Protein secondary structure assignement

Project presented by

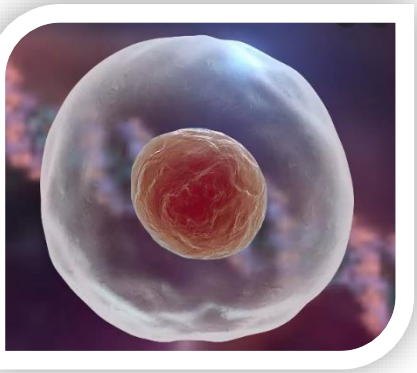
Maha GRAA

M2 Biologie Informatique

2022/2023

1. *INTRODUCTION*
2. *OBJECTIVES*
3. *MATERIAL AND METHODS*
4. *RESULTS AND DISCUSSION*

Introduction



Cell

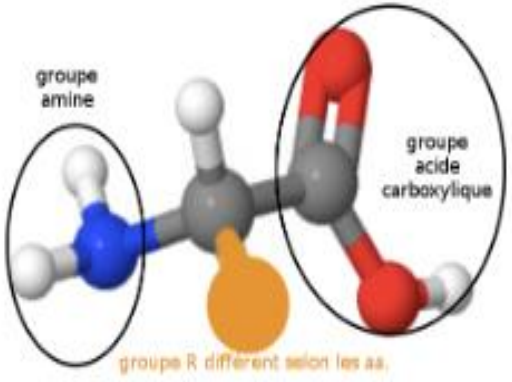
ARN

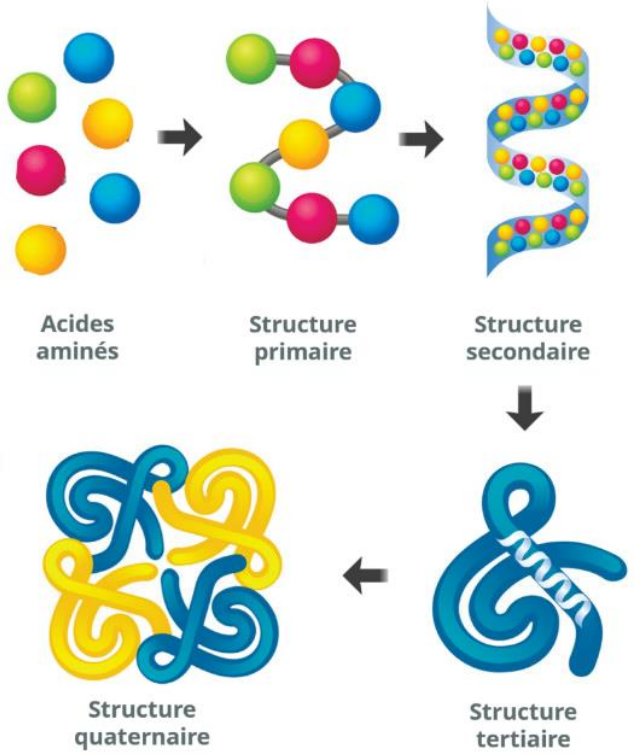
ADN

Protein

List of amino acids :

