

PRETET Maël

M1 BI-IPFB

Visualisation de données protéomiques sur les
modifications post-traductionnelles des cystéines dans
l'algue *Chlamydomonas Reinhardtii*

Stage encadré par M. Marc BAADEN

Laboratoire de Biochimie théorique

16 rue Pierre et Marie Curie

2019-2020

Remerciements

Je tiens à remercier M. Xavier MARTINEZ pour son aide lors de mon stage, notamment pour la compréhension du code de UnityMol, ainsi que lors des différentes visio-conférences pour les réflexions autour de l'avancée du stage. Je remercie également Mme Karen DRUART pour m'avoir fourni la base de données de travail, ainsi que son rapport préliminaire décrivant ses analyses et sur lequel j'ai pu me baser. Je remercie également l'ensemble de l'équipe du laboratoire organisant un symposium à distance nous permettant d'être informés sur les travaux de chacun lors de cette période malgré les contraintes engendrées par la situation exceptionnelle de ces derniers mois.

Je remercie enfin tout particulièrement M. Marc BAADEN, ayant été disponible lors de ces trois mois de stage et m'ayant conseillé et guidé pour le bon déroulement de celui-ci. Je le remercie également pour avoir relu ce rapport, et pour l'ensemble de son encadrement pendant cette période.

INTRODUCTION	4
MATERIEL ET METHODES	4
BASE DE DONNEES	4
METHODE K-MEANS.....	5
MATRICE DE DISTANCE ET METHODE CMDSCALE	6
TESTS ANOVA ET POST-HOC.....	6
UNITYMOL.....	7
RESULTATS.....	7
UTILISATION DE LA METHODE K-MEANS	8
ÉVALUATION DES DIFFERENCES ENTRE LES CLUSTERS DETERMINES	9
REPARTITION DU NOMBRE DE MODIFICATIONS DE CHAQUE CYSTEINE EN FONCTION DES CLUSTERS	12
REPARTITION DU NOMBRE DE CYSTEINES MODIFIEES DANS CHAQUE PROTEINE	12
REPRESENTATION VISUELLE DE L'ENVIRONNEMENT DES CYSTEINES	13
CONCLUSION ET DISCUSSION	14
REFERENCES BIBLIOGRAPHIQUES	15

Introduction

Chlamydomonas reinhardtii est un organisme modèle majeure de la biologie végétale mais également cellulaire et moléculaire : cette algue unicellulaire a permis depuis les années 1960 de nombreuses découvertes autour des organismes photosynthétiques, et reste aujourd'hui une référence essentielle pour de nombreuses études grâce à ses propriétés structurales et physiologiques. On peut ainsi citer des études très récentes envisageant la synthèse de plastique cellulaire durable [1] à partir de cet organisme eucaryote. On a ainsi un aperçu de l'importance fondamentale de l'étude de *Chlamydomonas reinhardtii* : on se penche ici sur son protéome, et particulièrement sur les modifications pouvant intervenir au niveau de ses cystéines.

On s'intéresse ici plus précisément à trois modifications post-traductionnelles possibles pouvant affecter les cystéines : la nitrosylation, la glutathionylation et l'oxydo-réduction.

Ces modifications post-traductionnelles (PTM) jouent un rôle essentiel dans l'organisme : la première intervient notamment dans le processus d'apoptose, la seconde dans le stress oxydatif, et donc dans l'apparition de certaines pathologies, comme la maladie d'Alzheimer ou de Parkinson, et enfin la dernière, liée à la formation de ponts disulfures, est ainsi associée aux structures protéiques. Ainsi, ces processus biochimiques sont majeurs dans les fonctions protéiques, et il est donc nécessaire de comprendre au mieux les mécanismes menant à ces modifications.

On cherche ici à orienter les analyses autour de la représentation visuelle. Il est parfois difficile de mettre en évidence uniquement avec des données statistiques des phénomènes biologiques qui pourraient en permettre une meilleure compréhension. Cependant, il est également complexe d'analyser des grands jeux de données protéomiques uniquement à l'œil humain. L'objectif sera donc ici d'associer analyse statistique et piste d'exploration visuelle pour permettre des représentations judicieuses de nos données, toujours axées autour des modifications post-traductionnelles subies par les cystéines.

