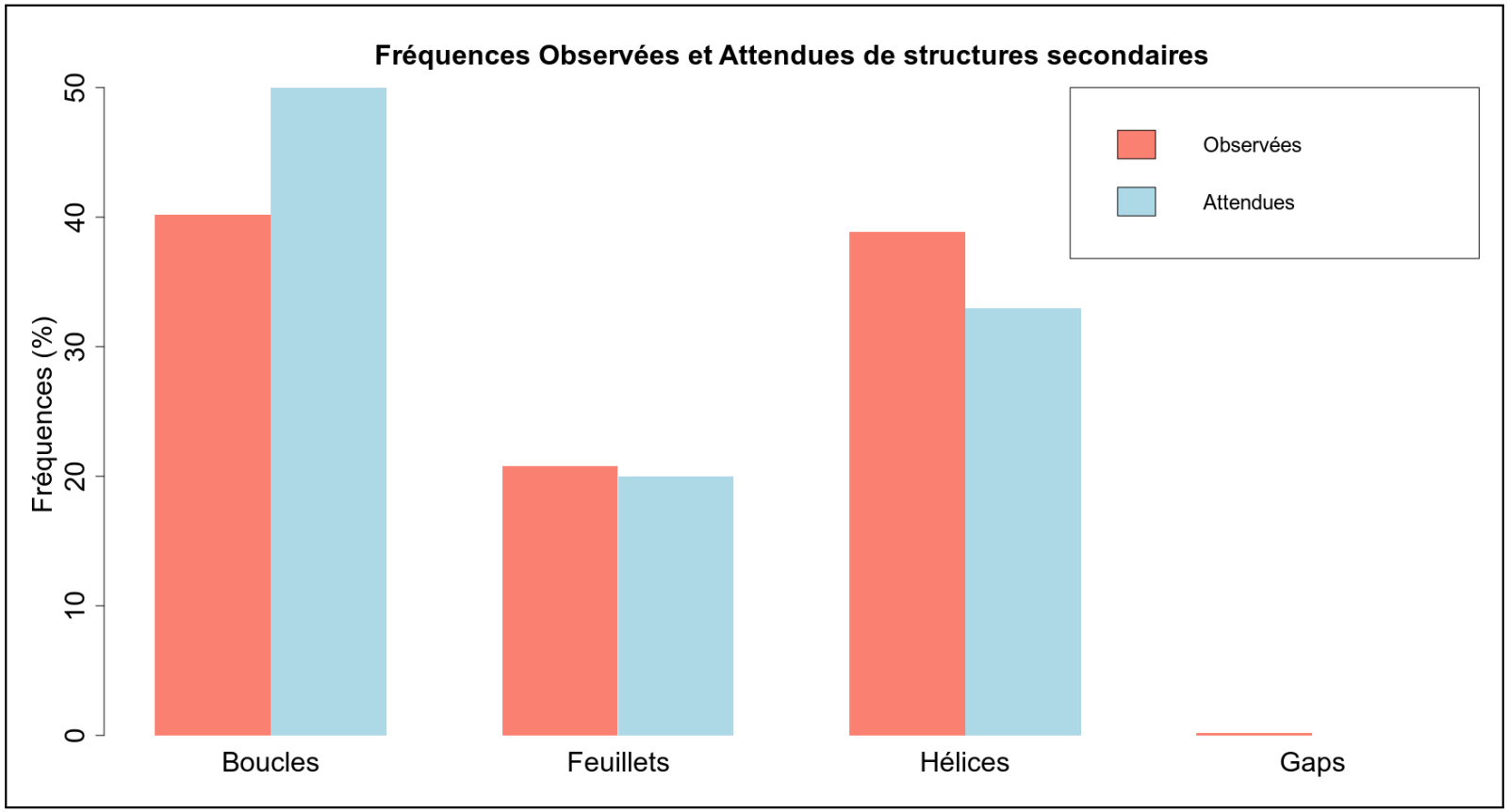
$$\alpha(t) = \frac{\alpha_0}{1+\frac{t}{n}} \quad \beta(t) = \exp -\frac{(r-r_{winner})^2}{2\eta^2}$$

