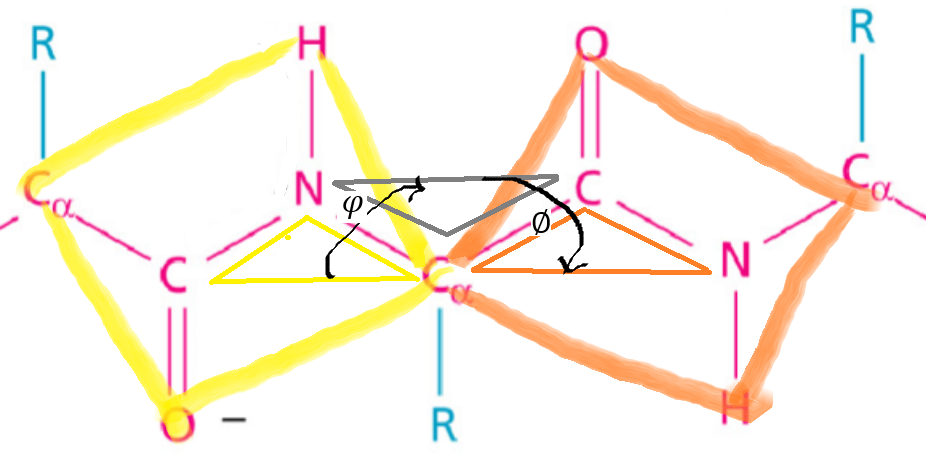
Notes:

Assigning secondary structure is different from structure predicton. Assigning a structure means we already know the angles between al atoms, we only need to figure out if these angles belong to a alpha helix or beta sheet.

# Calculating backbone dihedral angles using coordinates

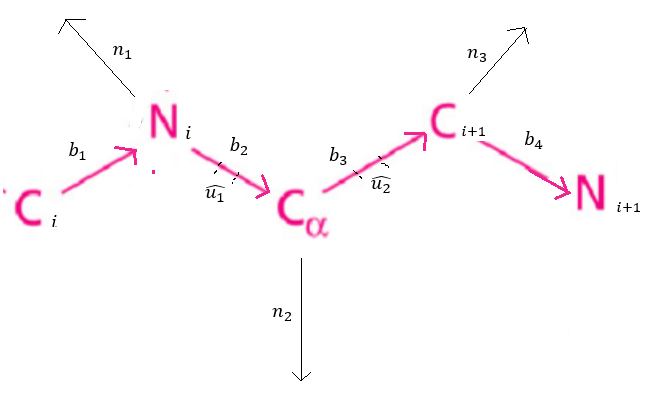
## Calculation of Phi and Psi dihedral angles



The image above shows 2 peptide bonds and the planes that can be created between the molecules of these 2 peptide bonds.

* Phi = dihedral angles between the yellow and grey plane
* Psi = dihedral angle between the grey and orange plane

To explain how to calculate these dihedral angles, we use the image below.



Between the atoms are vectors (b1, b2 and b3). These can be calculated by using the (x,y,z) coordinates of the atoms.

The normal vectors (n) are the vectors that make a 90 degree angle with the planes between the atoms and can be calculated using two vectors that determine the plane.

The unit vectors of b1 and b3 are calculated as follows:

To calculate the angle between the yellow and grey plates (phi) and the angle between the orange and grey plate (psi), we need to calculate cos and sin and then use these in the Atan2 function:

## Strategy for assigning secondary structure

So by knowing the phi and psi of the residue, we want to assign the secondary structure type (alpha, beta or loop). According to the Ramachandran plot (see figure), the following assignment can be done according to phi and psi values:

Alpha helix:

* -70 <= phi <= 40
* -160 <= psi <= -30

Beta sheet :

* 90 <= phi <= 180
* -180 <= psi <= -40

Loops (rest)

