Enzyme Science and Technology

Prof. Vishal Trivedi
Department of Biosciences and Bioengineering
Indian Institute of Technology, Guwahati

Module - II
Enzyme Structure
Lecture - 11
Molecular Modelling of Enzyme Structure (Part-II)

Hello everyone, this is Doctor Vishal Trivedi from the department of biosciences and bioengineering IIT Guwahati and in the course enzyme science and technology we are discussing about the different properties of the enzymes and its contribution in the development of science as well as the technology.

So far what we have discussed we have discussed about the enzyme classifications and enzyme nomenclature and in the previous module we have also discussed the history of the development of the field of enzymology and in the current module, we are discussing about the structure of the enzyme and in when we were discussing about the structure of enzyme? We have discussed about the primary structures.

We have also discussed about how to determine the primary structures and then subsequent to that we have also discussed about the secondary structure, we discussed about the tertiary structure right and when we were discussing about the secondary structure we discussed about the how to determine the secondary structure with the help of the C D spectroscopy and as well as the IR spectroscopy.

And in the previous lecture we have also discussed about the different methods to determine the tertiary structures and in that context we have discussed about the X-ray crystallography and as well as the N M R spectroscopy. So, in the today's lecture we are going to discuss about the computational methods to determine the protein structures. So, let us start todays lecture.

(Refer Slide Time: 02:19)

PROTEIN STRUCTURE

The ordered folding of polypeptide chain give rise to the 3-D conformation known as secondary structure of the protein such as helices, sheet and loops.
Arrangement of the secondary structure gives rise to the tertiary structure. α-helix and β-sheet are connected via unstructured loops to arrange themselves in the protein structure and it allows the secondary structure to change their direction. Tertiary structure defines the function of a protein, enzymatic activity or a nature of structural protein.

Different polypeptide chains are arranged to give quaternary structure.

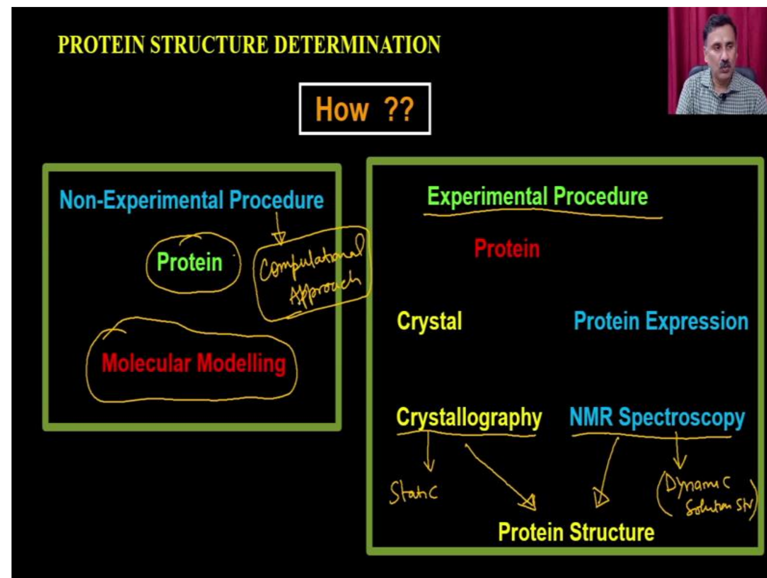
Primary Structure
Secondary Structure
Tertiary Structure
Quaternary Structure

Primary structure
Secondary structures
α-helix
β-sheet
Turn
Tertiary structure
Quaternary structure

So, what we are discussing in this particular module is we are discussing about the protein structures. So, as you can see that protein structure is made up of the or protein structure is been presented by the 4 stages or 4 levels right. So, protein structure is being represented by the 4 levels, primary structure, secondary structure, tertiary structure and some cases you may also have the quaternary structures.

Primary structure is the amino acid sequence of the proteins, secondary structure is the alpha helix beta sheets and turns and all these secondary structure when they come together and they will hold further they are going to give rise to the tertiary structures. So, while we were discussing about the structure determination of the tertiary structures in the previous module, we have discussed about the different methods different experimental methods. So, what are different experimental method we have discussed?

(Refer Slide Time: 03:14)

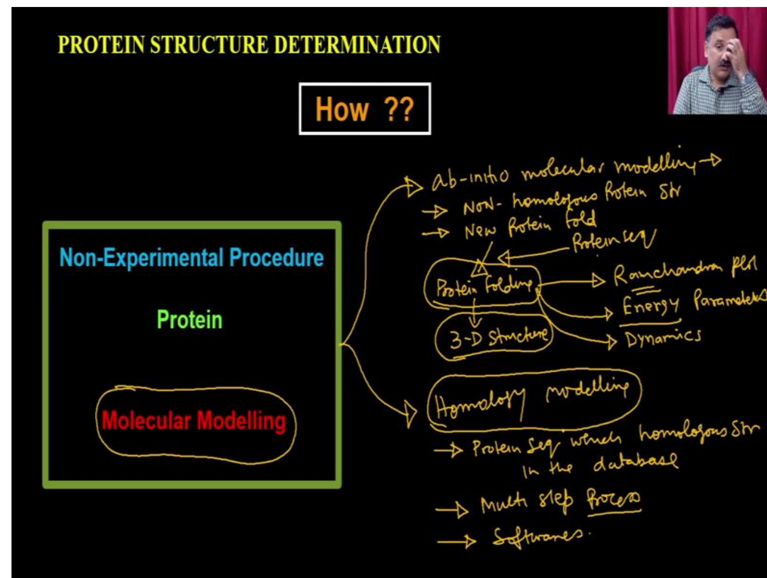


So, in the experimental method we have discussed about the X-ray crystallography and as well as the NMR spectroscopy and both of these methods are actually going to give us the protein structures, the structure what you are going to get from the X-ray crystallography is going to be a static structure or and whereas, in the case of NMR spectroscopy it is going to give you dynamic structure or solution structure. So, this is also called as the solution structures.

Now, in today's lecture we are going to discuss about the non experimental method. So, when you have a protein for which you would like to determine the protein structure you have the 2 options experimental procedures or the non experimental procedure. So, in when we say non experimental procedure; that means, we are actually going to talk about the computational approach what you can actually be able to use to determine a protein structure.

So, all these are actually being based on the determining or based on the data sets right which actually is going to be used to train the particular software and that is how you are actually going to get the protein structures. So, what we are discussing about is the molecular modelling of the protein structures using the computational tools.

(Refer Slide Time: 04:46)



So, when we talk about the molecular modelling, the molecular modelling can be of two types ok depending on the type of protein sequence what you are using it could be ab-initio or molecular modelling or it could be the template based homology modelling.

Under what condition you will do the ab-initio or the initio molecular modelling. So, when the protein is having no non homologous sequences or protein is going to have the or not available for the non homologous protein structure. So, it does not have any homologous structure then you are no option, but you are going to do the ab-initio or molecular modelling.

These are the places where you are actually working with the new proteins or new protein fold and in that case it does not have the any kind of homology and that actually is going to be very difficult to solve the get the protein structures. What are the things basically people do when they do the ab initio molecular modelling? Is they are actually going to use the help of the protein folding.

So, they are actually going to predict how the protein folding is going to occur, if this is the protein sequence. So, with they will looking at the protein sequence and then looking at the protein sequence they will actually going to predict what will be the protein folding which means what region of the protein is going to adopt what conformation and then they will actually going to cross verify by that with the help of the Ramchandran plot and all the other kinds of energy parameters.

Because, if you are actually going to get a good in a model, it is actually going to show you the low free energy right. So, that in the molecular going to be more and more stable. But if there are steric hindrance, if there are actually clashes between the side chains that is actually going to be get you will get that information from the Ramchandran plot and that also is going to be reflected in terms of the energy parameters.

So, both of these parameters can be used to determine whether the predicted protein fold can be correct or wrong and using this you can be able to go with the ab initio molecular modelling. Apart from this they are also going to use the dynamics. So, they are also going to you know unfold the protein and then again put it under the molecular dynamics stimulations and that is how they are actually going to use.

These are the three robust tools to predict the protein folds and once they predict the protein fold eventually they are actually going to get the 3D structures. So, in today's lecture we are not going to discuss about the ab-initio molecular modelling because that is not very popular, because it requires the extensive computational tools and you can imagine that if you have a protein of 100 amino acids, it is very difficult to you know predict the protein folds especially when it is not having the homologous sequences and it is actually a new protein fold.

So, in that case you might have to generate the protein folds using the Ramchandran plot and then you might have to determine what will be the free energy because ultimately the free energy of the system has to be lower down while you are it is going through the protein folding process.

So, then the second approach is the molecular modelling. So, molecular modelling under what condition you will do the molecular modelling? So, when you are working with a protein sequence which has homologous structures in the database that time you will use the homology modelling or so homology modelling is a multi-step process and it is actually going to use you are going to use the different process and you can be able to use the different types of software's to do the molecular modelling.

Then what we are going to do is we are going to first discuss about the how the molecular modelling is homology modelling is you know theoretically working and then we are actually going to show you a small demo.

(Refer Slide Time: 09:56)

So, what is the molecular modelling or homology modelling? So, this is a useful and fast structural solution method where the sequence similarity between the template and the target enzyme is used to model the 3 dimensional structure of the target enzyme.

The homology modelling exploits the idea that the amino acid sequence of a protein directs the folding of the molecule to adopt a suitable 3 dimensional conformation within the minimum free energy. So, what is the basic idea of doing the homology modelling? Homology modelling means that you are actually going to have the template structure right. So, template structure is actually going to give you two information.

One it is actually going to give you the information about the protein sequence which actually it is going to be right which is from the protein sequence from the template protein right and it is also going to give you the information about the protein folding right. How a particular sequence is actually getting folds into that particular thing? This means, once you know the protein folding you are actually going to get the 3D coordinates of the of the template residues right. So, right.

