Proteocross tutorial

This is a tutorial destined to help you use and undersand the Proteo cross app. This app is made to help you handle the results from two expriments of protein identification : cutting and digestion of SDS-PAGE gel bands after migration, and pull-down (capture and enrichment of target.s protein.s).

This code will help you if :

- You cut several bands from your protein migration gel and want to know if they share proteins of interest
- You made a pull-down to enrich your protein.s of interest and want to know if they match the ones identified in the cut band.s

It crosses the results obtained from the identification of proteins from the two techniques.

About the band.s :

- You will be asked how many bands were cut and to select as many excel file.s containing the identified proteins
- For each band, you will have to enter the minimal and maximal molecular weight of the protein of interest

About the pull-down :

- You will be asked to enter the lowest ratio that you consider significant for your experiment
- A protein will be considered "of interest" if its ratio is > than the one you entered and its p-value ≤ 0.05

If you have pull-down data, the app will plot the associated volcano, highlighting the proteins shared by the pull down and each band. You will also have a 'Top 15 Ratio' volcano plot and associated data : it highlights the 15 significants proteins having the best ratio.

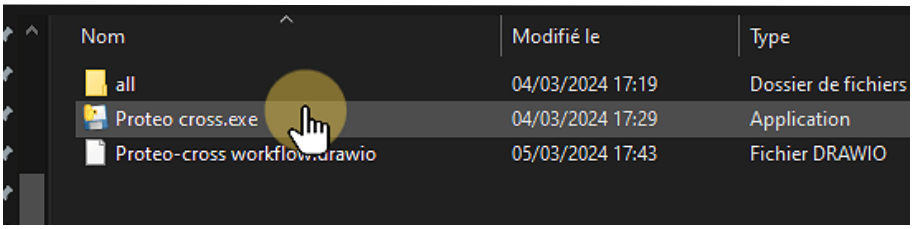
Results will be Excel files -if you have pull-down data,associated volcanos plot will be .png- and will be saved in the same folder as your data.

You can run as many analyses as you want or need one after the other, but make sure to have your different data in separate folders, otherwise the new Excel results files will overwrite the previous ones.

Table of content

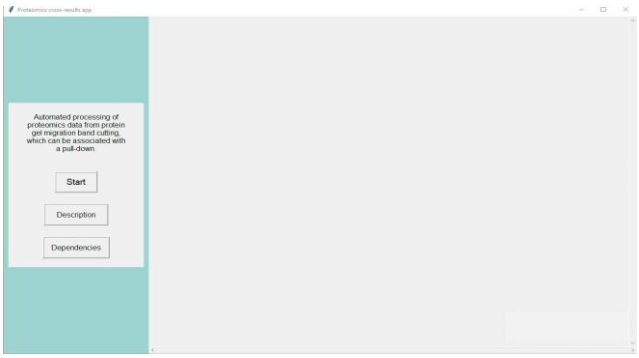
1. Launch the app

Click on 'Proteo cross'. You might not have the '.exe' visible on your computer, but the app will launch in the same way. On first launch, it may be slow to open.



2. The main window

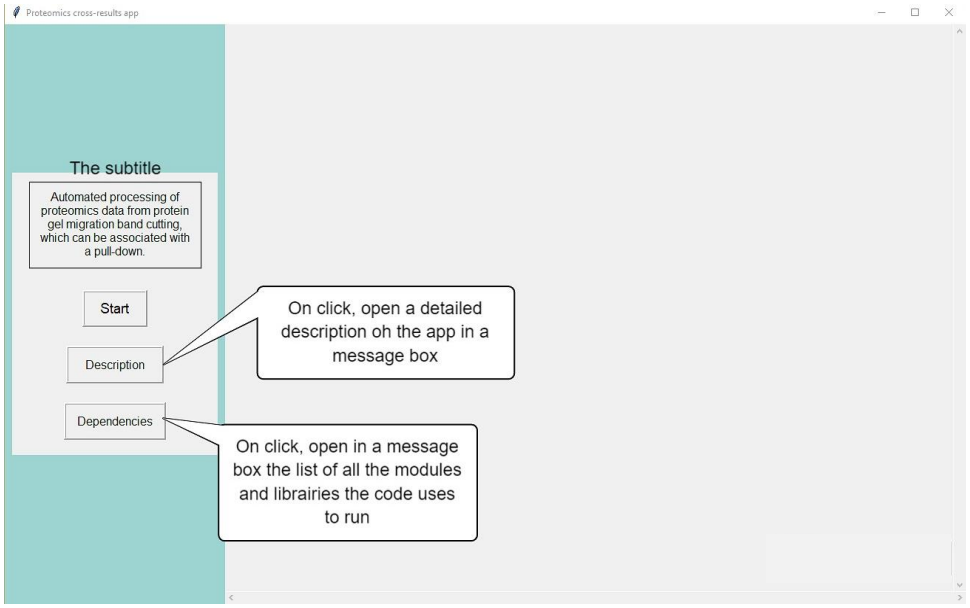
This is the window that will open. You have....



2.1 The informative objects

This left part of your screen will not change throughout the use of the application. It contains :

- The subtitle : a quick description of the app.
- Two informaitve buttons : Description and Dependencies.



About dependencies....

This message box is purely informative : you don't have to install yourself everything that's listed in it. On the first launch of the app, it will all be installed automatically.

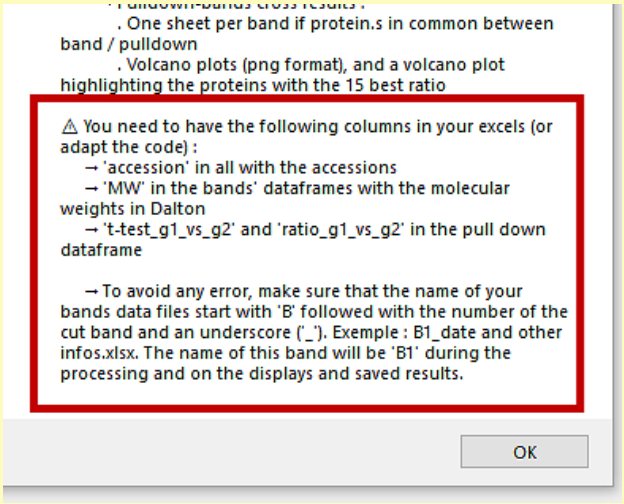
Important

As specified in the description, in order for the app to work you have to verify that your excel files contain thefollowing columns :

- accession (in both the bands and pulldown excels)
- MW (in the bands excels)
- t-test_g1_vs_g2 (in the pulldown excel)
- ratio_g1_vs_g2 (in the pulldown excel)

