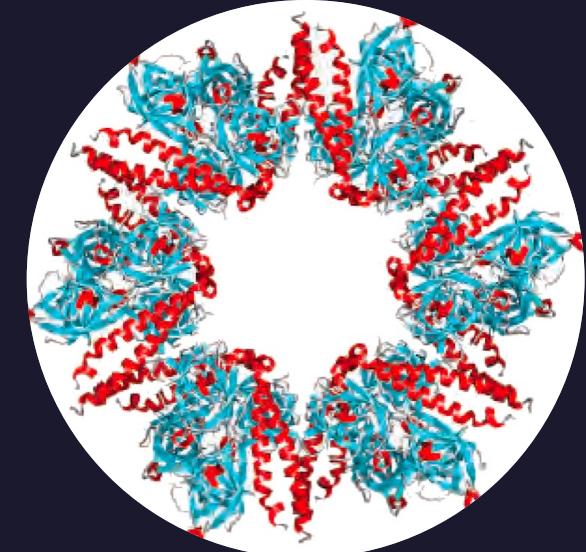


Determining accessible solvent area from pdb file with Python

Jane Schadtler Law

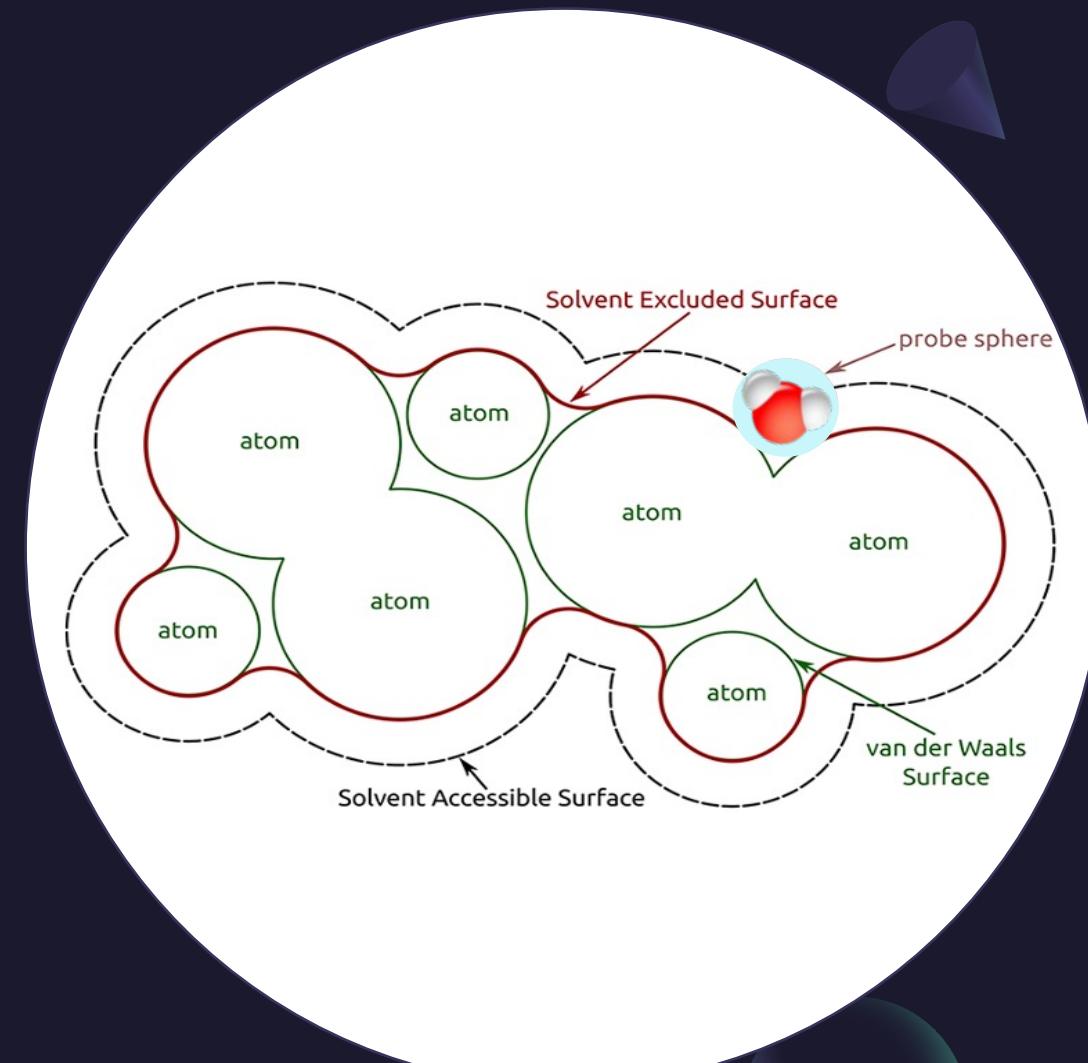
Contents

- What ASA is
- First steps of the python script
- Methodology adopted for the program
- Results
- Comparing with DSSP



Accessible surface area (ASA)

- Lee and Richards (1971).
- Key factor in understanding the protein's stability and energy.
- Addition of the solvent's radius in the calculation of the atom's surface.
- Determined from the atoms' coordinates.



Python script



First steps before coding the program's body

Extracting all usefull information:

Atoms id, coordinates, atoms' parent residues.

Calculating the distance matrix to identify of each atoms' neighbor.

Identifying the different elements on which to iterate

Parsing PDB

Filtering

Distance Matrix

Sphere Function

Body of the program

Filtering out all sorts of heteroatoms, disordered atom and modified residues.

Building a function able to return the coordinates of the evenly distributed 92 points of a sphere.
(Fibonacci Spiral Sphere)

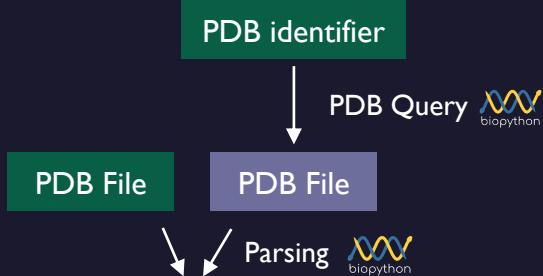


Legend

- Input
- Output
- List extraction
- Important conditions

The program's steps

Python 3.9



Structure object

- Keeping first model
- Filtering out heteroatoms
 - Atoms sequence
 - Residue sequence
 - Parent residue for each atom

Atoms' coordinates

Distance matrix

Loop on each atom

Identifying neighboring atoms (distance < 10 Å)
