CS123A Hands - On Exercise:

1. Extracting Bioactivity data for target protein from ChEMBL

GROUP MEMBER NAMES:

1. Jennifer Melendez , BS Molecular Biology

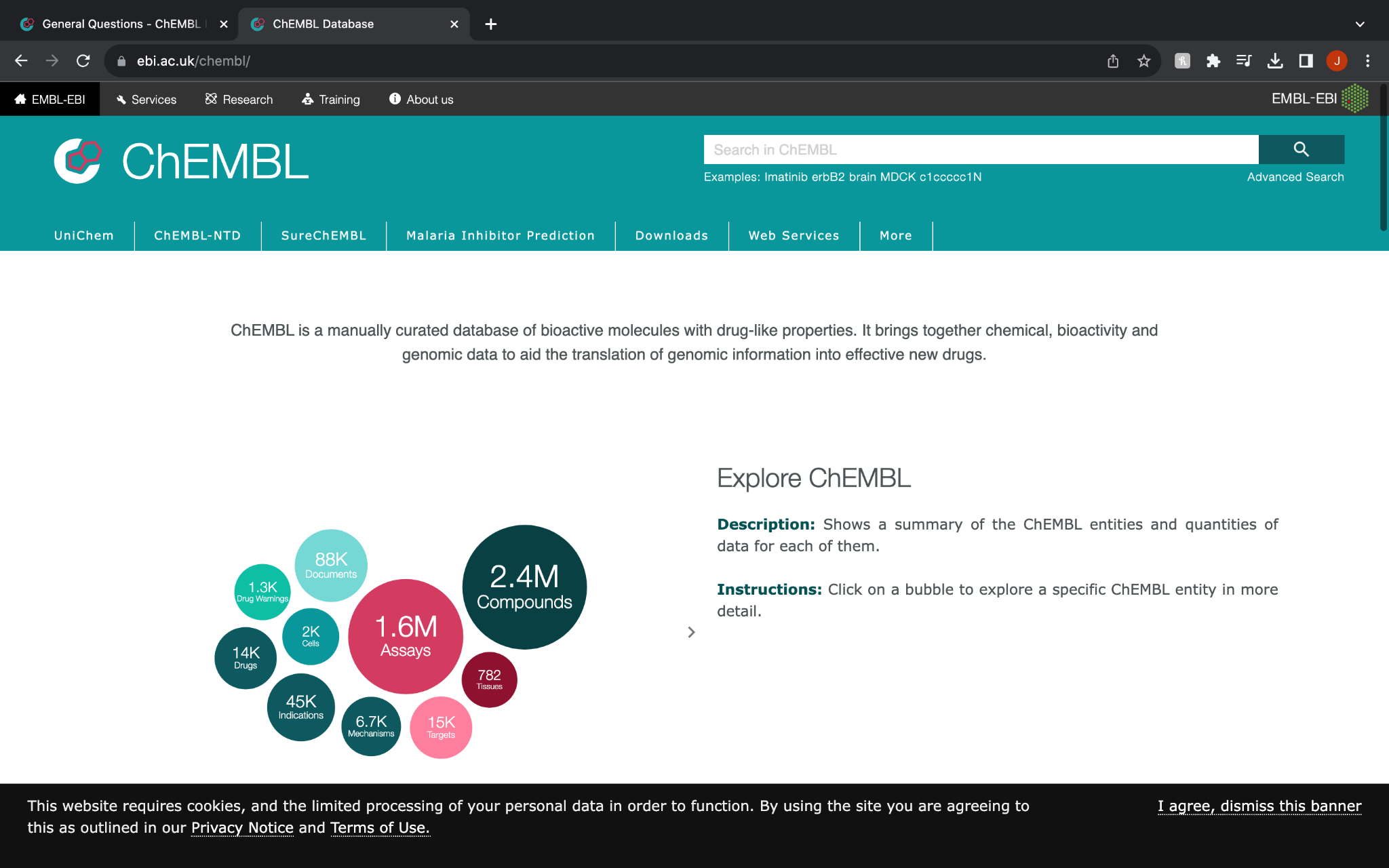
ENTER “COMPLETE” OR “INCOMPLETE” here \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

INTRODUCTION:

