# PHARMACOINFORMATICS

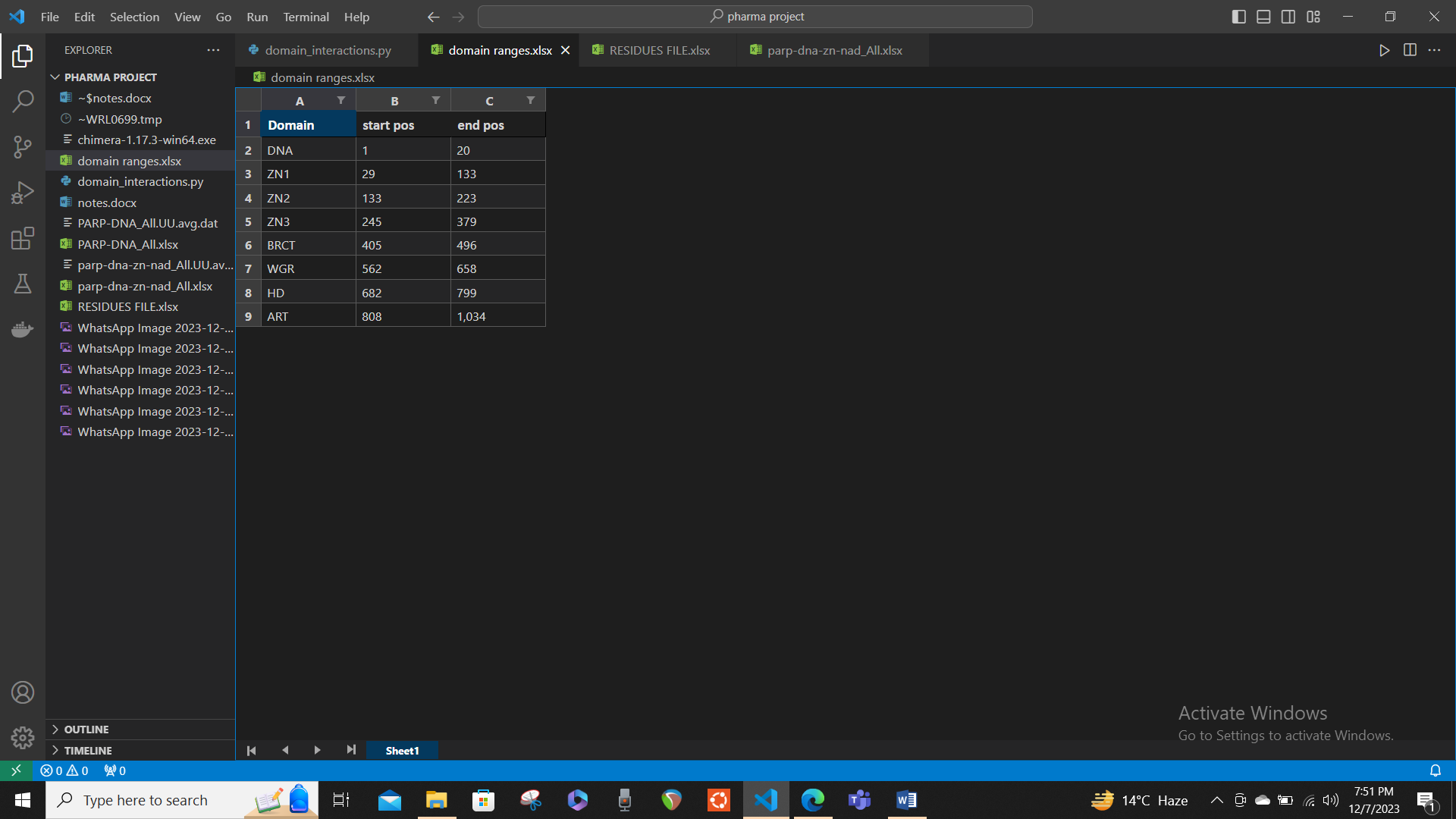
## Domain Interaction project (brain storming and outline)

