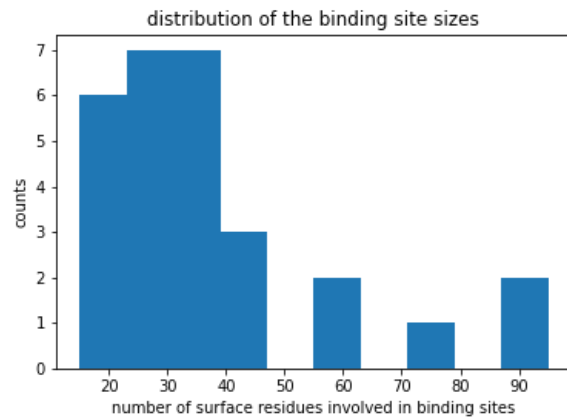


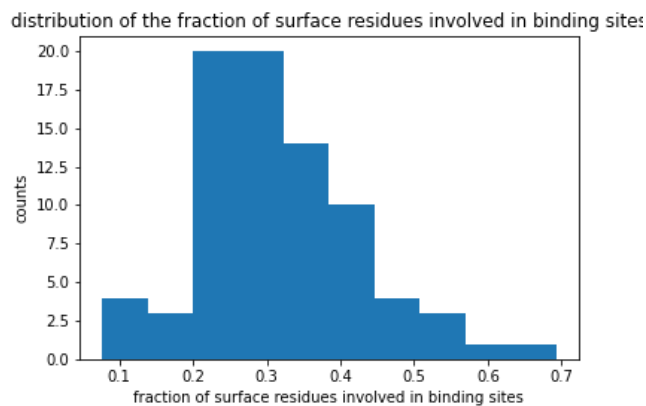
## Séances 1, 2 voir 3: caractérisation des sites d'interaction (BS binding sites) VS surface

- analyse de descripteurs choisis par les étudiants, qu'ils évalueront avec leurs scripts python sur un complexe (CV, taille des BS, propensions des 20 aa à être trouvés dans les BS vs surfaces, nb hydrophobes/chargés/polaires, autres idée ? On a parlé de fréquence de contacts, c'est peut-être un peu plus complexe à calculer et surtout il faut automatiser pour avoir suffisamment de données mais ce serait bien qd même, ou alors à aborder après pour commencer simplement déjà et qu'ils prennent en main les pdb, pymol et leurs scripts)
- Selon leur rapidité ils feront tourner leurs scripts sur un ou deux complexes ou sur l'ensemble des complexes en automatisant l'analyse mais comme l'a bien fait remarquer Marco, c'est pas si évident pour ceux qui ne sont pas encore très à l'aise avec python, la gestion des dir, des fichiers + le fait que cela demande un minimum d'organisation. Je ne me rendais pas compte au départ, mais en le faisant sur les complexes permanents et transitoires, je me suis en effet rendu compte que cela risque de leur prendre bcp de temps et ce n'est pas une UE de programmation. Du coup, l'idée est que tout le monde réussisse à évaluer au moins un petit set de descripteurs, vérifie leurs résultats sur ppymol et participe à la discussion en comprenant bien de quoi on parle puisqu'ils auront manipulé les concepts de façon concrète.
- En fin de séance 2 ou fin de séance 3, ils peuvent présenter en binôme leurs résultats via un petit ppt, les faire présenter régulièrement les aide bcp à formaliser et les rend actifs dans la discussion et dans l'ensemble du cours.
- Ensuite on peut présenter les résultats qu'on obtiendrait en lançant les calculs sur toute notre DB. Voilà slice suivant, les résultats obtenus sur le jeu de complexes dits « solubles »
- PS : je pense qu'en fait 3 séances seront nécessaires car pour définir les BS, il va falloir soit qu'ils utilisent la CV soit le calcul d'ASA via freesasa par ex et c'est important qu'ils prennent bien le temps de le faire en comprenant bien ce qu'ils font (montrer les vieux papiers de calcul d'ASA pour la définition de surface et interface ou le JMB2010 d'Emmanuel Levy). Puis qu'ils codent leur scripte, qu'ils valident en regardant sur pymol ce qui leur fera manipuler pymol. Ca peut faire l'objet de la première séance (2H environ) et ce sera assez ludique

