Secondary Structure Assignement

MESSAK Imane
EN-NAIMANI Saad

Bioinformatics master students, Paris University

September 16,2020

1 Introduction

DSSP is a database of secondary structure assignments (and much more) for all protein entries in the Protein Data Bank (PDB). DSSP is also the program that calculates DSSP entries from PDB entries.

1.1 General Purpose

The purpose of this project is to generate a pattern-recognition program that assigns secondary structure elements (SSE) to a protein by using the dictionary of protein secondary structures (dssp)1 fixed by Kabsch and Sander in the mid-80's. SSE are based on hydrogen-bonded and geometrical features extracted from atomic coordinates.

1.2 Secondary structure Assignment

As Kabsch and Sander described in their paper[1], the implemented dssp program assigns the SSE according to a specific priority order. First α-helices (**H**) are assign, follow by isolated β-bridges (**B**), extended strands (**E**), 3_{10}-helices (**G**), π-helices (**I**), H-bonded turns **T** and then bend (**S**). The assignment of a SSE depends mainly on H-bonded features.

2 Material & methods

2.1 GANTT Diagram

we used GANTT to plannify our time.

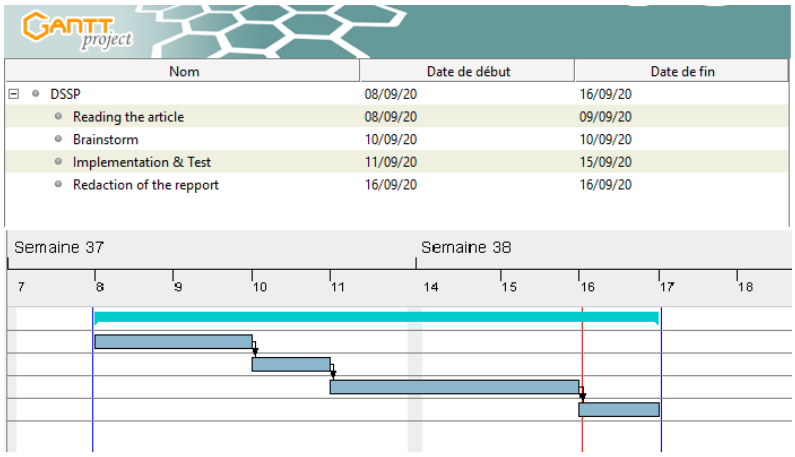

Figure 1: GANTT Diagram

2.2 Programming tools

All scripts are written in Python 3.6. PDB files are easily manage thanks to the BioPython[2] library. Adding hydrogens to a Protein DataBank (PDB) molecular structure file was done by Reduce program GitHub has been used for the development of this program. It is a web-based hosting service for version control using Git. Therefore, this project (its documentation and source code) is available online at the following address : `https://github.com/saad272/Projet1` .

2.3 Code structure

In addition to the main function (projet.py file), three modules are implemented : management (argument managment and display of results), classes (Residue class and its methods to calculate supplement parameters) and structures (Secondary structures assignment) (see Figure 1).

2.4 Input data

The input file must be in PDB (Protein Data Bank) format.We used three examples to test:

- - 1E9O Crystal structure of bovine SOD
- - 3OGB Sperm whale myoglobin mutant H64W deoxy-form (see Results section).

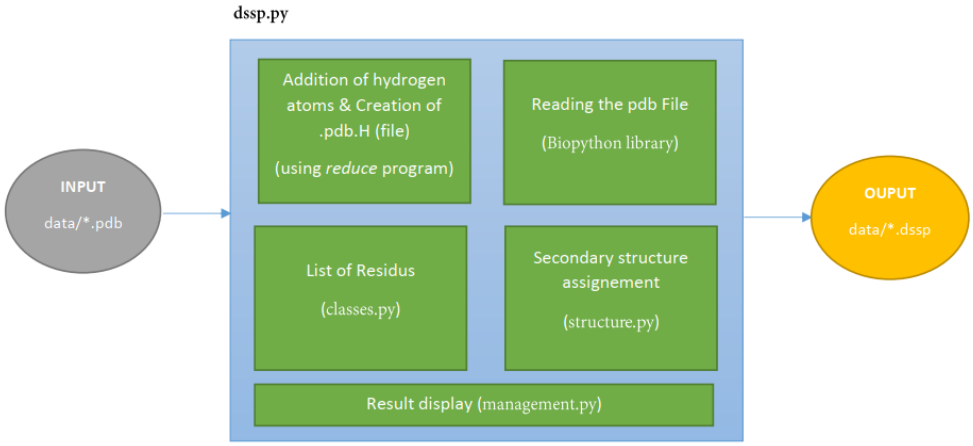

Figure 2: Schema of the implemented dssp program that assigns secondary structures from a pdb file.