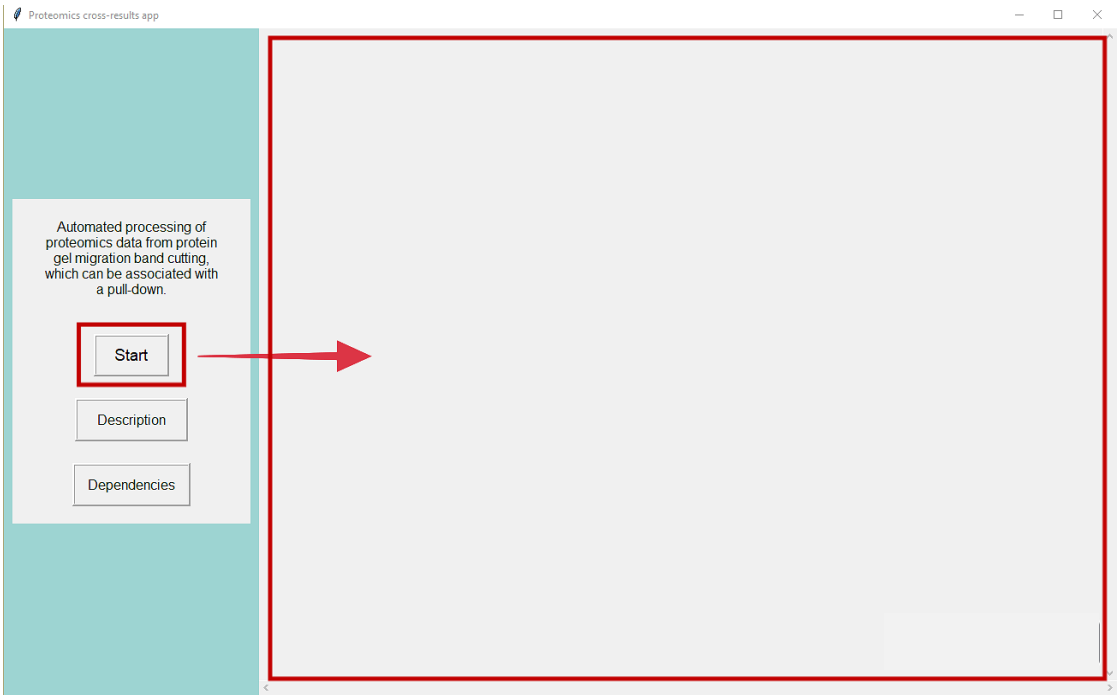
Be careful, as the writing is important, even the '_' between the words. Your columns names need to be identical tothese.

To avoid any errors, make sure your data files names for the bands start with B, followed by the number of the bandand an underscore (example : B1_date.xlsx).



2.2 The start button and the right frame

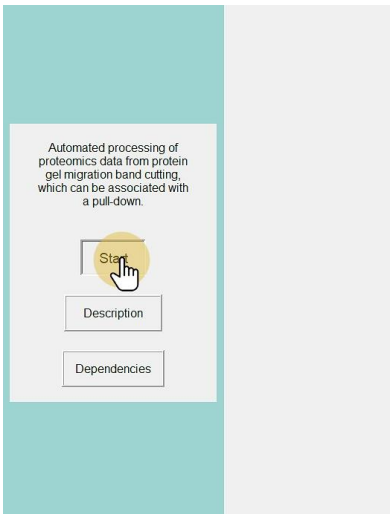
The Start button will start the data collection and treatment. All results will be displayed on the right frame as they are obtained.

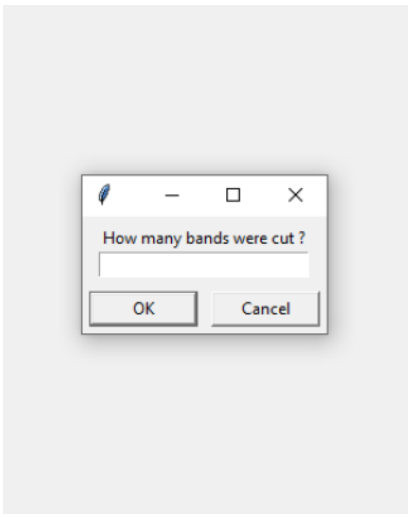


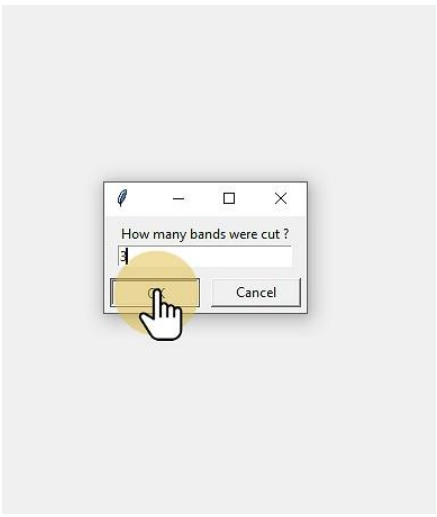
3. Band.s data processing

3.1 Enter the number of cut bands

Click on the Start button. It will open a question box asking you to enter how many bands were cut. It expects you to enter a number. In this example, I entered '3'. Write your number, then click on 'Ok'.





