

3i005: Statistique en bioinformatique

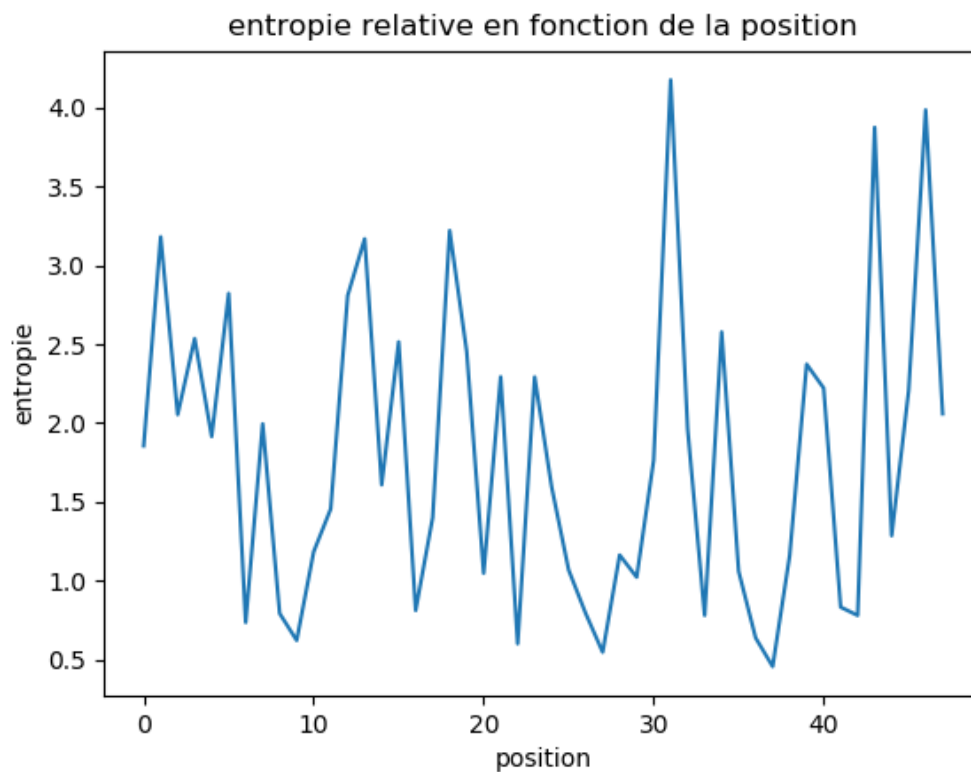
GOSSE-DUMESNIL tony / SAOUDET Ahmed

March 2018

1 Modelisation par PSWM

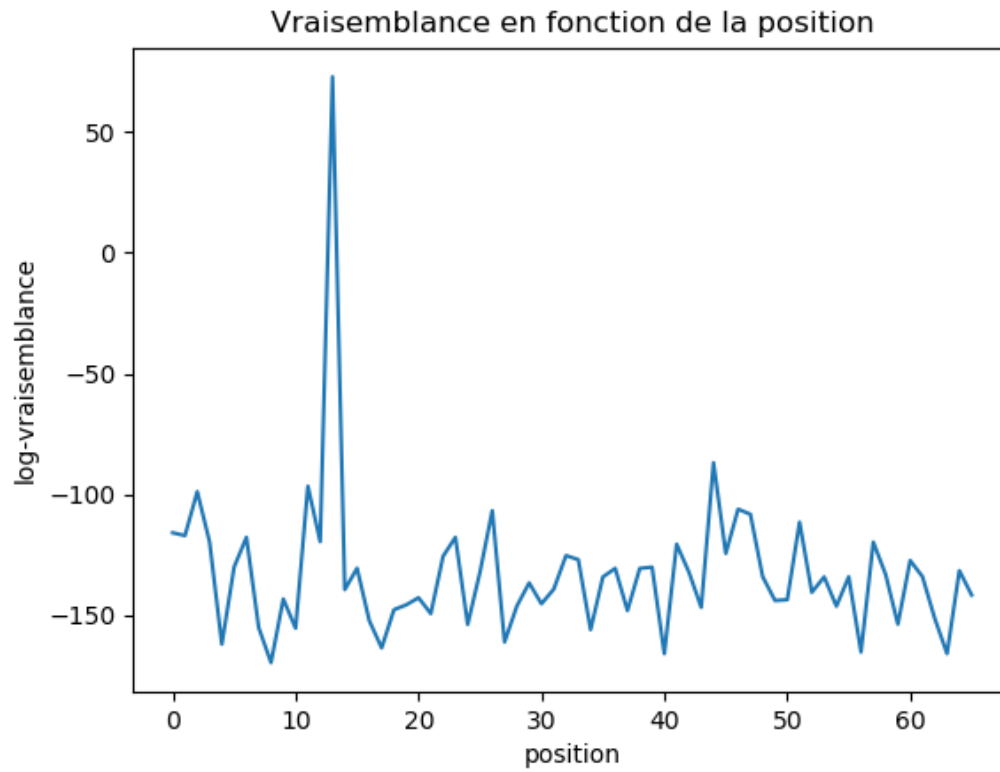
1.1 Acides aminées les plus conservées

Les 3 positions qui ont les meilleures entropies sont les positions 31, 46 et 43 dont leurs acides aminées les plus conservées sont respectivement les acides W, P et G.