2.5 Output data

Each line of a dssp output describes a residue and its assigned secondary structure .

- **RESIDUE :** Residue sequence number and chain identifier.
- **AA :** One letter amino acid code.
- **STRUCTURE** :
  - Secondary structure element : '**H**', '**B**', '**E**', '**G**', '**I**', '**T**', '**S**' or ' '
  - 3-turn/helix pattern : '>' (start), '<' (end), '**X**' (start/end), '**3**' (middle), or ' '
  - 4-turn/helix pattern : '>', '**X**', '<', '**4**', or ' '
  - 5-turn/helix pattern : '>', '**X**', '<', '**5**' or ' '
  - Geometrical bend : '**S**' or ' '
  - Chirality : '+' or '**-**'
  - Bridge label : '**a,b,c.**' for parallel β-bridge, '**A,B,C**.' for antiparallel β-bridge
- **BP1 BP2 :** Residue number of first and second bridge partner.
- **TCO :** Cosine of angle between C=O of residue i and C=O of residue i-1. TCO is near +1 for α-helices, TCO is near -1 for β-sheets.
- **KAPPA :** Virtual bond angle (bend angle) defined by the three C_α atoms of residues i-2, i, i+2. Used to define bend.

- **ALPHA :** Virtual torsion angle (dihedral angle) defined by the C_α atoms of residues i-1, i, i+1 and i+2. Used to define chirality.
- **PHI PSI :** Peptide backbone torsion angles (dihedral angles).
- **X-CA Y-CA Z-CA :** C_α atom coordinates.

2.6 Web server to calculate dssp files

The server is available at : https://www3.cmbi.umcn.nl/xssp/

3 Results

3.1 3OGB : Sperm whale myoglobin

3OGB is a is a dimensional structure with four -helices and does not contain β-sheets. the α helices are indicated in deferent colors, and the random ball regions are in white(see Figure 2) . this protein is the first whose structure was resoled by crystallography in 1958 by Max Perutz and John Kendrew, which earned them the nobel Prize in chemistry in 1962 . By comparing our dssp output (see Annexe section) with the result generated by dssp web (`https://www3.cmbi.umcn.nl/xssp/`), we notice the same assignment of the secondary structures for each residues except an isolated beta-bridge at residue number 59 that is not determined by the original program.

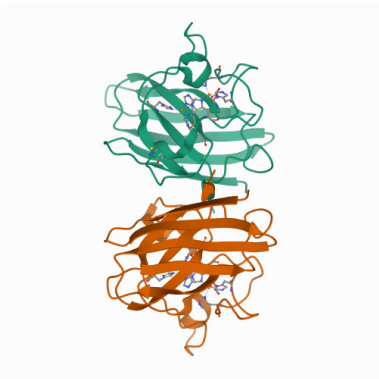

Figure 3: Structure of 1E90

3.2 1E9O: Crystal structure of bovine SOD

1E9O is a more complex example because the central part of these domains contains from 4 to more than 10 β strands the arrangement of these strands is most often antiparallel: the strands form 2 β sheets linked and compacted against each other The β sheets have a classic torsion and when 2 of these β

sheets are assembled, they form a barrel structure ("β barrel"). As before the main secondary structures (helices and strands) are well attributed. However, there are some differences/mistakes at some points.

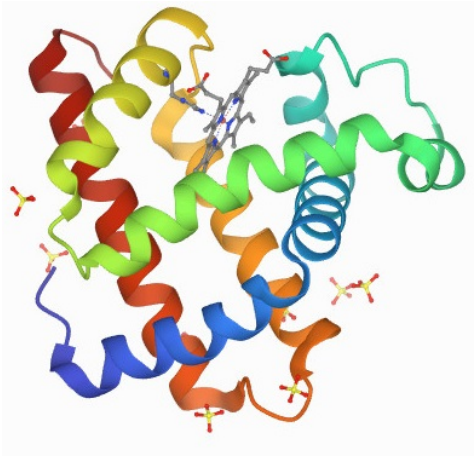

Figure 4: Structure of 3OGB

4 Conclusion

Our programme was inspired from the programme done by kabsch ans Sander in C++ , we tried to implement it in python to get assigns secondary structure elements for each residues, this means you do need to have a full and valid 3D structure for a protein to be able to calculate the secondary structure.