Matériel et méthodes

[Base de données](#)

On se repose ici sur une base de données de structures protéiques 3D obtenues à partir de séquences expérimentales. Celle-ci contenait 1282 modèles, mais seuls ont été retenues les structures contenant une cystéine modélisée avec un pourcentage d'identité d'au minima 30%. On travaille donc sur 286 modèles, possédant au total 501 cystéines ayant subi des modifications post-traductionnelles, qui seront donc notre échantillon lors des études statistiques. Chaque individu, donc chaque cystéine modifiée, possède 202 descripteurs ayant été calculés. On retrouve notamment des variables

quantitatives indiquant le pkA de la cystéine prédit, son taux de conservation, la surface accessible au solvant, ou encore la structure secondaire impliquée.

Cependant, on décide de s'intéresser dans les analyses suivantes à 180 de ces descripteurs uniquement : le décompte des acides aminés autour de la cystéine d'intérêt à x Å de distance, jusqu'à 11 Å. On choisit de transformer ces variables quantitatives discrètes. On somme pour chaque acide aminé son dénombrement jusqu'à 11 Å, puis on transforme ces sommes en fréquences.

On décide d'ajouter également 9 descripteurs selon certaines propriétés biochimiques regroupant les fréquences de plusieurs acides aminés. Les descripteurs supplémentaires sont donc les suivants (NB : on indique la propriété choisie suivie des acides aminés correspondants en code à trois lettres) :

- Aromatic : His, Phe, Trp, Tyr
- Polar : Asn, Asp, Arg, Cys, Glu, Gln, His, Lys, Ser, Thr, Trp, Tyr
- Aliphatic : Ile, Leu, Val
- Charged : Asp, Arg, Glu, His, Lys
- Negative : Asp, Glu
- Positive : Arg, His, Lys
- Hydrophobic : Ala, Ile, Leu, Met, Phe, Trp, Tyr, Val
- Small : Ala, Asn, Asp, Cys, Gly, Pro, Ser, Thr, Val
- Tiny : Ala, Cys, Gly, Ser

On notera ici que la base de données utilisée ne comporte ni NA, ni descripteur à variance nulle, ni valeurs aberrantes.

On conserve également la variable qualitative indiquant le type de modification subie par les différentes cystéines pour certaines analyses en particulier.

Méthode k-means

La méthode k-means est un algorithme de partitionnement. L'objectif est de minimiser la variance intra-groupe d'un nombre de classes donné k . L'algorithme choisit aléatoirement un nombre de points k comme centres initiaux et calcule ensuite les individus possédant les distances les plus faibles. On calcule ensuite les centres de ces groupes déterminés, puis on détermine de nouveau les individus possédant les distances les plus faibles pour réattribuer les partitions. Ces deux dernières étapes sont ensuite itérées jusqu'à ce que les partitions se stabilisent, ou jusqu'à ce que le nombre d'itérations maximales soit atteint. Cette méthode est réalisée via R : le nombre maximum d'itérations est à dix, et l'algorithme de calcul est celui de Hartigan-Wong¹.

¹ <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/kmeans>

Matrice de distance et méthode cmdscale

On calcule à partir de la fonction *dist()*² de R la matrice de distance entre tous les individus, qui calcule la distance euclidienne entre chaque ligne de la base de données initiales considérée comme un vecteur. Grâce à cette matrice, on peut mesurer les dissimilarités entre chaque cystéine à l'aide de la fonction *cmdscale()*³. Celle-ci retourne un nombre de valeurs pour chaque individu déterminé à l'avance, et reflétant les distances existantes calculées préalablement. En sélectionnant deux valeurs, on peut alors assimiler ces dernières à des coordonnées et les projeter sur un graphique à deux dimensions. On peut alors observer visuellement notre jeu de données et la répartition de nos individus les uns par rapport aux autres.

Tests ANOVA et post-hoc

- Test ANOVA

L'ensemble des tests sera réalisé en choisissant un risque alpha de 5%.

On effectue un test ANOVA pour chaque descripteur avec comme hypothèses :

H0 : Les moyennes de fréquence entre les k groupes sont égales.

H1 : Au moins l'une des moyennes est différente.

Si on trouve une p-value inférieure à 0.05, on rejette H0 et on doit poursuivre l'analyse via un test post-hoc.

