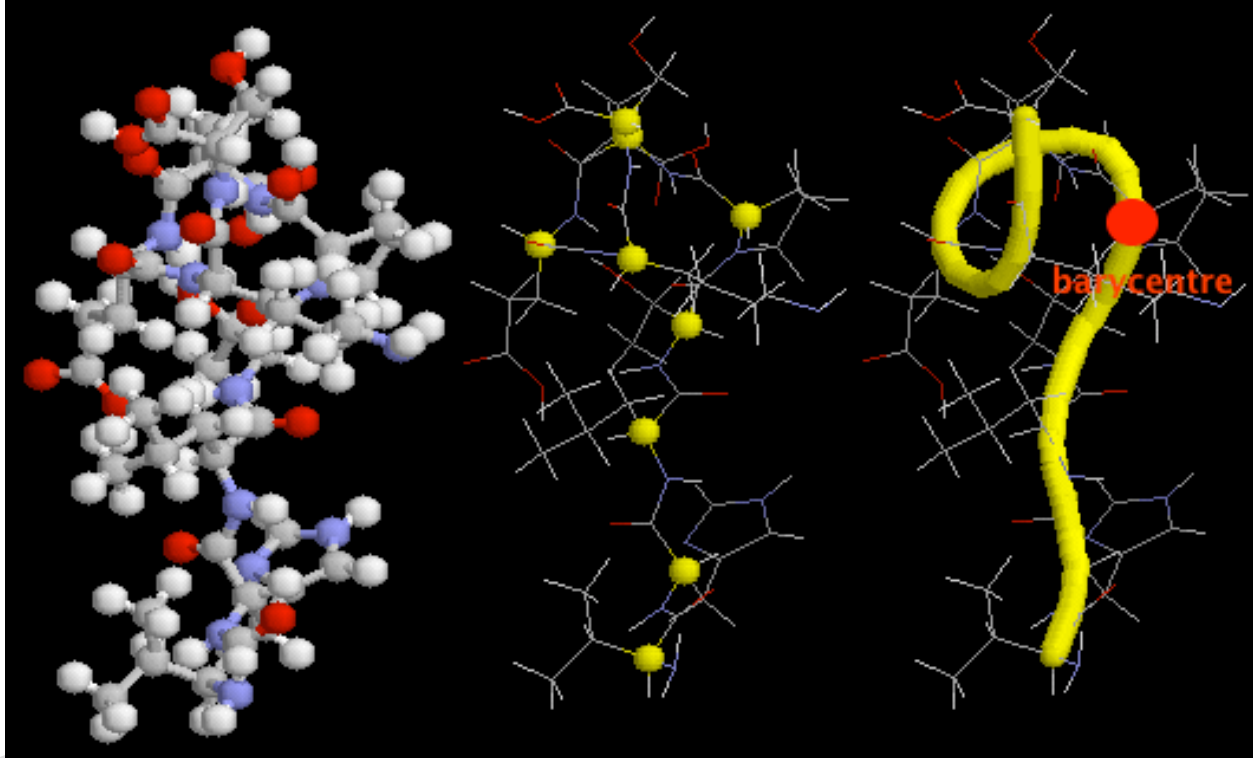


Calcul du barycentre des carbones alpha des acides aminés de la chaîne polypeptidique repliée d'une protéine



Le barycentre d'un ensemble est la position moyenne de cet ensemble.

Le **barycentre** d'une protéine est la moyenne des coordonnées [X, Y et Z] des carbones alpha de tous les résidus d'acides aminés de la chaîne polypeptidique.

Ce barycentre ne représente donc pas la structure 3D repliée d'une protéine.

Logique du programme

Générer l'URL (méthode « `urllib` ») d'un fichier de la base de données PDB (« *Protein Data Bank* »).

Consulter ce fichier pour récupérer les coordonnées [X, Y et Z] de tous les atomes des résidus d'acides aminés d'une chaîne polypeptidique dont on a déterminé la structure 3D.

Lire le fichier ligne par ligne et avec des expressions régulières :

- Repérer les lignes commençant par le mot-clé « ATOM ».
- Ne s'occuper que des lignes contenant l'indication « CA » (carbone alpha).
- Récupérer pour chacune de ces lignes les coordonnées X, Y et Z.
- Faire la moyenne pondérée de toutes les valeurs X, idem Y et idem Z.