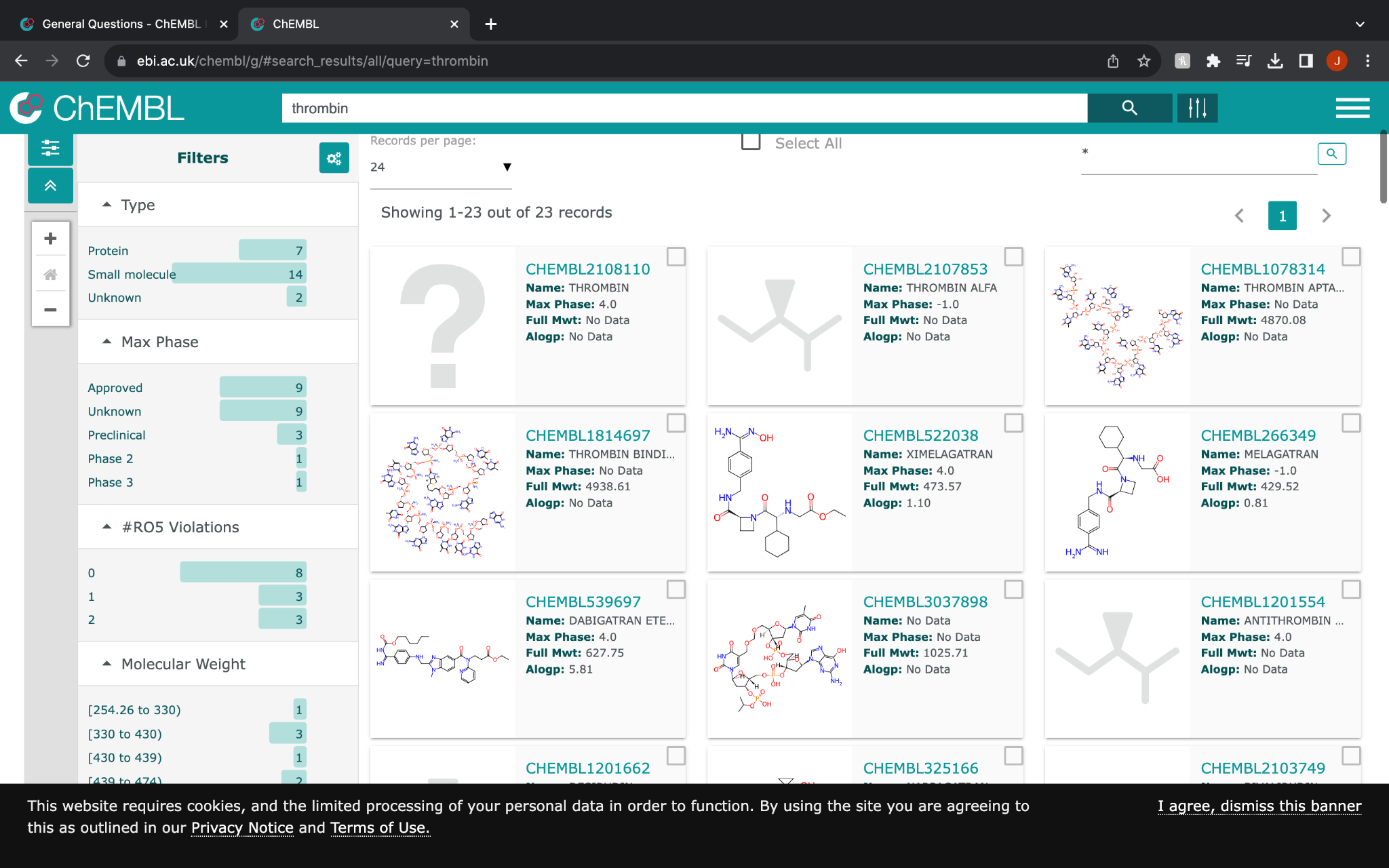
The ChEMBL Database is a database in which contains curated bioactivity data that has more than 2 million compounds. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs. It's a great tool to identify compounds for potential therapeutic targets. ChEMBL creates unique IDs that are assigned to compound, target, assays, document, tissue and cell types in ChEMBL.

