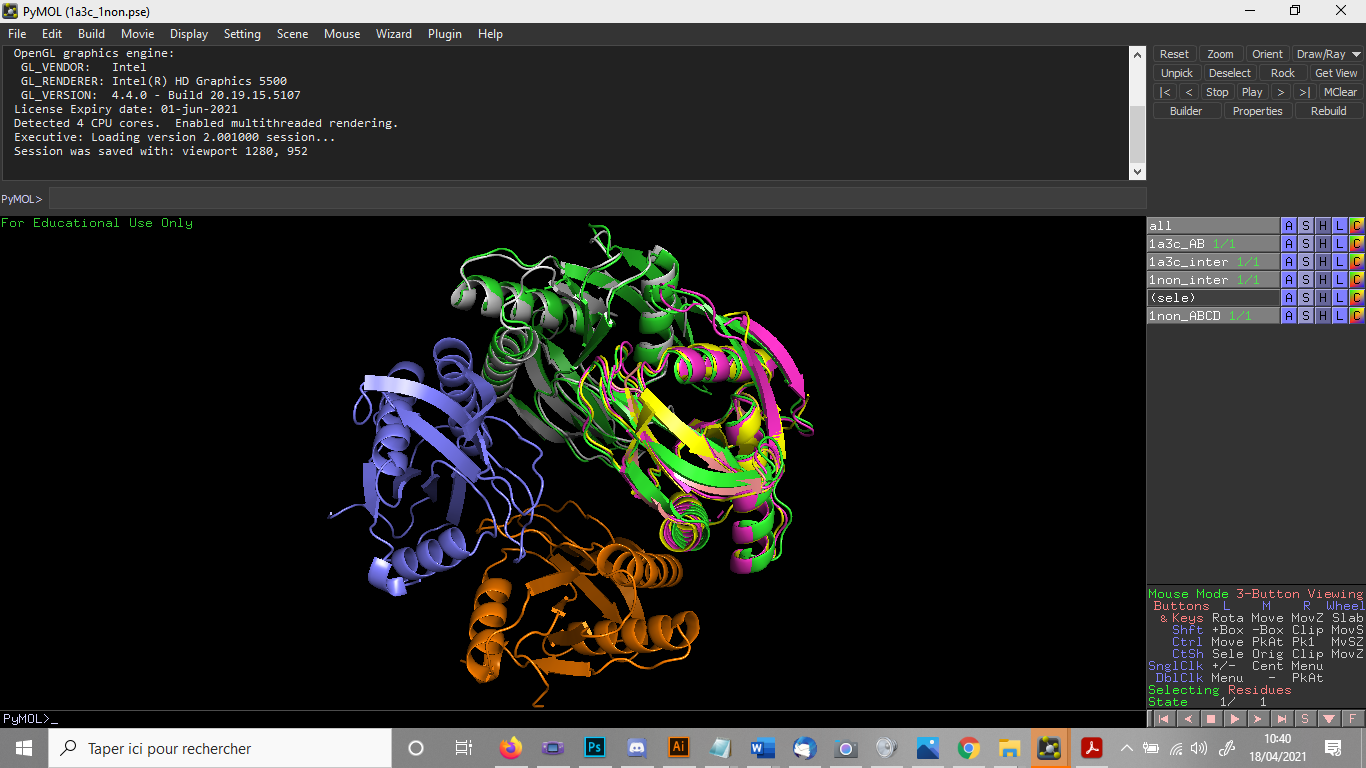