Frequencies



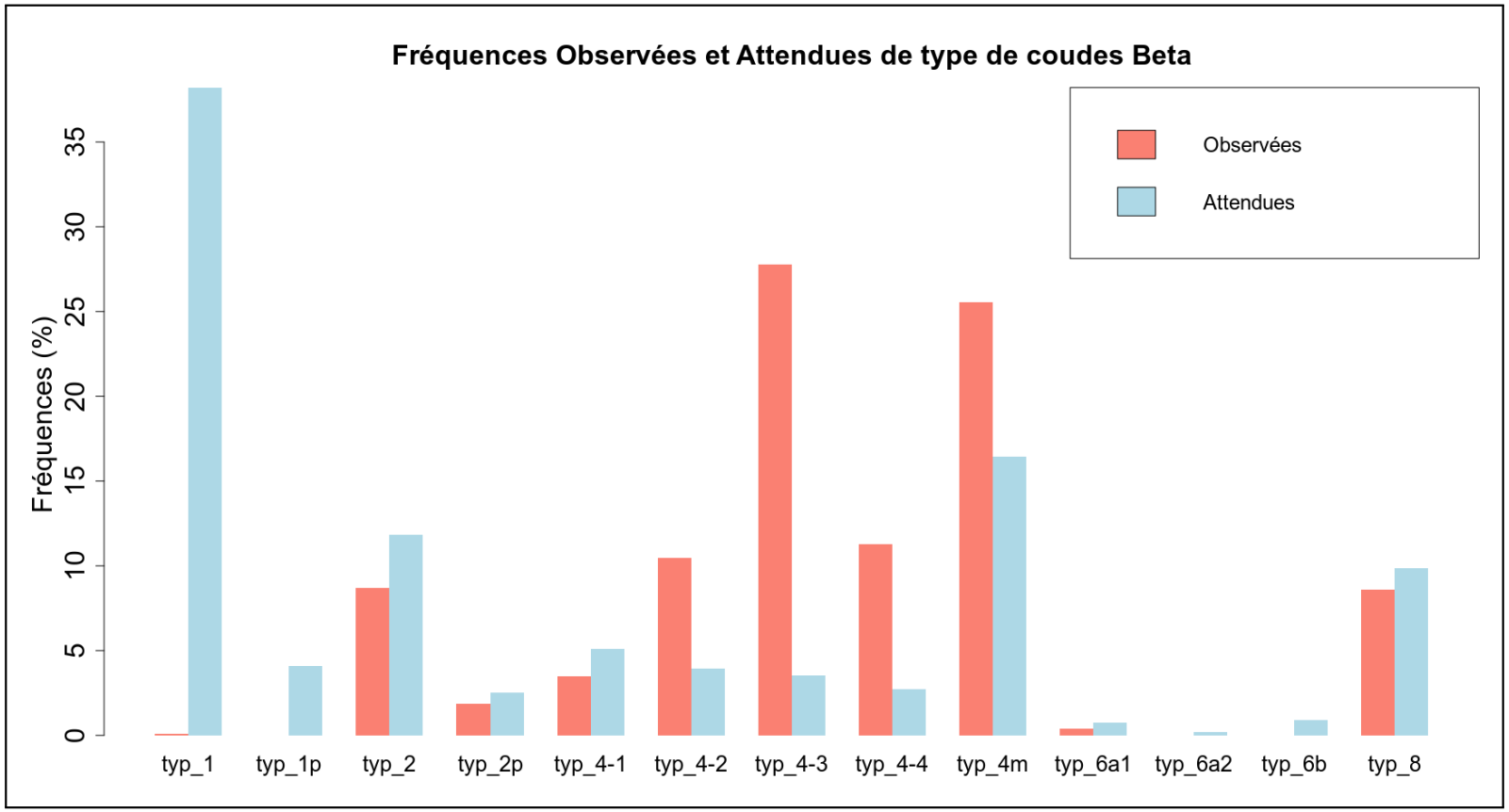
Observed frequencies almost as the expected ones except for Serines, Cysteine and Leucine.