## Distribution de la taille des sites d'interaction



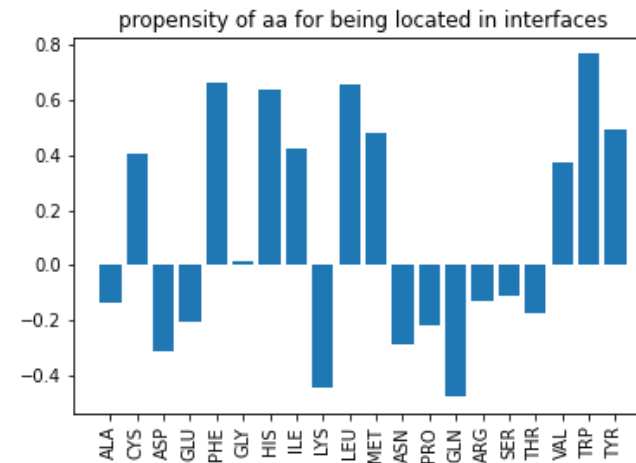
## Distribution de la taille des sites d'interaction (normalisée par le nb de résidus de surface)\*



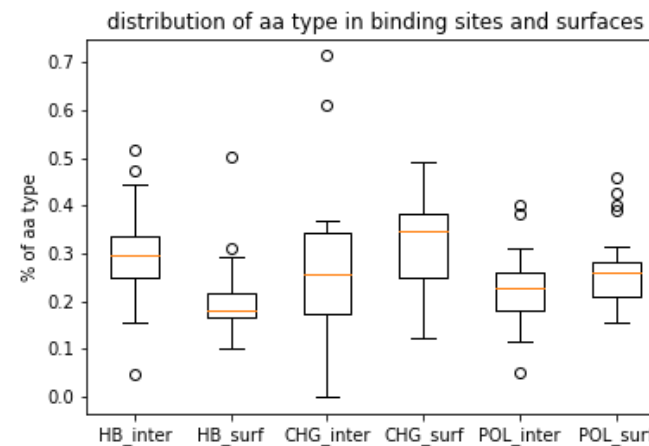
\*en y repensant on aurait pu normaliser par le nb de résidus totaux de la protéine (peut-être plus logique, mais l'allure sera globalement la même)

## Propensions des aa à être dans les BS

$$\text{Log}(\text{freq}_{\text{aainterface}}/\text{freq}_{\text{aasurface}})$$



## % des Hydrophobes (HB), chargés (CHG) et polaires (POL)



- Du coup, on peut commenter avec eux les résultats principaux (idéalement ajouter d'autres complexes pour introduire de la variabilité notamment niveau taille, lisser un peu la courbe).
- On voit clairement un effet de la CV, un effet sur les types d'aa et sur les propensions, tout cela à discuter et à mettre en rapport avec les papiers de Janin, Thornton, Jones etc ? On terminerait par une petite présentation de ces travaux et un récapitulatif des résultats obtenus en cours.

Séance suivante : 3 ou 4

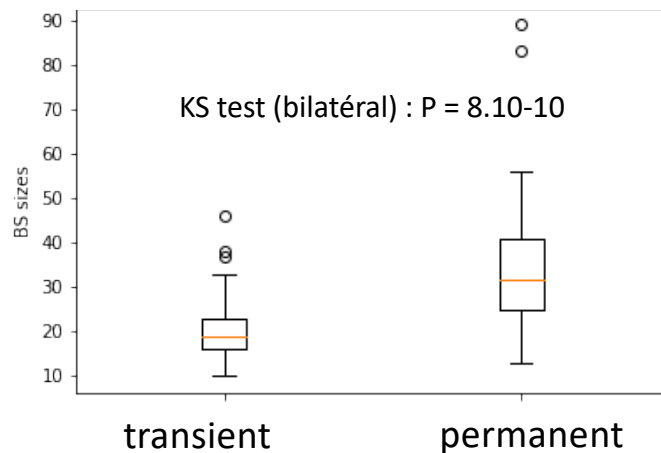
LA question est est-ce qu'on fait une séance ou demi séance pour appliquer les scripts à la comparaison de différents types de complexes (permanents VS transitoires et VS Ac-Ag ? Et ou autres ?).