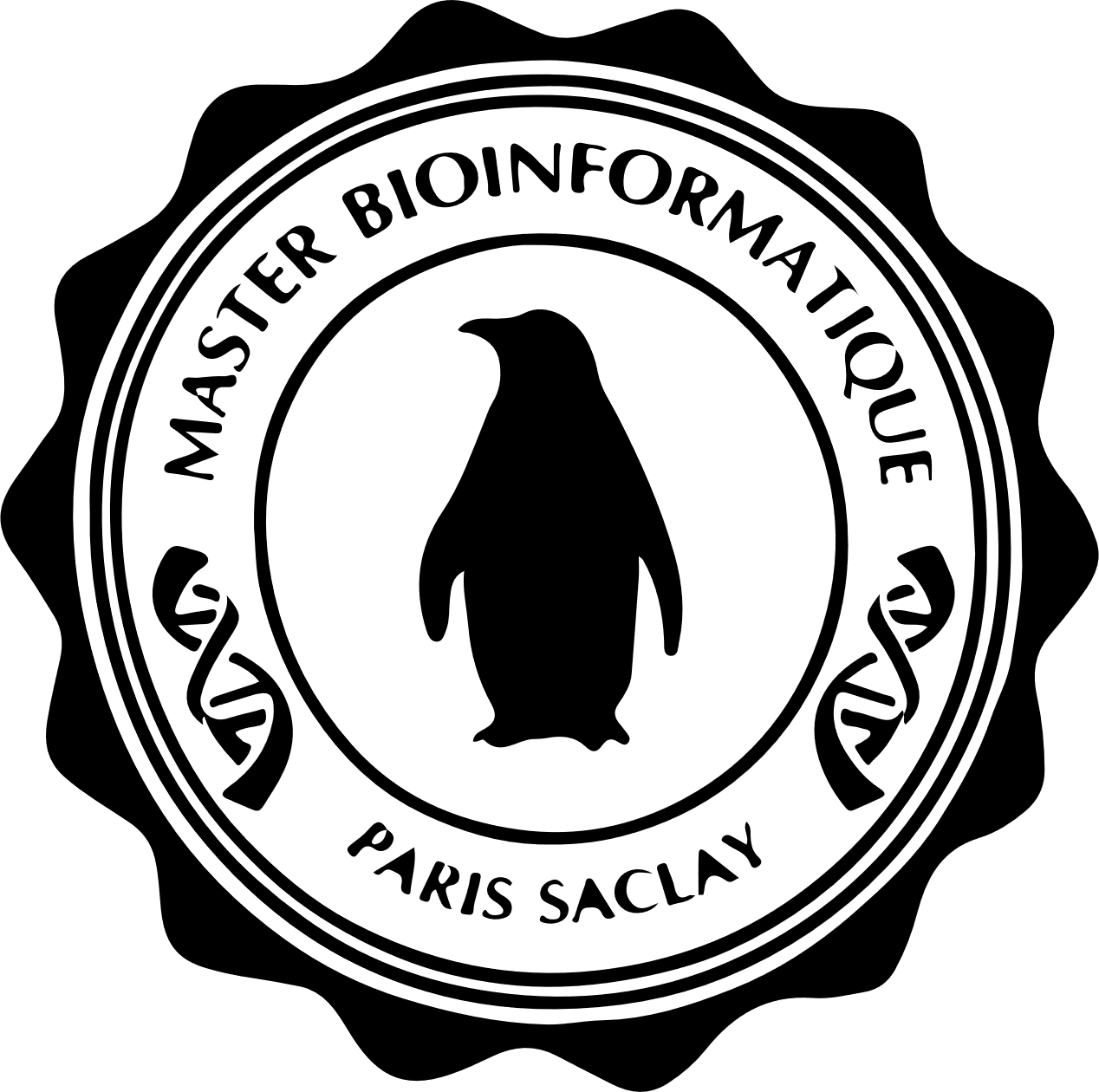
Projet Dimères/Tétramères