Frequencies



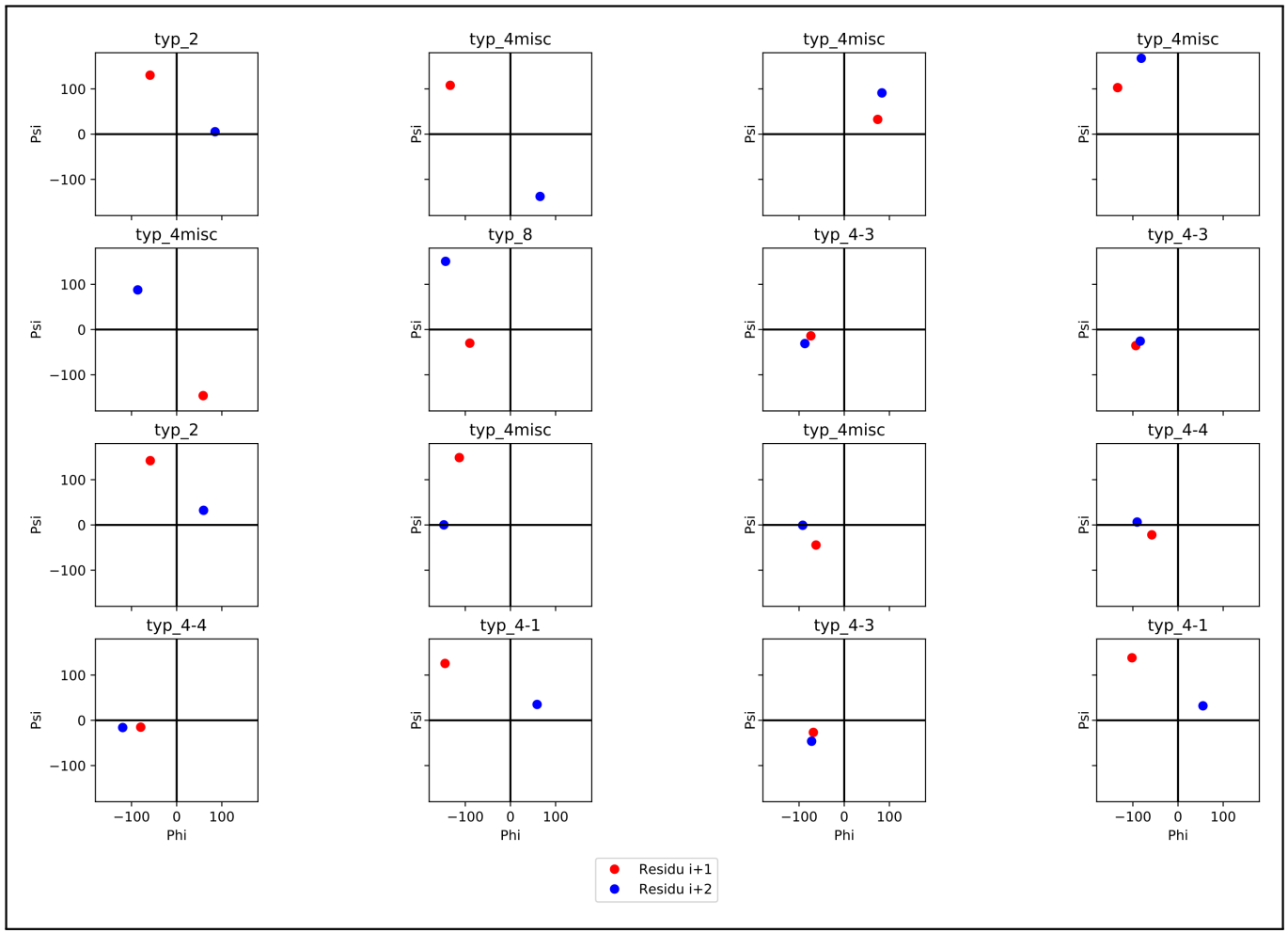
Observed frequencies fitted the expected frequencies.

