MASTER 2 Biologie Informatique

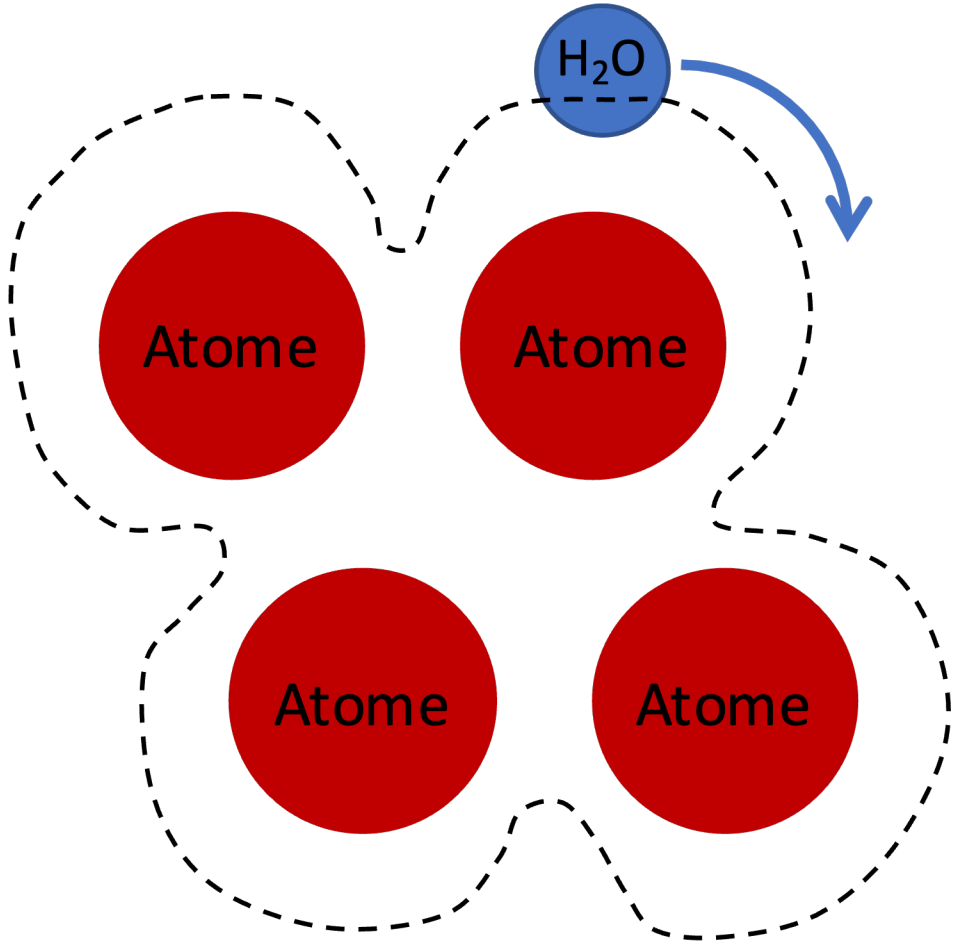
Université Paris Cité

Short project : Subject n°2

Calculation of the solvent accessible surface of a protein

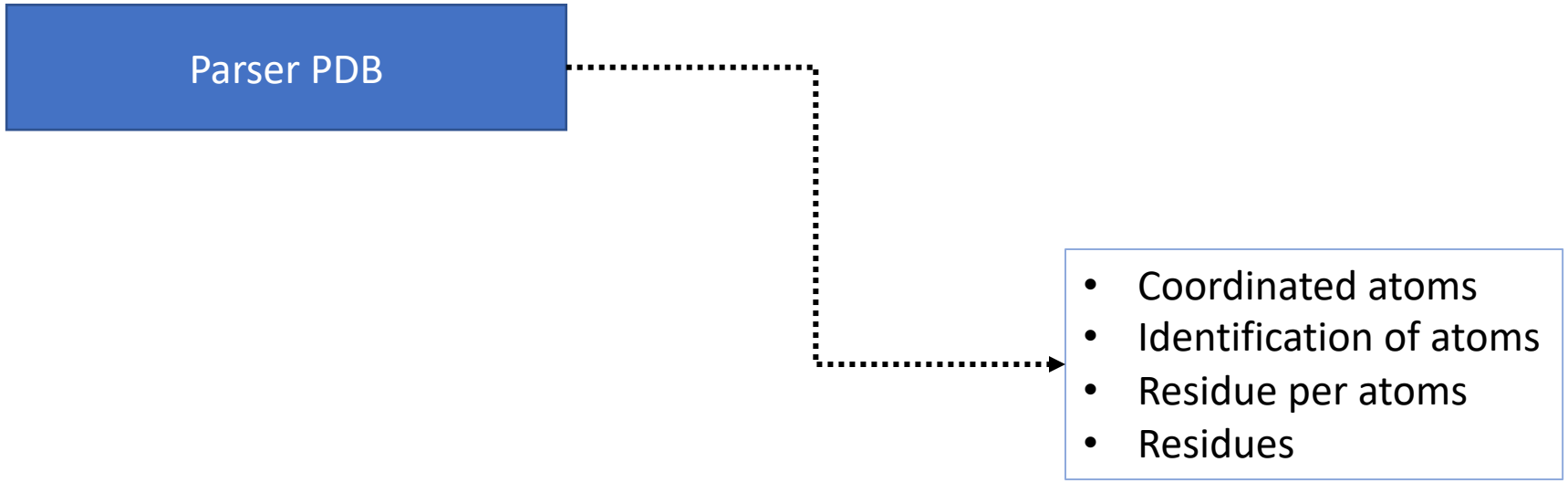
Malassigné Victor - Year 2022-2023

Objectif

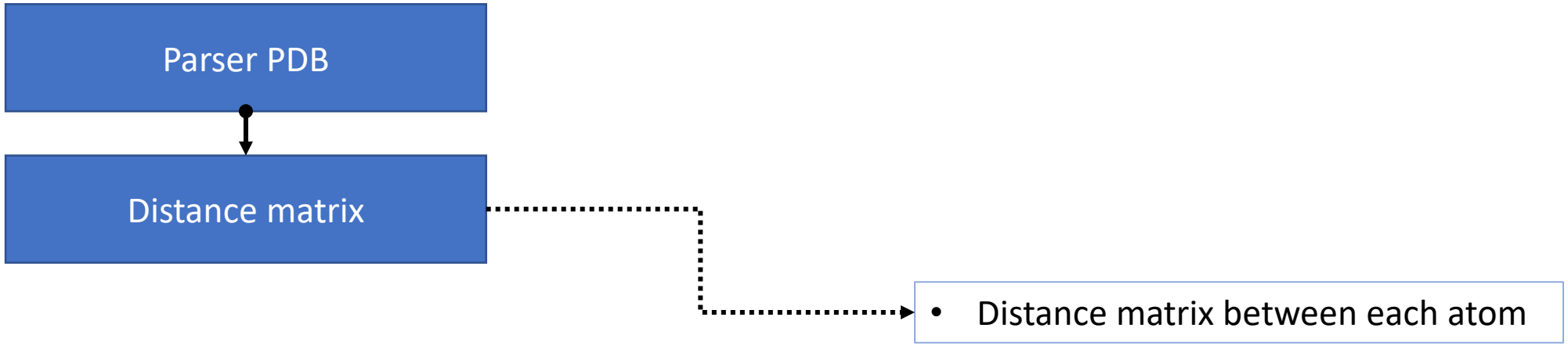


