

# Bio3d

Anne-Marie

03/12/2019

## Démonstration

La protéine d'intérêt de notre démonstration est le récepteur de canabinoïde CBR1. Son numéro d'accension uniprot est le *P21554*, le numéro d'accension de la structure est le *6N4B*. Il s'agit d'un récepteur couplé à une protéine G qui se lie avec différents ligands.

## Modèle d'analyse des modes normaux (NMA)

Référence : *Tutoriel sur le NMA*

### Analyse d'une seule structure

Acquisition et lecture de la protéine

```
pdb <- read.pdb("6N4B")
```

```
pdb
```

```
##
## Call: read.pdb(file = "6N4B")
##
## Total Models#: 1
## Total Atoms#: 8453, XYZs#: 25359 Chains#: 5 (values: A B C R S)
##
## Protein Atoms#: 8368 (residues/Calpha atoms#: 1121)
## Nucleic acid Atoms#: 0 (residues/phosphate atoms#: 0)
##
## Non-protein/nucleic Atoms#: 85 (residues: 3)
## Non-protein/nucleic resid values: [ CLR (2), KCA (1) ]
##
## Protein sequence:
## CTLSAEDKAAVERSKMIDRNLRDGEKAAREVKLLLLGAGESGKSTIVKQMTGIVETHF
## TFKDLHFKMFDVGGQRSEKRWIHCFEVTAIIFCVALSDYDLMNRMHESMKLFDSICNN
## KWFTDTSIILFLNKKDLFEKIKKSPLTICYPEYAGSNTYEEAAAYIQCQFEDLNKRKDT
## KEIYTHFTCATDTKNVQFVDAVTDVVIKNNLKDCLFELDQLRQ...<cut>...KLEL
##
## + attr: atom, xyz, seqres, helix, sheet,
## calpha, remark, call
```

Cette commande nous renvoie un objet de type `pdb`. La structure de protéine sauvegardée dans cet objet contient 8368 résidus et 5 chaines (A, B, C, R et S). Les différents attribut permettent d'accéder à divers niveaux d'informations sur la structure comme l'emplacement des résidus (xyz) ou les caractéristique des différents atomes.

Calcul des modes de la protéine

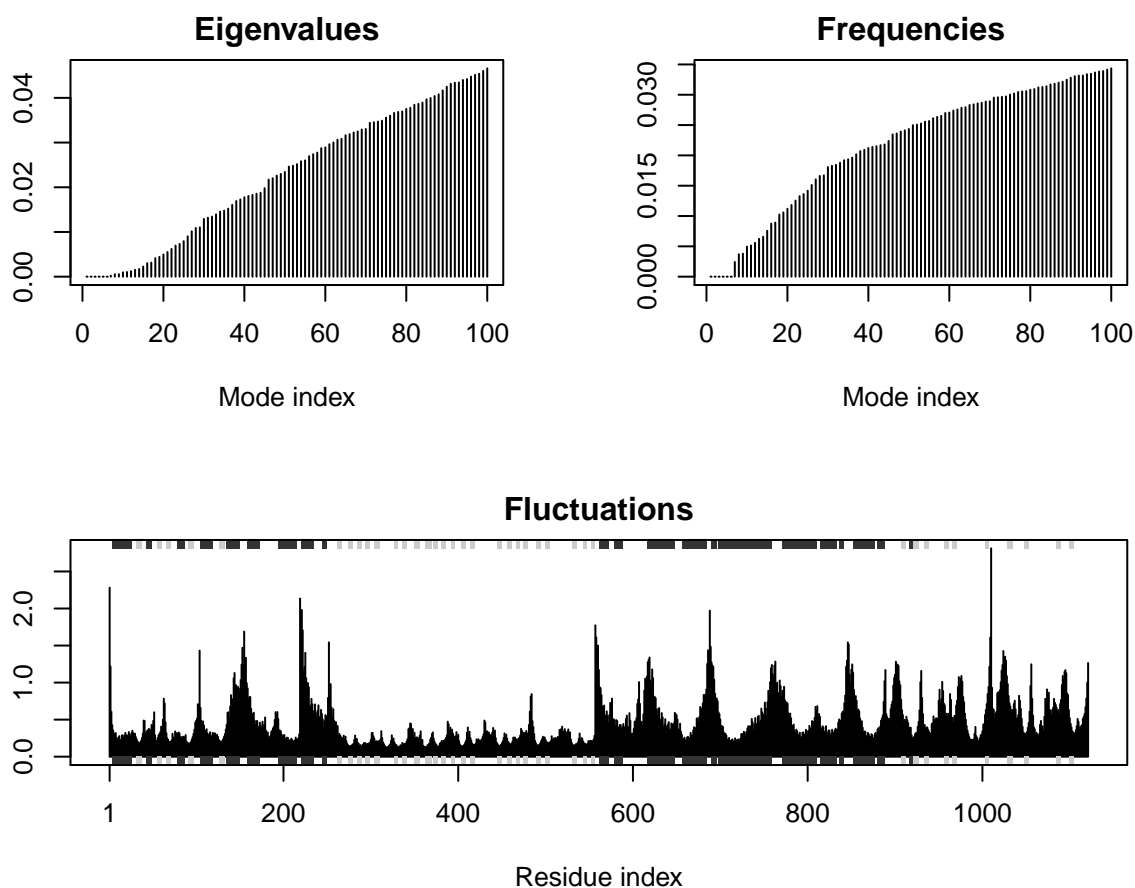
```
modes.pdb <- nma(pdb)
```

```
print(modes.pdb)
```

```
##
```

```
## Call:
##   nma.pdb(pdb = pdb)
##
## Class:
##   VibrationalModes (nma)
##
## Number of modes:
##   3363 (6 trivial)
##
## Frequencies:
##   Mode 7:    0.002
##   Mode 8:    0.004
##   Mode 9:    0.004
##   Mode 10:   0.005
##   Mode 11:   0.005
##   Mode 12:   0.006
##
## + attr: modes, frequencies, force.constants, fluctuations,
##         U, L, xyz, mass, temp, triv.modes, natoms, call
```

```
plot(modes.pdb, sse=pdb)
```