- L'intérêt est de discuter de la variabilité des complexes, propriétés différentes d'un point de vue fonctionnel, structural etc

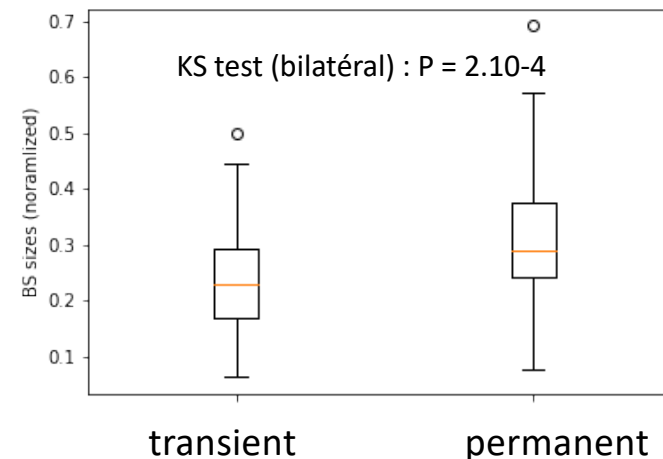
- l'autre intérêt est de leur faire manipuler un peu les plots et les stats pour comparer des distributions et faire le pont avec l'UE de stats (application concrète des stats, ils en voient l'intérêt et voient la facilité d'utilisation avec python) ?

Si on part là-dessus, ça donnera ça :

Distribution de la taille des sites d'interaction



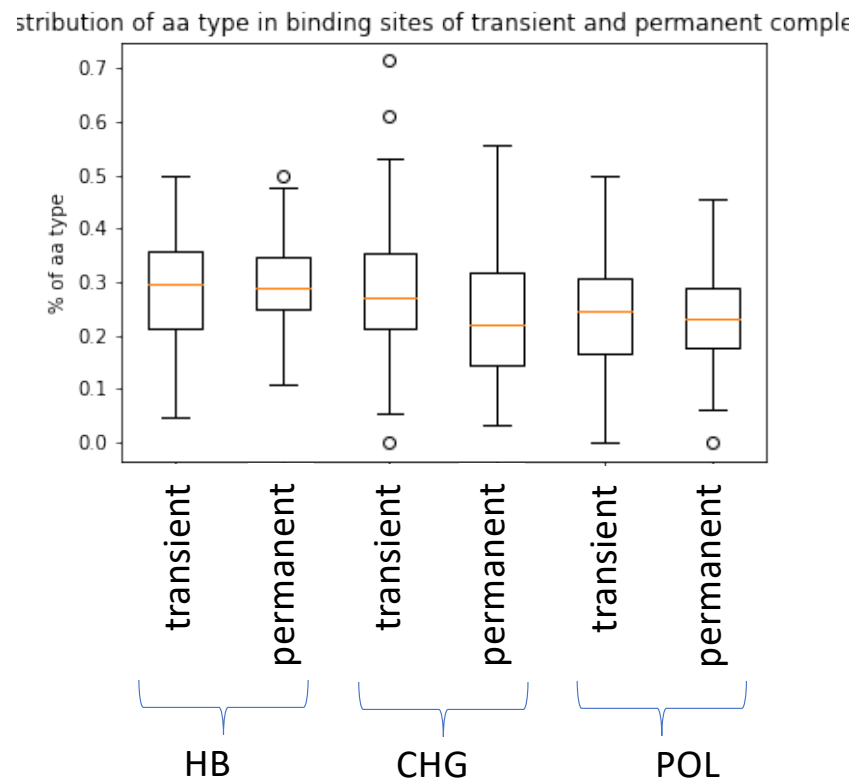
Distribution de la taille des sites d'interaction (normalisée par le nb de résidus de surface).