Conditions d'application :

- On admet que les données suivent une loi normale.
- On teste l'homoscédasticité avec un test de Bartlett en posant les hypothèses

H0 : L'homoscédasticité est respectée.

H1 : L'homoscédasticité n'est pas respectée.

Si on trouve une p-value inférieure à 0.05, on rejette H0 et les conditions d'application ne sont pas respectées.

- Tests post-hoc


On choisit de représenter les résultats dans un tableau de la manière suivante : chaque ligne contient le descripteur sur lequel on a effectué un test post-hoc, ainsi que les k moyennes de fréquence correspondant à chaque groupe. On range ces moyennes par ordre croissant sur chaque ligne, et on pose pour chaque test de comparaison de moyenne les hypothèses suivantes :

H0 : $\mu_1 = \mu_2$

H1 : $\mu_1 \neq \mu_2$

² <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/dist>

³ <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cmdscale>

Si on trouve une $p\text{-value} < 0.05$, on rejette H_0 , si la $p\text{-value}$ est supérieure à 0.05, on ne rejette pas H_0 au risque $\alpha = 5\%$. On cherche à déterminer s'il existe des groupes significativement différents pouvant être établis, on note donc les résultats significatifs seulement entre deux moyennes consécutives ordonnées avec le symbole suivant :  .

UnityMol

UnityMol est un logiciel de visualisation de structures 3D compatible avec de la réalité virtuelle, et permettant ainsi une meilleure immersion et des manipulations pouvant être simplifiées avec les modèles de protéines. On utilisera ici UnityMol dans sa version dite "desktop" pour visualiser les structures protéiques de l'étude, tout en cherchant à imaginer quels types de représentation pourraient être implémentées dans le logiciel pour nous permettre de manipuler au mieux les données.

Résultats

Avant la présentation des résultats, on rappelle au préalable les objectifs de l'étude et des différentes analyses. Des analyses avaient été effectuées et détaillées dans le rapport de Mme DRUART et montraient les difficultés à mettre en exergue l'implication d'un acide aminé en particulier dans le type de modification observé. Ces études permettent donc d'orienter les suivantes à travers différents points de réflexion. Tout d'abord, il pourrait être intéressant de considérer certains acides aminés de façon plurielle et dépendante, d'où l'idée de regrouper certains descripteurs de la base de données selon leurs caractéristiques physico-chimiques. Ensuite, étant donné le logiciel sur lequel sont observées les structures protéiques, le but est également de jongler entre analyses statistiques et représentation visuelle. Les analyses sont effectuées via R et donc purement calculatoires : les logiques algorithmiques devront donc également être mises en perspective par rapport à la biologie.

La première idée consiste à essayer de représenter les poches permettant d'accéder aux cystéines en "anti-surface". À partir de la représentation en surface calculée, l'objectif est de pouvoir visualiser uniquement l'environnement du solvant pour savoir si les formes de certaines poches pourraient être similaires. Pour vérifier la pertinence de cette idée, on visualise quatre protéines de notre jeu de données via UnityMol.

Les quatre protéines sont sélectionnées selon les critères suivants : on choisit des protéines présentant un pli de Rossmann et le même type de modification post-traductionnelle, à savoir une nitrosylation seule.

Lorsque l'on observe ces quatre protéines possédant une cystéine modifiée, on peut voir que seulement deux de ces cystéines sont exposées. Les deux autres sont complètement enfouies, il semble donc peu judicieux d'effectuer le type de représentation évoqué précédemment. On envisage

cependant l'hypothèse que l'on pourrait séparer préalablement les cystéines en différents groupes pour savoir s'il est possible de restreindre des analyses et des méthodes de visualisation sur un nombre de cystéines limité.

Figure 1 : Représentation des quatre protéines centrées sur la cystéine modifiée. On représente les protéines en surface et en couleur bleu-vert. Les cystéines d'intérêt sont colorées en magenta et représentées en hbond.