Sphere function: Sphere of 92 points around the atom

Loop on each point of the sphere

Loop on each neighboring atom

Calculates distance : point of the sphere - neighbor

IF distance < neighbor's radius + solvent radius :

This sphere's point is buried, do not continue the loop

IF NOT

This sphere's point is not buried with this neighbor, continue the loop on other neighbor

Calculates the number of « non-buried » points on the 92 points

Calculates ASA per atom

Calculates total area of solvated sphere

ASA

Accessibility %

$$\text{ASA} = \frac{\text{ASA per atm}}{\text{total area of solvated sphere}} \times 100$$

$$\text{rASA} = \sum_{i=1}^{\text{nb residues}} \left(\frac{\text{ASA}}{\text{MaxASA}} \right)^{\text{res } i}$$

$$\left(\frac{\text{ASA}}{\text{MaxASA}} \right)^{\text{res } 1} + \left(\frac{\text{ASA}}{\text{MaxASA}} \right)^{\text{res } 2} + \dots$$

Loop on each atom's parent

Create a dictionary to have the ASA for each residue

Loop the Dictionary

Divide each value by the residue corresponding MaxASA



rASA

$$\text{rASA} = \sum_{i=1}^{\text{nb residues}} \left(\frac{\text{ASA}}{\text{MaxASA}} \right)^{\text{res } i}$$

Results

Comparing with DSSP



Different outputs

Script run without output file

Bash command :

```
$ python src/main.py -n 6A5J
```

In terminal :

```
Downloading PDB structure '6A5J'...
For the protein 6A5J
  - The total accessible surface area is 1587.58 Å².
  - The relative accessible surface area is 4.32 Å².
  - The accessibility to solvent percentage is 10.33 %.
```