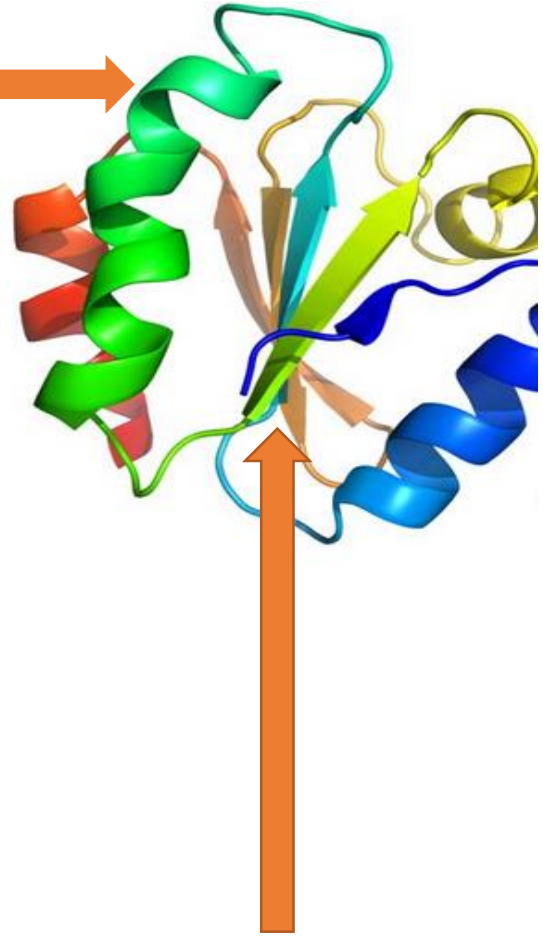
- Valine
- Leucine
- Isoleucine
- Lysine
- Thréonine
- Phénylalanine
- Méthionine
- Histidine
- tryptophane
- Glutamine
- Aspartate
- Glutamate
- Arginine
- Alanine
- Proline
- Cystéine
- Asparagine
- Sérine
- Glycine
- Tyrosine





We have two types of protein's secondary structure :

Alpha helix



Beta Sheet

This structure that we observe in PDB protein is obtained thanks to the technique of crystallography or RMN spectrometry

DSSP algorithm

DSSP attribue 7 or 8 cases :

H: hélice α (i-i+4 HB)
G: hélice 310 (i-i+3 HB)
I: hélice π (i-i+5 HB)
E: brin β (in a β sheet)
B : (isolated β bridge)
T: coude β (β-turn)
S: bend
C: everything else !

Objective :

We want to find an alternative method to assign the secondary structures of a protein by applying the conditions applied in the DSSP algorithm

Structure of the Protein of interest (From PDB)

« HBPLUS » program

File genrated by HBPLUS (it contains the Hbonds)

Protein secondary structure assignement Model :

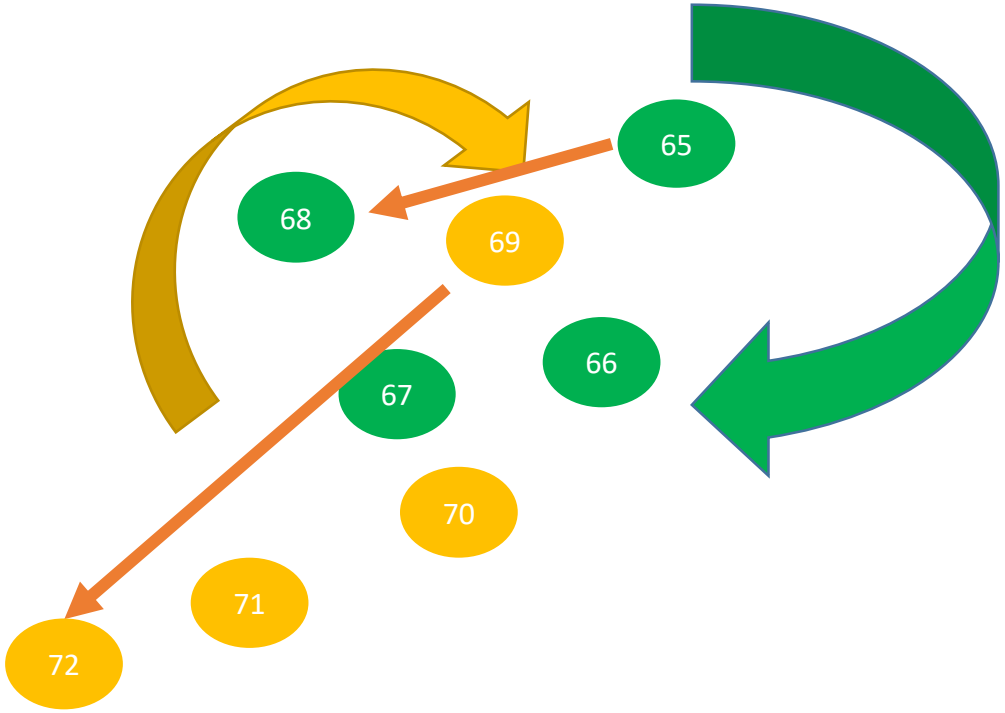
- **Reading the file generated by HBPLUS**
- **Creating the Hbond list**
- **helix fonction (application of DSSP condition)**
- **Sheet béta (application of DSSP condition)**