Figure 1.a : A8JH72

Figure 1.b : P09205

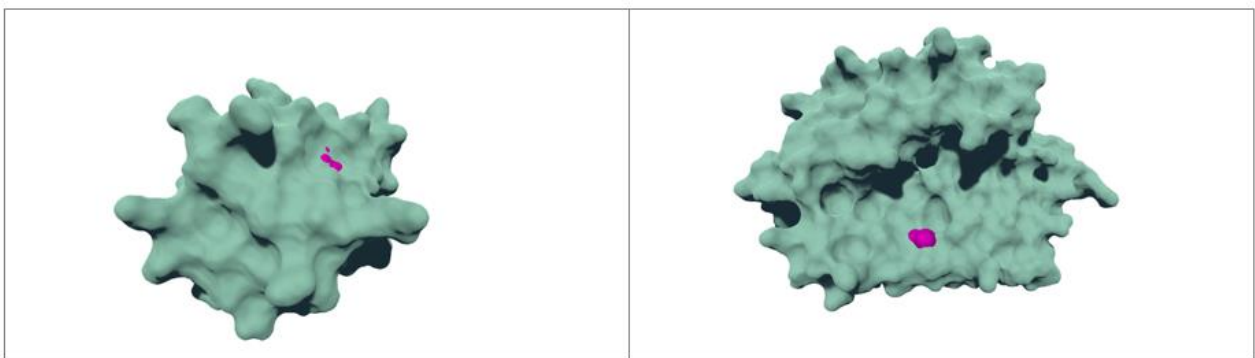


Figure 1.c : Q39580

Figure 1.d : Q4U0V8

Utilisation de la méthode k-means

On cherche donc ici à savoir s'il est possible de déterminer naïvement différents groupes entre les individus sur lesquels nous travaillons à partir de la méthodes k-means. On estime la somme totale

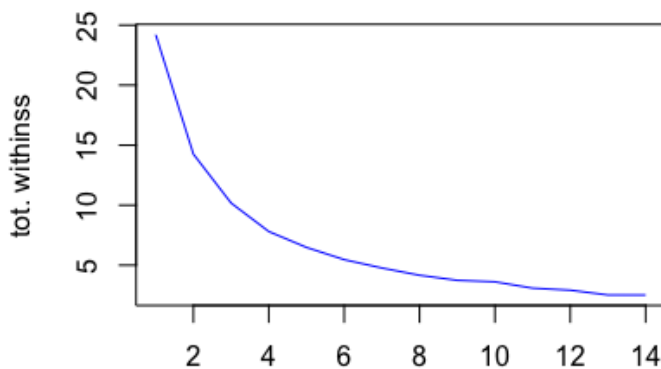


Figure 2 : Somme des carrés totale en fonction du nombre de groupes

des carrés (figure 5) ainsi que l'évaluation visuelle pour choisir le nombre de groupes calculé par l'algorithme. On choisit selon ces deux critères d'effectuer un k-means à 4 groupes.

Pour représenter les résultats, on applique la méthode cmdscale sur la matrice de distance calculée à partir de la base de données pour projeter les individus sur un graphique à deux dimensions. Après les avoir projetés, on les colore selon les clusters calculés pour visualiser si l'algorithme a pu efficacement distinguer les groupes.

On observe sur la figure 6 qu'on peut visualiser distinctement les quatre clusters calculés selon les descripteurs retenus. On cherche alors à savoir quels sont les descripteurs pour lesquels la moyenne est significativement différente entre ces groupes.

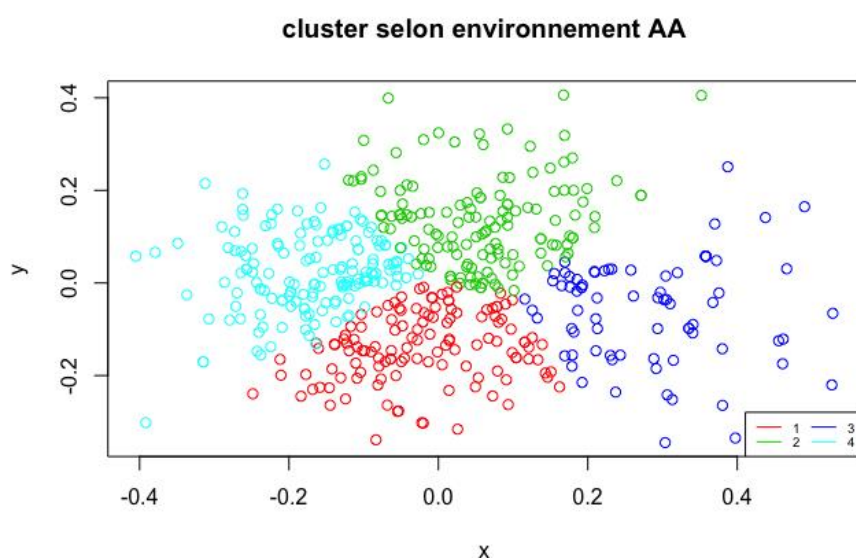


Figure 3 : Projection des individus colorés selon les clusters. Groupe 1 : 128 individus, groupe 2 : 139 idv. , groupe 3 : 76 idv. , groupe 4 : 158 idv.