Sommaire des modes (nombre total de modes =  $3N$  où  $N$  est le nombre d'atomes de la structure).  
 Les 6 premiers modes = mouvement de structure rigide avec une valeur de 0.

les premiers modes avec une valeur de plus de 0 sont représentatifs des mouvements de la protéine(?).

**Comment on peut savoir avec juste une structure? Plus d'info là-dessus à trouver**

Sur le graphique, on voit peu de fluctuation entre les résidus ~250 à ~550

*\*\* Voir si on peut faire un plus beau graphique*

### Analyse de plusieurs structures

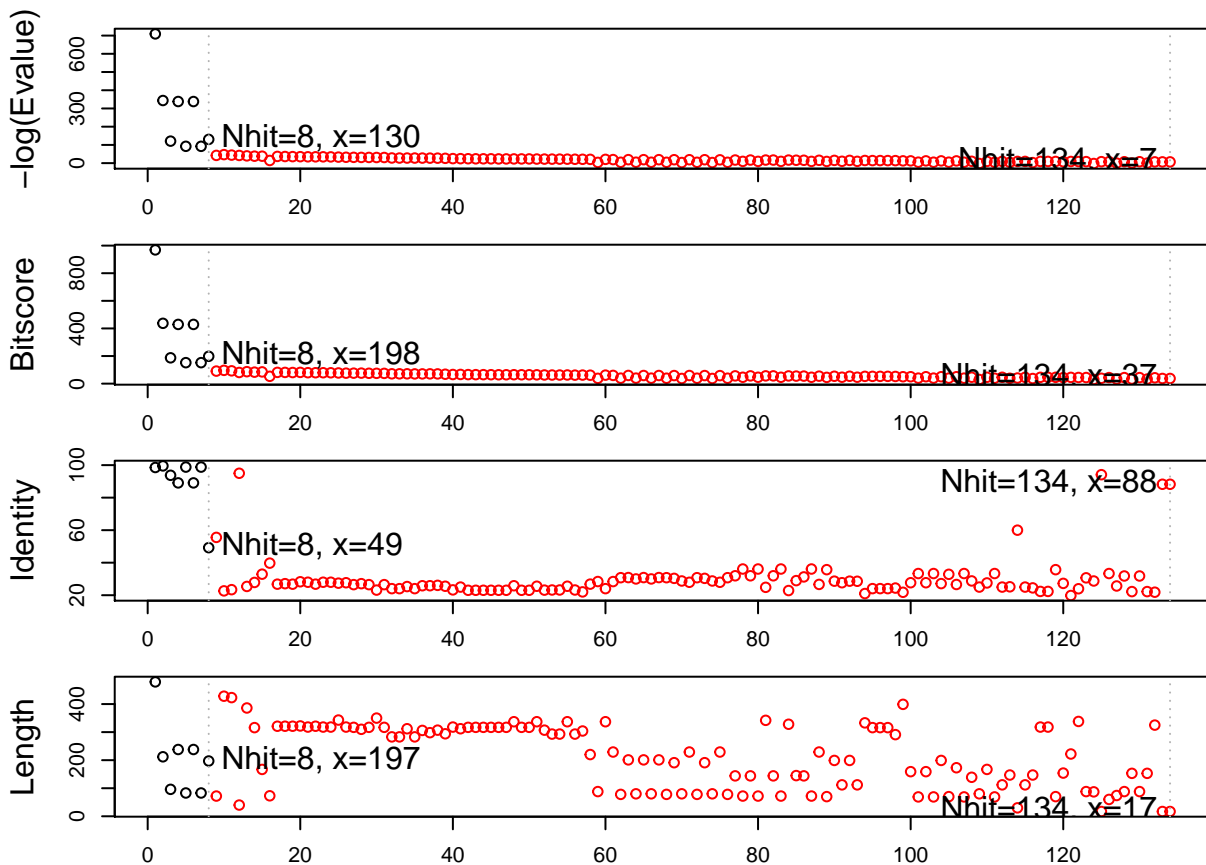
Nous allons maintenant utiliser la fonction blast pour rechercher les structures homologues à notre protéine d'intérêt. Ensuite, nous avons sélectionné un cutoff de 50, ce qui représente une valeur de *e-value*.

```
aa <- get.seq("P21554")
blast <- blast.pdb(aa)
```

```
## Searching ... please wait (updates every 5 seconds) RID = YDKKZR1Z01N
## .
## Reporting 134 hits
```

```
hits <- plot.blast(blast, cutoff = 70)
```

```
## * Possible cutoff values: 130 6
##           Yielding Nhits: 8 134
##
## * Chosen cutoff value of: 70
##           Yielding Nhits: 8
```



```
head(hits$ pdb.id)
```