Exercise 1: Extracting Bioactivity data for target protein from ChEMBL

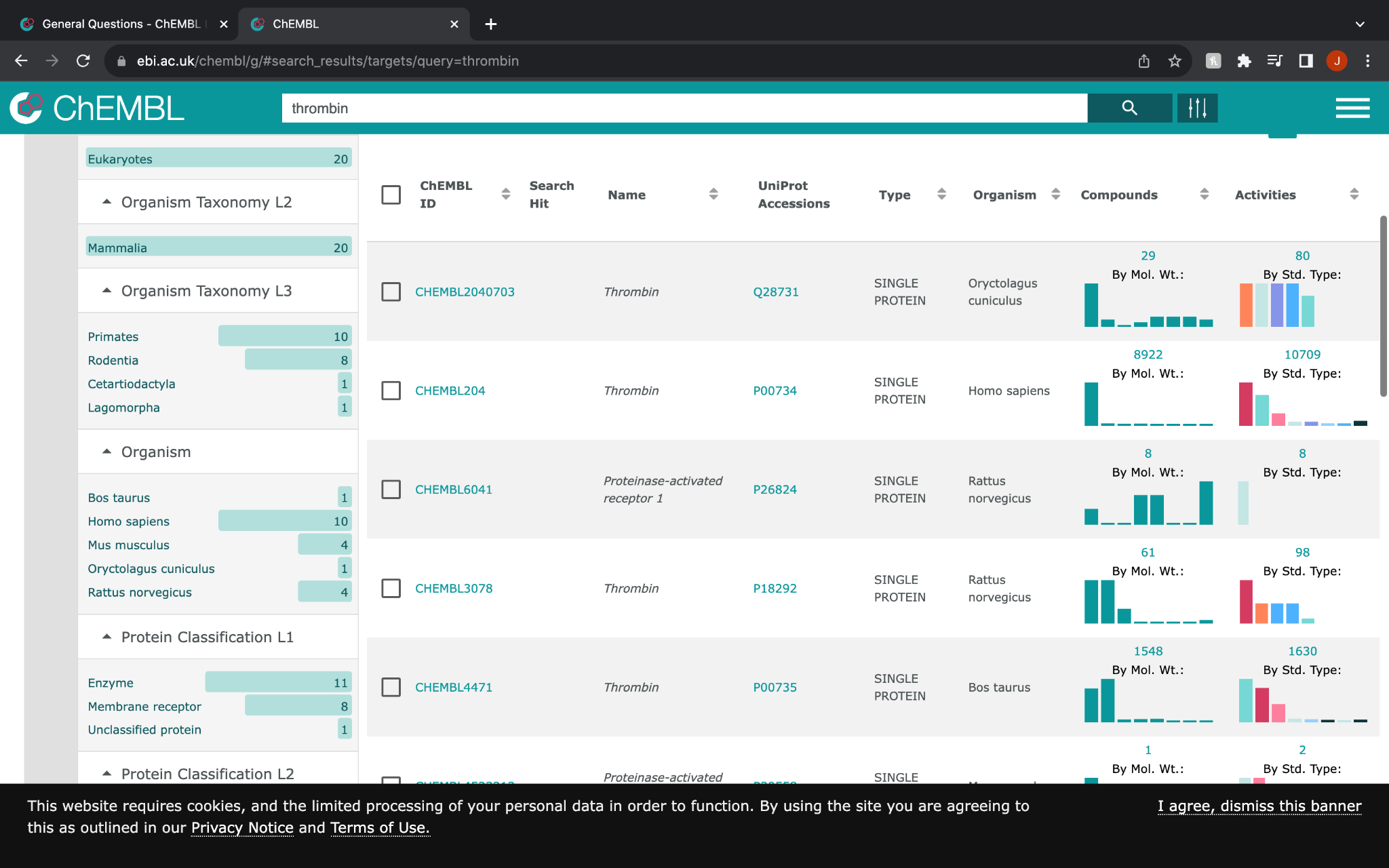
* First you're going to need to install the ChEMBL web service so that we can retrieve bioactivity data from ChEMBL. You can also hop onto this link: <https://www.ebi.ac.uk/chembl/> which can take you to the same page.