So, basically it is actually going to give you the x y z coordinates of the template residues. How the alanine glycine arginine everything is present in this particular template structure right. Now, if we have the target protein or if I have a target enzyme for which I am going to use I am going to determine the structure what I require is or what information I have is actually the protein sequence from the target enzyme.

Now, what I will do is I am going to take the 3D coordinates of the template residues ok or I will say the backbone because the residues are actually going to be same or different in some cases. So, since both of these the template structure and as well as target structures are homologous the many places the 3D coordinates of the template residues are actually going to be the same as it was present in the target enzyme.

But many places or few places the template residues are going to be different and because of that you are actually going to only take the 3D coordinates information of the template backbone which means you are also going you are only going to take the peptide bond information. You are going to only the peptide bond or the main chain. On side chain information you will not going to take because the side chain information may or may be same or may be different.

So, this is what is exactly what you have to do. But if you want to do this it has to you have to follow the multiple steps. So, that you can be able to achieve this target. So, you cannot just simply take the 3D coordinates you have to first determine and identify the templates and that is why this is the homology modelling is a multi-step process.

(Refer Slide Time: 13:46)

Homology Modelling

Steps in Homology Modelling?

- Step 1: Finding the Query Sequence
- Step 2: Selecting the Template
- Step 3: Alignment of Query with Template
- Step 4: Building the model
- Step 5: Structure Validation of the Models

Molecular Modelling: Principles and Applications
https://www.wiley.com/en-us/Molecular+Modeling%3A+Basic+Principles+and+Applications-p-9783527614769

DEMO

So, what are the steps? In the step one you are going to find the query sequence right. So, you are actually going to have you have to identify or isolate the query sequence. So, query sequence is nothing, but the amino acid sequence of the test enzyme. This you have to do from the NCBI server right. So, you can actually if you know the proteins

name or if you know the accession number you can be able to get the query sequence from the NCBI server right.

Then once you got the sequence then you are actually going to select the templates right because and that you are going to do with the help of T blast. So, what you are going to do is you are going to take this amino acid sequence and you are going to put it into the NCBI blast which is a program actually and NCBI blast once you put it and you will select the database as the protein structure database right or the PDB. So, when you do the PDB's database it is actually going to give you the templates.

So, it is only going to tell you that ok these are the templates through which your amino acid sequence is matching. So, these are the potential templates and then after that you are actually going to test these templates in the step 3. So, in the step 3 you are actually going to see the utility of the each template. Utility of each template right. So, that you are going to do by a pair wise alignment of the query with the template.

So, you are going to do like template 1 you are going to do like template 1 versus your test sequence right. So, you are going to do a pair wise clustal W and you are going to use the program which is called as clustal W. So, when you do the that it will actually going to tell you wherever you have the gaps and how much is actually it is having the identity.

If you have a very high number of gaps or if you have very small identity then it is actually not be suitable. So, in that case you will reject the template one. These are the things which can also be determined even by the some other kinds of scores. So, after once you select the template then you are going to use that template into the software and it is actually going to allow you to build the model. So, it is actually going to give you the 3D model.

And it is actually going to do the exactly the same phenomena right it is going to take the amino acid sequence from the target enzyme and it is going to take the 3D coordinates from the your template structures and then it is actually going to put them together and that is how it is going to do the final refinements and it will give you the 3D homology model. So, it is going to be called as homology model right.

Once you have prepared the model then it has to be validated with the help of the validation programs like Ramchandran plot or verify 3D or the errata plot right. So, all these are you are going to use and there are servers which are available to do this job ok.

So, I have given you a reference for this particular steps. So, you can actually be able to go through with this reference and the title this article and it is very good for determining or understanding the different steps. So, far what we have discussed? We have discussed the theoretical aspects of how to determine or how to perform the homology modelling. Now what I will do is I will take you to my lab where a students have prepared a small demo clip and they will be going to show you the different steps.

(Refer Slide Time: 17:54)

A demo on

Homology modelling

Using

Modeller

Demo by Alok Kumar Pandey
Graduate Student, IIT Guwahati, Assam

clideo.com

(Refer Slide Time: 17:58)

STEP 1: Finding the Query sequence

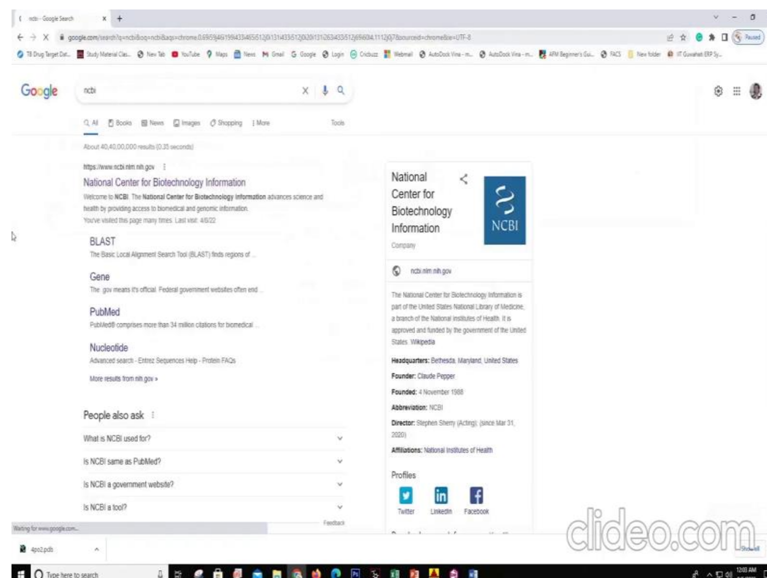
- The query sequence can be downloaded from NCBI(https://www.ncbi.nlm.nih.gov/)
- Make changes in the downloaded sequence in accordance with modeller requirements
  - Change the header code according to modeler script
  - Put a star sign(*) at the end of sequence
  - Save as name.ali file extension in "Modeller" folder in C drive
  - Copy modeller script for input file preparation in same folder in .py format
  - Run modeller script in modeller command line

```
>AAH08907.2 HSPA8 protein, partial [Homo sapiens]
FTTYSDNQPGVLIQVYEGERAMTKDNNLLGKFELTGIPPAPRGVPQIEVTFDID
ANGILNVSAVDKSTGKENKITITNDKGRLSKEDIERMVQEAEKYKAEDEKQRDK
VSSKNSLESYAFNMKATVEDEKLQGKINDEDKQKILDKCNEIINWLDKNQTAEK
EEFEHQQKELEKVCNPIITKLYQSAGGMPGGMPGGFPGGGAPPSGGASSGPTIE
EVD
```

clideo.com

In this video, I will be talking about the steps involved in the homology model modelling using modeller and parallelly I will be performing it on screen and explaining each step in detail. So, in step 1 we need to find out our query sequence which we need to model. So, that query sequence either we have the we can we have the sequence from the literature or we can download a our query sequence from NCBI website given here.

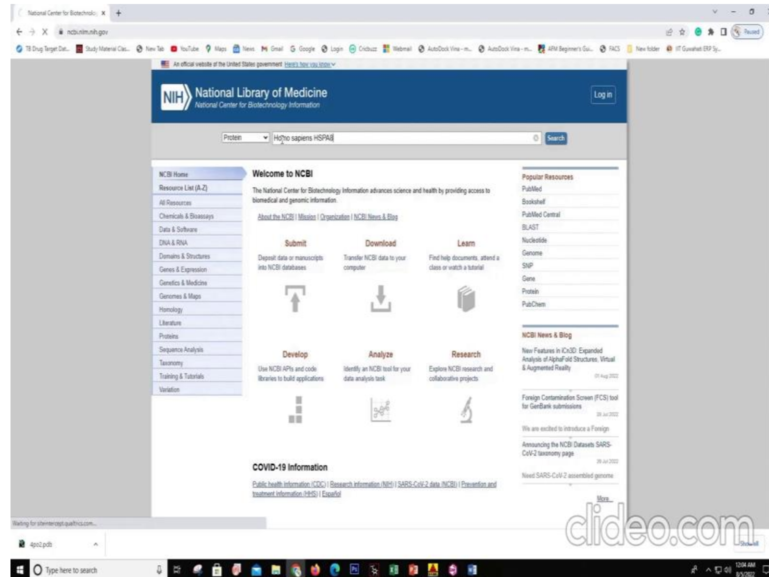
(Refer Slide Time: 18:44)



And then in the downloaded file we need to make some changes to make it compatible for working with modeller.

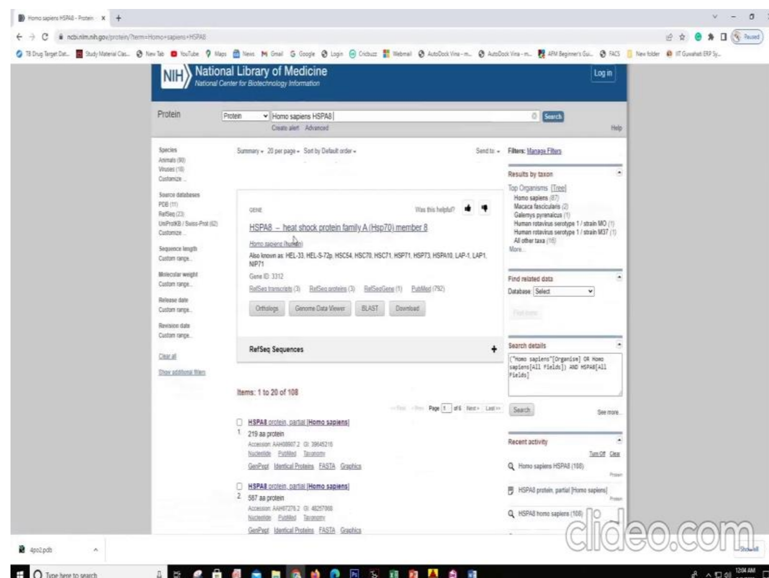
So, I will show you how to download and what changes we need to make. So, we will open up the NCBI website on the NCBI website on the NCBI website.