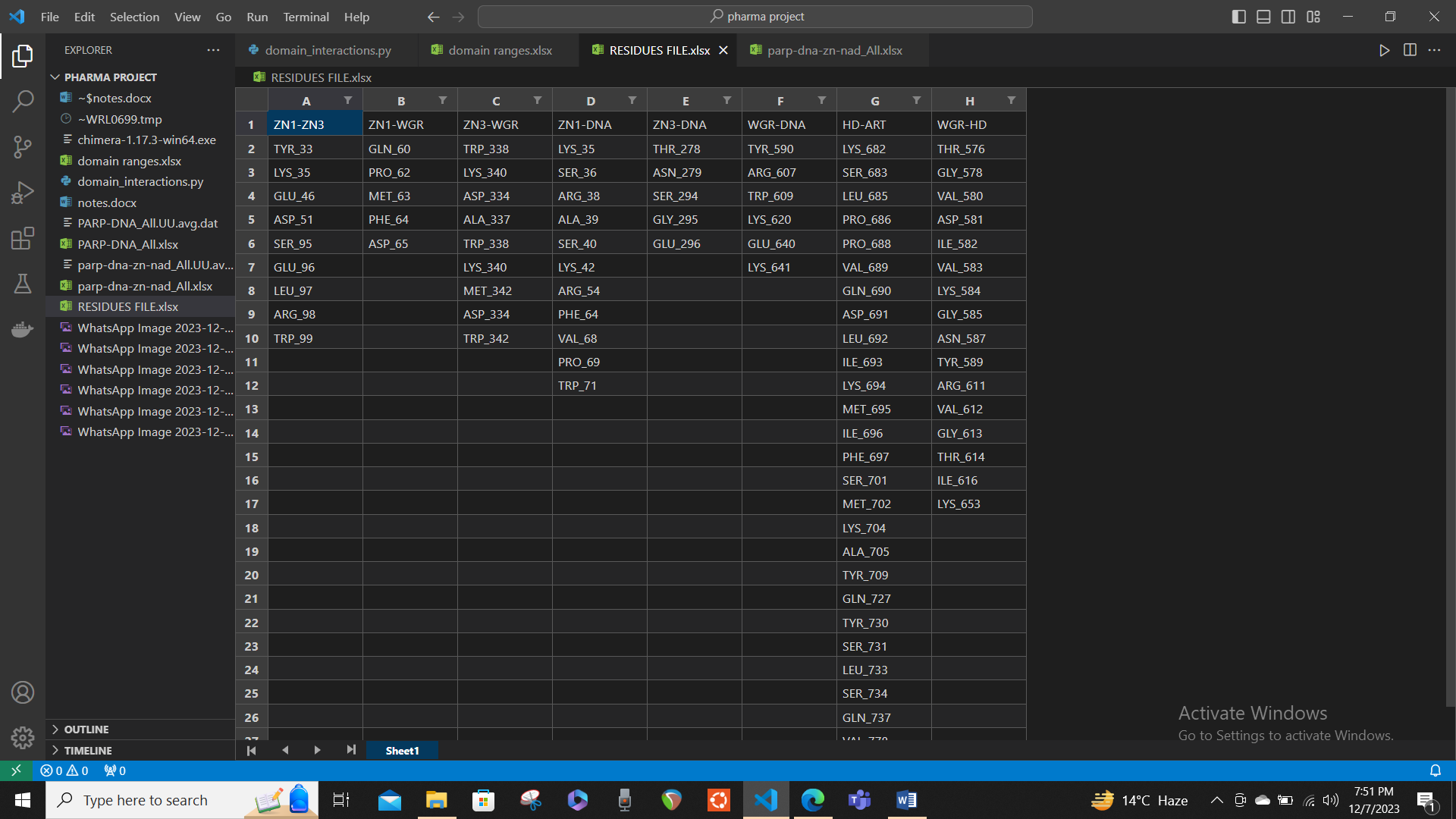
## BS Bioinformatics 7th

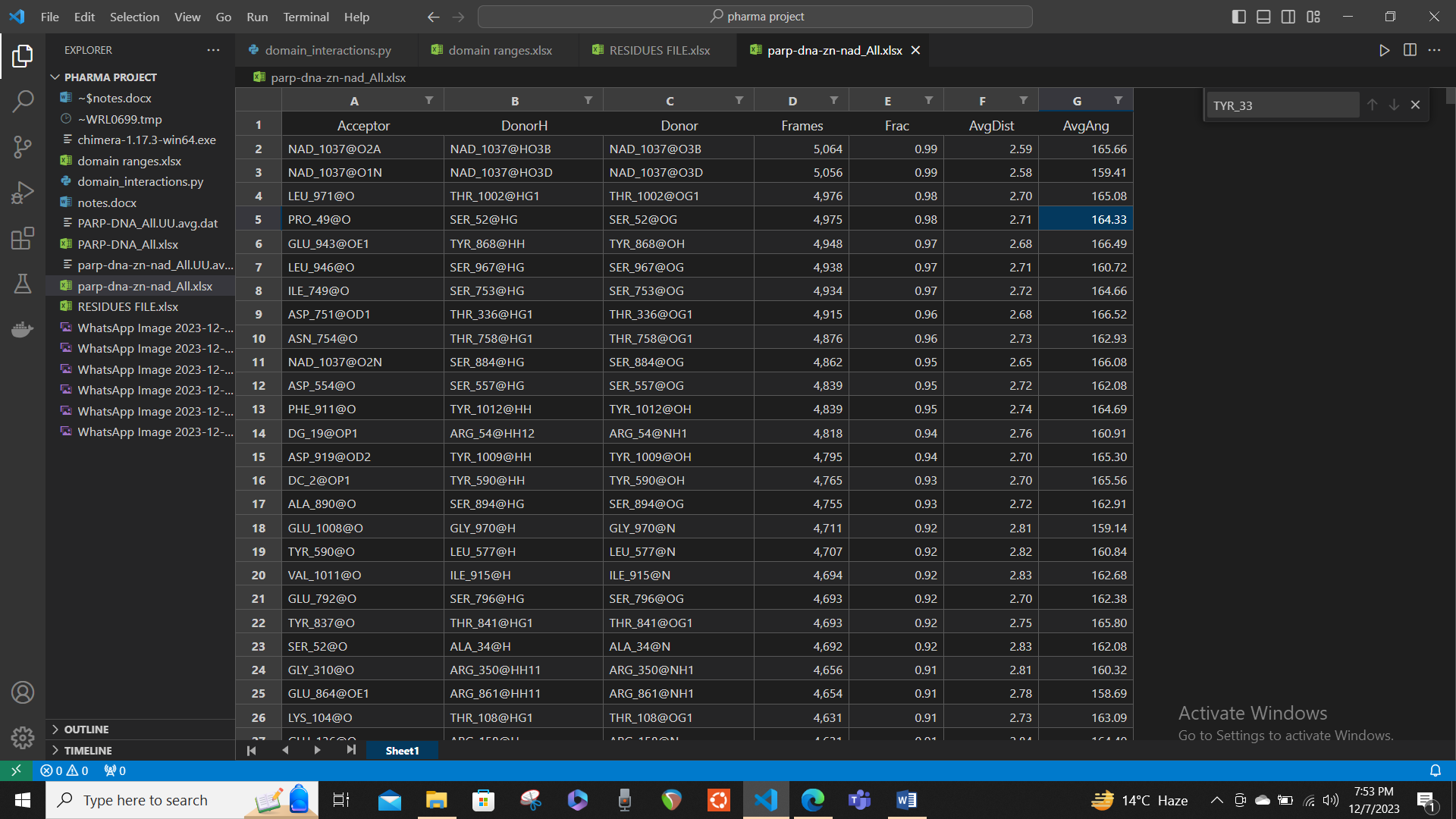
## NCB QAU

Given Input files:

1. Domain range xlsx file with data in the following way:



1. Interacting residue xlsx file with data in the following way: 
2. The cleaned MD simulation data in xlsx file with data in following way: One thing to note is that the program has to work on molecular level not on the atomic level so in the following data.