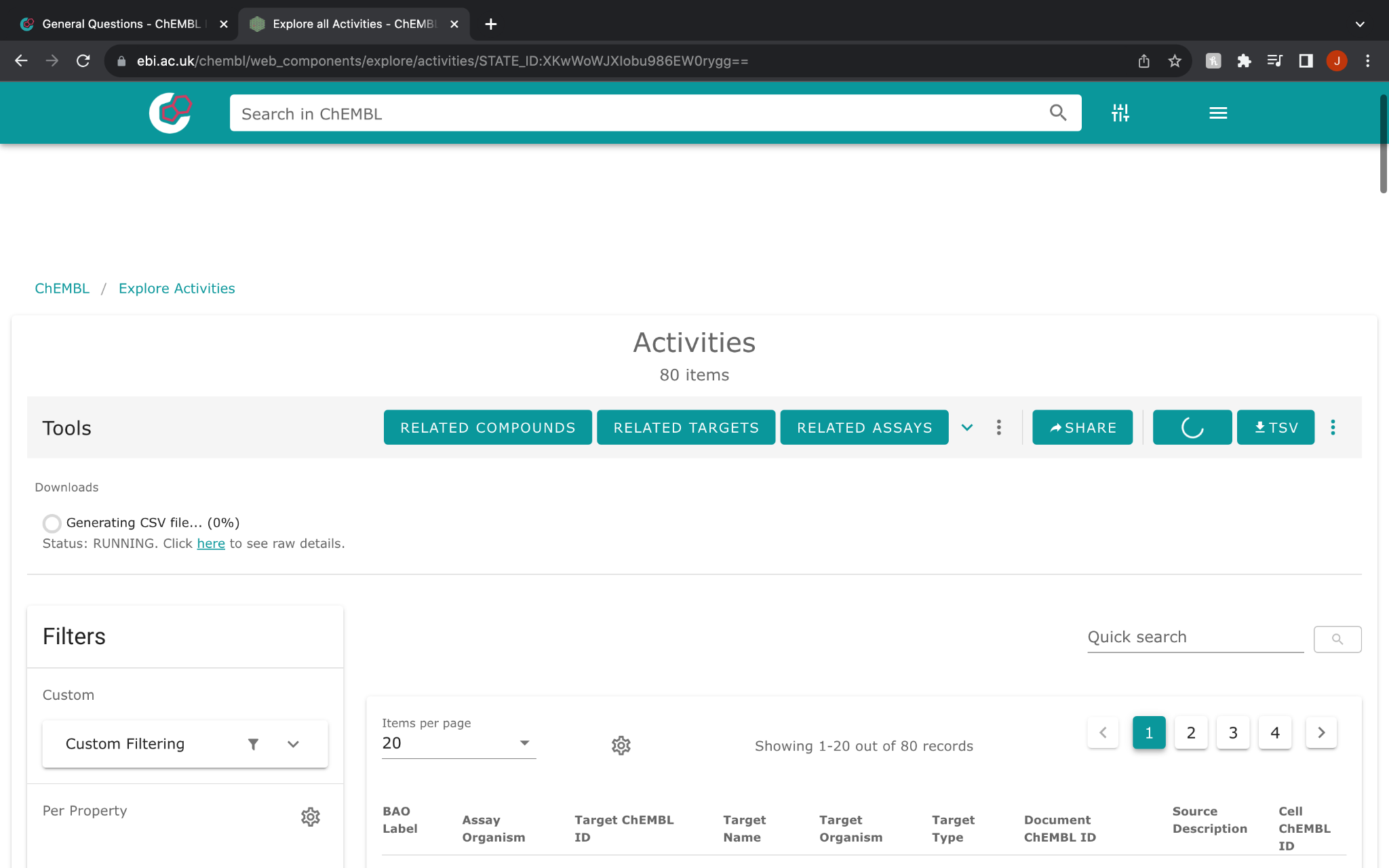
* Second step is searching up a protein or enzyme so we could get provided with bioactivity data. We are going to search for Thrombin because it's one of our main focuses in the project. A little background about thrombin, thrombin belongs to the trypsin family of serine proteases which is an enzyme in the human body that is encoded by the F2 genes. Thrombin is a unique enzyme which functions both as a procoagulant and anticoagulant. Thrombin is synthesized in the liver and it is secreted in the general circulation of an inactive form of zymogen which is a complex multi domain glycoprotein that is activated to yield thrombin at sites in which there is vascular injury.Thrombin is Na+ activated, allosteric serine protease thats plays opposing functions roles in blood coagulation.
* As you search for thrombin you are going to get various data that is being provided to you. In the compound you are going to be provided conical smiles. A good side note to remember.