(Refer Slide Time: 18:51)



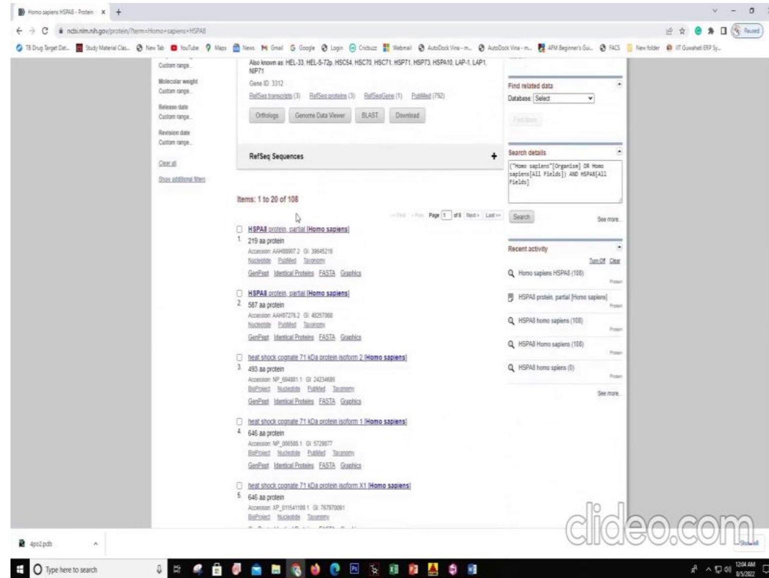
We will go to protein database we will select protein here and after that I will type the name of the sequence which we want to model.

(Refer Slide Time: 19:11)



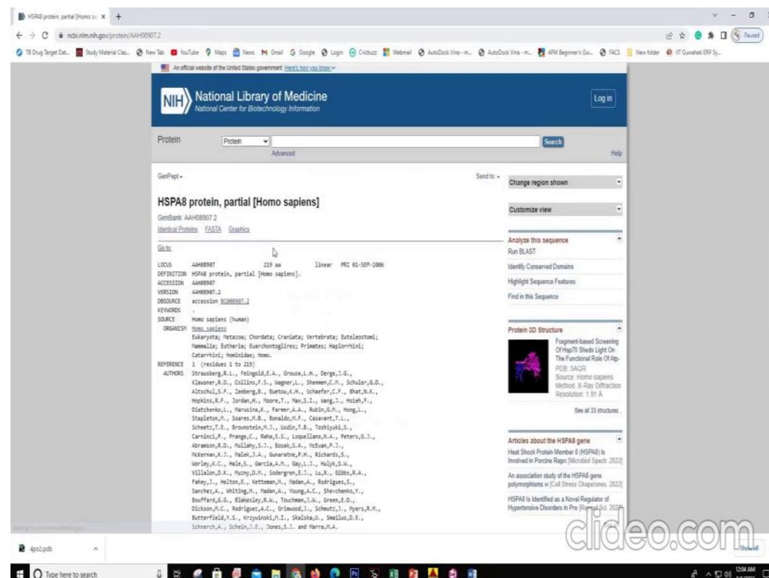
So, here it this is a list of related sequences.

(Refer Slide Time: 19:13)



So, for example, we need to model this first sequence HSPA 8 protein partial having 219 amino acids. So, we will click on this and after clicking on this it will open up the details of this protein.

(Refer Slide Time: 19:29)



Then we will click on FASTA and this will open up the amino acid sequence of the protein.

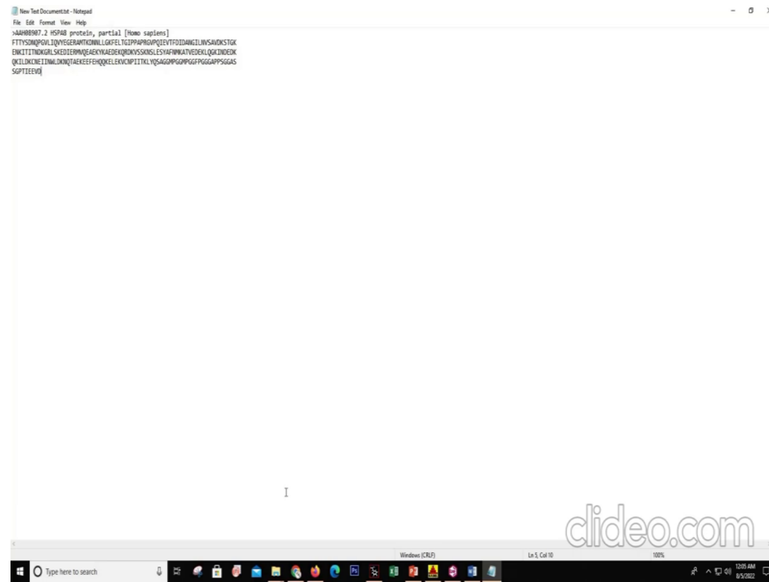
(Refer Slide Time: 19:37)

Now, we can just copy this sequence here and now we will go to our desktop.

(Refer Slide Time: 19:47)

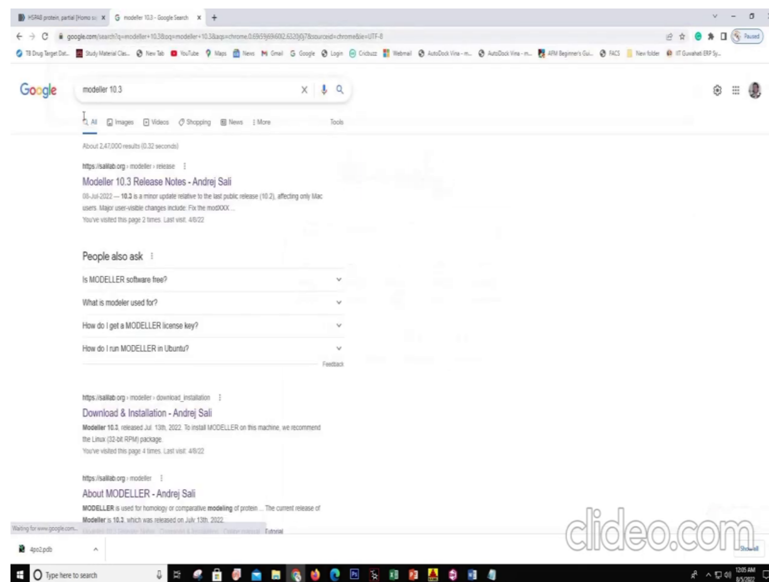
Where we can create a new text document file and this in this text document file we will paste our FASTA sequence.

(Refer Slide Time: 19:55)



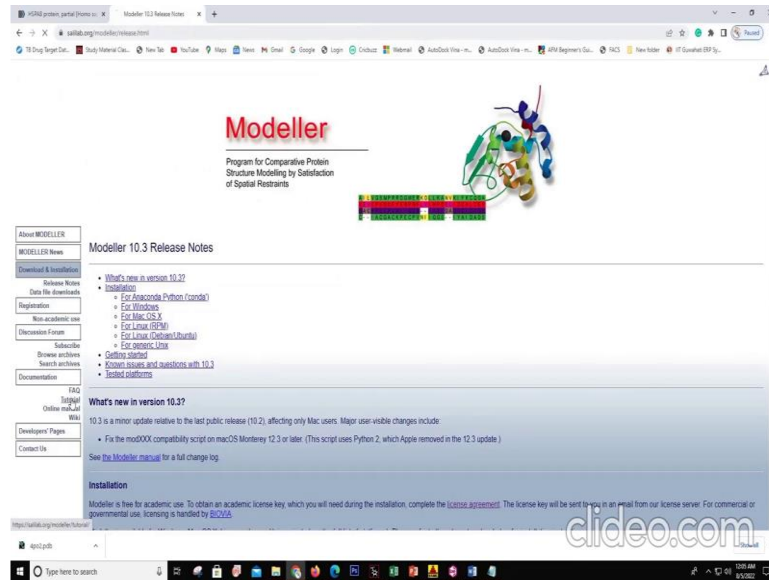
So, FASTA sequence has certain format. So, but this format is not compatible with modeller. So, we need to make some changes.

(Refer Slide Time: 20:17)



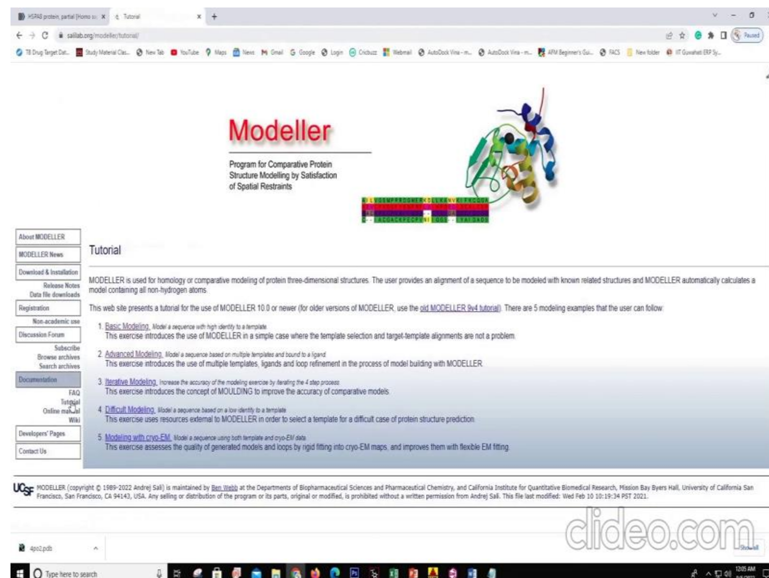
So, what changes we need to make? We can find it in the tutorial for modeller. So, for this demo we are using modeller 10.3. So, I will open tutorial for modeller 10.3. So, I will go to this website.