Generated file : hélices&feuillets.txt

* Amino acid

Applied method for helix structure assignment

- To assign a turn : If two amino acids that form an HBond are 3, 4 or 5 amino acids apart, we can say that they form a Turn.
- Minimum two successive turns of the same type are sufficient to say that there is a helix .

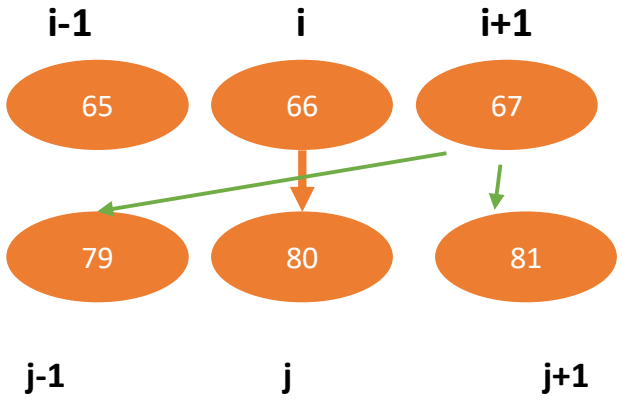


Feuillet beta condition

The condition for assigning the presence of a beta sheet :

If we have 3 successive amino acids that have HBond with other 3 amino acids, the condition we will assume to be true is the fact that the amino acid in the middle (i) has an Hbond with another amino acid (j) (figure) .

Then have to check the next amino acid (i+1) if it has an Hbond with the amino acid that comes just after or just before the J (that has an Hbond with in our initial condition true)



Results and discussion

Protein model used for results and discussion is TYROSINE PHOSPHATASE YOPH FROM YERSINIA PESTIS (1HUF)

Helix alpha from 8 to 9 position

Helix alpha from 12 to 13 position

Didn't detect Sheet

	type	helice	turn	bridge	leader	feuillet
0003	LEU	-	-	-	-	-
0004	SER	-	-	-	-	-
0004	SER	-	-	-	-	-
0005	LEU	-	-	-	-	-
0006	SER	-	-	-	-	-
0006	SER	-	-	-	-	-
0007	ASP	-	3	-	-	-
0008	LEU	H	4	B	-	-
0009	HIS	H	4	-	-	-
0009	HIS	-	-	-	-	-
0009	HIS	-	-	-	-	-
0010	ARG	-	4	-	-	-
0010	ARG	-	3	-	-	-
0011	GLN	H	4	B	L	-
0012	VAL	H	4	B	-	-
0013	SER	H	4	-	-	-
0013	SER	H	4	-	-	-
0013	SER	H	4	-	-	-
0014	ARG	H	4	-	-	-
0014	ARG	H	4	-	-	-
0014	ARG	-	-	-	-	-
0015	LEU	H	4	B	L	F
0016	VAL	H	4	B	L	F
0017	GLN	H	4	B	-	-
0018	GLN	H	4	-	-	-
0018	GLN	-	-	-	-	-
0019	GLU	-	3	B	-	-
0020	SER	-	5	-	-	-
0020	SER	-	-	-	-	-
0021	GLY	-	-	-	-	-
0023	CYS	-	3	-	-	-
0023	CYS	-	-	-	-	-
0024	THR	-	-	-	-	-
0024	THR	-	-	-	-	-
0025	GLY	-	-	-	-	-
0026	LYS	-	-	-	-	-
0026	LYS	-	-	-	-	-
0027	LEU	-	-	-	-	-
0030	ASN	-	-	-	-	-
0030	ASN	-	-	-	-	-
0031	VAL	-	-	-	-	-
0032	ALA	-	-	-	-	-
0033	ALA	-	-	-	-	-
0034	ASN	-	-	-	-	-
0034	ASN	-	-	-	-	-
0034	ASN	-	3	-	-	-
0035	LYS	-	-	-	-	-
0036	GLU	-	-	-	-	-
0037	THR	-	3	-	-	-
0037	THR	-	-	-	-	-
0038	THR	-	-	-	-	-