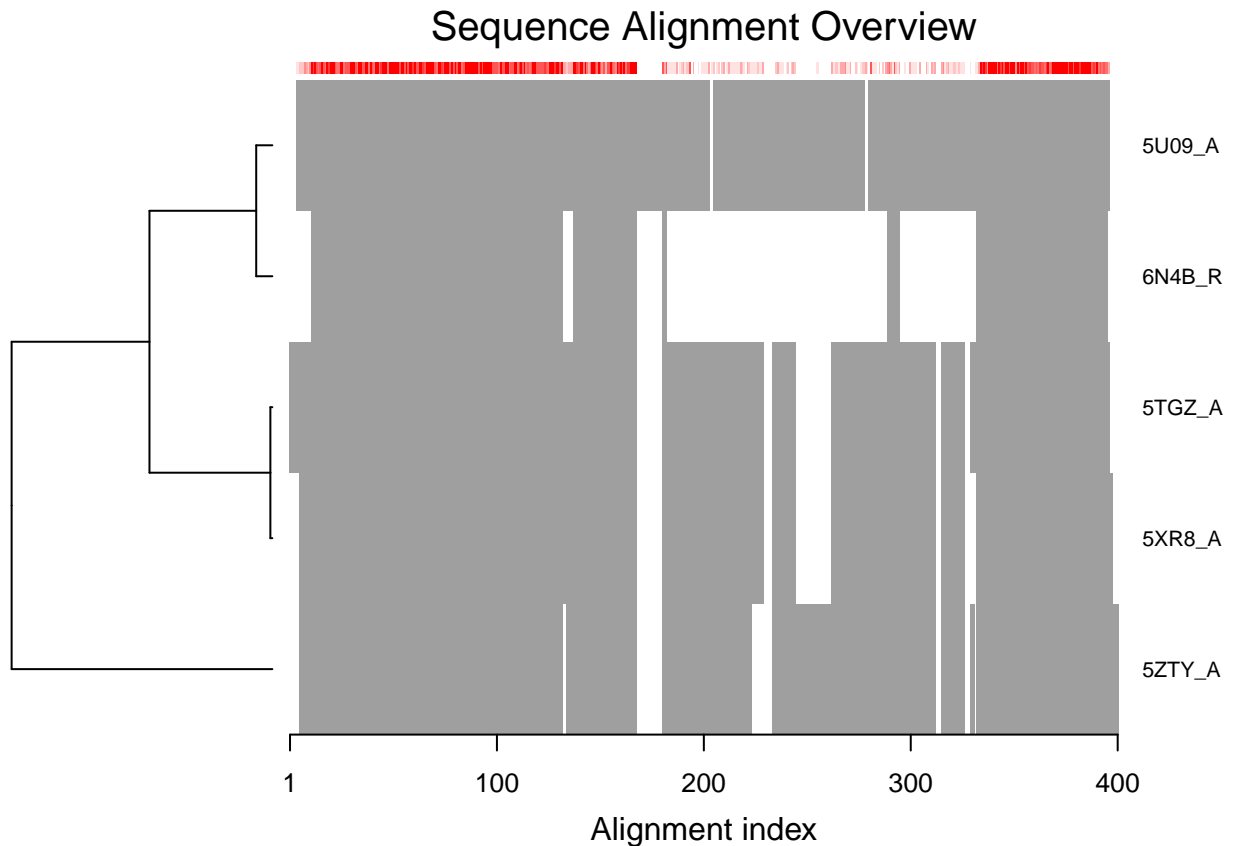
```
## [1] "6N4B_R" "5U09_A" "5U09_A" "5XR8_A" "5XR8_A" "5TGZ_A"
```