Évaluation des différences entre les clusters déterminés

- Test ANOVA

On trouve une p-value < 0.05 pour les descripteurs suivants : Arg, Asn, Asp, Cys, Glu, Gln, Gly, His, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val, charged, negative, positive, hydrophobic, small, tiny.

Pour ces descripteurs ne respectant pas les conditions d'application, on effectue un test de kruskal wallis non paramétrique en gardant les mêmes hypothèses de départ, et on conserve le test ANOVA pour les autres. On reporte les p-value de ces tests dans le tableau 2. Les p-values inférieures à 2.2×10^{-16} sont notées 0, et les descripteurs pour lesquels les tests sont significatifs sont colorés en rouge.

Tableau 1 : P-value des tests pour chaque descripteur

A	4.5x10 ⁻¹⁴	H	4.2x10 ⁻³	T	4x10 ⁻⁵	Negative	0
R	9.1x10 ⁻¹⁴	I	0	W	0.93	Positive	0
N	0.11	L	0	Y	0.01	Hydrophobic	0
D	3.8x10 ⁻⁹	K	6.7x10 ⁻¹³	V	1.4x10 ⁻¹⁸	Small	0
C	0.21	M	3.9x10 ⁻⁵	Aromatic	8.8x10 ⁻⁵	Tiny	0
E	0	F	7.7x10 ⁻⁹	Polar	0		
Q	0	P	5.6x10 ⁻³	Aliphatic	0		
G	0	S	6.5x10 ⁻⁹	Charged	0		

Pour les descripteurs dont les tests ont été significatifs, on effectue des tests post-hoc de comparaison de moyenne deux à deux (paramétriques pour les descripteurs sur lesquels on a effectué un test ANOVA, et non paramétriques lorsque l'on a choisi d'opérer un test de kruskal wallis).

- Tests post-hoc

Tableau 2 : Résultats des tests post-hoc

A	1 0.075	3 0.082	2 0.103	4 0.121
R	4 0.029	2 0.041	1 0.050	3 0.070
D	4 0.033	1 0.049	2 0.052	3 0.081
E	4 0.025	2 0.035	1 0.049	3 0.068
Q	4 0.027	2 0.032	1 0.043	3 0.060
G	1 0.054	3 0.055	4 0.077	2 0.115
H	4 0.016	2 0.020	1 0.022	3 0.030
I	2 0.050	3 0.051	4 0.080	1 0.088
L	3 0.066	2 0.075	1 0.123	4 0.124
K	4 0.027	2 0.035	1 0.052	3 0.062
M	3 0.016	2 0.026	1 0.029	4 0.033
F	2 0.035	3 0.036	4 0.051	1 0.059
P	1 0.030	4 0.039	2 0.048	3 0.049

S	1 0.038	4 0.050	3 0.050	2 0.071
T	1 0.045	4 0.050	3 0.062	2 0.070
Y	2 0.025	1 0.032	4 0.033	3 0.038
V	3 0.048	2 0.091	1 0.099	4 0.121
Aromatic	2 0.089	4 0.107	3 0.114	1 0.123
Polar	4 0.355	1 0.442	2 0.457	3 0.596
Aliphatic	3 0.166	2 0.216	1 0.310	4 0.325
Charged	4 0.130	2 0.183	1 0.223	3 0.312
Negative	4 0.059	2 0.087	1 0.098	3 0.149
Positive	4 0.072	2 0.095	1 0.125	3 0.163
Hydrophobic	3 0.348	2 0.415	1 0.515	4 0.570
Small	1 0.442	3 0.491	4 0.549	2 0.617
Tiny	1 0.186	3 0.209	4 0.272	2 0.313

On peut dégager quelques différences significatives au niveau des clusters déterminés :

- Les quatre clusters sont tous significativement différents les uns des autres en ce qui concerne les propriétés d'hydrophobicité, de charges globales et positives, ainsi que pour les résidus de petite taille (small).
- Il pourrait ainsi être intéressant d'exploiter certaines de ces propriétés pour visualiser de manière différentielle les cystéines, en s'appuyant sur les propriétés biochimiques évoquées précédemment, mais également en revenant peut-être dans un second temps sur certains acides aminés en particulier dont les fréquences ont été calculées comme statistiquement significative entre les différents groupes (par exemple les résidus arginine ou glutamate).