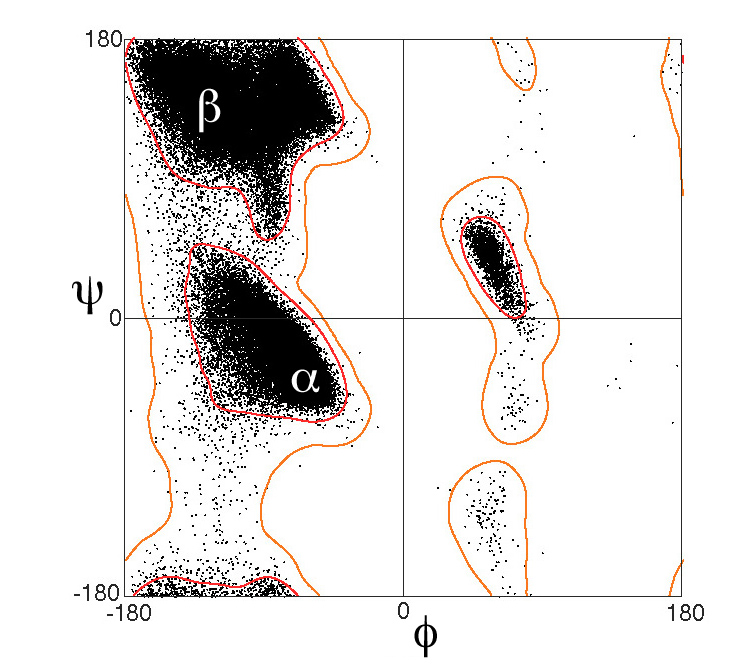


Figure 1. from https://en.wikipedia.org/wiki/Ramachandran\_plot

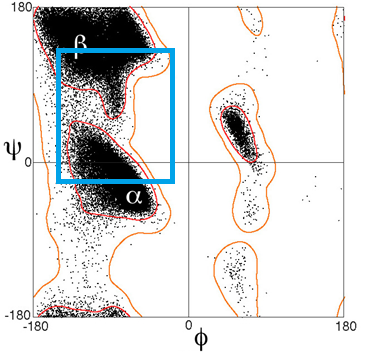
## Implement your strategy for assigning secondary structure

See CodeGrade

## Discussion of secondary structure assignment strategy

The 3D structure of the protein with uniProtID 1TIM (<https://www.rcsb.org/structure/1tim>), starts with a loop at position 1, a beta sheets starts at position 6 and ends at 12. Then a loop starts and ends at 16. Then a helix starts and ends at 31. My assignment of secondary structures missed the first beta sheed (pos 6-12). My program assigned these aa as loops. In the rest of my results, I only assigned a aa as beta sheet 5 times in the A part of the protein. While in the 3D structure, multiple beta sheets are shown.

From position 159 (LYS) till 167 (VAL), the 3D structure shows a beta sheet. My calculations in this region of 8 aa’s have a phi from -79 till -159 and a psi from -12 till 147. This region is shown in the figure with a blue square. Probably, the phi and psi coordinates of the 8 aa’s are just outside the thresholds I gave in the program for the assignment of the ss. To solve this problem, maybe the thresholds to assign the secondary structures can be made bigger. Or the thresholds should be based on a Ramachandran plot made for only the 1TIM protein.

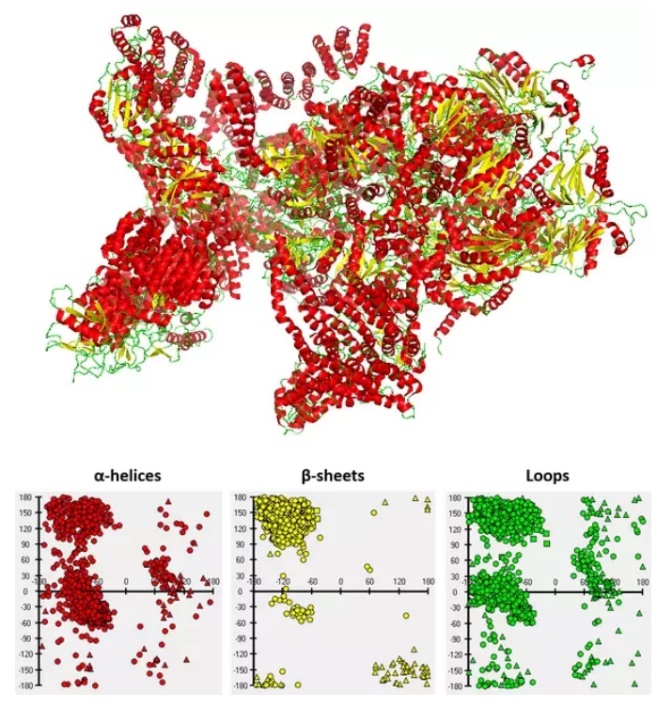
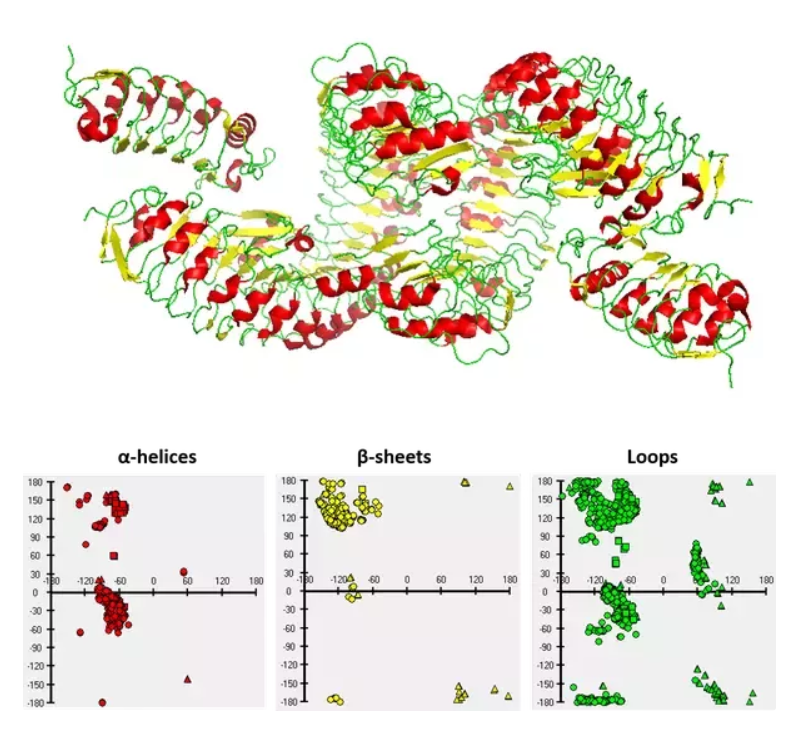


