**The algorithmic flow of filterDCA:**

1. Starts with laoding the MSA file: where we are given the dca scores.

There are given the dca score of 40 proteins interaction with 104 proteins, including themselves.

1. Then the program goes to the function create\_dca\_matrix

It makes the msa file as the given format:

Pic

There are 40 rows and 103 columns

1. Then we select the filter (lets assume filter 69)
2. We load the npy file for that appropriate filter in the variable named “liste\_mat\_filtre”

The variable has 6 filters each element is made up by 69x69 matrix. The liste\_mat\_filter[0] looks like following:

Pic

1. Then we call the function pattern\_computation with the dca\_matrix and liste\_mat\_filter
2. We make another object named df, rearranging the dca matrix in the following way

Pic

1. For 6 filters we get into a loop and go to another function named compute\_matrix\_filter\_for\_one\_filter with the input of dca\_matrix and the filter which we call mat\_f in the new function

THE LOOP:

Lets assume for protein 0 and protein 0 and filter 69

* i\_centre , j\_centre == (0, 0)
* **sous\_matrix = we are taking a part of matrice\_dca**

**for (0, 0) proteins row 0 to 35 rows**

**and column 0 to 35 columns**

* we are making the (34,34) term in the filter nan
* we are formatting mat\_f. for (0, 0) proteins we have

row of 34: 69

column 34:69

* we are making a index indic\_flatt for protein (0, 0) it is 0
* some error checking
* flatting the filter and deleting the term of that index indic\_flatt
* deleting the same index term form sous matrix
* finding the correlation between sous\_matrix and filter

Lets assume for protein 0 and protein 1 and filter 69

* i\_centre , j\_centre == 0, 1
* sous\_matrix = we are taking a part of matrice\_dca

for (0, 0) proteins

row 0 to 35 rows

and column 0 to 36 columns

* we are making the (34,34) term in the filter nan
* we are formatting mat\_f. for (0, 1) proteins we have

row from 34th row to 69th row

column from 33th row to 69th row

we are making a index indic\_flatt for protein (0, 0) it is 0

* some error checking
* flatting the filter and deleting the term of that index indic\_flatt
* deleting the same index term form sous matrix
* finding the correlation between sous\_matrix and filter

Lets assume for protein 0 and protein 33 and filter 69

* i\_centre , j\_centre == 0, 33
* sous\_matrix = we are taking a part of matrice\_dca

for (0, 0) proteins

row from 0th to 35th rows

and column 0 to 69 columns

* we are making the (34,34) term in the filter nan
* we are formatting mat\_f. for (0, 33) proteins we have

row from row no 34 to row no 68 row

column from colm no 0 to clmn no 68

we are making a index indic\_flatt for protein (0, 0) it is 0

* some error checking
* flatting the filter and deleting the term of that index indic\_flatt
* deleting the same index term form sous matrix
* finding the correlation between sous\_matrix and filter