On cherche cependant à savoir également si cette séparation en différents clusters pourrait être corrélées avec notre première variable d'intérêt, à savoir le type de modification, et donc si les propriétés de la composition de l'environnement des cystéines pourraient être corrélées avec le type de modification observable.

Répartition du nombre de modifications de chaque cystéine en fonction des clusters

On cherche à savoir si les variables "types de modification" et "clusters" sont indépendantes. Pour cela on réalise un test du Chi 2 d'indépendance au risque $\alpha = 5\%$. On pose les hypothèses suivantes :

H0 : Les deux variables sont indépendantes.

H1 : Il n'y a pas d'indépendance.

Pour respecter les conditions d'application (tous les $C_{ij} > 5$), sur les sept catégories de modification post-traductionnelle (NRG, NRg, NrG, Nrg, nRG, nRg, nrG), il est nécessaire d'en regrouper certaines. On décide pour que le test reste cohérent biologiquement de regrouper les modifications uniques d'un côté (Nrg, nRg et nrG), et les modifications multiples de l'autre (NRg, nRG, NrG et NRG). On reformule donc la question biologique posée, à savoir : est-ce que les variables "nombre de modification (unique ou multiple)" et "clusters" sont indépendantes ?

On trouve une p-value de 0.0004, donc inférieure à 0.05. On peut donc rejeter H0 au risque $\alpha = 5\%$: les deux variables ne sont pas indépendantes et il existe des différences de proportion entre les catégories. On visualise la matrice de résidus :

Modifications \ groupes	1	2	3	4
Modifications unique	-0.036	-1.087	-0.017	1.092
Modifications multiples	0.094	2.814	0.042	-2.826

On peut ainsi noter que le groupe 4 contient une proportion de cystéines ayant subi une seule modification plus importante qu'attendue, et une proportion de cystéines ayant subi plusieurs modifications moins importantes, et qu'on observe l'inverse pour le groupe 2.

Répartition du nombre de cystéines modifiées dans chaque protéine

On travaille ici sur le nombre de cystéines modifiées par protéine. On distingue d'un côté les cystéines étant les seules de la protéine à avoir été modifiées, et de l'autre les cystéines présentes dans des protéines possédant plusieurs cystéines modifiées. On essaye de savoir si la variable "cluster" est indépendante de cette première variable. On effectue donc de nouveau un test du Chi 2 d'indépendance en posant les mêmes hypothèses que précédemment.

H0 : Les deux variables sont indépendantes.

H1 : Il n'y a pas d'indépendance.

Tous les C_{ij} sont supérieurs à 5. On calcule donc la p-value, et on trouve une valeur de 0.0342, inférieure à 0.05. On rejette H0 au risque $\alpha = 5\%$, et on observe la matrice de résidus :

Cystéines par protéines \ groupes	1	2	3	4
Uni-prot	1.560	0.081	0.404	-1.760

Multi-prot		-1.123		-0.058		-0.291		1.267
------------	--	--------	--	--------	--	--------	--	-------

Le groupe 4 contient une proportion de cystéines présentes dans une protéine possédant plusieurs cystéines modifiées plus importante, et moins importante en ce qui concerne les cystéines ayant été les seules à être modifiées dans la protéine à laquelle elles appartiennent. On note l'inverse dans le groupe 1.