(Refer Slide Time: 20:27)



And here we can see here it is tutorial I will click on the tutorial.

(Refer Slide Time: 20:31)



And then this basic modelling.

(Refer Slide Time: 20:34)

So, before making the changes I would we need to download this zip format files. So, that because it contains some in files which are required to create other input files using modeller. So, I have already downloaded this. So, I will just go to downloads.

(Refer Slide Time: 20:59)

(Refer Slide Time: 21:02)

And in downloads this is the; this is the folder which I have downloaded and extracted here. So, from this folder.

(Refer Slide Time: 21:08)

We need to copy this PDB underscore 95 dot pr file and we need to paste this file in C drive where our modeller is installed you open C drive.

(Refer Slide Time: 21:19)

And then go to program files.

(Refer Slide Time: 21:23)

And then modeller 10.3.

(Refer Slide Time: 21:26)

(Refer Slide Time: 21:27)

Then bin and here I will create new folder and I will name it modelling and then I will paste that file here.

(Refer Slide Time: 21:40)

Now, we will use we will be using this folder for our further files and for further operations. So, I will minimize this and now here we can see this is the format for the query sequence in for modeller.

(Refer Slide Time: 21:51)

In this the starting the header code this one is different from what we seen in our FASTA sequence. So, we need to copy this code and we need to paste it in place of the header code of this FASTA sequence. So, I have pasted it.

(Refer Slide Time: 22:19)

And now we will we see we can see that at the end of the sequence there is a star. So, that star is also required for FASTA query sequence. So, I will put a star here and now in this sequence in the tutorial we can see they are using TvLDH for the tutorial they are using TvLDH, but our sequence is different. So, we have to we can give certain name to our sequence in place of TvLDH.

So, I will open the FASTA the downloaded sequence and here in place of TvLDH we can give just we will replace all the TvLDH wherever it is written TvLDH with query and then we will replace all.

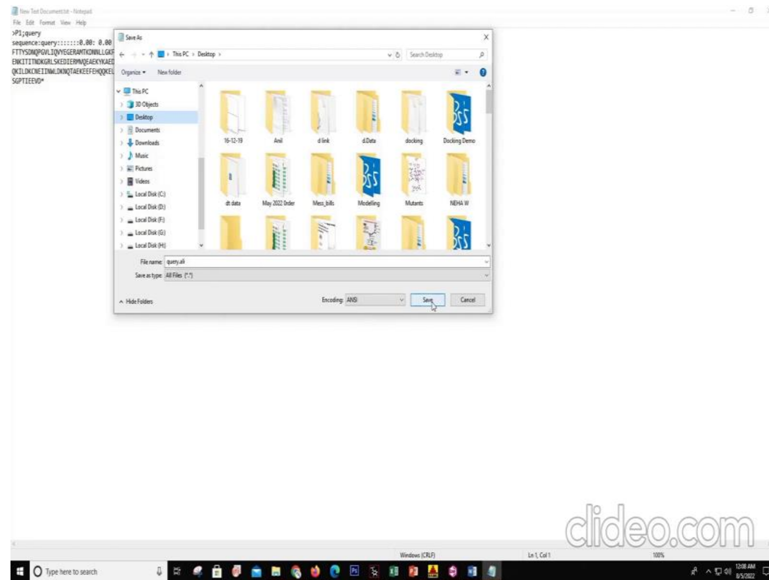
(Refer Slide Time: 23:12)

So, we can see the TvLDH it is replaced with query. Now our query sequence file is ready now we will go to file and save it save it as query dot ali that dot ali extension is very important.

(Refer Slide Time: 23:21)

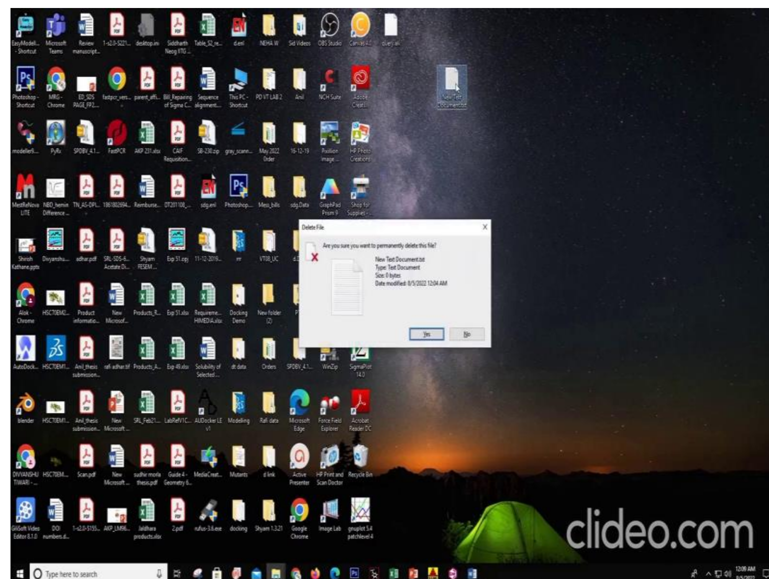
So, we need to save it as query dot ali.

(Refer Slide Time: 23:24)



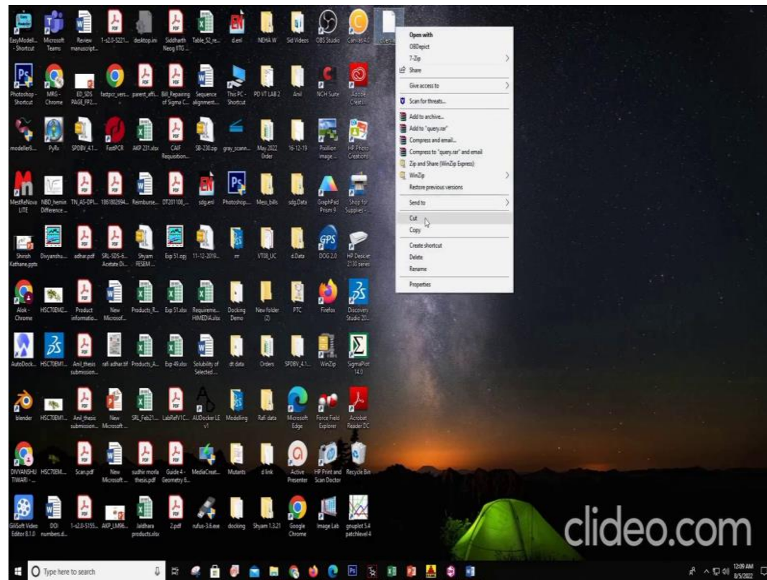
And we will remove the text extension dot dot txt extension and first we will save this on desktop because if we directly save it in the folder we created it will ask for administrative privileges. So, I will save it on desktop first it save on saved on desktop I will close it, I will go to desktop and here is the query dot ali file.

(Refer Slide Time: 24:03)



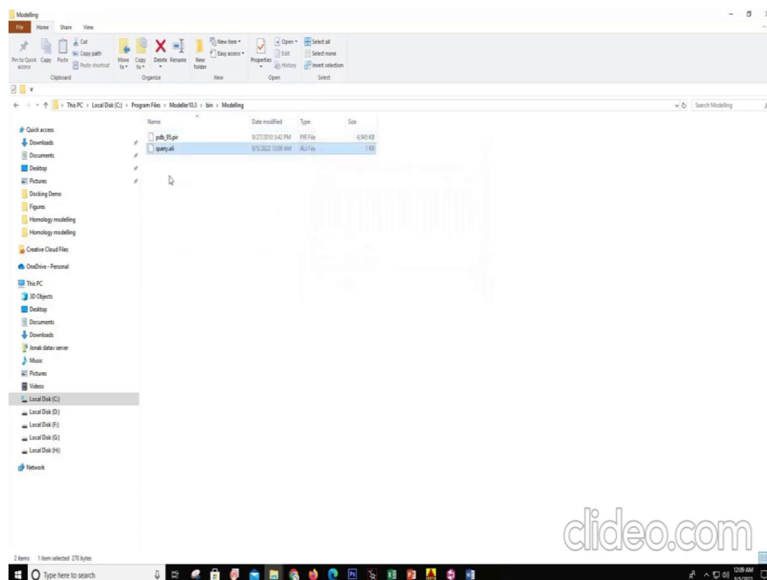
Now, I will delete this previous new text document file.

(Refer Slide Time: 24:06)



And this one I will query dot ali I will cut and I will paste it in the folder which we have created in the bin folder earlier.

(Refer Slide Time: 24:21)



So, I will paste it here. Now it is pasted. Now we will go to again go to the tutorial and the tutorial.