Script run with output file

Bash command :

```
$ python src/main.py -n 6A5J -o
```

In terminal :

```
Downloading PDB structure '6A5J'...
For the protein 6A5J
  - The total accessible surface area is 1587.58 Å².
  - The relative accessible surface area is 4.32 Å².
  - The accessibility to solvent percentage is 10.33 %.
```

+ In working directory :



results_6A5J.txt

→ with ASA per residue

Overview of the output file generated :

To save the output in a file

Format : results_proteinID.txt

```
● ● ● results_6A5J.txt  
For the protein 6A5J  
- The total accessible surface area is 1587.58 Å².  
- The relative accessible surface area is 4.3204 Å².  
- The accessibility to solvent percentage is 10.33 %.  
  
Residues on surface :  
Residues      ASA  
1    ILE    298.8166556575444  
2    LYS    178.61889729914552  
3    LYS    159.47817037501295  
4    ILE    115.7715684687561  
5    LEU    102.81920539005873  
6    SER    58.650966932004586
```

Same as in
the terminal

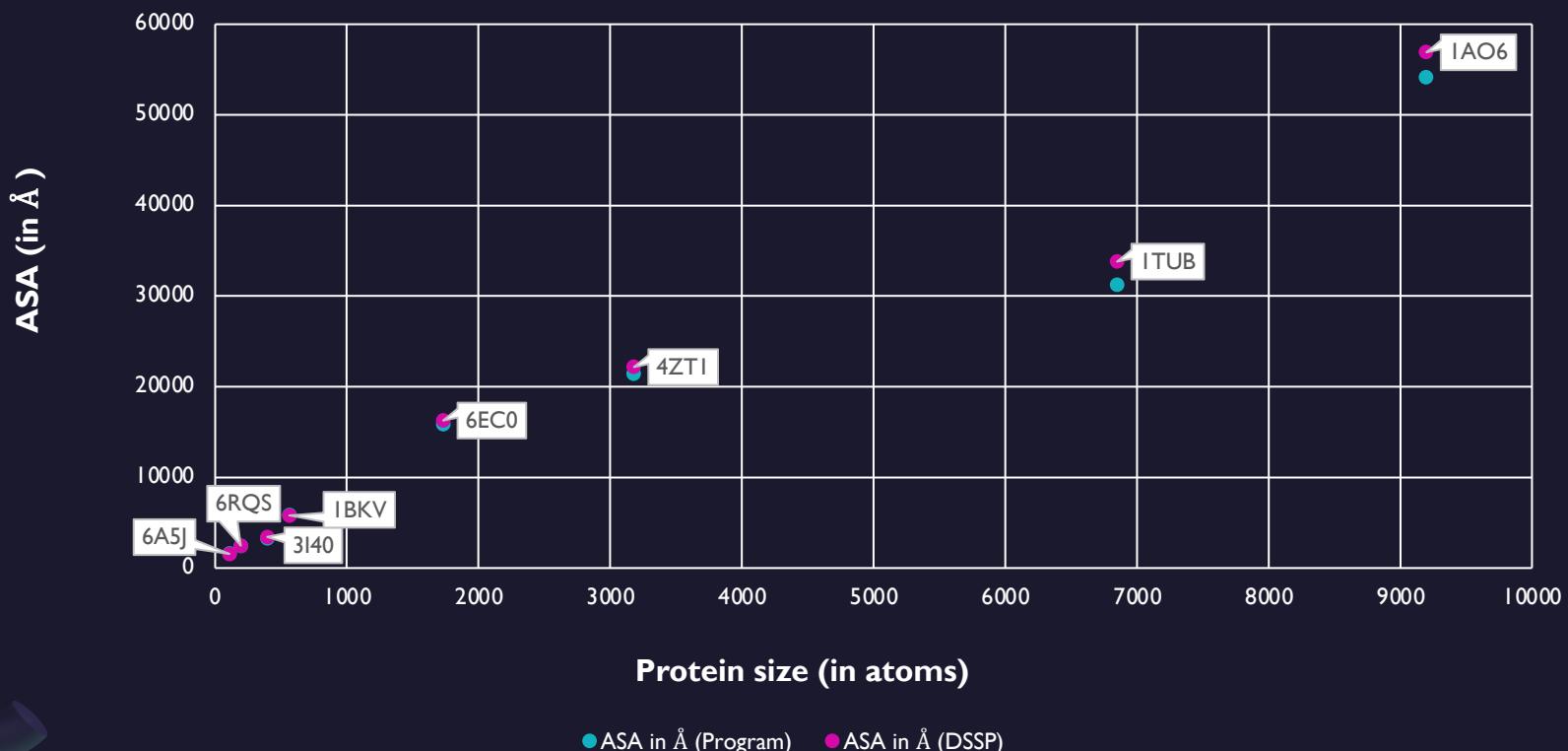
Bonus : Residue nb, id, ASA

Comparing results with DSSP

Define Secondary Structure of Proteins (Kabsch and Sander)

Over 8 different proteins

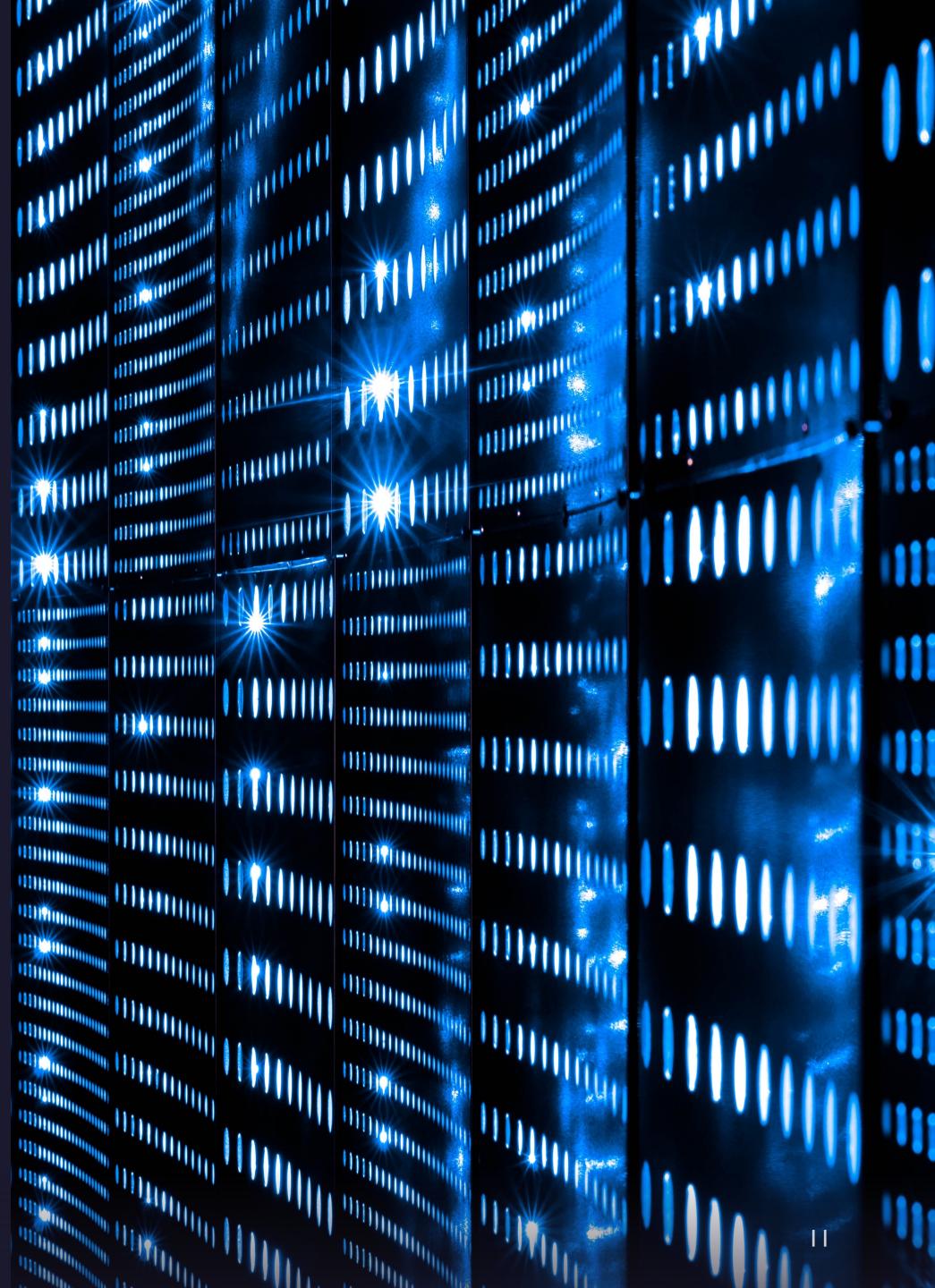
Accessible solvent area of the selected proteins according to their size.