Représentation visuelle de l'environnement des cystéines

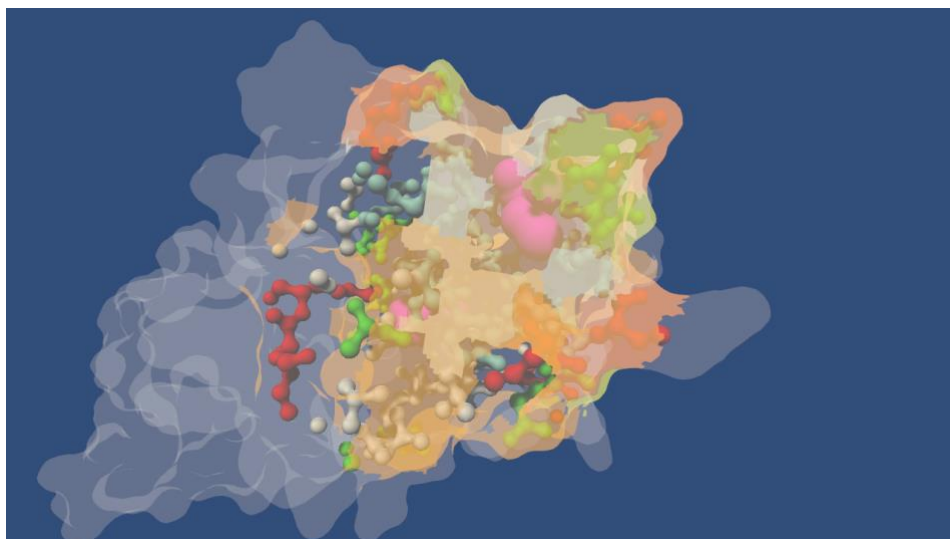


Figure 4 : Protéine A8IX81, la cystéine modifiée étant le résidu 164

On propose ici un type de représentation à l'aide de UnityMol de l'environnement de la cystéine à 11 Å de distance :

- Les résidus chargés positivement sont représentés en sticks rouges et en surface transparente rouge
- Les résidus chargés négativement sont représentés en sticks verts et en surface transparente verte
- Les résidus hydrophobes sont représentés en sticks bleus et en surface transparente bleue
- Les autres acides aminés sont représentés en sticks blancs et en surface transparente orange
- Le reste de la protéine est représentée en surface transparente blanche.

L'objectif est à partir de ce type de représentation d'aligner l'ensemble des environnements des cystéines les uns avec les autres. Cependant, les premières expérimentations alignant ces environnements selon les groupes auxquels appartiennent les cystéines se sont avérées illisibles visuellement, allant donc à l'encontre de l'objectif initial. Il pourrait donc être intéressant pour des analyses futures de réajuster ce type de représentation pour permettre une meilleure exploitation. Cependant, même si dans un premier temps la première idée serait de visualiser distinctement chaque type de descripteur physico-chimique indépendamment, on doit discuter de la pertinence de cet arrangement : serait-il toujours aussi judicieux de représenter isolément ces descripteurs quand l'objectif initial était de les rassembler pour que l'algorithme puisse déterminer des clusters distincts ?

Conclusion et discussion

On a ainsi pu noter au cours de ces différentes analyses qu'à partir de la détermination de clusters naïves via des méthodes algorithmiques, on a pu faire ressortir certaines différences de propriété que l'on peut ensuite corrélés indirectement à des différences de répartition des types de modification. On notera notamment le groupe 4, ressortant dans les différentes analyses. Les cystéines de ce cluster semblent ainsi posséder un environnement plus hydrophobe, et ainsi moins polaire et moins chargé. On note également que les cystéines de ce groupe subissent plus souvent un seul type de modifications post-traductionnelle, et qu'elles sont impliquées de manière plus importante dans des protéines possédant plusieurs cystéines modifiées. Ainsi, même si quatre groupes distincts ont été calculés par l'algorithme, il semble apparaître que l'un d'entre eux se détache particulièrement des autres en termes d'intérêt biologique. Il pourrait donc être intéressant d'approfondir les études autour de ce cluster en particulier en l'opposant à l'ensemble des autres cystéines, ou au contraire en le considérant individuellement et en élargissant les analyses possibles lorsqu'elles sont effectuées sur une catégorie de cystéine précise.

Ces différentes pistes ont ainsi permis d'explorer des possibilités de représentation visuelle comparatives des différents environnements des cystéines. Cette réflexion se basait au départ sur le fait que les acides aminés pris indépendamment autour des cystéines ne permettaient pas d'orienter de manière suffisamment précise les caractéristiques pouvant provoquer une modification post-traductionnelle. On s'est donc intéressé à l'environnement dans sa globalité pour dégager de nouvelles observations. Cependant, il est important de notamment garder à l'esprit deux points capitaux lors de ces analyses. Tout d'abord, les algorithmes utilisés, notamment la méthode k-means, possède une étape se basant avant tout sur l'aléatoire. Même si celui-ci est prévu pour garder une certaine stabilité lors de la reproductibilité de ce type d'expérience, une part de hasard inhérente au processus peut provoquer une certaine variabilité. On doit également noter qu'on a pris en compte ici la fréquence totale des acides aminés présents dans l'environnement des cystéines : les notions de positionnement dans l'espace et d'orientation ne sont donc pas prises en compte dans ces calculs, mais doivent par conséquent être approfondies, notamment lors de l'étude visuelle.