5 References and webography

1. Kabsch, W., and C. Sander. "Secondary structure definition by the program DSSP." Biopolymers 22 (1983): 2577-2637.
2. Peter JA, et al. "Biopython: freely available Python tools for computational molecular biology and bioinformatics." Bioinformatics 25.11 (2009): 1422-1423.
3. https://www3.cmbi.umcn.nl/xssp/
4. https://www.rcsb.org/structure/1E9O
5. https://www.rcsb.org/structure/3OGB

6 Annexe

```
(base) imane@imane-HP-Laptop-15-da0xxx:~/DSSP$ ./dssp.py -i data/1e9o.pdb
------------------ Secondary Structure Assignment using DSSP method ------------------------
  #  RESIDUE AA STRUCTURE BP1 BP2     TCO  KAPPA ALPHA   PHI   PSI    X-CA   Y-CA   Z-CA
    1    1 A A              0   0   0.000 360.0 360.0 360.0 146.8   -5.1   30.0   58.2
    2    2 A T        +     0   0   0.453 360.0  35.2-108.3 -18.2   -7.8   28.6   60.5
    3    3 A S  E     +A   21   0  -0.999  62.7 174.2-144.5 143.5   -6.1   25.3   61.1
    4    4 A A  E     -B  149   0  -0.913   8.9-161.2-143.7 164.4   -2.5   24.1   61.5
    5    5 A V  E     -C  148   0  -0.991   2.0-167.2-147.0 156.3   -0.6   20.9   62.3
    6    6 A C  E     -D   18   0  -0.987  11.1-151.4-148.5 133.6    2.8   20.0   63.4
    7    7 A V  E     -D   17   0  -0.958  20.3-145.9-107.0 121.4    4.6   16.6   63.6
    8    8 A L  E     +D   16   0  -0.728  27.3 162.6 -92.0 125.6    7.2   16.4   66.3
    9    9 A S  E     +E  144   0  -0.988   4.8 164.1-138.6 154.6   10.3   14.4   65.7
   10   10 A G         -     0   0  -0.787  53.2 -83.4-149.5-167.2   13.7   14.1   67.3
   11   11 A D  S     S+     0   0   0.508 110.1  48.3 -82.7  -7.1   16.9   12.1   67.6
   12   12 A G  S     S-     0   0  -0.248 110.4 -75.3-112.5-154.6   15.3    9.9   70.4
   13   13 A P  S     S+     0   0   0.358  86.8 116.9 -90.2   2.0   11.9    8.1   70.7
   14   14 A V         +     0   0  -0.639  36.0 154.3 -76.0 128.5    9.8   11.2   71.4
   15   15 A Q  E     + F   0  34  -0.958   8.3 156.1-146.1 162.4    7.2   11.8   68.7
   16   16 A G  E     -DF   8  33  -0.984  33.4-125.8-176.4 167.2    3.8   13.5   68.4
   17   17 A T  E     -DF   7  32  -0.998  23.6-171.1-131.8 126.5    1.1   15.2   66.5
   18   18 A I  E     -DF   6  31  -0.968   3.7-162.1-120.1 134.3   -0.2   18.6   67.5
   19   19 A H  E     -CF   5  30  -0.839   2.7-161.2-116.0 150.8   -3.3   20.3   65.9
   20   20 A F  E     -AF   4  29  -0.998   8.4-179.7-133.2 138.1   -4.1   24.0   66.1
   21   21 A E  E     -AF   3  28  -0.986  25.1-130.5-143.4 127.8   -7.5   25.6   65.5
   22   22 A A  E     + F   0  27  -0.547  31.6 171.2 -72.5 132.9   -8.5   29.3   65.6
   23   23 A K  E >   - F   0  26  -0.886  60.8 -41.6-147.0 110.0  -11.6   29.8   67.7
   24   24 A G  T 3   S-    0   0  -0.530 121.8 -29.9  71.8-132.6  -12.7   33.3   68.6
   25   25 A D  T 3   S+    0   0  -0.061 127.5  82.4-108.0  30.2   -9.6   35.3   69.5
   26   26 A T  E <   S-F  23   0  -0.846  70.8-132.8-123.1 167.3   -7.9   32.1   70.8
   27   27 A V  E     -FG  22  99  -0.987  18.2-148.8-122.4 125.6   -5.9   29.2   69.4
   28   28 A V  E     -FG  21  98  -0.865  11.9-162.2 -95.6 124.7   -6.9   25.7   70.6
   29   29 A V  E     +FG  20  97  -0.937  18.5 162.8-112.4 119.2   -4.0   23.3   70.7
   30   30 A T  E     +FG  19  96  -0.897  16.4  91.6-131.2 166.3   -4.7   19.6   70.9
   31   31 A G  E     -FG  18  95  -0.931  55.8 -93.7 144.9-170.2   -2.8
```