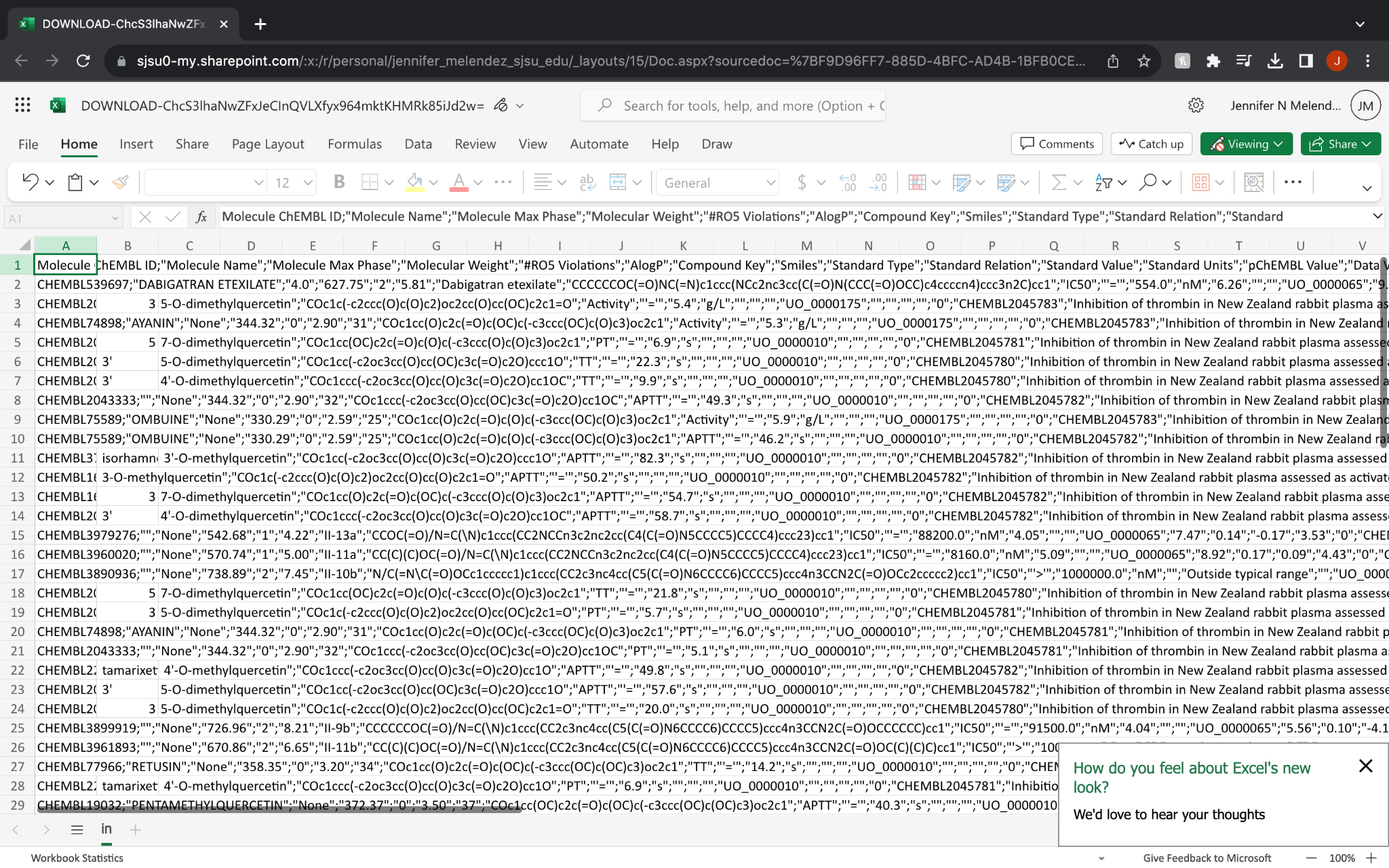


* In the tool box on top you have an option called “targets”. You are going to find yourself being displayed two panels one of the left and one on the right. The one on the left is the characteristic of the targets, protein classification organism and other subjects . While on the right it displays the target compound itself, the chEMBL ID and the proteins that are being provided.