The values before @ are considered for each term like this. NAD\_1037 NAD\_1037 NAD\_1037 5,064 0.99 2.59 165.66

So, the program should first ignore the @ and the values after it and use this kind to filter out the residues. The content including @and afterwards will be written but not considered will comparison with the residues from residue file since the residue file only have NAD\_1037 format, not having NAD\_1037@O2A.

Filter out the residues from the cleaned MD simulation file for all the 8 interacting domains in the residue files, separately in one file.

To filter out the residues for all the 8 domain interactions you have to keep following things in mind:

1. While filtering, if you take the 1st domain interaction out of the 8 (let's say the first domain interaction is ZN1-ZN3) and you have the interacting residues as TYR\_33 LYS\_35 GLU\_46 ASP\_51 SER\_95 GLU\_96 LEU\_97 ARG\_98 TRP\_99 listed among the two domains then to filter the file for all of the residues from md file for ZN1-ZN3, you have to make sure that all of the residues fall in the range of these domains that are interacting ZN1-ZN3.

2. Let suppose the range of domain ZN1 is 29-133 and the range the domain ZN3 is 245-379. If you want to filter out the residues and let’s suppose the first residue under ZN1-ZN3 column is ‘TYR\_33’, and if TYR\_33 is from the acceptor column then from the donor column the interacting residues should fall under the range of the other domain that is interacting, which in this case in ZN3 and its domain range is 245\_379. hence to filter out TYR\_33 for ZN1-ZN3 domain interaction, a possible residue in the corresponding column of the same row can be VAL\_250; because 250 is in the domain range of ZN3.

3. Note that in the residue files the residues that are mentioned under the domain interaction are from the first domain in the given pair of interacting domains. For example; the domain interaction ZN1-ZN3 has the residues as (TYR\_33 LYS\_35 GLU\_46 ASP\_51 SER\_95 GLU\_96 LEU\_97 ARG\_98 TRP\_99) all these are the residues of ZN1 domain (as we can see by the number attached to the residues, because the domain range of AZ1 is 29-133 and the positions of residues are also in range of 29-133). So these residues are from the ZN1 domain and while filtering we need to make sure that the other interacting residues should fall in the range of the other domain’s range (in this case ZN3).

4. So, the filtering will be like; TYR\_33 and LEU\_300 will be allowed but TYR\_33. But for example, for GLU\_96 and HIS\_240 won’t be allowed since the number with HIS (which is the residue of supposedly ZN3 domain have the residue number outside of the range of ZN3 which is 245\_379; and that also clearly confirms that the residue is not the part of ZN3 domain because the residual position is outside the domain range).

5. Another thing to consider while filtering is that the interacting residues that are listed among the two domains to filter can be under any column. It is not necessary that it’s only under the acceptor (1st column). It can be under the donor (2nd column) as well. Then the other residue will be under the range of the respective other domain. For example: TYR\_33 and LEU\_300 will be allowed and LEU\_300 and TYR\_33 will also be allowed. This shows that a can be acceptor or donor relative to the other residue it is interacting with of the other domain (which in this case is ZN3).

6. Hence the filtering will only take place if the residue in the 1st column and the residue in the 2nd column both follows the domain range of the interacting domain.

7. The process in repeated until all the filtration is done for all 8 interacting residues that are present in the residue file. Desired output file: the output should be 1 xlsx file with data in the organized way for all the domain interactions like

ZN1-ZN3 (the name of the domain is written then under it the filtered rows are written for this interaction)

All the filtered residues for this domain interaction are in the format of md simulation file with all the headers as following.

Acceptor DonorH Donor Frames Frac AvgDist AvgAng

(also note that during filtration process, in the md file, the @ and values after it are ignored for the comparison of the residues, from the residue file, but while writing it to the output file the values are written as such. Important note: the filtration is only using Acceptor DonorH Donor so the other columns are ignored and added as such to the respective row.)

Then the next domain interaction is appended to the same output xlsx file

ZN1-WGR

Then down here, all the filtered residues for this domain interaction goes and so on and so forth for all the 8 domain interactions

------------------------------------------------ to be continued