- Calculation of the protein’s surface incorporating the water molecule.
- This method uses the principle of rolling a water molecule over the atoms to determine the accessible solvant surface.

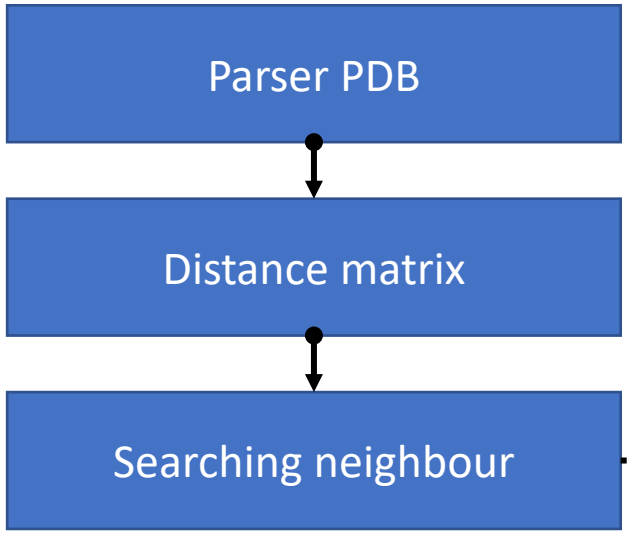
Pipeline


Parser PDB
• Coordinated atoms
• Identification of atoms
• Residue per atoms
• Residues