# UE BIOINFO- STRUCTURALE II

### MASTER BIBS

## Julia von Grafenstein et Virginie Noël







# INTRODUCTION

Un homéomère peut être décrit comme étant une structure quaternaire protéique formée par un groupe d’une protéine se liant à des copies d’elle-même. On peut observer ce phénomène dans les 60% des structures quaternaire protéiques connues. Malgré cette prévalence, tout n’est pas connu sur les mécanismes d’évolution menant à la création des homéomères.

Un article publié par Levy et all en 2008 a tenté d’explorer cette dernière question avec une analyse de plus de 2000 structures différentes. Entre autres, elle traite des tétramères, qui sont des homéomères composés de 4 sous-unités et des dimères, avec 2 sous-unités.

Les auteurs de l’article proposent l’hypothèse d’une évolution à temps biparti des tétramères, le modèle de la création des deux types d’interfaces est donc successive et non simultanée. On peut alors se demander si ces deux interfaces, qu’on nommera ancestrale et récente selon l’ancienneté de leur apparition, ont des propriétés différentes et dans ce cas, permettent-elles de comprendre les déterminants structuraux ayant permis l'émergence des premiers dimères puis des tétramères ?

On tentera au cours de ce rapport d’explorer ces interrogations, avec tout d’abord une description des données utilisées, les manipulations mises en œuvre sur ces données pour montrer leurs propriétés et comment ces dernières ont été analysées. Ensuite, avec une description des résultats de l’analyse, suivie de leur interprétation et finalement une conclusion.

# MATERIELS ET METHODES

## Données

Le jeu de données utilisé est composé de structures en format PDB avec 21 couples d’homéomères, composés chacun d’un dimère et de son tétramère homologue. En plus des données structurales, sont aussi fournis des fichiers PDB dans lesquels le bfactor prend les valeurs 0, 1 et 2 selon que les acides aminés appartiennent à une interface et si c’est le cas, selon le type d’interface récente (2) ou ancestrale (1).

[capture d’écran]

## Manipulation des données

Les scores de conservation des acides aminés (score consurf) ont été obtenus pour la chaîne A de chaque dimère et tétramère grâce au site <https://consurf.tau.ac.il/>. Les paramètres n’ont pas été modifiés. Le script Python colour\_from\_consurf\_scores.py a été utilisé pour obtenir les commandes permettant de visualiser les scores consurf de chaque homéomère sur Pymol.

Le script Python écrit lors des cours pour générer les fichiers csv de la séance 2 a été remanié et réutilisé pour obtenir des fichiers similaires pour chaque monomère. Les fichiers obtenus contiennent des informations sur la taille de la surface et de l’interface, leur composition en acides aminés et le nombre des types d’acides aminés hydrophobes, polaires ou chargés, ainsi que les données des scores moyens consurf des interfaces.

[Calcul propriete]

## Analyse stastique

Ensuite nous avons traiter les données pour faire différents graphiques et fait des analyses et tests stastiques avec le language de programmation R.

Vu que on s’interesse à si il y a des differences entres les propriétés des interfaces ancenstrales et les interfaces récentes nous avons décider de faire des test stastiques de comparaison de moyennes.

Ces tests sont tous sous la forme :

Avec , les moyenne de deux variables aletoires X1 et X2 . On rejette H0 lorsque on p-value 0.05. Lorsque X1 et X2 suivent tout les deux une loi gaussienne et que leurs deux variances sont inconnues mais supposés égales, ont peut emploie le test Student, qui se fait avec t.test() sur R. Si X1 et/ou X2 ne suit pas une loi gaussienne, un test de comparaison non parametriques peut être alors fait un test de Wilcoxon, qui se fait avec wilcoxon.test() sur R.

Afin de savoir quels test pour chacune de nos variables, nous observerons les variances observes des échantillons x1 et x2, aussi nous ferons un test de normalite, ou test de Shapiro ,sur X1 puis X2.Si les variances de x1 et x2 sont proches et que chacun passe le test de Shapiro, nous saurons qu’on pourra faire le test de Student sur x1 et x2. Sinon nous ferons le test de wilcoxon.

Le test de Shapiro à comme H0 : la distribution X ne devie pas de la loi normale, h1 : la distribution X devie de la loi normale, , les moyenne de deux variables aletoires X1 et X2 . On rejette aussi H0 lorsque on p-value 0.05.