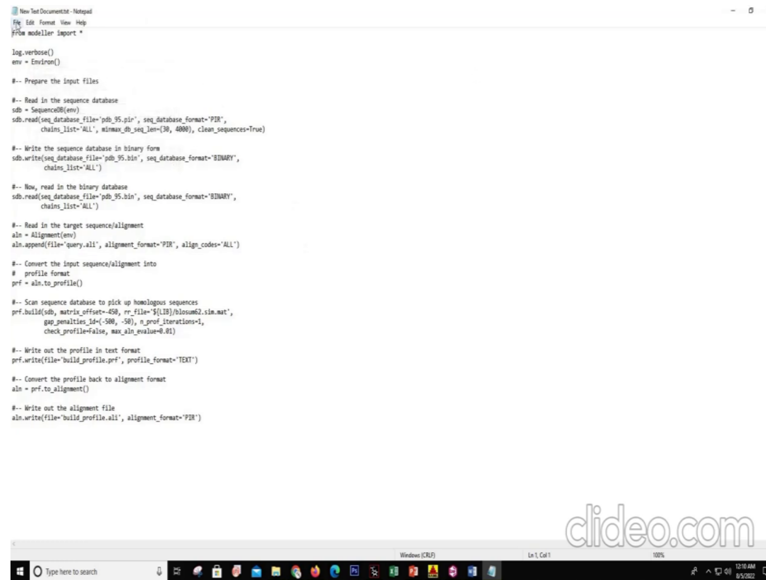
(Refer Slide Time: 24:29)

Now we need to run a script which will create certain input files for the modeller. So, this is the script this complete is the script again we will do the same thing we will copy this script and then we will go to desktop.

(Refer Slide Time: 24:52)

We will open new text document and we will paste our script here.

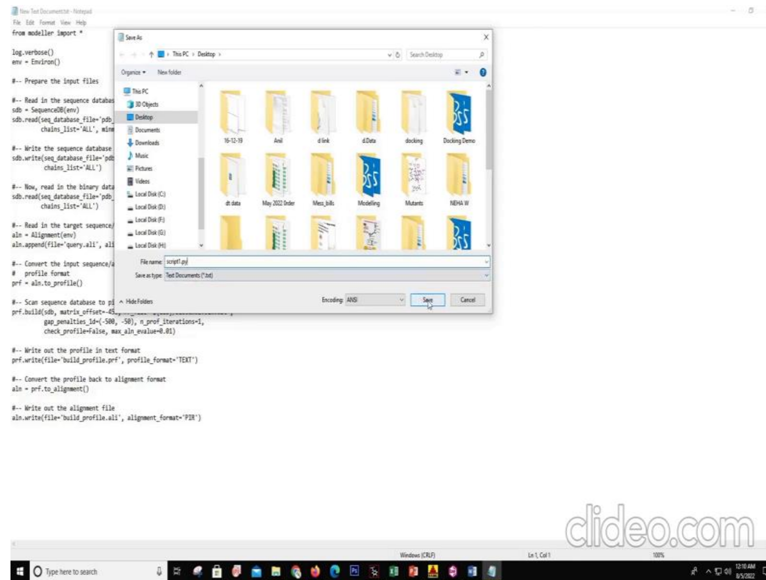
(Refer Slide Time: 24:56)



And now in this script also we need to make certain changes and we can see in the script they have used TvLDH here. So, just we will TvLDH dot ali, but our file is query dot ali [FL] we will replace this we will replace this TvLDH with query. So, we will replace also that wherever it is TvLDH it will be replaced with query. So, no other changes are required.

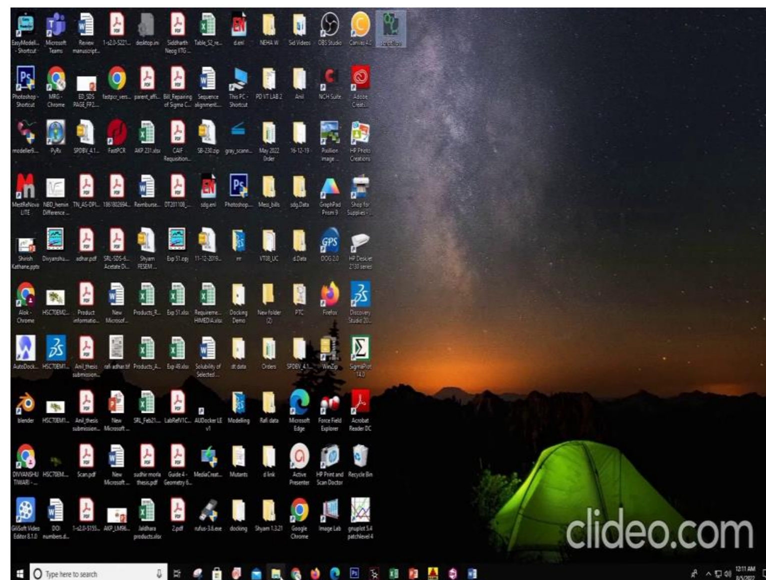
So, we will just go to file and save as. Now we will save this on desktop and we will we can give any name like script 1, but we need to give the file extension as.py script 1 dot py. The py extension is very important for to be make it readable to modeller.

(Refer Slide Time: 25:40)



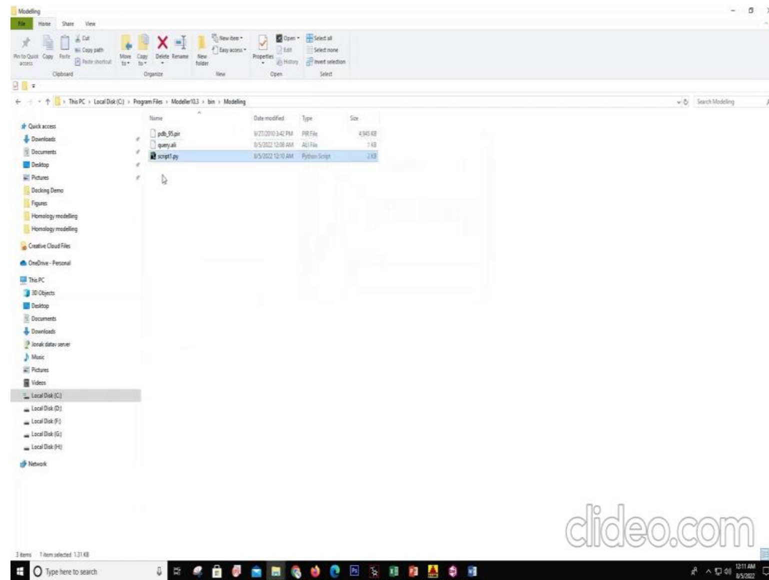
Now we will save it I will close it.

(Refer Slide Time: 26:02)

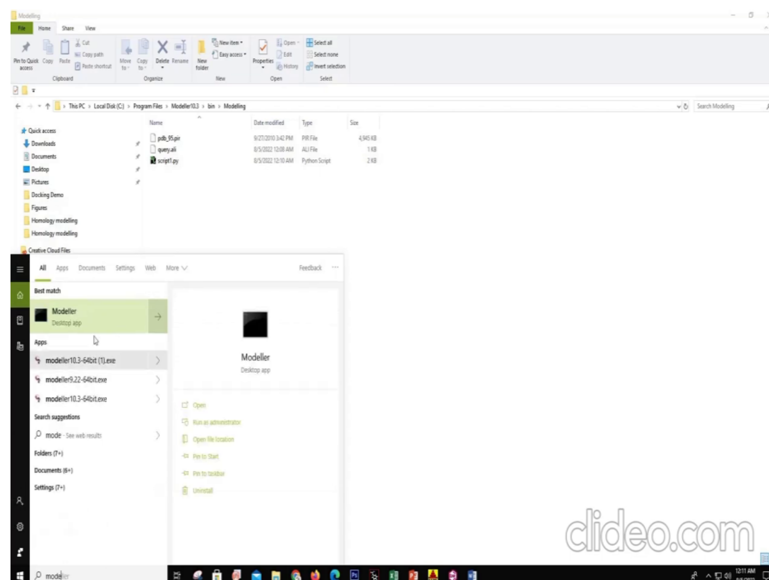


And we can see on the desktop this file script dot 1 dot py is created I will delete this older file and then I will cut this file and now I will paste this in the folder modelling folder which we created.

(Refer Slide Time: 26:20)

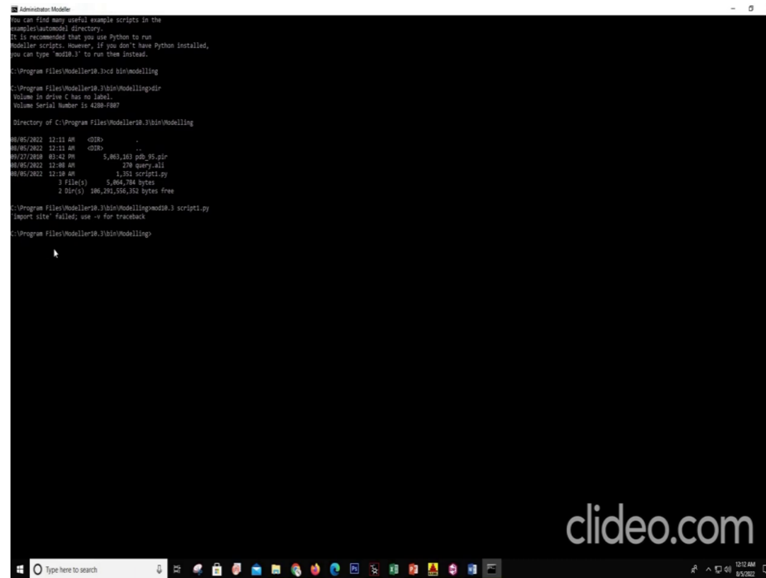


(Refer Slide Time: 26:22)



So, now, we can see it is pasted here. Now, we need to run this script to create other input files for modeller. So, for running the script we need to go to modeller. So, I will open modeller and one thing is very important we need to run modeller as administrator. So, that we do not get any error and so I will run it as administrator. So, this is the modeller command line.