Pipeline

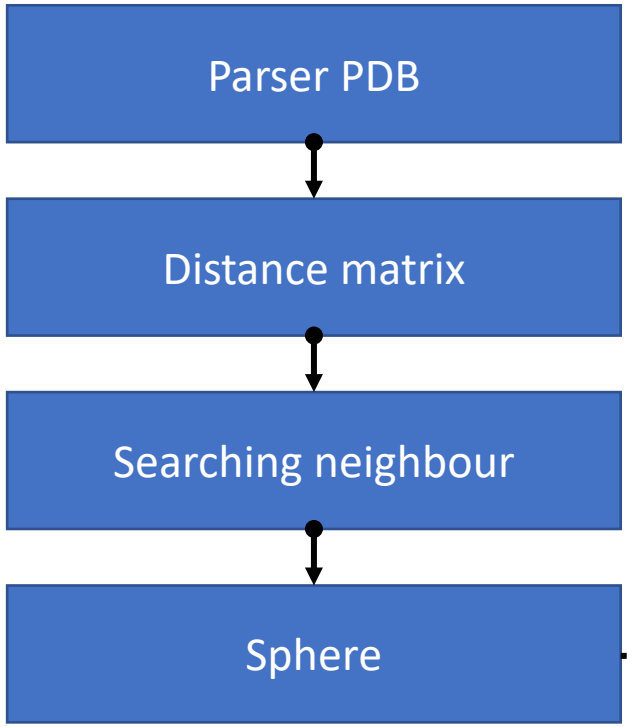

Parser PDB
Distance matrix
• Distance matrix between each atom

Pipeline



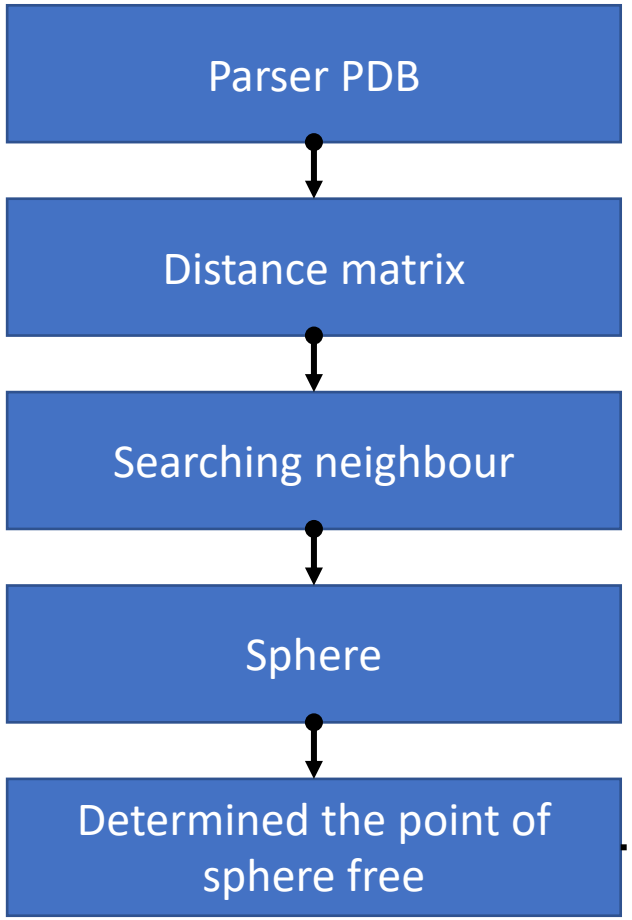
- Data in the distance matrix
- For each atom we selected :
  - all atoms with a distance <10 Å