	16.4	70.2														
35	32	32	A	S	E	–FG	17	94	−0.972	17.2	−158.4	−146.4	151.2	−0.6		
	13.7	71.9														
36	33	33	A	I	E	–FG	16	93	−0.998	15.6	−168.2	−130.5	130.7	3.0		
	13.2	72.4														
37	34	34	A	T	E	+F	15	0	−0.813	57.4	47.7	−118.4	167.4	4.5		
	9.8	73.2														
38	35	35	A	G	S	S+	0	0	0.667	76.5	136.7	85.2	14.7	7.9		
	8.4	74.4														
39	36	36	A	L		−	0	0	−0.610	61.0	−108.2	−92.9	162.3	8.3		
	10.9	77.2														
40	37	37	A	T		−	0	0	−0.560	54.9	−90.3	−82.2	141.5	9.5		
	10.3	80.8														
41	38	38	A	E	S	S+	0	0	−0.183	91.0	45.0	−56.3	149.9	6.6		
	10.6	83.2														
42	39	39	A	G	E	S–H	87	0	−0.512	97.3	−27.0	110.6	178.4	6.0		
	14.1	84.5														
43	40	40	A	D	E	–H	86	0	−0.353	48.5	−170.6	−72.8	143.9	5.9		
	17.7	83.2														
44	41	41	A	H	E	–H	85	0	−0.995	29.5	−111.0	−137.0	133.1	7.8		
	18.8	80.0														
45	42	42	A	G	E	–H	84	0	−0.312	31.3	−164.4	−64.6	148.8	8.2		
	22.4	78.8														
46	43	43	A	F	E	+I	117	0	−0.937	19.1	158.8	−141.3	101.3	6.3		
	23.1	75.6														
47	44	44	A	H	B	–J	116	0	−0.937	43.2	−121.6	−132.6	157.9	7.1		
	26.2	73.6														
48	45	45	A	V	E	–J	115	0	−0.834	34.7	−155.5	−95.9	118.7	6.9		
	27.9	70.2														
49	46	46	A	H	B	–K	114	0	−0.720	21.6	−111.0	−94.1	157.6	10.4		
	28.7	69.1														
50	47	47	A	Q	S	S+	0	0	0.775	87.8	52.2	−63.6	−47.5	11.1		
	31.5	66.6														
51	48	48	A	F		−	0	0	−0.828	60.3	−146.3	−106.0	150.2	12.4		
	29.9	63.5														
52	49	49	A	G		+	0	0	−0.315	60.8	129.5	−98.0	48.0	10.9		
	27.1	61.3														
53	50	50	A	D		+	0	0	−0.927	27.9	168.8	−112.9	111.0	14.5		
	26.0	60.6														
54	51	51	A	N	>	+	0	0	0.043	42.8	115.4	−105.1	23.4	15.3		
	22.3	61.1														
55	52	52	A	T	T 3	S+	0	0	0.818	90.2	30.5	−63.3	−29.9	18.7		
	22.4	59.3														
56	53	53	A	Q	T >	S−	0	0	−0.481	120.1	−104.2	−119.5	51.0	20.5		
	21.5	62.6														
57	54	54	A	G	T <	S−	0	0	−0.293	82.0	−27.8	62.9	−139.7	17.7		
	19.5	64.2														
58	55	55	A	C	T >	S+	0	0	0.666	129.2	76.9	−81.6	−17.3	15.8		
	21.4	66.9														
59	56	56	A	T	G X	S+	0	0	0.871	87.8	58.5	−62.9	−36.4	18.8		
	23.5	67.8														
60	57	57	A	S	G 3	S+	0	0	0.553	81.7	85.6	−75.3	0.4	18.3		
	25.7	64.7														
61	58	58	A	A	G <	S−	0	0	0.649	86.2	−149.5	−75.4	−13.7	14.8		
	26.7	65.8														
62	59	59	A	G	<	−	0	0	−0.347	33.4	−30.4	81.8	−165.1	16.3		
	29.5	67.8														
63	60	60	A	P	S	S−	0	0	−0.095	86.4	−58.2	−83.1	−175.5	14.9		
	31.0	71.1														
64	61	61	A	H	B	–K	46	0	−0.410	64.1	−96.8	−68.0	146.0	11.3		
	31.3	72.2														
65	62	62	A	F		+	0	0	−0.457	58.8	163.0	−63.7	111.5	9.1		
	33.3	69.9														
66	63	63	A	N	>	+	0	0	−0.516	13.4	158.0	−136.3	72.1	9.2		
	36.7	71.6														
67	64	64	A	P	T 3	S+	0	0	0.764	80.2	51.1	−68.4	−22.7	8.0		
	39.5	69.3														
68	65	65	A	L	T 3	S−	0	0	0.323	107.7	−123.7	−97.1	5.7	7.2		