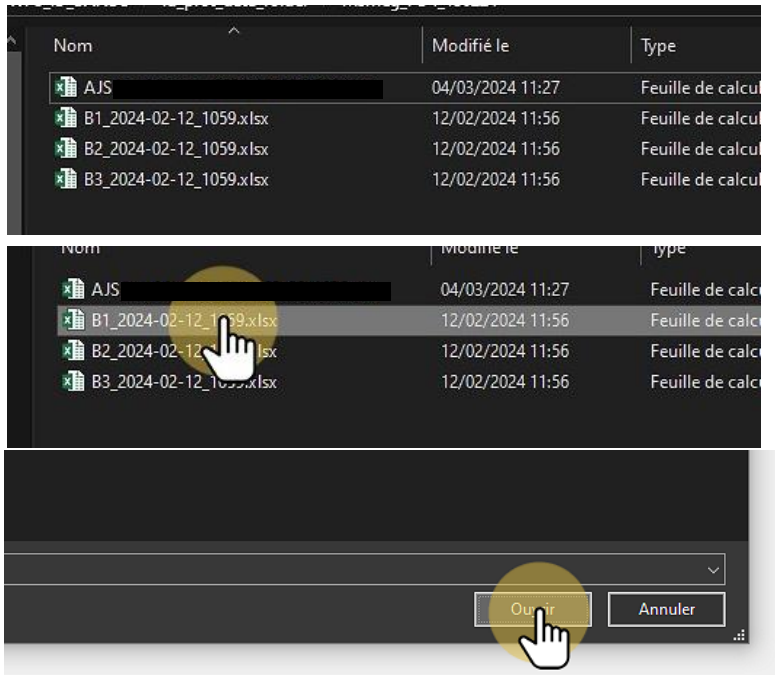


3.2 Share the data for each band

3.2.1 Select your data files

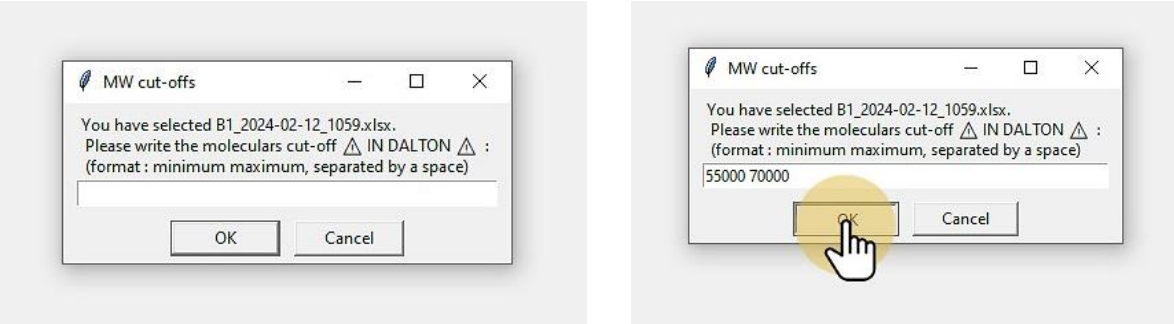
A file selection window opens. Move to the folder containing your data and select your file of interest. You can either :

- double click on your file,
- select your file and press the enter key,
- select your file and click on 'Open'.



3.2.2 Specify the molecular weights cut-offs for the selected band

Once you have selected an excel file corresponding to proteins identified behind a cut band, you will be asked to enter the molecular weight (MW) cut-offs. The app expects two numbers separated by a space. You can either put the minimum or the maximum in first.
Once you have typed your threshold molecular wiehgts ***in Dalton***, you can either click on 'Ok' or press Enter.

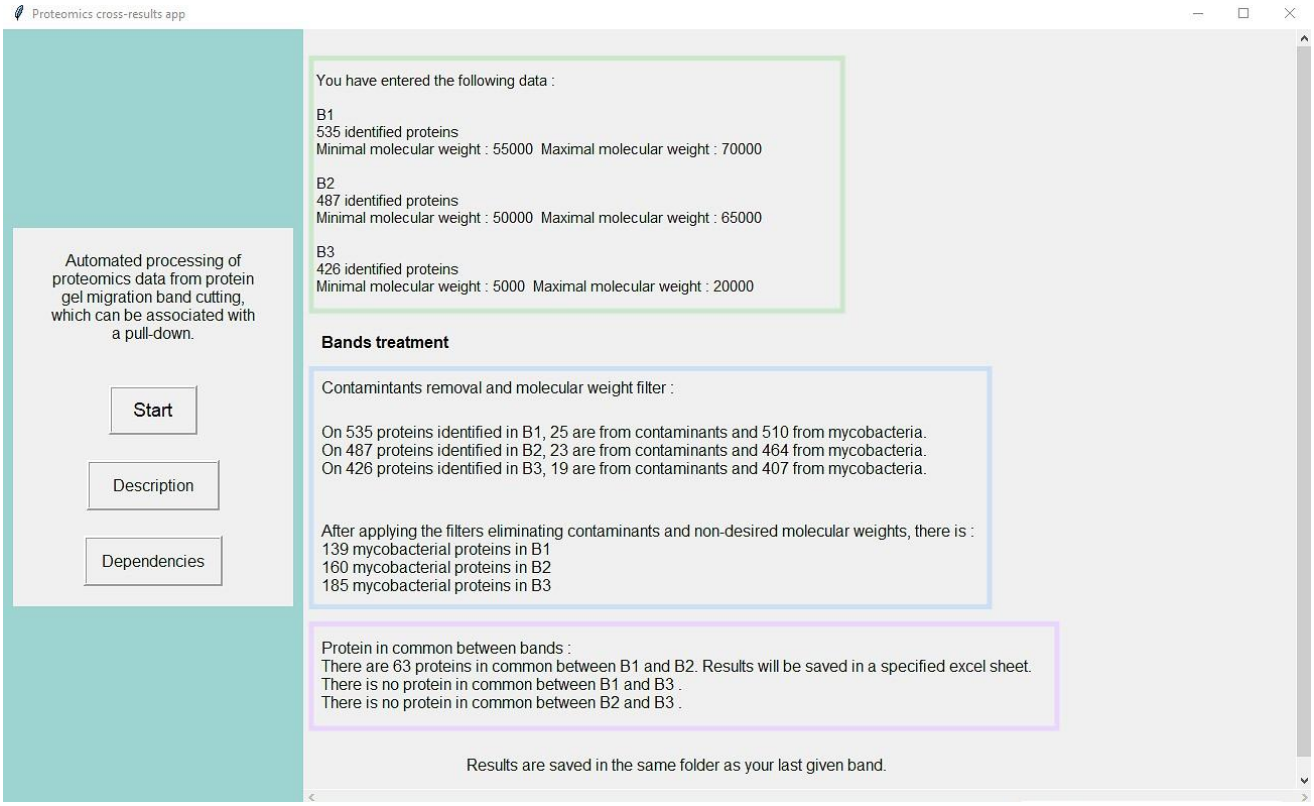


This step will repeat as many times as the number of bands you entered.

3.3 Your results are displayed…

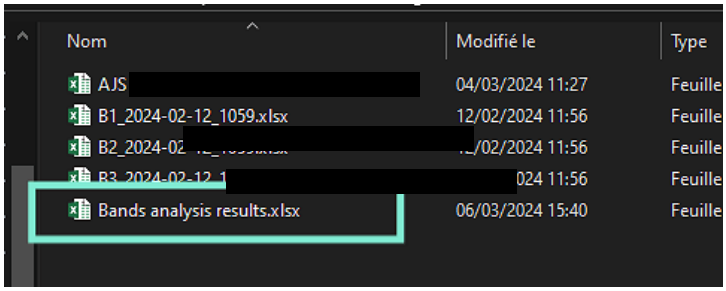
Congratulation, you have your results ! Some info is displayed on your screen. Here, you have :

- The data you entered
- The number of proteins that made it through the filters (contaminant removal and molecular weight cut- off)
- The number of shared proteins by pair of bands