Frequencies



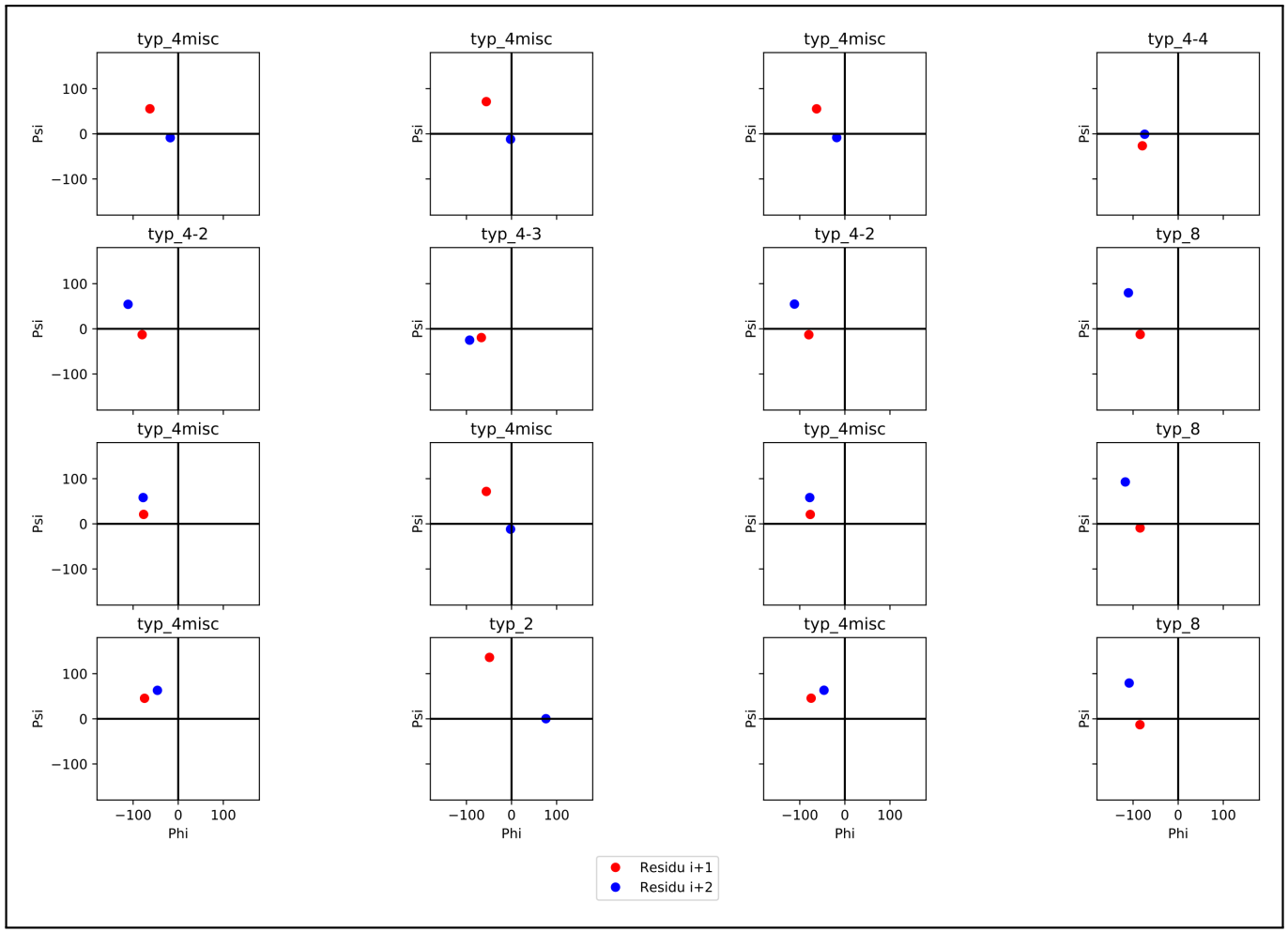
Observed frequencies not following the expected frequencies.

Pre-learning SOM classification



There are types: *II*, IV_1, IV_3, IV_4, IV_{misc} and *VIII*.

Post-learning SOM classification



There are types: II, IV_2, IV_3, IV_4, IV_{misc} and $VIII$.

Discussion

- Differences in types frequences can be due to programmation errors, errors in files or proteins chaîns with shifted frequencies.
- Learning because initialisation values are different from post learning values.
- Few types diversity at the end => may be due to the differences in types proportions in the file
- May be interesting to compare with Fisherman's rule learning method.

References


Ramachandran, G. N. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* 7, 95 – 99.


King, J. L., & Jukes, T. H. (1969). Non-darwinian evolution. *Science*, 164(3881), 788 – 798.


de Brevern, A. G. (2016). Extension of the classical classification of β-turns. *Scientific reports*, 6, 33191.