	41.8	72.3												
69	66	66	A	S	<	+	0	0	0.879	56.0	150.3	53.0	55.6	10.6
	41.3	73.9												
70	67	67	A	K		−	0	0	−0.506	48.7	−99.1	−106.0	176.1	9.4
	40.1	77.3												
71	68	68	A	K		−	0	0	−0.510	52.3	−84.0	−87.6	166.6	10.9
	37.8	79.8												
72	69	69	A	H		+	0	0	−0.381	60.3	152.6	−70.1	139.0	9.8
	34.1	80.0												
73	70	70	A	G		−	0	0	−0.810	45.5	−79.7	−147.6	−171.3	6.7
	33.2	82.0												
74	71	71	A	G	>	−	0	0	−0.549	53.0	−98.0	−91.9	171.0	3.9
	30.7	82.2												
75	72	72	A	P	T	S+	0	0	0.779	122.9	51.1	−64.1	−25.5	0.9
	30.9	79.9												
76	73	73	A	K	T	S+	0	0	0.502	91.4	102.6	−89.9	−0.2	−1.3
	32.8	82.4												
77	74	74	A	D	<	−	0	0	−0.529	67.1	−144.2	−82.8	145.4	1.4
	35.4	83.1												
78	75	75	A	E	S	S+	0	0	0.863	97.4	59.8	−73.5	−32.3	1.3
	38.9	81.7												
79	76	76	A	E	S	S+	0	0	−0.833	83.7	112.2	−99.6	96.9	5.1
	38.8	81.4												
80	77	77	A	R		−	0	0	−0.975	69.4	−97.2	−154.3	165.7	5.7
	35.9	79.1												
81	78	78	A	H	>	−	0	0	−0.672	35.6	−119.2	−80.5	159.3	6.9
	35.1	75.6												
82	79	79	A	V	T	S+	0	0	0.913	117.2	53.4	−61.4	−38.2	4.2
	34.6	72.9												
83	80	80	A	G	T	S+	0	0	0.416	78.4	113.7	−79.8	4.8	5.6
	31.1	72.5												
84	81	81	A	D	B	+J	44	0	−0.648	27.9	158.9	−80.2	93.1	5.4
	30.2	76.2												
85	82	82	A	L		−	0	0	0.261	41.4	−136.0	−100.8	10.9	2.7
	27.5	76.4												
86	83	83	A	G		−	0	0	−0.506	45.3	−27.4	80.7	−139.4	3.8
	26.1	79.8												
87	84	84	A	N	E	–H	42	0	−0.935	39.2	−150.8	−120.5	152.6	4.1
	22.3	80.4												
88	85	85	A	V	E	–H	41	0	−0.791	24.8	−133.9	−108.1	156.0	2.5
	19.2	79.0												
89	86	86	A	T	E	–H	40	0	−0.945	12.2	−157.0	−119.3	122.0	2.3
	16.2	81.4												
90	87	87	A	A	E	–L	93	0	−0.738	12.8	−140.3	−94.2	142.5	3.2
	12.6	80.5												
91	88	88	A	D	>	−	0	0	−0.327	41.3	−79.0	−86.3	−176.3	1.8
	9.6	82.3												
92	89	89	A	S	T	S+	0	0	0.739	131.8	54.3	−57.2	−24.7	3.6
	6.4	83.4												
93	90	90	A	N	T	S−	0	0	0.369	120.9	−109.3	−90.0	3.9	3.3
	5.0	79.8												
94	91	91	A	G	S	S+	0	0	0.651	72.4	141.1	78.8	19.5	5.0
	8.2	78.5												
95	92	92	A	V		−	0	0	−0.799	32.4	−166.8	−101.7	119.9	1.9
	9.5	76.9												
96	93	93	A	A	E	–L	87	0	−0.909	7.7	−153.2	−104.5	130.7	1.1
	13.2	77.1												
97	94	94	A	I	E	–G	32	0	−0.843	13.7	−149.0	−100.1	124.8	−2.3
	14.5	76.2												
98	95	95	A	V	E	+G	31	0	−0.853	32.4	148.3	−99.8	131.9	−2.3
	18.1	74.9												
99	96	96	A	D	E	+G	30	0	−0.674	16.3	151.3	−161.4	88.4	−5.2
	20.5	75.4												
100	97	97	A	I	E	–G	29	0	−0.969	29.4	−157.8	−121.1	141.5	−4.4
	24.2	75.8												