The screenshot shows the RCSB PDB website interface. At the top, there's a navigation bar with links like 'RCSB PDB', 'Deposit', 'Search', 'Visualize', 'Analyze', 'Download', 'Learn', 'More', 'Documentation', and 'MyPDB'. Below this is a search bar with the text 'Enter search term(s)' and a search icon. The main content area is divided into tabs: 'Structure Summary', '3D View', 'Annotations', 'Experiment', 'Sequence', and 'Genome'. The 'Structure Summary' tab is selected, showing details for entry 2NBT: NEURONAL BUNGAROTOXIN, NMR. A red arrow points from the text box on the left to the 'FASTA Sequence' download option in the 'Display Files' dropdown menu. The page also includes a 3D view of the protein structure, a table of validation metrics (Clashscore, Ramachandran outliers, Sidechain outliers), and a list of deposition authors.

Metric	Percentile Ranks	Value
Clashscore		62
Ramachandran outliers		13.5%
Sidechain outliers		18.7%

Script du calcul du barycentre des carbones alpha des acides amines de la chaîne polypeptidique repliée d'une protéine

Script inspire de http://julien.maupetit.free.fr/teachingFiles/2007/M1SPGF_Python_TD_corr.pdf

```
#!/usr/bin/env python
# -*-coding:Utf-8 -*
```

```
##### Emulation des bibliothèques Python :
```

- pour générer l'URL du fichier et le consulter (« urllib »)
- pour utiliser les expressions régulières (« regular expressions » ou « re »)

```
import urllib
import re
```

```
##### Initialisation variables somme des coordonnees X, Y et Z de tous les carbones alpha (ou CA)
```

```
sommeCoordX = 0
sommeCoordY = 0
sommeCoordZ = 0
```

```
##### Generation automatique de l'URL recherche du fichier PDB

CodePDB = raw_input("Entrer le code PDB : ")

URLPDB = "https://files.rcsb.org/view/" + CodePDB + ".pdb »
# symbole + pour concatener des chaines / Exemple : https://files.rcsb.org/view/2NBT.pdb

FichierPDB = urllib.urlopen( URLPDB )
# fonction urlopen lit le texte complet de la page web / comme pour un fichier texte

##### Affichage du nom de la proteine dont on consulte le fichier

for LigneEnCours in FichierPDB.readlines():

    if re.match(r"TITLE", LigneEnCours):
        LigneEnCours = LigneEnCours.replace('TITLE','')
        print "Proteine :", LigneEnCours

FichierPDB.close()
```

```
# fonction « urlopen » lit le texte complet de la page
web consultée
```

```
FichierPDB = urllib.urlopen( URLPDB )
```

```
# Selection des lignes des carbones alpha (CA)
```

```
ligneCA = [ligne for ligne in FichierPDB if
ligne.split()[0] == "ATOM" and ligne.split()[2] == "CA"]
```

```
# "ligneCA" est un objet de type liste d'ou la methode
enumerate
```

```
for chaqueLigne in enumerate(ligneCA):
```

```
# Conversion de la liste en une serie de chaines de
caracteres ("string »)
```

```
conversionChaine = str(chaqueLigne)
```

```
# Utilisation des expressions regulieres pour enlever
les caractères ne correspondant pas aux coordonnées
```

```
conversionChaine = re.sub(r" \ ( ( \d ) + ,
\s' ", "", conversionChaine)
conversionChaine = re.sub(r"\d.( \d ) +( \s ) + \d.( \d ) +( \s ) +.
( \s ) + \\n \\ ' ) ", "", conversionChaine)
```

```
# Exemple : la chaine de caracteres (ATOM 182 CA THR A
13 -11.268 -9.843 -1.187 1.00 0.51 \n')
```

```
# devient : ATOM 182 CA THR A 13 -11.268 -9.843 -1.187
```