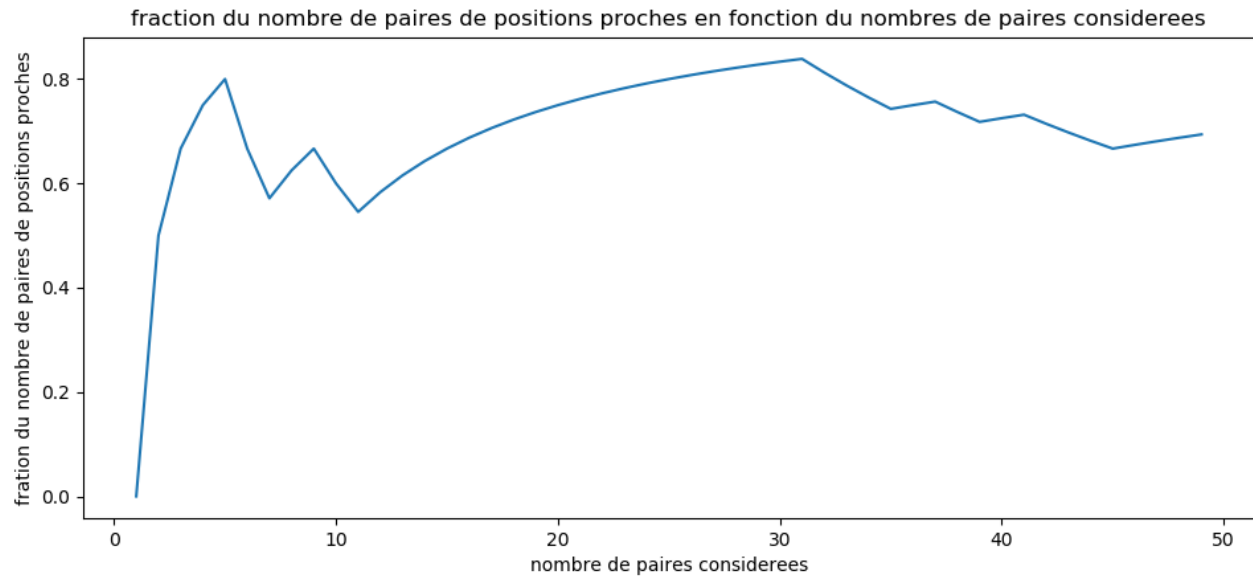
On voit bien sur le graphique qu'il y a des pics maximales en position 31, 43 et 46

1.2 Familiarité de sequence

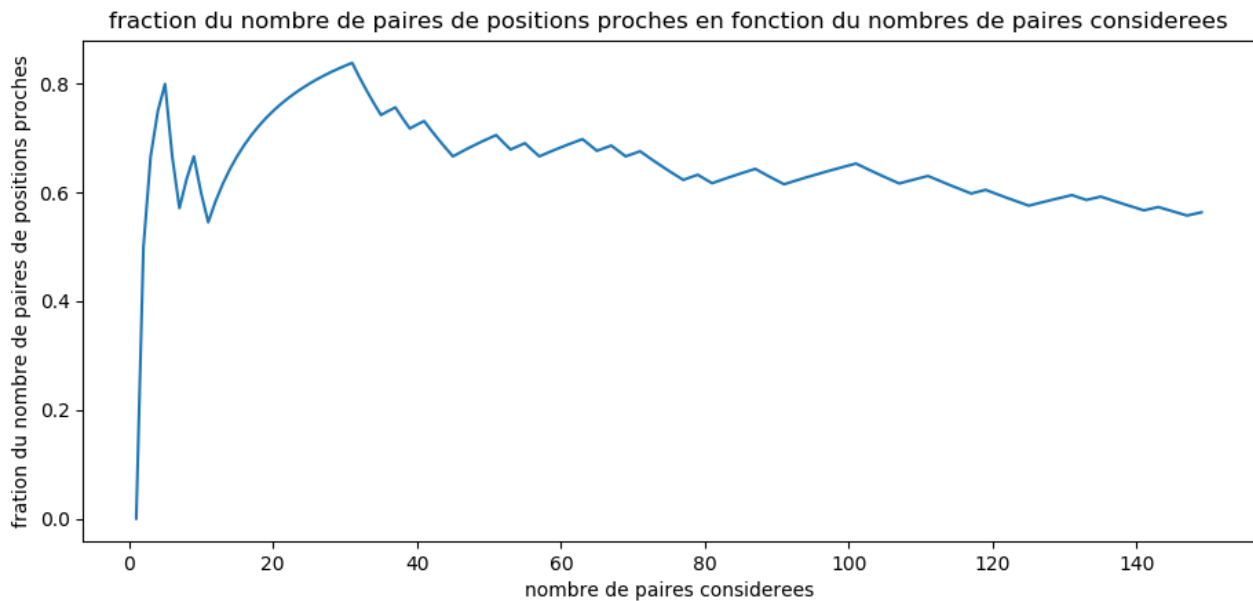


On remarque qu'il existe une sous-séquence de la famille définie par D_{train} pour $i = 13$

2 Co-évolution de résidues en contact



On remarque sur le graphique que de 1 à 30 paires de positions considérées la proportion de paires en contact est très élevée et cette même proportion diminue après 30 paires considérées.



Ce graphique élargi jusqu'à 150 paires considérées montre mieux cette diminution de la proportion de paires en contact à partir de 30 paires considérées. Sachant qu'il y a $48 \times 48 = 2304$ paires différentes et que nous considérons les paires qui sont les plus corrélées, on en déduit que les paires les plus corrélées ont une probabilité élevée d'être en contact.