Références bibliographiques

- [1] Nakanishi, A., Iritani, K., Sakihama, Y. *et al.* Construction of cell-plastics as neo-plastics consisted of cell-layer provided green alga *Chlamydomonas reinhardtii* covered by two-dimensional polymer. *AMB Expr* **10**, 112 (2020). <https://doi.org/10.1186/s13568-020-01046-y>
- [2] Maes A, Martinez X, Druart K, et al. MinOmics, an Integrative and Immersive Tool for Multi-Omics Analysis. *J Integr Bioinform*. 2018;15(2):20180006. Published 2018 Jun 21. doi:10.1515/jib-2018-0006
- [3] Raldúa, D., Casado, M., Prats, E. *et al.* Targeting redox metabolism: the perfect storm induced by acrylamide poisoning in the brain. *Sci Rep* **10**, 312 (2020). <https://doi.org/10.1038/s41598-019-57142-y>
- [4] Xiong Y, Uys JD, Tew KD, Townsend DM. S-glutathionylation: from molecular mechanisms to health outcomes. *Antioxid Redox Signal*. 2011;15(1):233-270. doi:10.1089/ars.2010.3540
- [5] Dou, X., Li, S., Hu, L. *et al.* Glutathione disulfide sensitizes hepatocytes to TNF α -mediated cytotoxicity via IKK- β S-glutathionylation: a potential mechanism underlying non-alcoholic fatty liver disease. *Exp Mol Med* **50**, 7 (2018). <https://doi.org/10.1038/s12276-017-0013-x>
- [6] Munkanatta Godage, D.N.P., VanHecke, G.C., Samarasinghe, K.T.G. *et al.* SMYD2 glutathionylation contributes to degradation of sarcomeric proteins. *Nat Commun* **9**, 4341 (2018). <https://doi.org/10.1038/s41467-018-06786-x>
- [7] Savojardo C, Fariselli P, Alhamdoosh M, Martelli PL, Pierleoni A, Casadio R. Improving the prediction of disulfide bonds in Eukaryotes with machine learning methods and protein subcellular localization. *Bioinformatics*. 2011;27(16):2224-2230. doi:10.1093/bioinformatics/btr387
- [8] Ozawa, K., Tsumoto, H., Miura, Y. *et al.* DJ-1 is indispensable for the S-nitrosylation of Parkin, which maintains function of mitochondria. *Sci Rep* **10**, 4377 (2020). <https://doi.org/10.1038/s41598-020-61287-6>

Résumé en français

Le rapport présente ici une étude sur le protéome de l'organisme *Chlamydomonas reinhardtii*, et plus précisément sur les modifications post-traductionnelles subies par les cystéines. Celle-ci se base sur des études préliminaires effectuées sur des modèles 3D calculés, et tente d'explorer et d'approfondir certaines des pistes ayant été présentées. On cherche plus spécifiquement ici à étudier l'environnement dans un rayon de 11 Å de ces cystéines, pour déterminer si certaines de ses propriétés pourraient être corrélées aux types de modification observées. L'objectif est dans un second temps de déterminer si ces dites propriétés pourraient être intéressantes à mettre en évidence de façon visuelle pour permettre un comparatif visible à l'œil nu, et plus précisément dans un environnement de réalité virtuelle.

English summary

This report presents a study on the proteome of *Chlamydomonas reinhardtii*, and more specifically on the post-translational modifications in its cysteines. It is based on preliminary studies carried out on 3D computed models, and attempts to explore some of the approaches that have been presented. More precisely, the aim is to study the environment within 11 Å of these cysteines, to determine if some of its properties could be correlated with the types of modification observed. The objective is then to determine if these properties could be interesting to highlight visually to allow a comparison visible to the naked eye, and more precisely in a virtual reality environment.