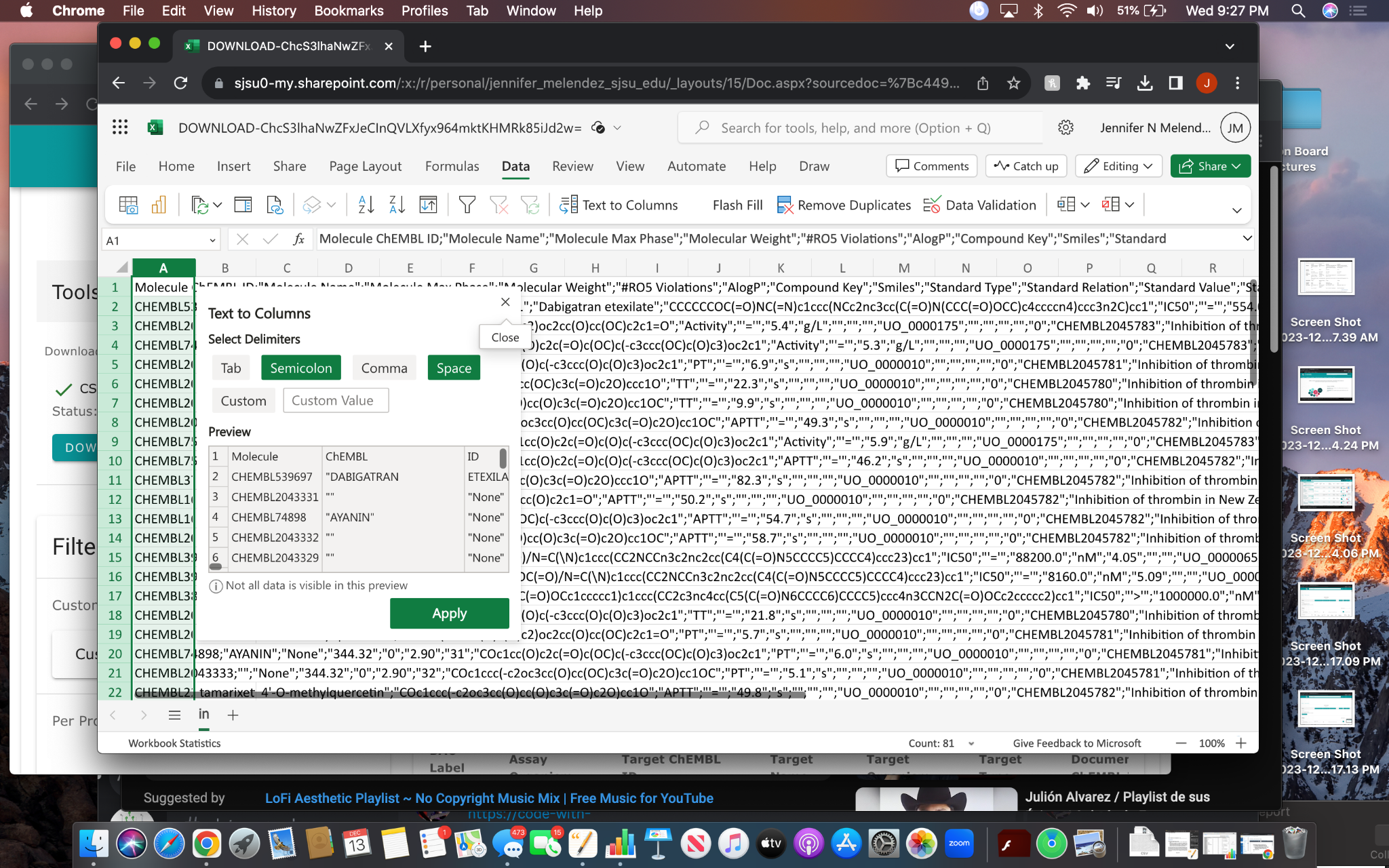


* Once you get to this placement you click on the activity number on the first option that is provided. Which will take you to a different page in which you can download all the bio data by pressing onto CSV so the file can be downloaded.