(Refer Slide Time: 26:51)

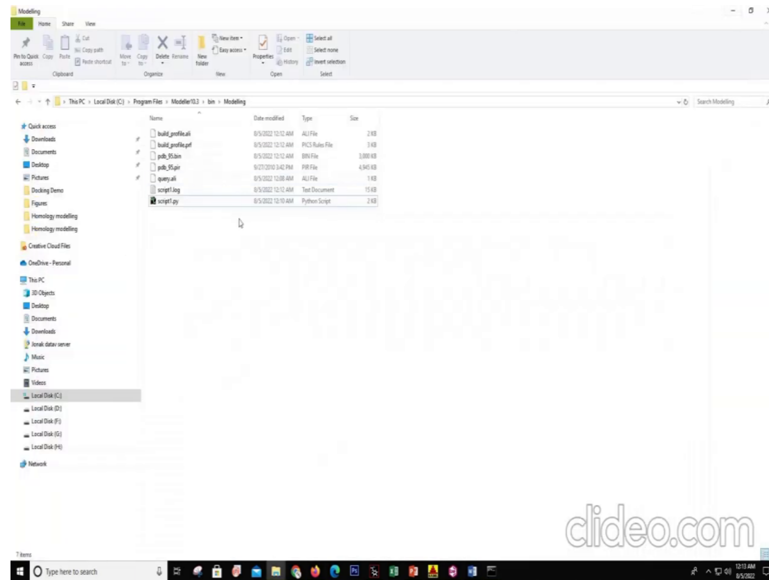


Now, we will the folder active folder is modeller 10.3 we will move to our folder which we created [FL] cd space bin backslash modelling this is the folder we created. Now we have opened that folder here in modeller command line and now we can type dir to see the files present in this folder.

So, we can see pdb underscore 9 5 which we downloaded from the tutorial and this is query dot ali which we created the script 1 dot py which we created. So, we need to run this script. So the we will type the command to run this script that is mod 10.3 is the version of modeller we are using space script 1 dot py.

Now, we press enter and the command will run and after the command is run we will see certain files are automatically created in the fold in the modelling folder.

(Refer Slide Time: 28:06)



So, we can see now this these three files has have already are created like script 1 dot log build profile dot ali build profile dot pr prf and build pdb underscore 95 dot bin these are the new files created which will be used by modeller vector. So, now, our first step is done. So, we will move to the next step which is the selecting the template. So, in selecting the template.

(Refer Slide Time: 28:38)

STEP 2: Selecting the template

- Run protein blast for your sequence using the pblast server on NCBI(https://blast.ncbi.nlm.nih.gov/Blast.cgi)
- Select top three to four sequences with maximum blast score.
- Download pdb structures of the selected templates in .pdb format
- Save in same folder as query sequence
- Make changes in the modeller script for selecting template and save in same folder in .py format
- Run modeller script in modeller command line

- Sequences selected for the demo
- Temp1- PDB Id - 4PO2
- Temp1- PDB Id – 5GJJ
- Temp1- PDB Id – 5FPN
- Temp1- PDB Id - 7KW7

clideo.com

First we need to run a protein blast in NCBI itself for our query sequence and then we will get certain sequences which have which have matching sequence identity. So, we

will select the top 3 to 4 template from those based on their score and then we will download the pdb structure of those templates. And we will save those in the same folder which we created and then we will run the script for selecting the best template.

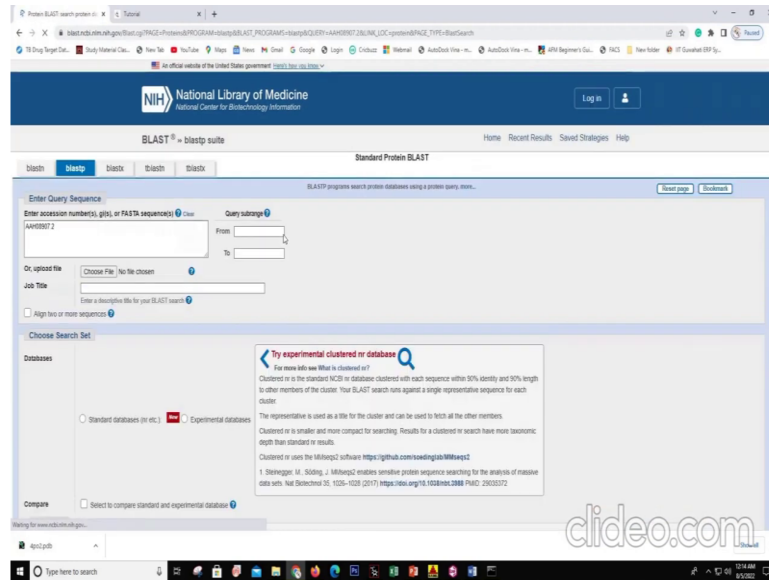
(Refer Slide Time: 29:19)

So, I will show it now. So, we will go to the NCBI site again.

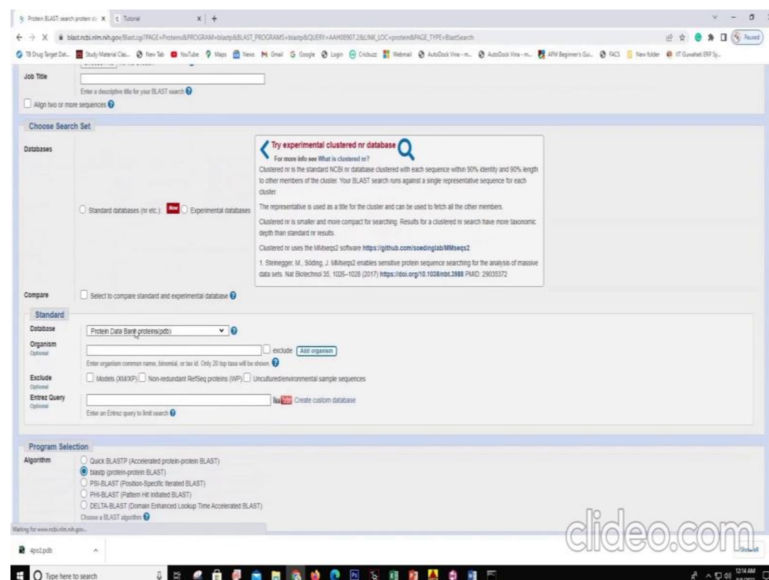
(Refer Slide Time: 29:21)

And here we can see our sequence which we are modelling is already there. So, I will click here on run blast.

(Refer Slide Time: 29:31)

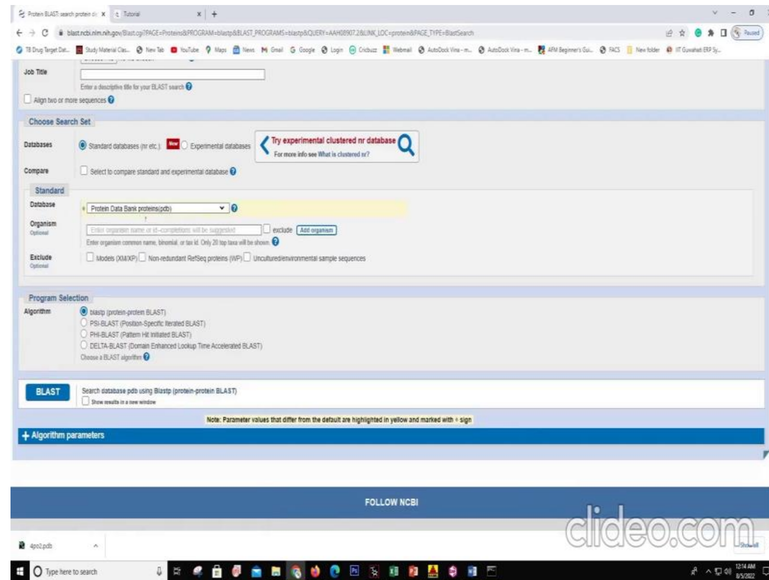


(Refer Slide Time: 29:33)



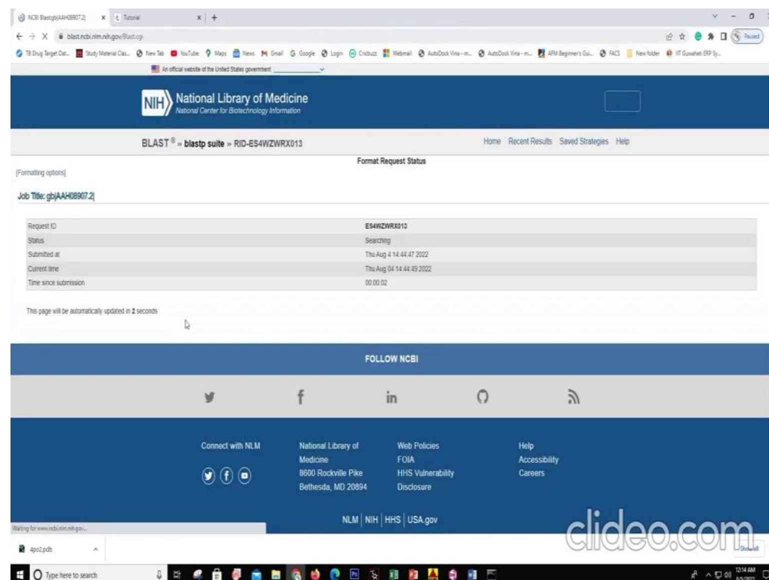
So, after running blast it will a page will open where it will ask for certain parameters.

(Refer Slide Time: 29:37)



So, here we will select the database as protein data bank proteins because we need the pdb files of the templates. So, I will click here blast now.

(Refer Slide Time: 29:49)



And now it will take 2 3 seconds to give the results and here now ok.