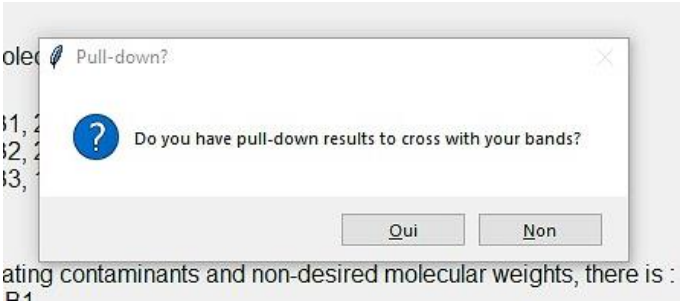
3.4 … And saved as an Excel file

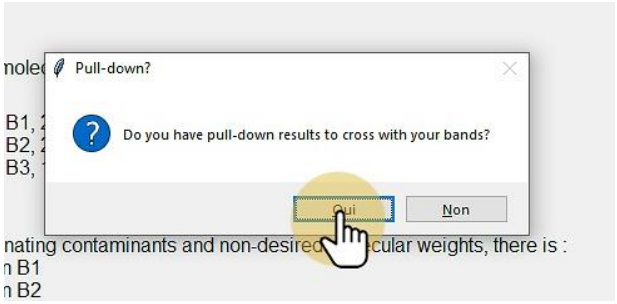
You also have an Excel file containing them, it the same folder as your last given band. It will be named ‘Bands analysis results’ (xlsx format).



4. Pull-down data processing

Once the band.s data are processed, the app will automatically ask you if you have assicated pull-down data. If you do, click on 'Yes'. If you don't have some, simply click 'No', then you can close the app whenever you want.



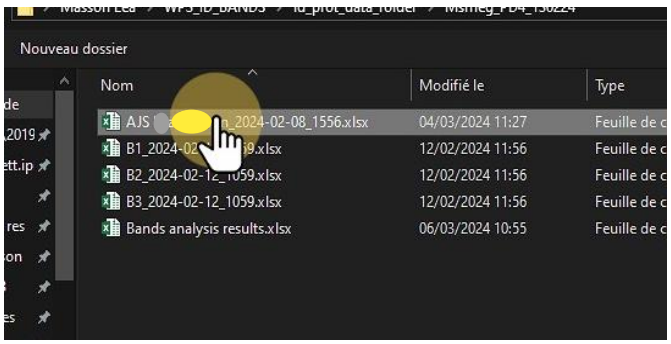


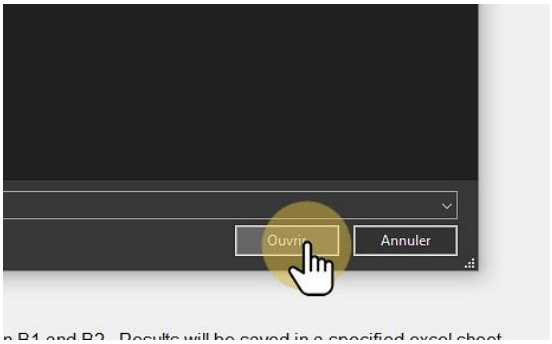
4.1 Enter your pull-down data

4.1.1 Select the data file

Just as before, a file selection window will open and you will be able to select the Excel file containing you pull down data. In the same way, to select the file, you can either :

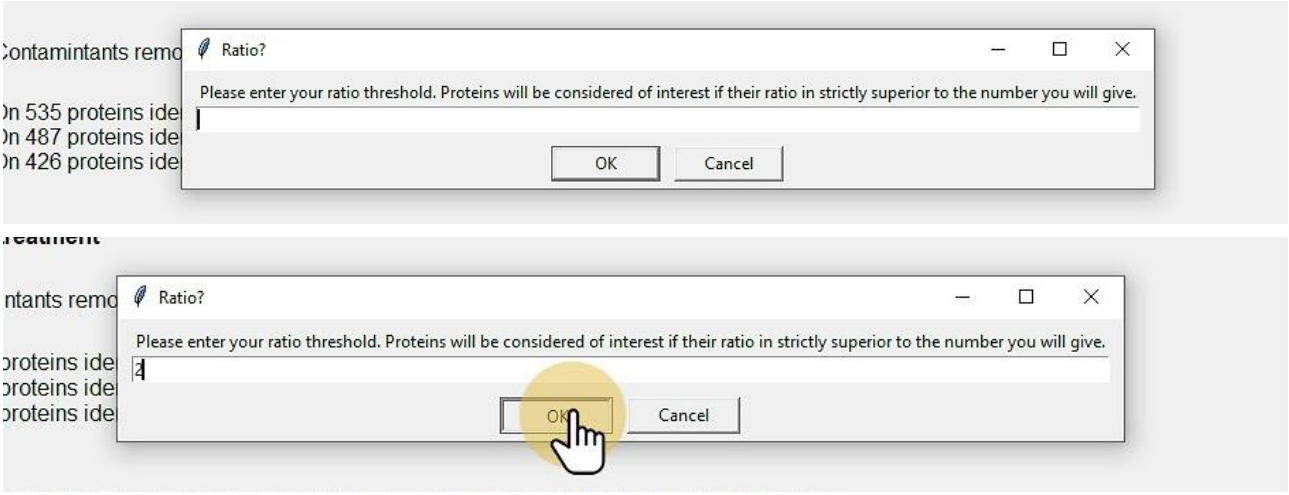
- double click on your file,
- select your file and press the enter key,
- select your file and click on 'Open'.





4.1.2 Enter your minimal ratio

You will then be asked to enter the lowest ratio at which proteins should be considered. Keep in mind that all proteins will be used to plot the volcanos, but only the ones that are significant and show a ratio strictly higher than that entered will be search in the bands. Once you have entered your number, click on 'Ok'.

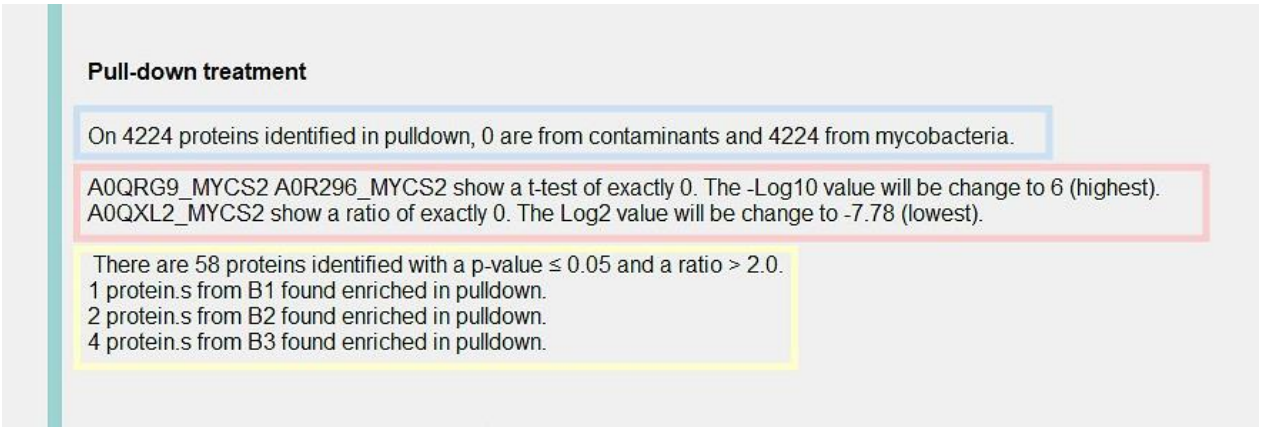


4.2 Your results are displayed…

4.2.1 Textual information

Once your pull-down data is treated, you will be able to scroll thought the displayed results. There are some interesting points to note :