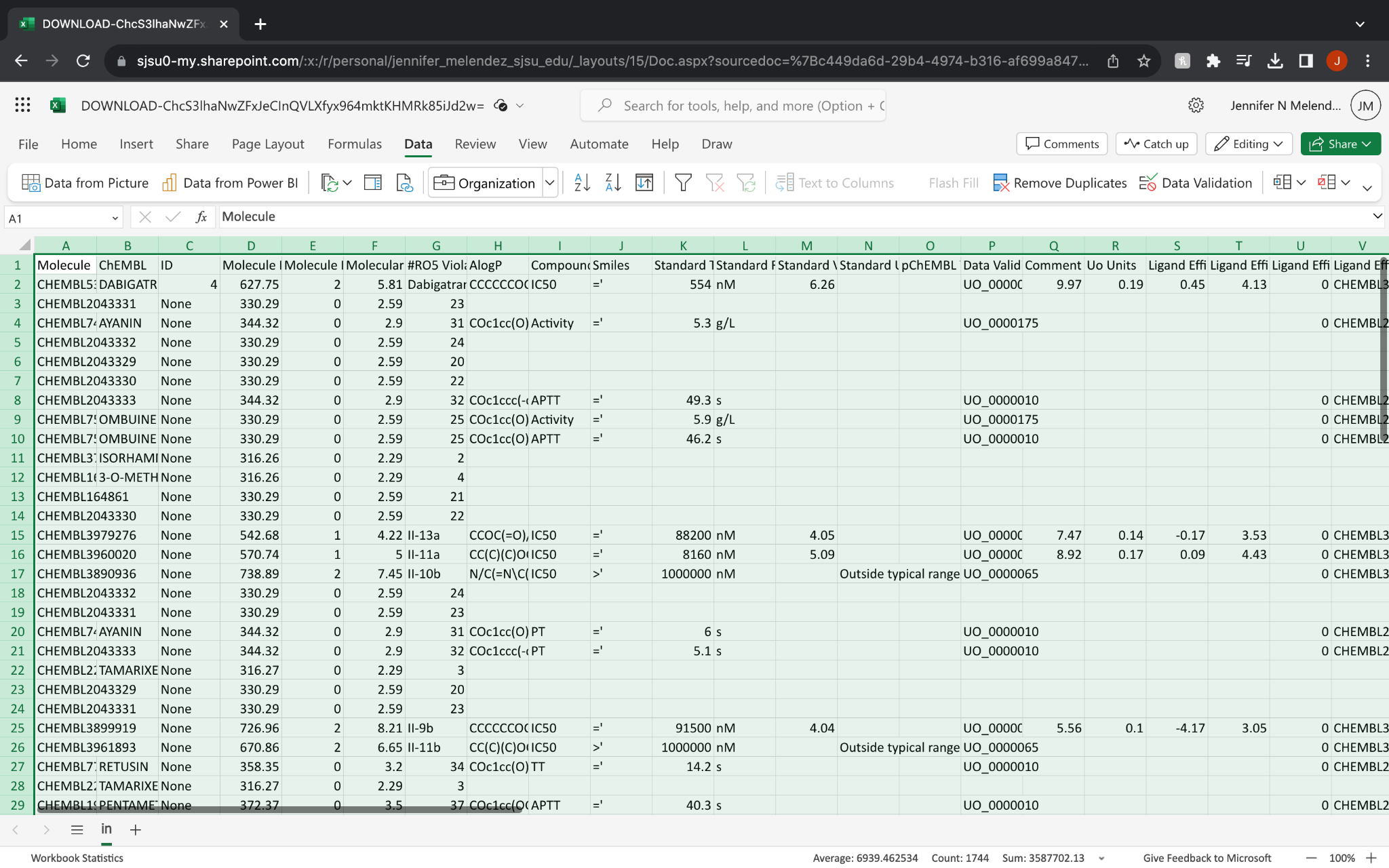


Once the file is downloaded, you can open the file by hopping onto Microsoft Excel.

The numbers can be intimidating but once the file is on microsoft excel you click on Data on the excel tabs and then you press column A. As you press column A you look for the text column in the “data section” on your tab which leads you to a pop up which asks you how you would like to layout your data.



You press semicolon as an option and then you proceed to agreeing to the changes which lead you to an outlined bioactivity data.



You're all set with the given information!