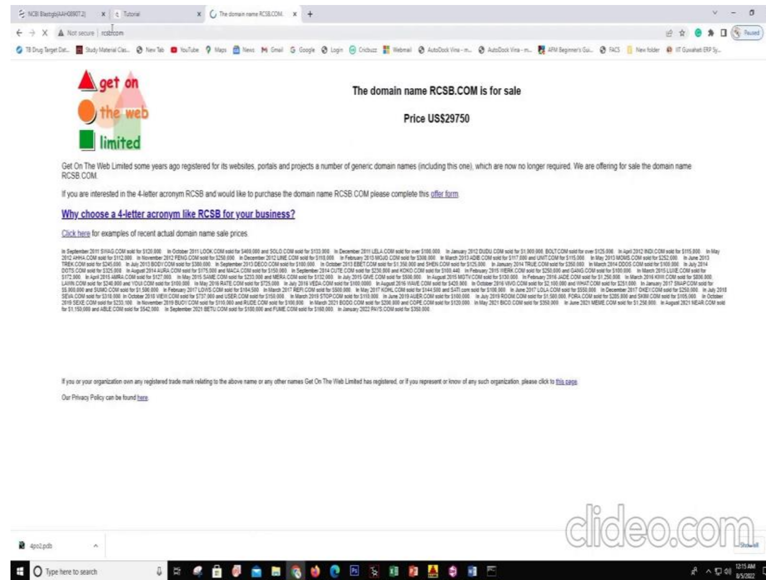
(Refer Slide Time: 29:54)

(Refer Slide Time: 29:56)

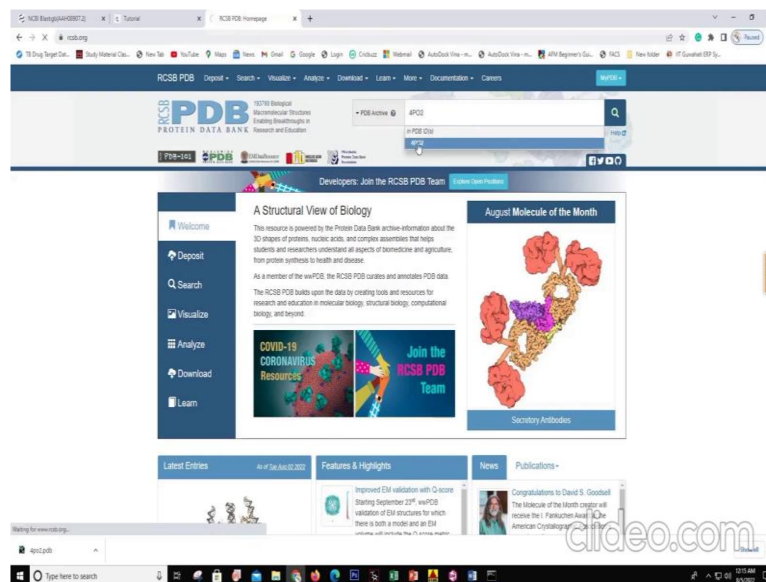
We can see here that the top 3 to 4 sequences have like query cover is more than 84 percent their percent identity is above 80 percent. So, these are significant values [FL] we will select this top 4 structures.

So, now we need to download these 4 structures. So, we will go to the pdb website that is rcsb dot org dot org and then.

(Refer Slide Time: 30:27)



(Refer Slide Time: 30:39)



So it is rcbi dot org and it open and then we will type the pdbid and this 4PO 2 here. So, here I will type 4PO2 and select this. So, it will open the page for.

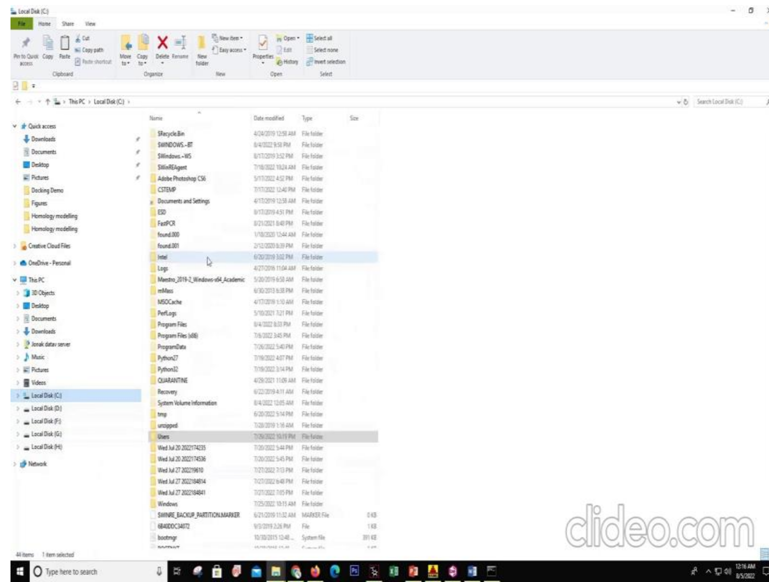
(Refer Slide Time: 31:01)

Now we will can go to this download files and here we can click on pdb format. So, that we will get the pdb format of this file.

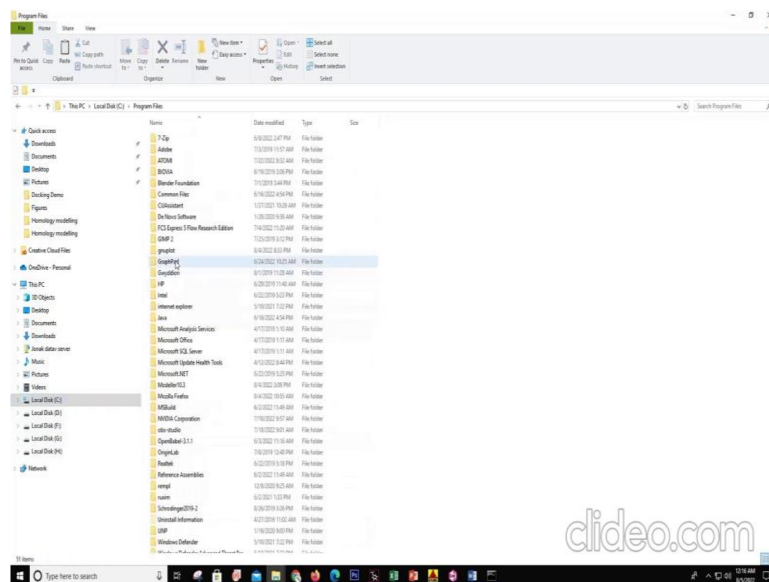
(Refer Slide Time: 31:18)

So, now we will copy and paste this file to our folder. So, before copying and before pasting it there we can just write some other name suppose temp 1 template 1.

(Refer Slide Time: 31:41)



(Refer Slide Time: 31:43)



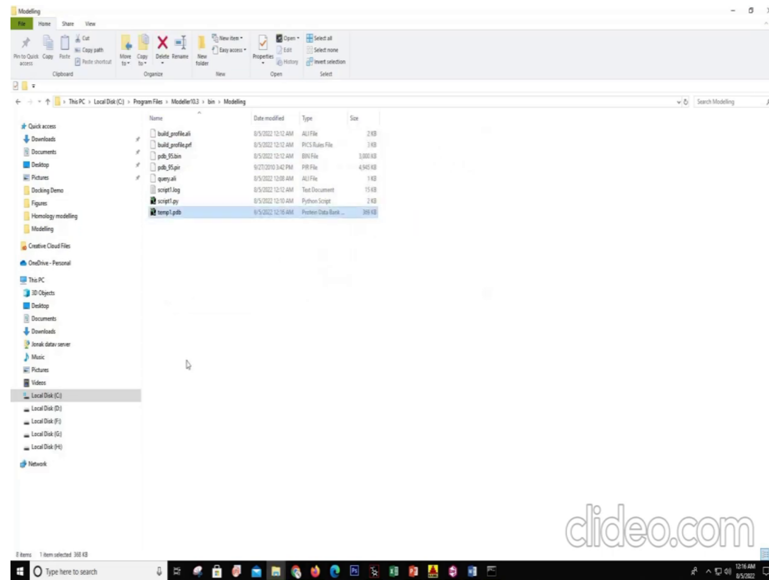
So, for our convenience and now we can just cut it and paste it in the folder we created in modeller 10.3.

(Refer Slide Time: 31:46)

So, I will paste this here.

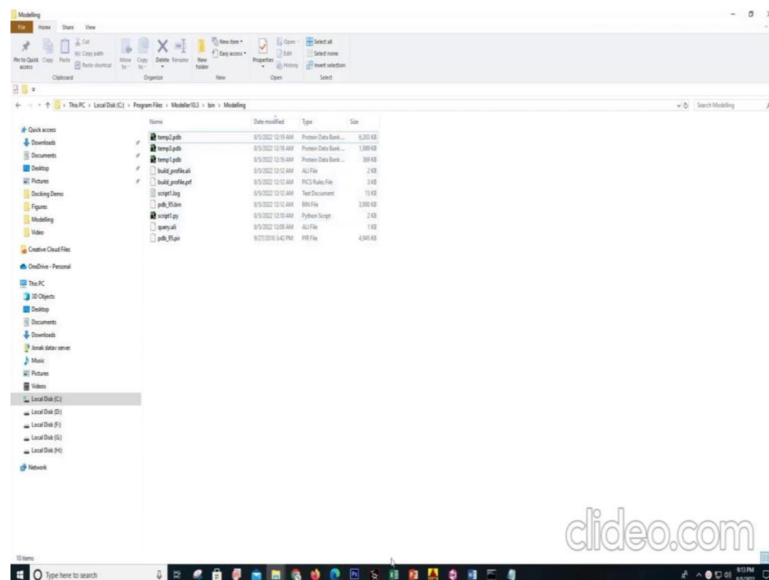
(Refer Slide Time: 31:47)

(Refer Slide Time: 31:48)



Similarly we will download other files and change their name names and paste in the in this folder. So, I have already downloaded the other 2 more sequences.

(Refer Slide Time: 32:05)



And I have pasted them in the modelling change their name and pasted them in the modelling folder. These are temp 1 already we I showed you how to download this then temp 2 and temp 3 I have downloaded and pasted it here. Now, after this we need to go to the tutorial again.