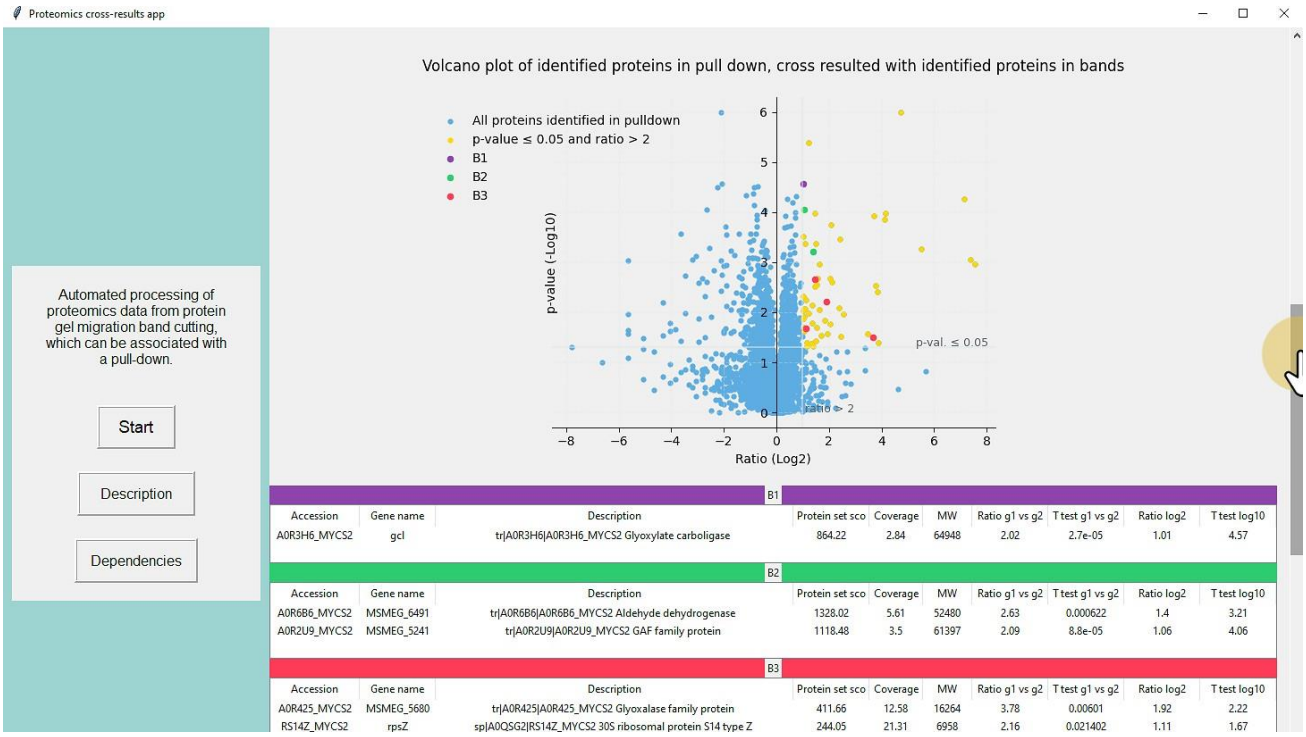
- The number of proteins after the contaminant removal
- Changed information in order to plot every identified proteins. It is possible that, in your data file, you have some identified proteins with a p-value and/or a ratio strictly equal to 0. The volcano plot requires to calculate the -Log10 value of the t-test and the Log2 of the ratio. As we can't do Logs on 0, a p-value of 0 will have its -Log10 set on 6 (highest significance score), and a ratio of 0 will get a Log2 of -7.78 (lowest ratio possible). The accession number of the corresponding proteins are specified.
- Your total proteins of interest and the number of shared proteins between the pull-down data and each band. Don't worry, their accessions will be displayed and saved too.



4.2.2 Volcano plots

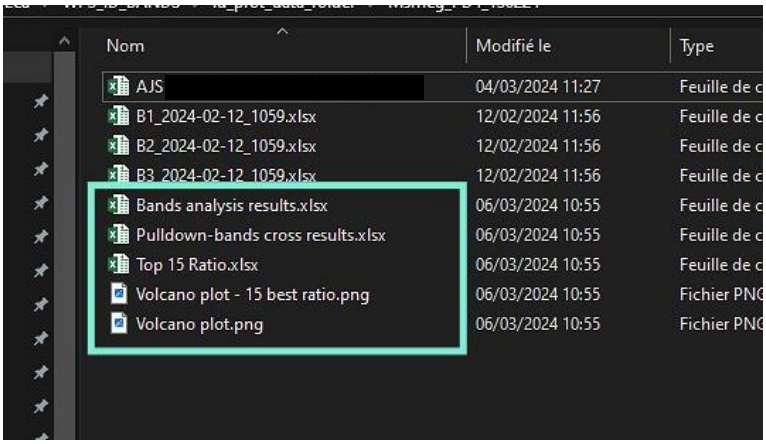
Just after, your main volcano plot will be displayed. All proteins identidied in pull-down and that are not contaminants will be displayed. This volcano will not only highlight all proteins with a *p*-value ≤ 0.05 and a ratio > 2, but also the ones shared between each band. For these, you will also have a table -one for each band- diplaying their accession number, gene name, description, score, coverage, MW, ratio and *p*-value with their Logs.

Beneath this, you will have another volcano plot bringing out the 15 significant proteins with the best ratio. You will see a table below showing the same list of information as above for these 15 proteins. The proteins shared with bands will still be colored.



4.3 … And saved with your pull-down data

The cross-results, the 'Top 15 Ratio' and the two volcano plots are saved in the same folder as your pull-down data.