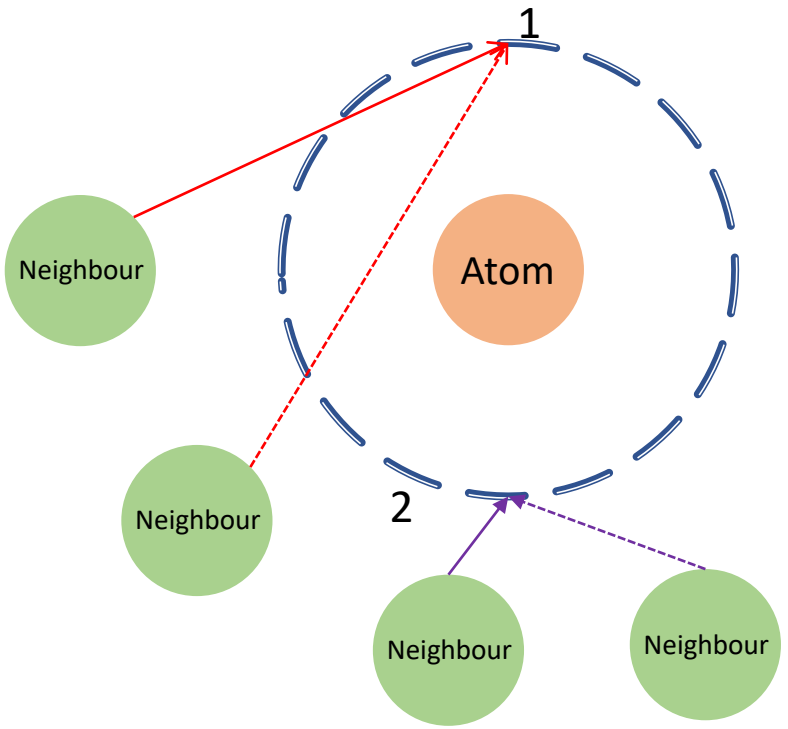
Pipeline



- Creating the sphere around the atom.
- Sphere consists of 92 points
- The radius of the sphere is composed of the Wan der Waals radius of the water and the Wan der Waals radius of the atom.