| Protein | Error Percentage |
|---------|------------------|
| 6A5J | 2,0% |
| 6RQS | 0,7% |
| 3I40 | 4,1% |
| IBKV | 0,2% |
| 6EC0 | 2,4% |
| 4ZTI | 3,3% |
| ITUB | 7,5% |
| IAO6 | 4,9% |

Summary

- A little less accurate than DSSP
- Should execute time-consuming calculations simultaneously in multiple processors Tried with multiprocessing module...
- Increase number of points composing the sphere to the same as DSSP.



The background of the slide features a complex, abstract network structure composed of numerous small, glowing blue and red dots (nodes) connected by thin lines (edges). The nodes are more densely packed on the right side of the slide, creating a sense of depth and connectivity.

Merci

Jane Schadtler-Law

2022

Bibliography

- [1] B. Lee et F. M. Richards, « The interpretation of protein structures: estimation of static accessibility », *J. Mol. Biol.*, vol. 55, n° 3, p. 379-400, févr. 1971, doi: 10.1016/0022-2836(71)90324-x.
- [2] « Biopython · Biopython ». <https://biopython.org/> (consulté le 14 septembre 2022).
- [3] « NumPy ». <https://numpy.org/> (consulté le 14 septembre 2022).
- [4] R. P. D. Bank, « RCSB PDB - 1BKV: COLLAGEN ». <https://www.rcsb.org/structure/1BKV> (consulté le 14 septembre 2022).
- [5] R. P. D. Bank, « RCSB PDB - 1TUB: TUBULIN ALPHA-BETA DIMER, ELECTRON DIFFRACTION ». <https://www.rcsb.org/structure/1TUB> (consulté le 14 septembre 2022).
- [6] R. P. D. Bank, « RCSB PDB - 6EC0: Crystal structure of the wild-type heterocomplex between coil 1B domains of human intermediate filament proteins keratin 1 (KRT1) and keratin 10 (KRT10) ». <https://www.rcsb.org/structure/6ec0> (consulté le 14 septembre 2022).
- [7] R. P. D. Bank, « RCSB PDB - 1AO6: CRYSTAL STRUCTURE OF HUMAN SERUM ALBUMIN ». <https://www.rcsb.org/structure/1ao6> (consulté le 14 septembre 2022).
- [8] R. P. D. Bank, « RCSB PDB - 4ZT1: Crystal structure of human E-Cadherin (residues 3-213) in x-dimer conformation ». <https://www.rcsb.org/structure/4zt1> (consulté le 14 septembre 2022).
- [9] R. P. D. Bank, « RCSB PDB - 6A5J: solution NMR Structure of small peptide ». <https://www.rcsb.org/structure/6A5J> (consulté le 14 septembre 2022).
- [10] A. Shrike et J. A. Rupley, « Environment and exposure to solvent of protein atoms. Lysozyme and insulin », *J. Mol. Biol.*, vol. 79, n° 2, p. 351-371, sept. 1973, doi: 10.1016/0022-2836(73)90011-9.
- [11] M. Z. Tien, A. G. Meyer, D. K. Sydykova, S. J. Spielman, et C. O. Wilke, « Maximum Allowed Solvent Accessibilités of Residues in Proteins », *PLoS ONE*, vol. 8, n° 11, p. e80635, nov. 2013, doi: 10.1371/journal.pone.0080635.
- [12] « DSSP ». <https://swift.cmbi.umcn.nl/gv/dssp/> (consulté le 14 septembre 2022).
- [13] W. Kabsch et C. Sander, « Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features », *Biopolymers*, vol. 22, n° 12, p. 2577-2637, déc. 1983, doi: 10.1002/bip.360221211.