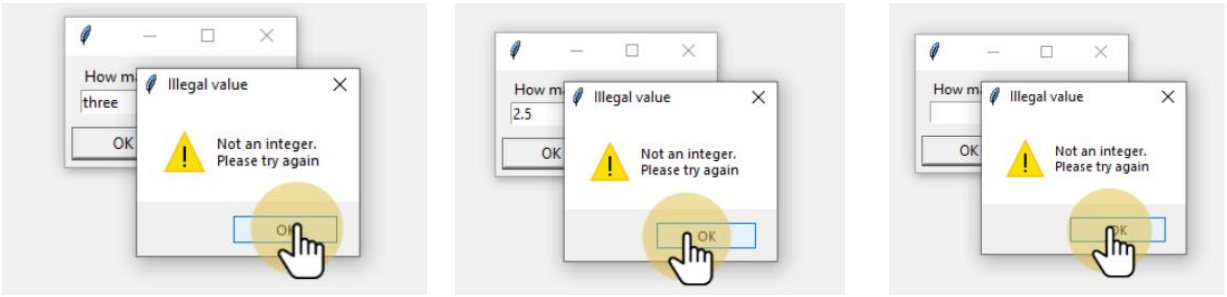


5. Potential errors managment

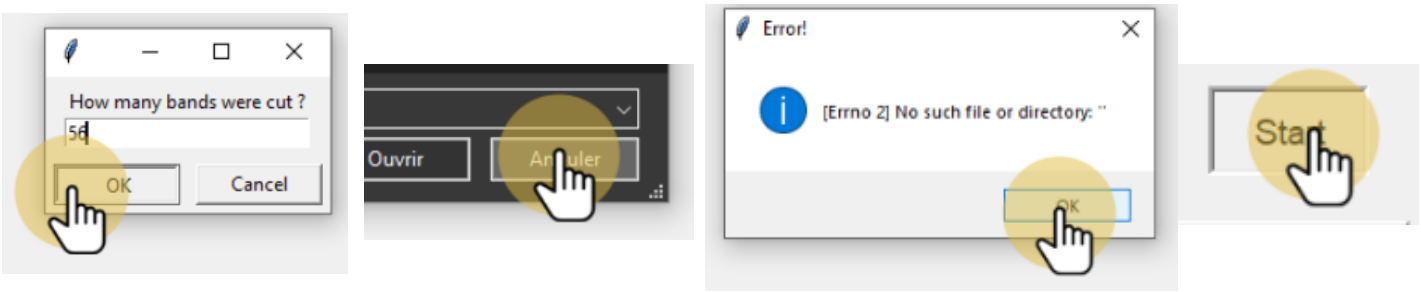
Here is how to manage potential problems while running the app.

5.1 Entering your number of cut bands

You didn't write a whole number while entering your number of cut bands. If you put something other than a number, the app won't allow it. It will display a message informing you that your entry is not what was expected, as it will accept only a whole number. Click 'Ok' and the app will ask you again for the number of cut bands.

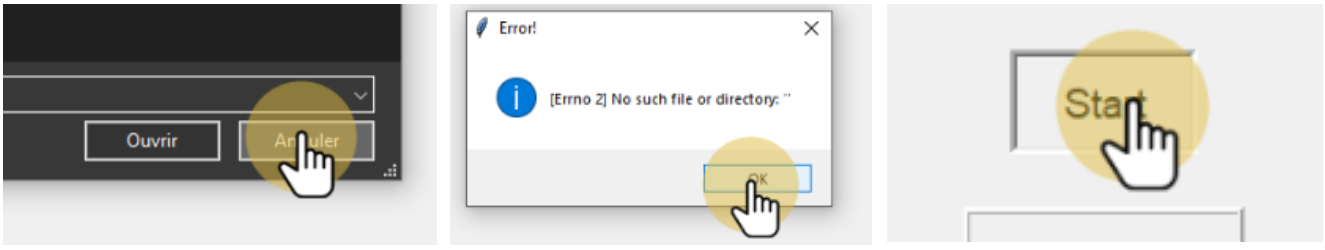


You entered the wrong number of cut bands. When you click 'Ok' after writing your number of cut bands, the file selection window will open. If you think your entry is not the right one, press 'Cancel' instead of 'Open' when selecting a file. An 'Error !' window will open, as the script can't find the non-selected file. Click on 'Ok' to close the pop-up, then click again on the 'Start' button to restart.

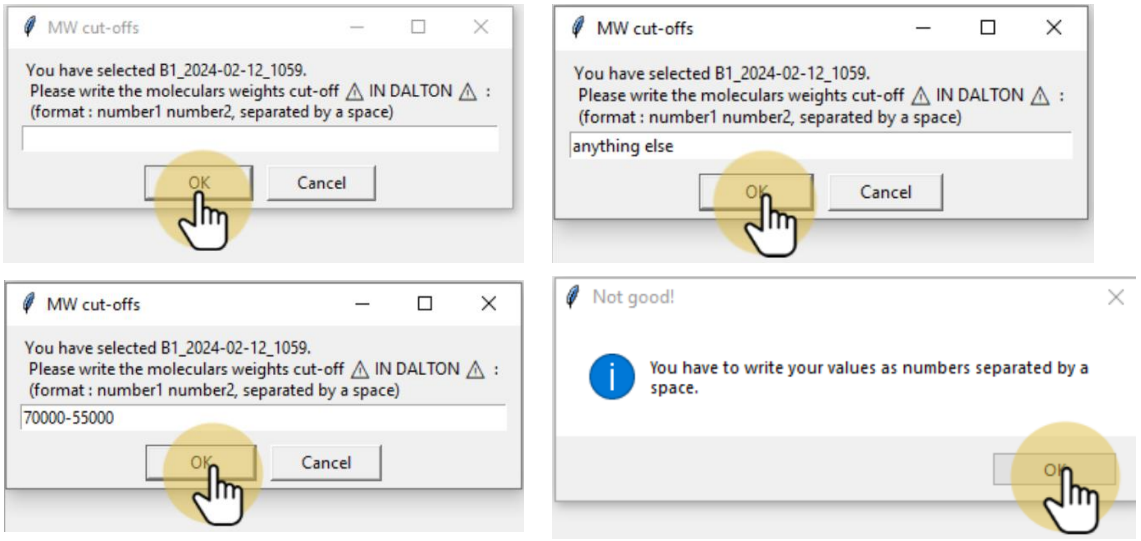


5.2 Entering the associated molecular weights

You wrote your molecular weights in kiloDalton (kDa). As this app needs to be helpful for all molecular weights and they are mainly wrote in Da in your data files, it was decided to run the script in Dalton. If you wrote your MW in kDa, you will have to cancel the next file selection, close the 'Error!' pop-up by clicking on 'Ok' and click on the 'Start' button to re-enter all of your data.