Feedback op ingeleverde assignment 2020:

You should have looked at your own assignment strategy and results . Also, you should have provided an improvement strategy

## Dihedrals versus hydrogen bonds

No currently available secondary structure assignment program (e.g. Stride or DSSP), relies solely on dihedral angles for their assignment. Describe the pros and cons of using hydrogen bonds versus dihedral angles for secondary structure assignments. Please use max 150 words.



<https://www.quora.com/Which-regions-of-Ramachandran-plot-are-loop-regions>

A con of using dihedral angles for secondary structure assignment is the variability of phi and psi values belonging to the secondary structures. Per protein, there can be a big difference on how the ramachandram plot looks. This means, per protein, there can be a big difference which phi and psi angles belong to which secondary structure (see figure).

# Propensities for amino acids to be buried

Every amino acid has a preference for a specific type of SS. By calculating the amino acid propensities, we quantify this preference.

A amino acid residue makes the AA unique of all other AA’s. It’s features, such as how it interacts with water, help guide the structure of the protein. So all the 20 amino acids are similar, expect for there residues.

## Implement propensities to be buried

See code

## Special amino acids

The code skips a line in de DSSP file, when it this line has a aa\_type property that is NOT present in the AccUnfold.data file.

## Table with propensities to be buried

|  |  |  |
| --- | --- | --- |
| AA | Propensity | Side chain characteristics according to Dyte-Doolitle scale |
| A | 1.3820626154972455 | Hydrophobic |
| C | 1.9566886944625534 | Hydrophobic |
| D | 0.3884944635847414 | Hydrophilic |
| E | 0.26495086462651446 | Hydrophilic |
| F | 1.7618183600385922 | Hydrophobic |
| G | 0.9781452179377049 | l |
| H | 0.7119939691389363 |  |
| I | 1.8388484035311152 | Hydrophobic |
| K | 0.137090339015454 | Hydrophilic |
| L | 1.708750658345777 | Hydrophobic |
| M | 1.4835092978209217 | Hydrophobic |
| N | 0.5209149651889374 | Hydrophilic |
| P | 0.6590159477807598 |  |
| Q | 0.4024891105888512 | Hydrophilic |
| R | 0.30924850196692033 | Hydrophilic |
| S | 0.806363315390888 |  |
| T | 0.8687042801192236 |  |
| V | 1.7716673461715893 | Hydrophobic |
| W | 1.4445410738461344 | Hydrophobic |
| Y | 1.1607944544164481 |  |
|  |  |  |

## Discussion on propensities

The propensities show how favourable the amino acid is to be buried. This means the following:

Propensity < 1 –-> hydrophilic

Propensity > 1 –-> hydrophobic

Propensity = 1 –-> neutral

Each amino acid can also be put in a hydrophobic, neutral and hydrophilic class.

* Hydrophobic (A, C, I, L, M, F, W, V).
* Neutral (G, H, P, S, T, Y).
* Hydrophilic (R, N, D, Q, E, K).