DSSP model

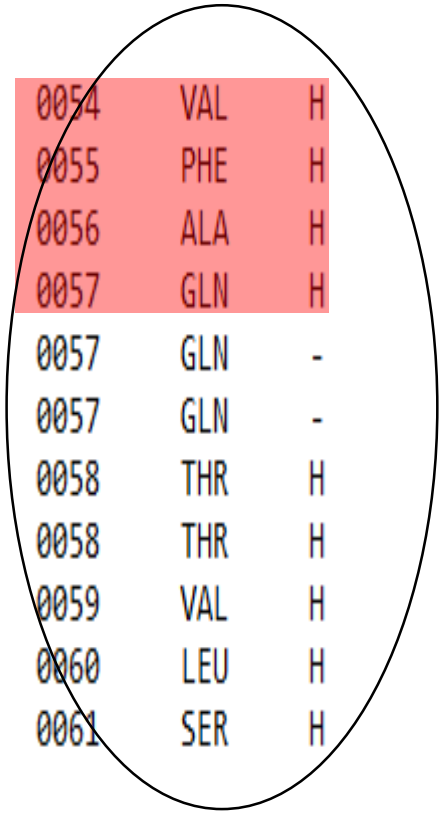
#	RESIDUE	AA	STRUCTURE
1	3	A	L
2	4	A	S > -
3	5	A	L H > S+
4	6	A	S H > S+
5	7	A	D H > S+
6	8	A	L H X S+
7	9	A	H H X S+
8	10	A	R H X S+
9	11	A	Q H X S+
10	12	A	V H X S+
11	13	A	S H X S+
12	14	A	R H X S+
13	15	A	L H <>S+
14	16	A	V I ><>S+
15	17	A	Q I 3<5S+
16	18	A	Q I 3<5S-
17	19	A	E I < 5S+
18	20	A	S I > < +
19	21	A	G T 3 <S+
20	22	A	D T 3 S+
21	23	A	C S < S-
22	24	A	T B -A
23	25	A	G E -B
24	26	A	K E -B
25	27	A	L E +B
26	28	A	R S S-
27	29	A	G S S+
28	30	A	N S S-
29	31	A	V E -B
30	32	A	A E -B
31	33	A	A E -B
32	34	A	N +
33	35	A	K +
34	36	A	E S S+
35	37	A	T S S-
36	38	A	T -
37	39	A	F +
38	40	A	Q -

Helix alpha from 5 to 9 position.

Sheet from 25 to 27 position

Helix alpha from
54 to 57 position

Helix alpha from
58 to 61 position



0054	VAL	H
0055	PHE	H
0056	ALA	H
0057	GLN	H
0057	GLN	-
0057	GLN	-
0058	THR	H
0058	THR	H
0059	VAL	H
0060	LEU	H
0061	SER	H

54	A	V	H
55	A	F	H
56	A	A	H
57	A	Q	H
58	A	T	H
59	A	V	H
60	A	L	H
61	A	S	H
62	A	H	H

helix alpha from
54 to 62 position

Helix count

My model	DSSP	PDB protein Database
From 8 to 18	Frome 5 to 15	From 4 to 18
From 43 to 46	From 16 to 20	From 49 to 63
From 54 to 64	From 44 to 46	From 69 to 84
From 69 to 83	From 50 to 62	From 109 to 125
From 113 to 125	From 70 to 83	
	From 110 to 124	

Sheets count

My model	DSSP	PDB protein Database
From 15 to 16	From 25 to 27	From 25 to 33
From 54 to 55	From 31 to 33	From 88 to 103
From 80 to 81	From 88 to 94	
From 115 to 116	From 79 to 103	