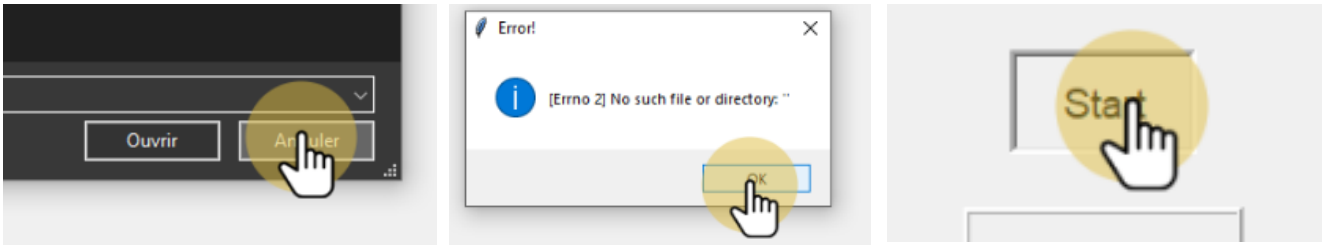
You : input letters, didn't separate the two values by a space, or put nothing at all. Any of these will display the same 'Error !' pop-up. Click on 'Ok' to close it, you will then be able to change your values.



5.3 Files importation and treatment

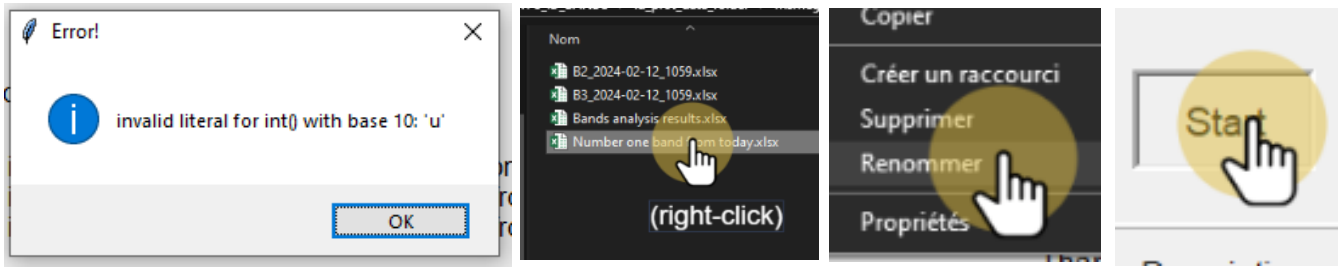
You associated the wrong molecular weights cut-offs to a band file.

- You still have files to enter : you will have to press 'Cancel' on the next file important window, click on 'Ok' to close the 'Error !' pop-up that will appear, then click on the 'Start' button again to re-enter all of your data.
- You don't have any more files to enter : click on the 'Start' button and re-enter all of your data.



Your band files are not named 'Bx…'. To store and keep the number associated with each one of your bands, the app retrieve it from the file names. This is why your Excel files should be name 'Bx...', x being any number. 'B' can be any other letter of your choice, or even a number. If the name of your files is different, it can raise this 'Error !' pop-up : "*invalid literal for int() with base 10 : '[something in your file name]'* ". This means that the app wasn't able to find the number associated with your band in the name of the file. In this example, I named a file "Number one band". The app tells me that the 'u' (from N**u**mber) is not a number allowing it to discriminate the bands data among the others.

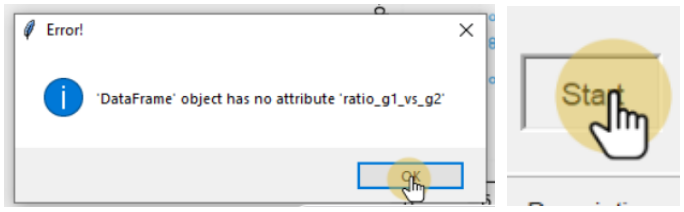
If you encounter this error, click 'Ok' to close the pop-up. You should rename your files and make sure they all start with 'B' (or any other letter) and a number. You can then run again the app by clicking on the 'Start' button and enter your newly-named files.



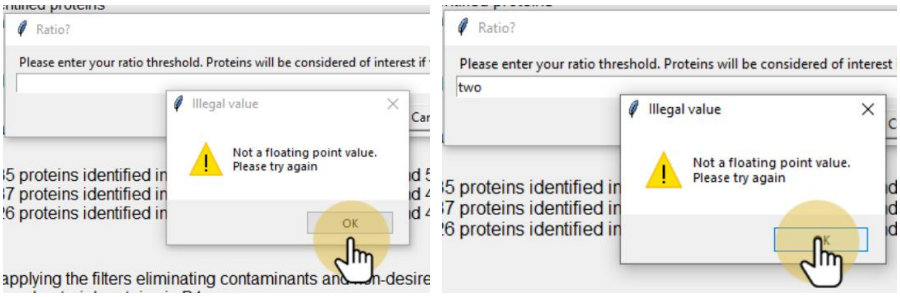
5.4 Entering the pull-down data

You selected the wrong file. If you notice too late that you selected the wrong file for your pull-down data and input something entirely different, it is highly probable that you will see an 'Error !' pop-up appear to inform you that your file doesn't contain a column named 'ratio_g1_vs_g2'. Click on 'Ok' to close it, then on the 'Start' button. You will have to re-enter all of your data.

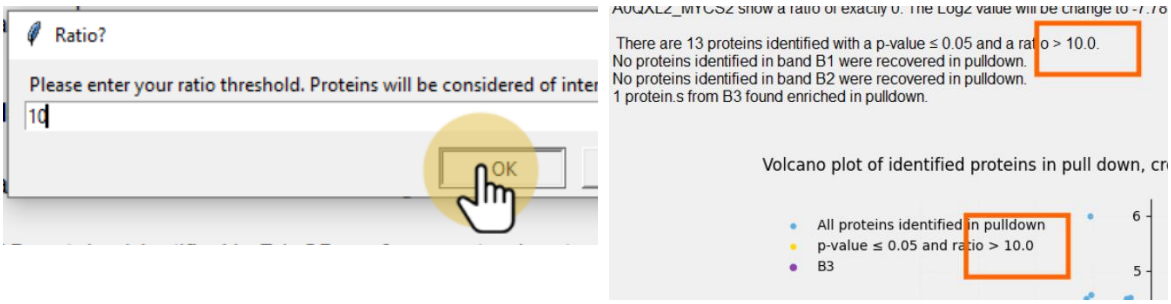
If you entered a pull-down file but not the one associated with your bands, no error will be raised but your results will be false. Click on the 'Start' button and re-enter all of you data to re-do the cross-result using the right data.