# RESULTATS

## Scores consurf

[parler visuelement ce qu’on voit dans les pymol, wow en general interface plus conserve que surface]

Nous nous sommes interesse à la moyenne des scores consurf des interfaces ancencestrales et les interfaces récentes.

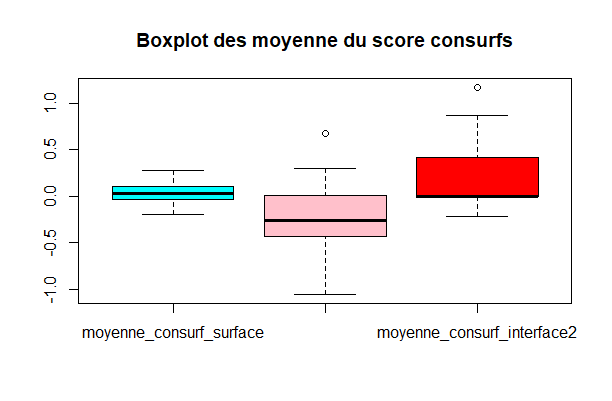


Figure 1 : Distributions sur la forme de boite à moustache des moyennes des scores consurf. En cyan, les surface, en rose les interfaces ancestrales et en rouges les interfaces récentes.

Visuellement nous observer que la moyenne pour les interfaces ancestrales est petite que celles pour les interfaces récentes.

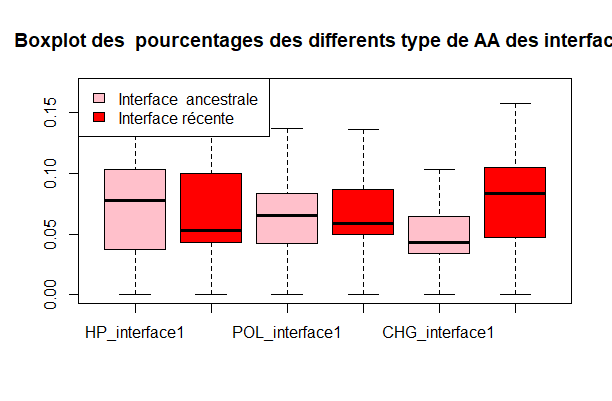
Tableau 1 :

|  |  |  |
| --- | --- | --- |
|  | Interfaces ancestrales | Interfaces récentes |
| Variance | 0.119045138 | 0.102997607 |
| P-value test Shapiro | 0.4692 | 8.882e-06 |
| P-value test Wilcoson | 1.047e-06 | |

## Tailles

|  |  |  |
| --- | --- | --- |
|  | Interfaces ancestrales | Interfaces récentes |
| Variance | 0.014831556 | 0.003191318 |
| P-value test Shapiro | 8.042e-06 | 0.5427 |
| P-value test Wilcoson | 6.528e-05 | |

## Types acide amines



### Acides amines polaires

|  |  |  |
| --- | --- | --- |
|  | Interfaces ancestrales | Interfaces récentes |
| Variance | 0. 0012393572 | 0. 0012330457 |
| P-value test Shapiro | 0.2404 | 0.4001 |
| P-value test Student | 0.7411 | |

### Acides amines chargés

|  |  |  |
| --- | --- | --- |
|  | Interfaces ancestrales | Interfaces récentes |
| Variance | 0. 0011850941 | 0. 0056838331 |
| P-value test Shapiro | 0.004664 | 0.0003848 |
| P-value test Wilcoson | 0.04105 | |

### Acide amines hydrophobes

|  |  |  |
| --- | --- | --- |
|  | Interfaces ancestrales | Interfaces récentes |
| Variance | 0.0021910020 | 0.0021910020 |
| P-value test Shapiro | 0.2215 | 6.684e-06 |
| P-value test Wilcoson | 0.5448 | |

## Ponts Salins

# CONCLUSION

# ANNEXE