

Analyse et visualisation du protéome de *Chlamydomonas reinhardtii*

Maël Pretet

Directeur de stage : Marc Baaden

Introduction



Objectifs du stage

—→ Analyser les données protéomiques des cystéines ayant subi une modification post-traductionnelle pour en dégager des pistes de représentation visuelle à travers UnityMol.

Question biologique :

Quels sont les facteurs permettant de déterminer les cystéines modifiées après traduction ?

Chlamydomonas reinhardtii

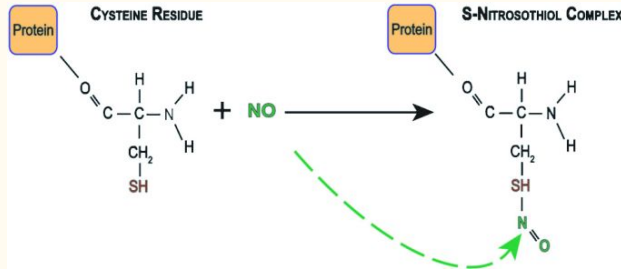
- Organisme modèle
- 1282 structures protéiques
- 501 cystéines modifiées provenant de 286 modèles



<https://eol.org>

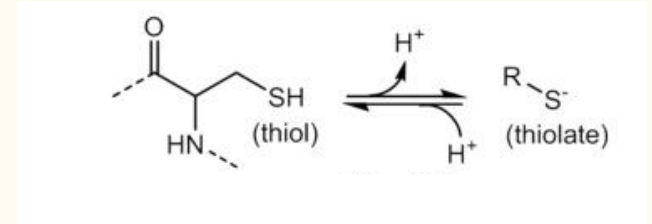
Modifications post-traductionnelle (PTM)

<https://www.researchgate.net>



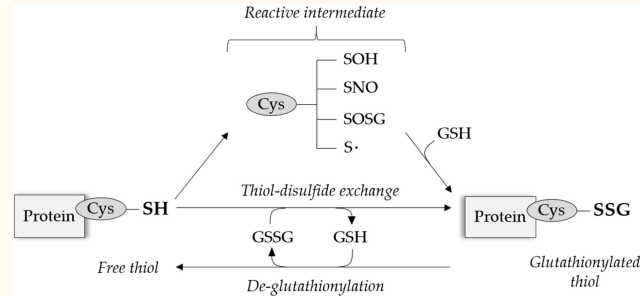
Nitrosylation

<https://www.sciencedirect.com>



Oxydo-réduction

Glutathionylation



<https://www.mdpi.com>

Choix des descripteurs

- **Descripteurs initiaux :**

Access / cys_pos / **modif** / PDBcode / p_id / qmean / cys_modelled / PDB_blast /
p_id_blast / SS / SS_pos / pKa / ASA_cys / ASA_SH / Rossman-like / scop_access /
scop_cys / nb_A / nb_B / patch_score / patch_name / **residue up to 11 Å**

- **Descripteurs supplémentaires :**

Aromatic / polar / aliphatic / charged / negative / positive / hydrophobic / small / tiny

Présentation des résultats



Partitionnement des données

Algorithme k-means

- Minimiser la variance intra-groupe
- 4 centres initiaux

Représentation graphique

- Matrice de distance
- Fonction cmdscale
- Coloration en fonction du cluster

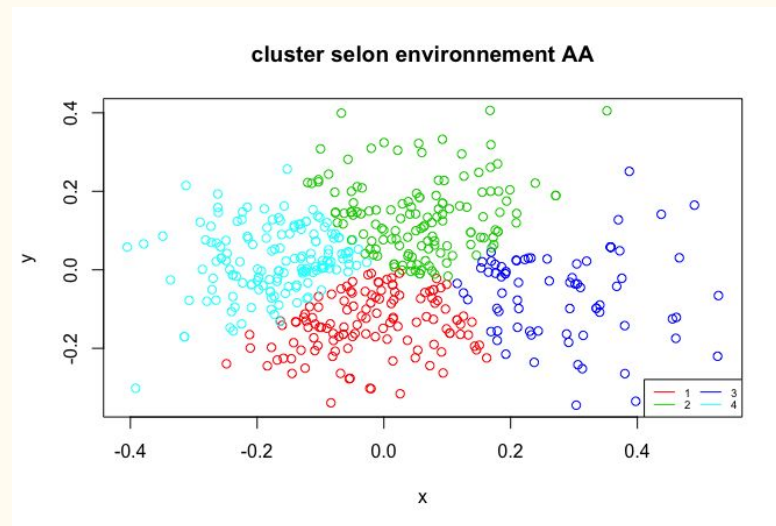


Figure 1 : Projection des individus colorés selon leur cluster

Répartition des PTM dans les clusters

Le nombre de modifications est-il indépendant des clusters ?