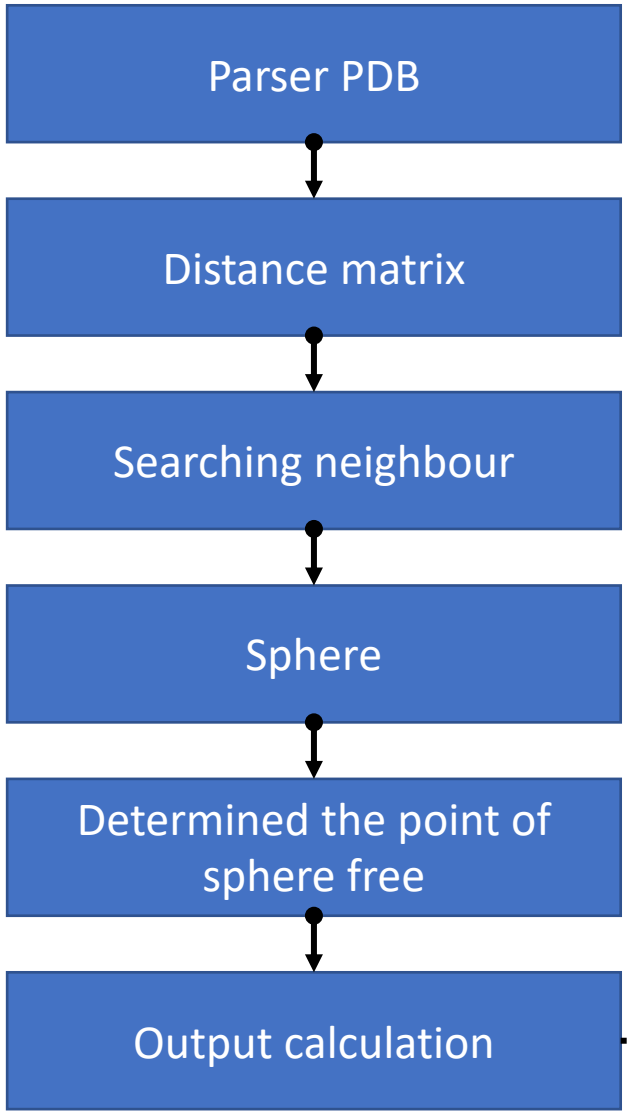
Pipeline





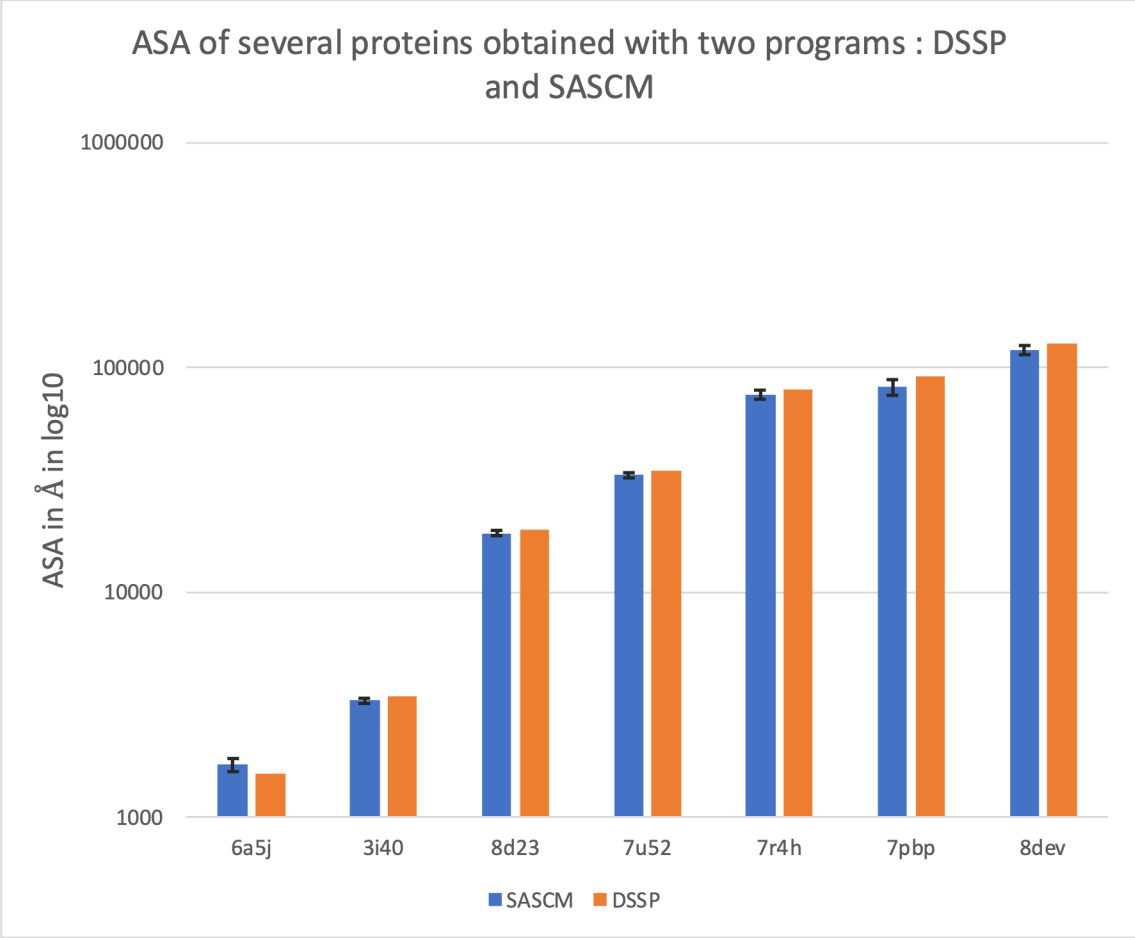
- Calculation of the distance between a point on the sphere and the neighbouring atoms.
- distance < radius of neighbouring Wan der Waals + water
- 1 : free point
- 2 : not free point