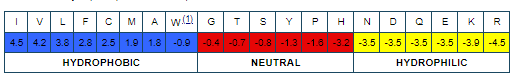


Image from <http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/IMGTclasses.html#:~:text=Amino%20acids%20are%20ordered%20from,%2DDoolitle%20scale%20%5B2%5D>. Scale used in image is the Kyte-Doolitle scale (Kyte, J. and Doolittle, R.F., J. Mol. Biol., 157, 105-132 (1982).)

Each amino acid has a propensity that correspond to the class it is in.

## Discussion on database used

The databased used to calculate the propensities contains atomic coordinates for a collection of proteins. The quality of these calculated propensities depends on the execution of the experiments which are used to obtain the coordinates and also the diversity of the proteins in the database. A low quality experiment can affect the quality of atomic coordinates and a low diversity of proteins can bias the total number of residues in a specific structure.

# Ramachandran plots

## Contour lines

The contour lines are derived from psi and phi calculations over a huge set of PBD files containing atomic coordinates (probably around 1 million proteins). So plotting the million phi and psi values is knowledge based. Then a line is drawn along the super dense data points, which is not knowledge based. So overall, the contour lines are not drawn knowledge based.

## Increasing the sampling

One would expect the contour lines to form a smaller circle/cluster. This is because the higher quality structures give less bias and there are less outliers in the datapoints. Those outliers are present now (with a low quality db) and are considered valuable data.

# Extra

**PDB file:**

CA is always part of the backbone. All other atoms with two letters are part of a residue!

Calculate by hand for residue PRO:

ATOM 2 CA ALA A 1 43.888 10.862 -6.231 1.00 0.00 C

ATOM 3 C ALA A 1 44.791 11.378 -5.094 1.00 0.00 C

ATOM 4 O ALA A 1 44.633 10.992 -3.937 1.00 0.00 O

ATOM 5 CB ALA A 1 44.722 10.051 -7.240 1.00 0.00 C

ATOM 6 N PRO A 2 45.714 12.244 -5.497 1.00 0.00 N

ATOM 7 CA PRO A 2 46.689 12.815 -4.561 1.00 0.00 C

ATOM 8 C PRO A 2 46.042 13.601 -3.411 1.00 0.00 C

ATOM 9 O PRO A 2 46.030 13.141 -2.267 1.00 0.00 O

ATOM 10 CB PRO A 2 47.640 13.732 -5.359 1.00 0.00 C

ATOM 11 CG PRO A 2 47.006 13.820 -6.760 1.00 0.00 C

ATOM 12 CD PRO A 2 46.056 12.615 -6.882 1.00 0.00 C

ATOM 13 N ARG A 4 45.521 14.773 -3.763 1.00 0.00 N

ATOM 14 CA ARG A 4 44.872 15.621 -2.730 1.00 0.00 C

1. **Make small yellow plane for atoms 3,6,7**

Eerst 2 vectoren maken van punt 3 naar 6 en van punt 6 naar 7.

V36 = 6 – 3 =(45.714 , 12.244 , -5.497) – (44.791 11.378 -5.094) = (45.714-44.791 , 12.244-11.378, -5.497- - 5.094) = …

V67 =7 – 6 = (46.689 12.815 -4.561) - (45.714 , 12.244 , -5.497)

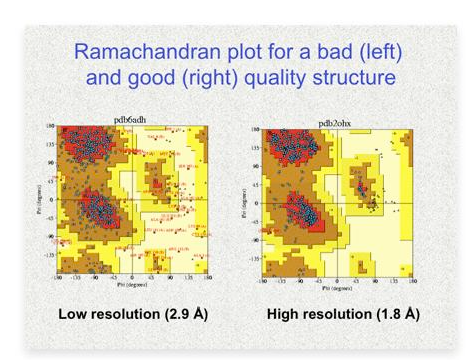
Plane = V36 x V67

1. **Make small orange plane for atoms 7,8,13**
2. **Make small grey plane for atoms 6,7,8**
3. Make big yellow plane for atoms 2,4,6,7
4. Make big orange plane for atoms 7,9,13,14

Coordinates of the backbone atoms from a PDB file: column 7, 8,9 are x,y,z?

Angle of the planes can only go until 180 degrees. That’s why atan2() is there.

<https://proteinstructures.com/Structure/Structure/Ramachandran-plot.html> :



The PDB files contains data obtained from the average structure formation.. The actual structure in real life “breathes”, so the structure formation changes a little bit.

Also the PDB files does not contain information about the active or inactive form. 