

BIOINFORMATIQUE, Unité libre 2010
Master TPV, Partie "Statistiques"
(à rédiger sur une feuille simple séparée)

Dans leur article *Strong hydrophobic nature of cysteine residues in proteins* de 1999 paru dans **FEBS Letters**, N. Nagano, M. Ota et K. Nishikawa K. présentent deux échelles d'hydrophobicité, celle de Rose-et-al et celle d'Ota-Nishikawa dont nous reproduisons ici les valeurs.

| Elt | Rose | Ota | | Elt | Rose | Ota |
|--------|------|------|--|--------|------|------|
| Cys_SS | 0.83 | 6.12 | | Cys_SH | 0.89 | 6.82 |
| Ala | 0.78 | 5.62 | | Arg | 0.66 | 4.71 |
| Asn | 0.63 | 4.22 | | Asp | 0.58 | 3.89 |
| Gln | 0.61 | 4.34 | | Glu | 0.55 | 3.80 |
| Gly | 0.71 | 4.61 | | His | 0.73 | 5.45 |
| Ile | 0.88 | 7.03 | | Leu | 0.87 | 6.91 |
| Lys | 0.56 | 3.71 | | Met | 0.84 | 6.60 |
| Phe | 0.87 | 6.79 | | Pro | 0.66 | 4.42 |
| Ser | 0.68 | 4.50 | | Thr | 0.71 | 4.81 |
| Trp | 0.85 | 6.47 | | Tyr | 0.81 | 6.34 |
| Val | 0.87 | 6.66 | | | | |

En admettant qu'on puisse bien considérer les colonnes **Rose** et **Ota** comme des variables quantitatives, étudier la relation linéaire entre ces deux colonnes. On pourra utiliser les fonctions du logiciel R nommées **cor**, **cor.test**, **lm** et **anova** ou la fonction **anaLin** du fichier **statgh.r**; les données peuvent être lues à l'aide de `hyd <- lit.dar("k:/stat_ad/hydropho.dar")`. Elles sont également disponibles sur Internet à l'adresse :

<http://forge.info.univ-angers.fr/~gh/Datasets/hydropho.dar>

On séparera soigneusement la partie *présentation (succincte) des résultats numériques* de la partie *rédaction en direction d'un public de biologistes*.