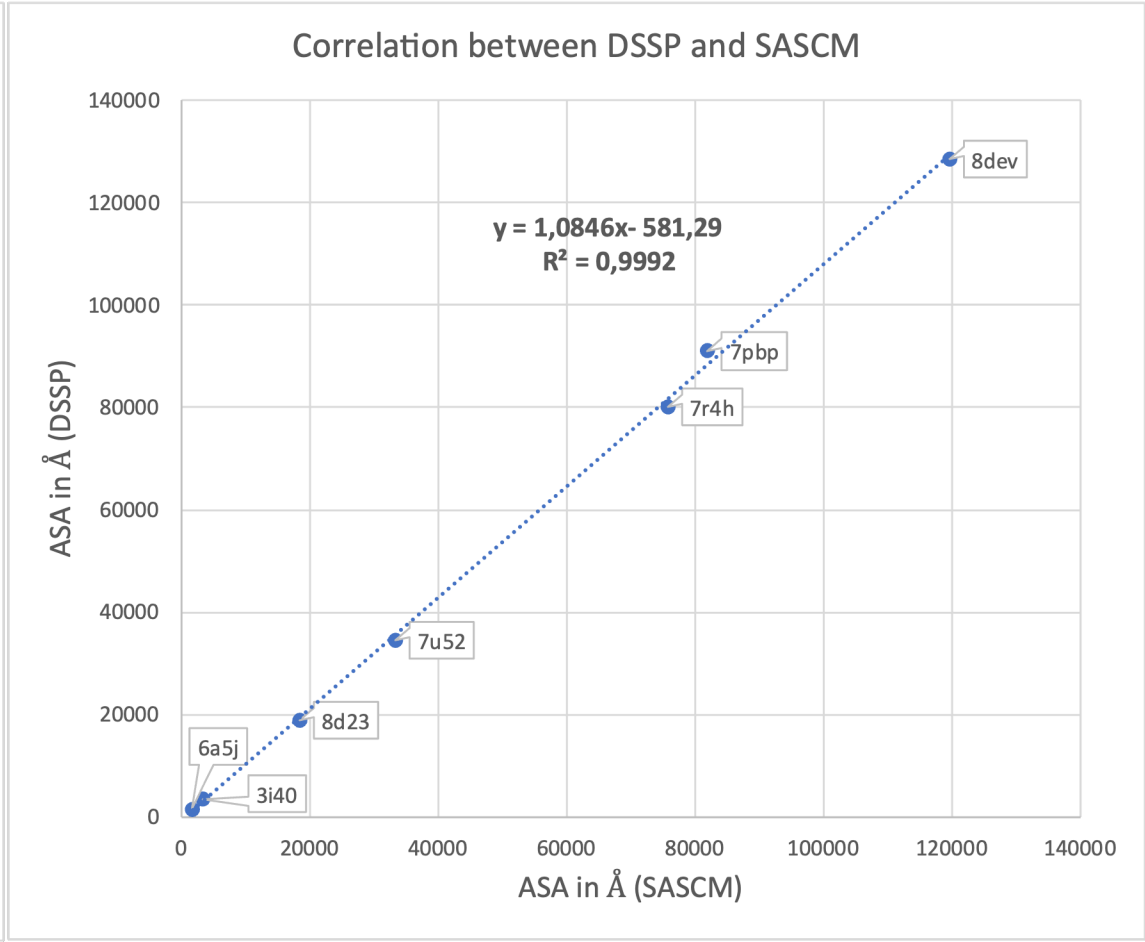
Pipeline



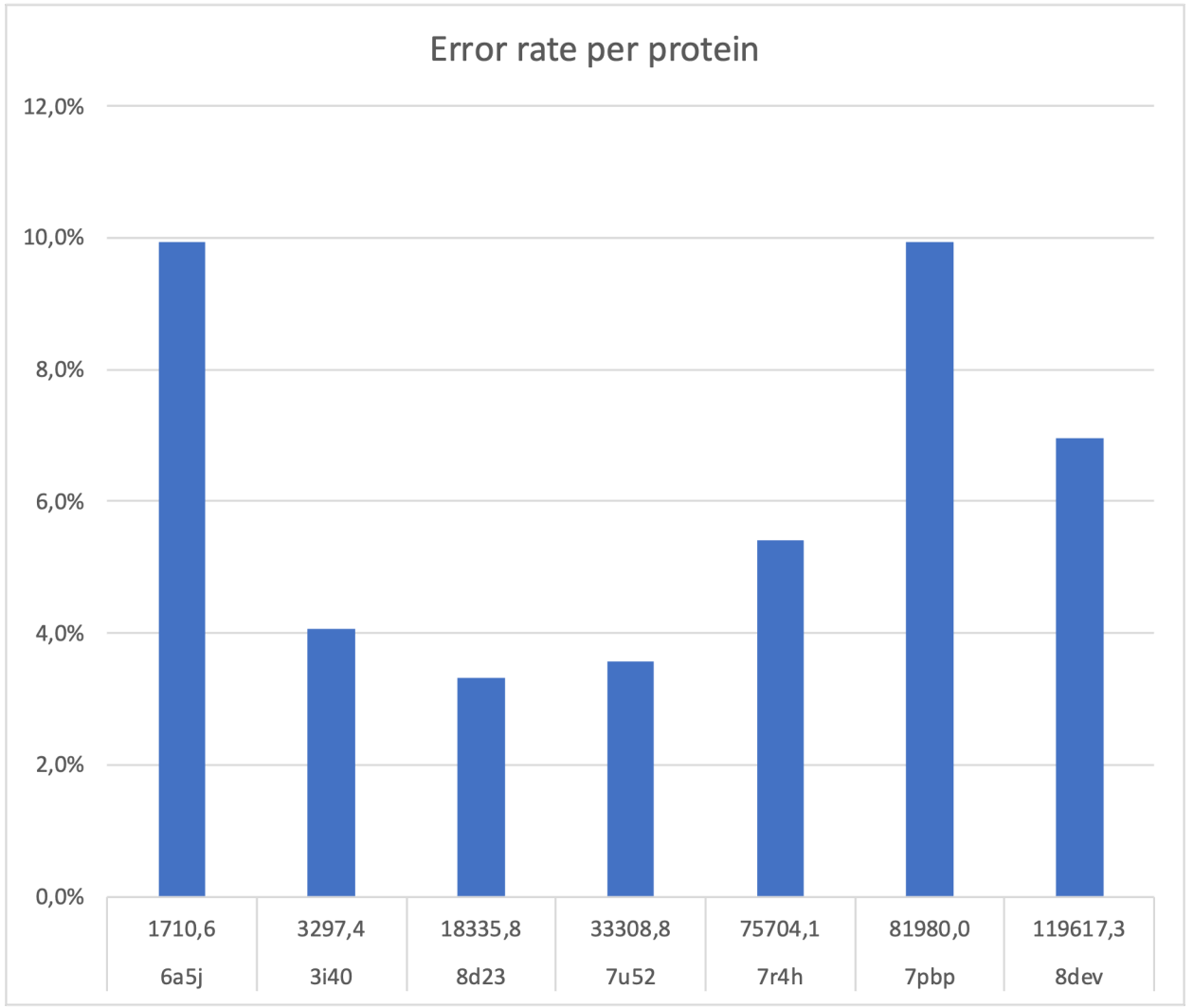
- Calculation the ASA
- Calculation the relative ASA
- Calculation the accessiblity percentage

Results


ASA of several proteins obtained with two programs : DSSP and SASCM
1000000
100000
10000
1000
ASA in Å in log10
6a5j
3i40
8d23
7u52
7r4h
7pbp
8dev
SASCM
DSSP


Correlation between DSSP and SASCM
140000
120000
100000
80000
60000
40000
20000
0
ASA in Å (DSSP)
y = 1,0846x- 581,29
R² = 0,9992
8dev
7pbp
7r4h
7u52
8d23
6a5j
3i40
0
20000
40000
60000
80000
100000
120000
140000
ASA in Å (SASCM)

Results


Error rate per protein
12,0%
10,0%
8,0%
6,0%
4,0%
2,0%
0,0%
1710,6
6a5j
3297,4
3i40
18335,8
8d23
33308,8
7u52
75704,1
7r4h
81980,0
7pbp
119617,3
8dev

Conclusion

- Similar results between DSSP and SASCM with an error percentage varying between 3 and 10%.
- Longer time of execution : a few seconds to a few minutes depending on the size of the protein.
- To save time → implement parallelization.