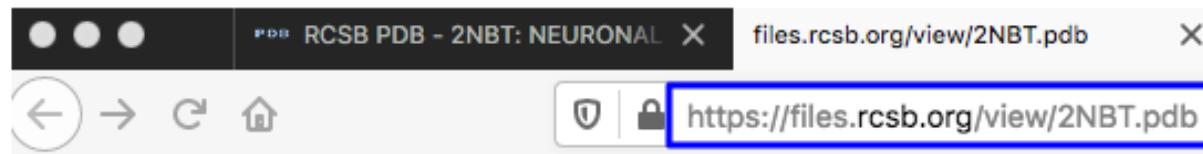


CONTENU PARTIEL DU FICHIER PDB 2NBT (affichage HTML page Web)



URL du fichier PDB généré par le début du script
=> URLPDB = "http://files.rcsb.org/view/" + CodePDB + ".pdb"

```
HEADER      TOXIN                      29-OCT-97   2NBT
TITLE       NEURONAL BUNGAROTOXIN, NMR, 10 STRUCTURES
COMPND      MOL_ID: 1;
COMPND      2 MOLECULE: NEURONAL BUNGAROTOXIN;
COMPND      3 CHAIN: A, B
SOURCE      MOL_ID: 1;
SOURCE      2 ORGANISM_SCIENTIFIC: BUNGARUS MULTICINCTUS;
SOURCE      3 ORGANISM_COMMON: MANY-BANDED KRAIT;
SOURCE      4 ORGANISM_TAXID: 8616;
SOURCE      5 SECRETION: VENOM
KEYWDS      TOXIN, VENOM, NEUROTOXIN
EXPDTA      SOLUTION NMR
```

A screenshot of a PDB file viewer interface. The main content area displays a list of disulfide bonds (SSBOND) between cysteine residues (CYS) in chains A and B. The bonds are numbered 1 through 10. A pink box highlights the first five bonds (1-5) and the second five bonds (6-10). To the right of this box, pink text reads: "Position et chaîne des CYS impliquées dans les (2 x 5) ponts disulfures". Below the SSBOND list, there is a search bar with the text "SSBOND" entered, highlighted by a red box. To the right of the search bar are several buttons: "Tout surligner", "Respecter la casse", "Mots entiers", and "Occurrence 1 sur 10".

| SSBOND | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | CYS A | CYS A | CYS A | CYS A | CYS A | CYS B | CYS B | CYS B | CYS B | CYS B |
| | 3 | 14 | 27 | 46 | 59 | 3 | 14 | 27 | 46 | 59 |
| | CYS A | CYS A | CYS A | CYS A | CYS A | CYS B | CYS B | CYS B | CYS B | CYS B |
| | 21 | 42 | 31 | 58 | 64 | 21 | 42 | 31 | 58 | 64 |

CTRL-F : recherche dans la page web du terme "SSBOND"
=> identique à une expression régulière

1. Import des bibliothèques et initialisation des variables

```
import re
import urllib.request
PONTnumero = 1
NombrePONTS = 0
```

2. Génération de l'URL consultée à la « volée » - concaténation

```
CodePDB = input("Entrer le code PDB : ")
URLPDB = "http://files.rcsb.org/view/" + CodePDB + ".pdb"
```

3. Ouverture du fichier HTML consulté (fichier PDB)

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

4. Conversion de chaque ligne (« readlines ») du fichier html ("byte-like object") en une chaîne de caractères avec "decode"

```
for LigneEnCours in FichierPDB.readlines():
    LigneEnCours = LigneEnCours.decode('utf-8')
```

5. Recherche des lignes du fichier PDB commençant par « SSBOND »

```
if re.search(r"^(SSBOND)", LigneEnCours):
```

6. Sélection des caractères correspondant aux positions des CYS

```
print ("numero ", PONTnumero, " => ", LigneEnCours [10:40:+1])
```

| | | | | | | | |
|--------|----|-----|---|----|-----|---|----|
| SSBOND | 9 | CYS | B | 46 | CYS | B | 58 |
| SSBOND | 10 | CYS | B | 59 | CYS | B | 64 |

Expression régulière pour rechercher la position et la chaîne polypeptidique de chaque CYS impliquée dans un pont disulfure :

```
re.search(r"SSBOND(\s)+(\d)+\sCYS\sB(\s)+(\d)+(\s)+CYS\sB(\s)+\d", LigneEnCours)
```

SSBOND : les 6 lettres majuscules SSBOND exactement

\s : caractère « espace » (s comme « space »)
=>(\s)+ signifie « 1 ou n espace(s) »

\d signifie « n'importe quel chiffre »
=>(\d)+ signifie « 1 ou n chiffre(s) »

CYS : les 3 lettres majuscules CYS exactement

B : la lettre majuscule B (chaîne polypeptidique)

Rappel : r est le caractère d'échappement dans le cas de RE contenant des anti-slash (\)