Dans le résultat du code, on voit les identifiants des 6 structures retenues (2 des structures sont présentes en double, donc on obtient 6 structures uniques)

On peut maintenant télécharger et faire l'alignement de ces structures

```
files <- get.pdb(hits$ pdb.id, path = "pdbs", split = TRUE, gzip = TRUE)
pdbs <- pdbaln(files, fit=TRUE, web.args=list(email="anmayroy@gmail.com"))
ids <- basename.pdb(pdbs$id)
```

```
plot(pdbs, labels=ids)
```



Le graphique permet d'évaluer l'alignement général des différentes structures.

Avec la commande ci-dessous, on peut extraire les annotations associées avec chaque structure et afficher les espèces sources.

```
anno <- pdb.annotate(ids)
print(unique(anno$source))
```

```
## [1] "Homo sapiens" "Homo sapiens,Pyrococcus abyssi"
## [3] "Desulfovibrio vulgaris,Homo sapiens" "Homo sapiens,Escherichia virus T4"
```

NMA sur toutes les structures

```
modes <- nma(pdbs)
```

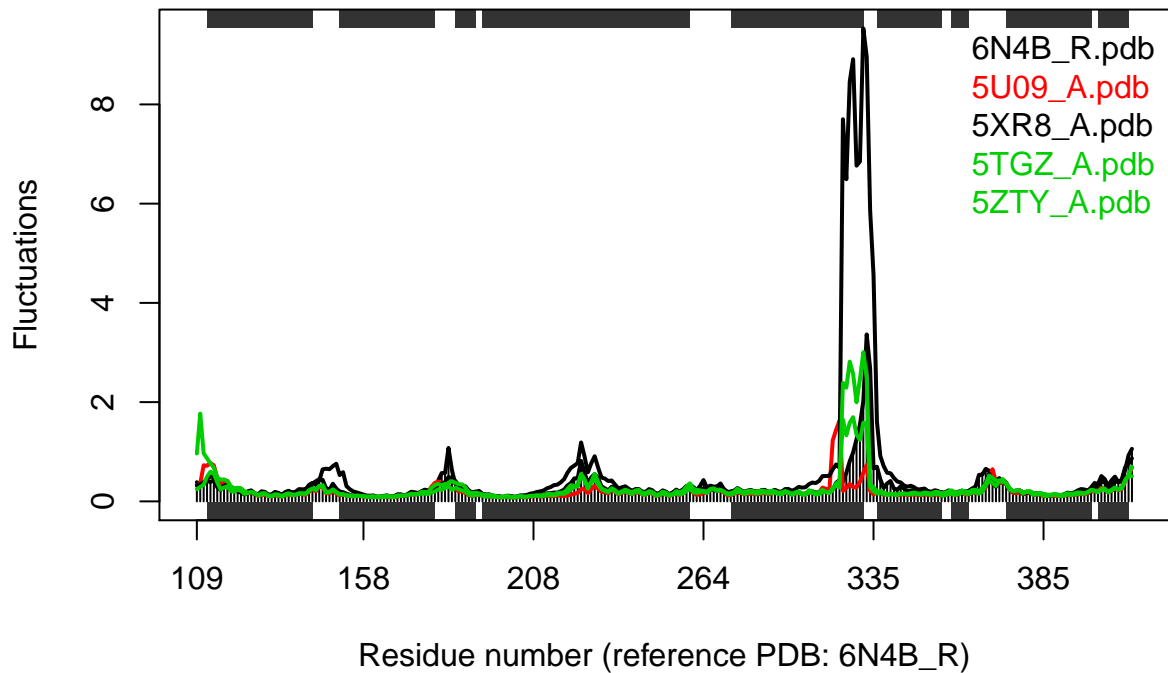
Structure-based clustering

```
rd <- rmsd(pdbs)
hc <- hclust(as.dist(rd))
grps <- cutree(hc, k=3)
```

Pour obtenir un graphique de fluctuation avec trois clusters

```
plot(modes, pdbs, col=grps)
```

```
## Extracting SSE from pdbs$sse attribute
```



Dans ce dernier graphique, on peut observer plusieurs choses:

1- On ne conserve que les résidus 109 à ~400 dans l'analyse des fluctuations. 2- On observe une importante fluctuation environ au résidu 330 dans presque toutes les structures.

### À FAIRE:

Trouver comment l'analyse est différente entre 1 vs plusieurs structures.

Trouver qu'est-ce qui se passe au résidu 330 pour cause une si grande fluctuation.

Pourquoi c'est si différent de l'analyse de la structure seule