- Test du χ^2 , $\alpha = 5\%$

H0 : Facteurs indépendants

H1 : Pas d'indépendance

p-value = 0.0004, H0 rejetée

PTM \ Cluster	Unique	Multiples
1	-0.036	0.094
2	-1.087	2.814
3	-0.017	0.042
4	1.092	-2.826

Tableau 1 : Matrice des résidus

Différences statistiques entre les clusters

Le facteur cluster a-t-il une influence sur les descripteurs ?

- Test ANOVA, $\alpha = 5\%$

H0 : Toutes les moyennes sont égales

H1 : Une moyenne au moins est différente

Différences statistiques entre les clusters

Si conditions d'application non respectées

- Test kruskal-wallis, $\alpha = 5\%$

H0 : Toutes les médianes sont égales

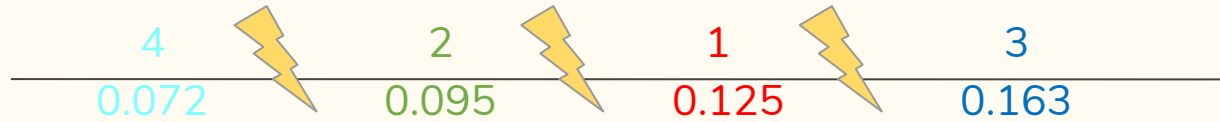
H1 : Une médiane au moins est différente

Tableau récapitulatif des p-value

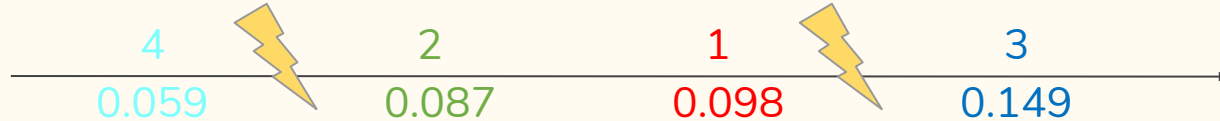
A	4.5×10^{-14}	H	4.2×10^{-3}	T	4×10^{-5}	Negative	0
R	9.1×10^{-14}	I	0	W	0.93	Positive	0
N	0.11	L	0	Y	0.01	Hydrophobic	0
D	3.8×10^{-9}	K	6.7×10^{-13}	V	1.4×10^{-18}	Small	0
C	0.21	M	3.9×10^{-5}	<u>Aromatic</u>	8.8×10^{-5}	Tiny	0
E	0	F	7.7×10^{-9}	<u>Polar</u>	0	<ul style="list-style-type: none"> Bleu : significatif <u>Souligné</u> : ANOVA Non souligné : kruskal-wallis 	
Q	0	P	5.6×10^{-3}	<u>Aliphatic</u>	0		
G	0	S	6.5×10^{-9}	<u>Charged</u>	0		

Résultats post-hoc d'intérêt

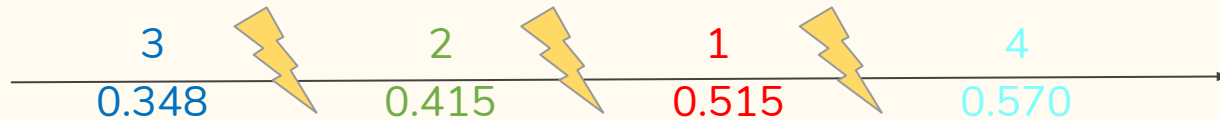
Résidus chargés positivement



Résidus chargés négativement



Résidus hydrophobes



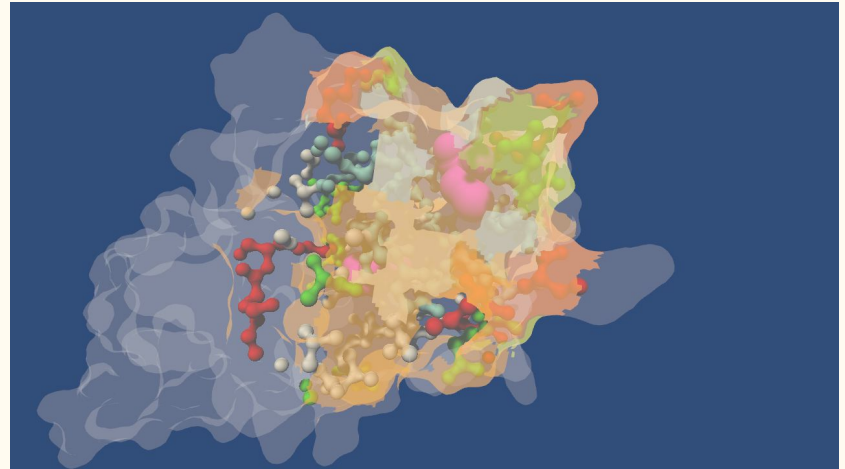
Représentation des propriétés d'intérêt

Choix des couleurs des résidus

Hydrophobe : Bleu

Positifs : Rouge

Négatifs : Vert



Protéine A8IX81, cystéine 164 modifiée

Conclusion

- Difficultés de lier les PTM aux descripteurs de manière directe
- Clustering pour pallier à ce problème
- Permet de pistes de représentation et de visualisation comparative