ATOM	163	HD3	PRO	A	11	-14.548	-7.357	-9.502	1.00	0.62
ATOM	164	N	GLN	A	12	-11.506	-9.296	-5.371	1.00	0.49
ATOM	165	CA	GLN	A	12	-10.694	-10.458	-4.920	1.00	0.51
ATOM	166	C	GLN	A	12	-10.622	-10.549	-3.389	1.00	0.52
ATOM	167	O	GLN	A	12	-9.880	-11.352	-2.860	1.00	0.57
ATOM	168	CB	GLN	A	12	-9.302	-10.228	-5.502	1.00	0.52
ATOM	169	CG	GLN	A	12	-8.829	-11.517	-6.179	1.00	0.61
ATOM	170	CD	GLN	A	12	-7.453	-11.292	-6.804	1.00	0.58
ATOM	171	OE1	GLN	A	12	-6.440	-11.558	-6.186	1.00	0.76
ATOM	172	NE2	GLN	A	12	-7.374	-10.806	-8.013	1.00	0.67
ATOM	173	H	GLN	A	12	-11.691	-8.550	-4.761	1.00	0.52
ATOM	174	HA	GLN	A	12	-11.100	-11.371	-5.324	1.00	0.52
ATOM	175	HB2	GLN	A	12	-9.340	-9.429	-6.229	1.00	0.56
ATOM	176	HB3	GLN	A	12	-8.618	-9.966	-4.710	1.00	0.55
ATOM	177	HG2	GLN	A	12	-8.768	-12.309	-5.444	1.00	0.77
ATOM	178	HG3	GLN	A	12	-9.531	-11.797	-6.949	1.00	0.71
ATOM	179	HE21	GLN	A	12	-8.192	-10.589	-8.510	1.00	0.79
ATOM	180	HE22	GLN	A	12	-6.498	-10.661	-8.427	1.00	0.76
ATOM	181	N	THR	A	13	-11.359	-9.749	-2.656	1.00	0.50
ATOM	182	CA	THR	A	13	-11.268	-9.843	-1.187	1.00	0.51
ATOM	183	C	THR	A	13	-12.270	-10.829	-0.625	1.00	0.53
ATOM	184	O	THR	A	13	-12.680	-11.772	-1.274	1.00	0.58
ATOM	185	CB	THR	A	13	-11.573	-8.460	-0.655	1.00	0.48
ATOM	186	OG1	THR	A	13	-11.284	-7.494	-1.646	1.00	0.45
ATOM	187	CG2	THR	A	13	-10.728	-8.188	0.588	1.00	0.52
ATOM	188	H	THR	A	13	-11.952	-9.093	-3.057	1.00	0.50
ATOM	189	HA	THR	A	13	-10.295	-10.124	-0.911	1.00	0.53
ATOM	190	HB	THR	A	13	-12.615	-8.419	-0.392	1.00	0.51
ATOM	191	HG1	THR	A	13	-12.004	-7.495	-2.268	1.00	0.43
ATOM	192	HG21	THR	A	13	-9.945	-8.925	0.658	1.00	0.56
ATOM	193	HG22	THR	A	13	-10.289	-7.207	0.515	1.00	0.54
ATOM	194	HG23	THR	A	13	-11.353	-8.239	1.469	1.00	0.57
ATOM	195	N	CYS	A	14	-12.666	-10.615	0.597	1.00	0.53
ATOM	196	CA	CYS	A	14	-13.646	-11.529	1.222	1.00	0.55
ATOM	197	C	CYS	A	14	-13.762	-11.270	2.730	1.00	0.55
ATOM	198	O	CYS	A	14	-13.564	-12.165	3.527	1.00	0.64

Recuperation de toutes les coordonnees X, Y et Z des CA

```
listeChaine = conversionChaine.split()
```

```
# Chaque ligne "conversionChaine" devient une LISTE
```

```
#Exemple : ['ATOM', '561', 'CA', 'ASP', 'A', '35', '-5.897', '-6.886', '-8.018'] => La coordonnee X  
(-5.897) est le 7e element de la liste
```

```
# On recupere les coordonnees X, Y et Z de chaque ligne CA (donc de chaque liste) :
```

```
coordX = listeChaine[6] # La coordonnee X est le 7e element de la liste donc [6] car on commence a 0
```

```
coordY = listeChaine[7] # La coordonnee Y est le 8e element de la liste
```

```
coordZ = listeChaine[8] # La coordonnee Z est le 9e element de la liste
```

```
##### Transformation de la chaine de caracteres en un reel ("float »)
```

```
coordX = float(coordX)
coordY = float(coordY)
coordZ = float(coordZ)
```

```
##### Calcul de la somme des coordonnees X, Y et Z => opérateur +=
```

```
sommeCoordX += coordX
sommeCoordY += coordY
sommeCoordZ += coordZ
```

```
##### Calcul des coordonnees du barycentre
```

```
barycentreX = sommeCoordX / len(ligneCA)
barycentreY = sommeCoordY / len(ligneCA)
barycentreZ = sommeCoordZ / len(ligneCA)
```

```
print "Nombre d'acides amines : ",len(ligneCA)
```

```
##### Affichage des coordonnees X, Y et Z du barycentre
```

```
print "Coordonnees du barycentre de",CodePDB," : X =%8.3f - Y =%8.3f - Z =%8.3f" %
(barycentreX,barycentreY,barycentreZ)
```

```
FichierPDB.close()
```