Le signal est clair : les BS intervenant dans les assemblages « permanents » sont plus grandes

Séance suivante : 3 ou 4

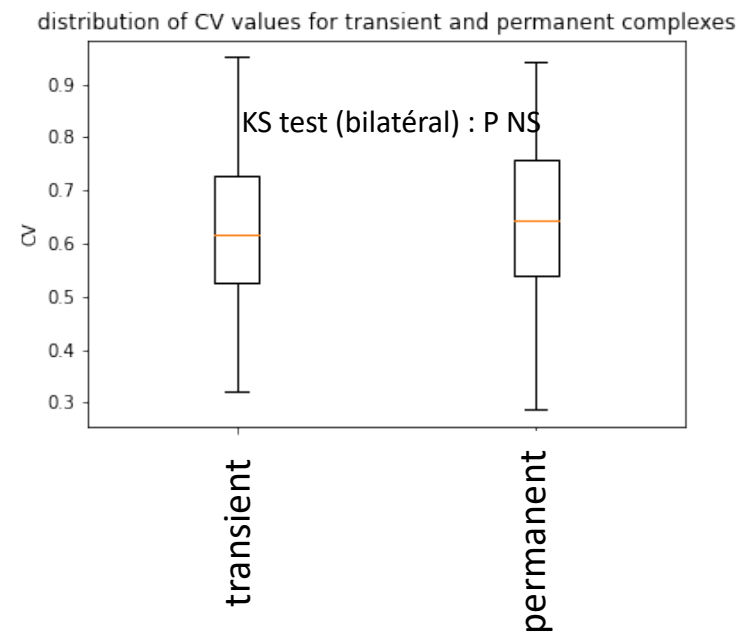
Par contre les résultats ne sont pas probants pour les %d'HB, POL et CHG (je ne l'explique pas alors qu'on sait que les assemblages permanents sont enrichis en aa hydrophobes (HB) par rapport aux complexes transitoires. Une idée ? Un autre jeu de données à proposer ?



KS test (bilatéral) : P NS pour chaque type d'aa

Séance suivante : 3 ou 4

IDEM pour la CV



Séance suivante : 3 ou 4

Ensuite plusieurs possibilités en fonction du temps qu'auront pris les parties précédentes et de ce qu'on veut faire.

- Soit on pousse un peu plus la comparaison des différents types de complexes mais il faudrait trouver des jeux de données plus satisfaisants. On peut ajouter le calcul des freq de contacts etc Dans tous les cas il faut viser de ne pas dépasser la fin de séance 4 pour pouvoir commencer en séance 5, la modélisation par homologie. Cette solution est moins satisfaisante car redondante avec les autres.
- Soit, (perso j'aime bien l'idée mais ça demande un peu de travail de notre côté pour préparer les jeux de données), leur introduire les notions d'évolution et que les interfaces sont plus conservées. Du coup, on leur prépare des alignements et ils calculent la conservation (juste avec le calcul de l'entropie des colonnes en bouclant sur les colonnes de chaque alignement), et le nb de gaps par ex et regardent la différence entre les colonnes associées aux sites d'interaction et celles associées à la surface (peut-être à comparer avec les colonnes correspondant aux résidus enfouis). Mais la difficulté c'est le mapping entre les colonnes et les positions dans le pdb.
- Raphaël, je me rappelle sinon que tu avais parlé d'un exemple avec des couples interologues je crois qui avaient des interfaces différentes et qu'on pouvait essayer de leur faire retrouver les vraies paires de partenaires. Est-ce que cela te semble jouable en une séance ?
- Sinon j'ai pas encore évoqué le docking mais ça peut intervenir là. La question est : jusqu'où va-t-on avec le docking ? On leur montre comment lancer un calcul. Possibilité on lance le même calcul avec différents algos et ils regardent si les meilleures solutions se recoupent ? Notion de RMSD, comparaison visuelle et automatique. On les pousse vers l'idée de faire un consensus des x algos ? Ça fait une séance complète. La version courte c'est une demi séance où ils lancent le calcul mais sans la comparaison des résultats entre algos.
- On peut aussi choisir docking sur une séance complète + un des 3 premiers points ce qui décale la séance d'homologie en séance 6 et la séance 7 pour l'abondance (avec données mass spec) VS aa propensities et abondance vs propriétés qu'ils auront vues en début d'UE.

Séance suivante : 5 ou 6

Modélisation par homologie

Séance 6 et 7 ou uniquement 7

Abondances VS différentes propriétés