You didn't enter a ratio number. If you are too fast and click 'Ok' althought nothing was entered, of if you wrote letters instead of a number, an 'Error!' pop-up will be raised. Close it by clicking on 'Ok' and you will be able to change your ratio value.



You entered the wrong ratio number. No confirmation is needed when you enter your minimal expected ratio. If you're not sure you entered the right number, you can verify it on the data displayed on your scree : written as applied filter and in the volcano legend.



6. Script diagram

Next page is a diagram of how the application script works. It is made of three interconnected parts:

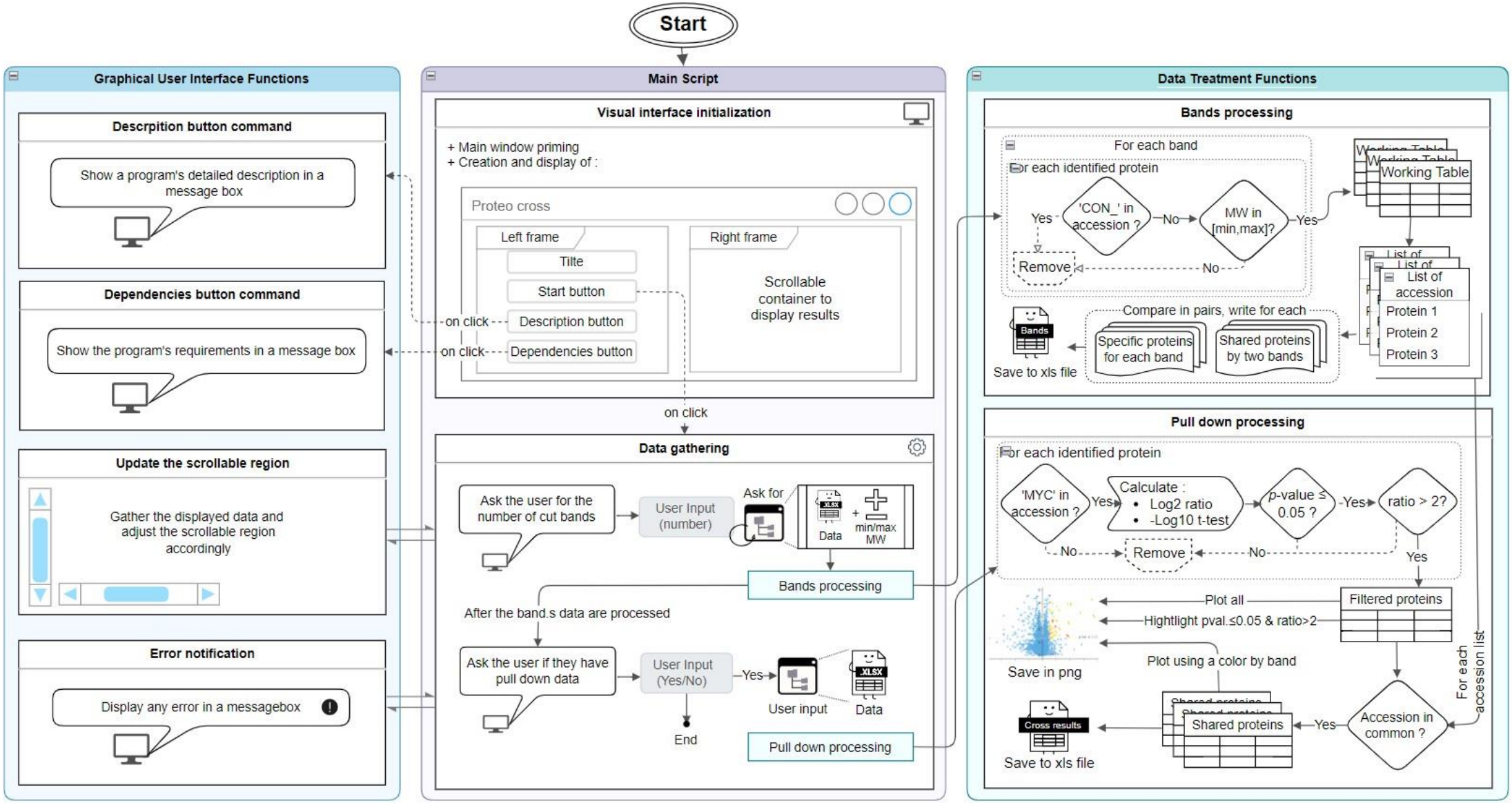
- the link with the visual interface (left)
- all the functions for processing spreadsheets (right)
- the main part in the centre, which manages the application window (linked to the left-hand part) and sends the data entered by the user to the processing functions (the right-hand part).

Textual informations, such as those displayed on button-click, and directly stored in the app. No further files are necessary.

The python scripts were compiled into an executable file using PyInstaller 6.4.0, so no install should be necessary.

List of dependencies :

- python 3.11.2
- tkinter Tcl/Tk 8.6
- PIL (Image, ImageTK)
- pandas 1.5.3
- itertools
- numpy 1.24.2
- matplotlib.pyplot 3.7.1
- traceback


Start
Graphical User Interface Functions
Descrpition button command
Show a program's detailed description in a message box
Dependencies button command
Show the program's requirements in a message box
Update the scrollable region
Gather the displayed data and adjust the scrollable region accordingly
Error notification
Display any error in a messagebox
Main Script
Visual interface initialization
+ Main window priming
+ Creation and display of :
Proteo cross
Left frame
Tilte
Start button
Description button
Dependencies button
Right frame
Scrollable container to display results
on click
on click
on click
Data gathering
Ask the user for the number of cut bands
User Input (number)
Ask for
Data
min/max MW
Bands processing
After the band.s data are processed
Ask the user if they have pull down data
User Input (Yes/No)
Yes
User input
Data
End
Pull down processing
Data Treatment Functions
Bands processing
For each band
For each identified protein
'CON_' in accession ?
Yes
No
MW in [min,max]?
Yes
No
Remove
Working Table
List of accession
Protein 1
Protein 2
Protein 3
Compare in pairs, write for each
Specific proteins for each band
Shared proteins by two bands
Bands
Save to xls file
Pull down processing
For each identified protein
'MYC' in accession ?
Yes
Calculate :
• Log2 ratio
• -Log10 t-test
p-value ≤ 0.05 ?
Yes
ratio > 2?
Yes
No
Remove
No
Filtered proteins
Plot all
Hightlight pval.≤0.05 & ratio>2
Save in png
Plot using a color by band
For each accession list
Accession in common ?
Yes
Shared proteins
Cross results
Save to xls file