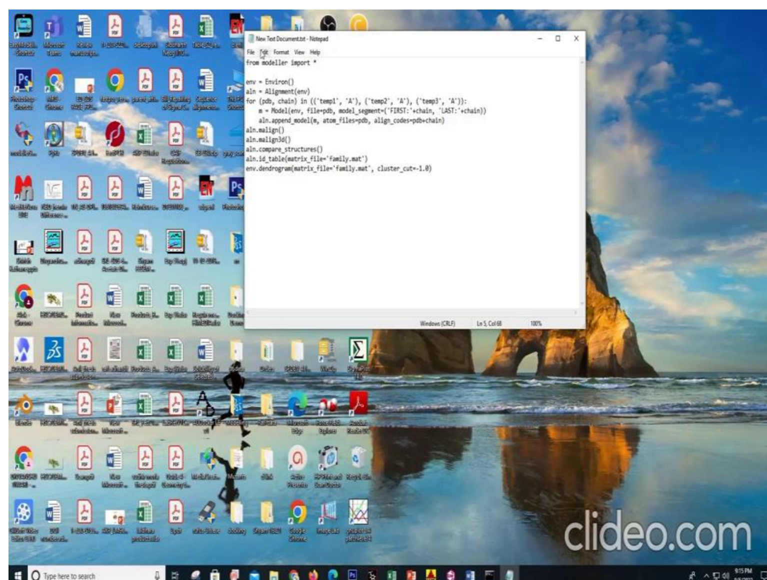
(Refer Slide Time: 32:23)

(Refer Slide Time: 32:25)

And from here we will go to this topic 2 selecting a template and this is the script for selecting the template. So, I will copy this script and again I will paste it on desktop. So, in a new text document file.

(Refer Slide Time: 32:45)

(Refer Slide Time: 32:48)



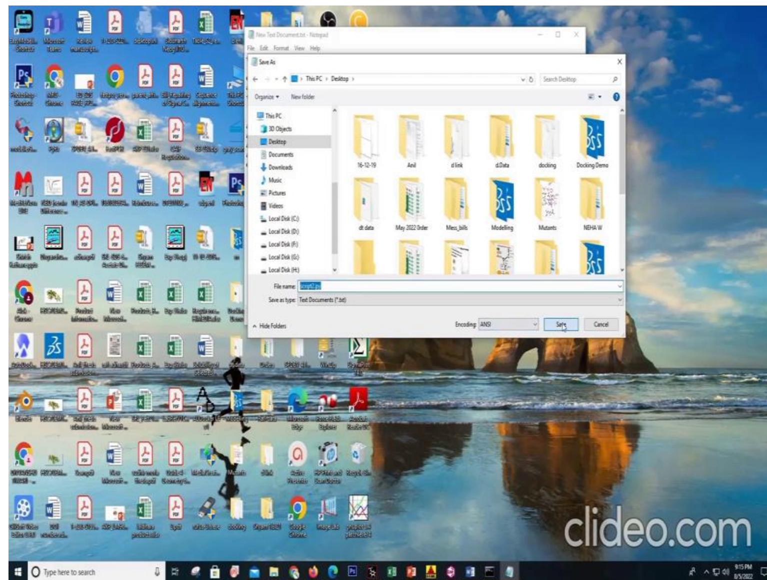
So, I have pasted it here and in this we will make changes accordingly here we can see they have given the name of the sequences which they have used in the tutorial. So, this we will replace with our template sequences which with the names which we have given.

So, this I will replace with temp 1, 1 bdm I will replace with temp 2, one civ I will replace with temp 3 and 5 mdh I will replace with temp 4. Now, they have used more than we are using 3 sequences only. So, this temp 4 is not required. So, rest of the

sequences which they have used in the tutorial we can delete it. So, we need to delete this ok.

Now, this is done. So, there are no more changes required. So, I will save this file in py format.

(Refer Slide Time: 33:50)



So, I will save it on desktop I will write it at this script 2 dot py and save it on desktop first. Now we can go to the desktop and that script 2 dot py file is here. We can delete this text document file first and then we will cut this script file and paste it in the modelling folder.

(Refer Slide Time: 34:22)

Now, the script is here. So, we can now we need to run this script we will go to modeller command line.

(Refer Slide Time: 34:26)

And here ok. So, now, I will type the command for running the script mod 10.3 space script 2 dot py. So, now, script 2 has run. So, we will go to the to our modelling folder.

(Refer Slide Time: 34:58)

And here we can see few more files are generated. So, these script 2 dot log and this family file this has been created by modeller. So now, we need to go to the tutorial again.

(Refer Slide Time: 35:23)

And now we will move to before that we should go to the modelling folder and we can open this script 2 dot log file.

(Refer Slide Time: 35:24)

We can open if it notepad and at the end of this file.

(Refer Slide Time: 35:37)

We can see a comparison between the three templates which we used. So, here the template one is the best template this 2.0 it shows its crystallographic structure resolution. So, we will select this template 1 from this 3 sequences for further operations. So, now, we can close this script file.

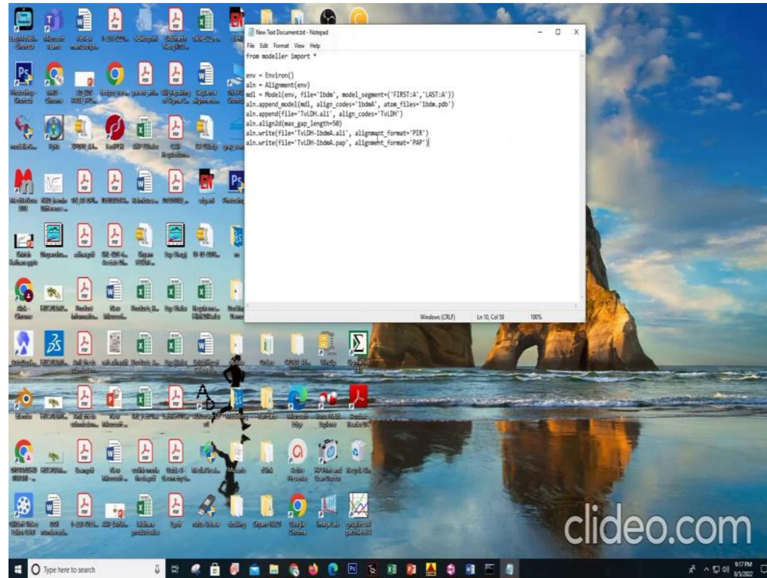
(Refer Slide Time: 36:09)

And now we will go to the next step that is aligning our sequence with the template. So, this is the script for alignment of our query sequence with our template sequence. So, the now I will copy this script again we will paste it in a new text document file.

(Refer Slide Time: 36:30)

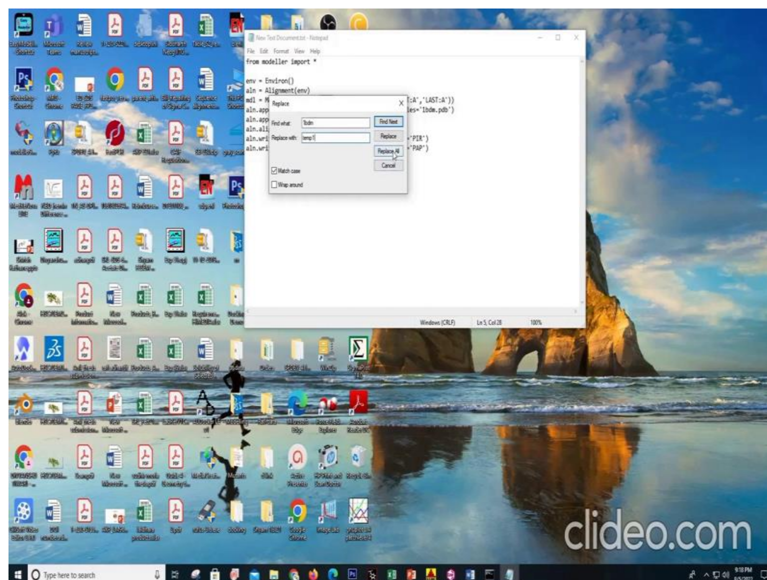
And we will make required changes.

(Refer Slide Time: 36:34)



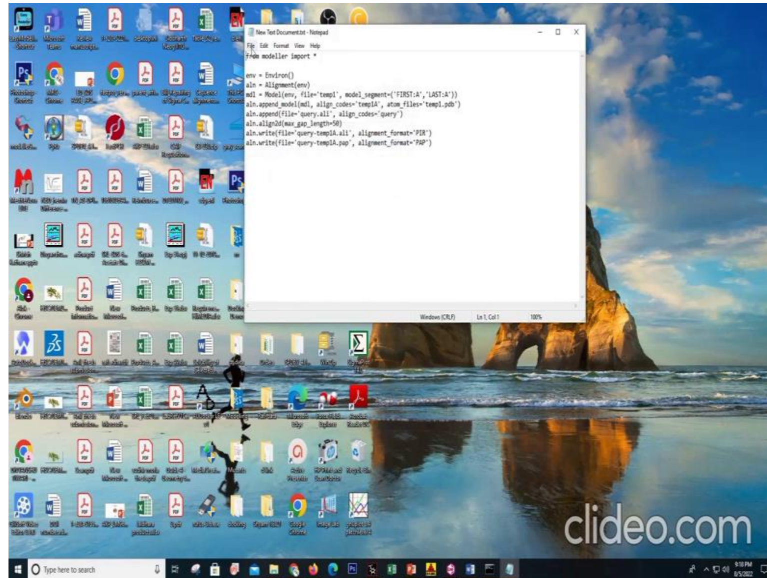
So, here we can see here it is 1 bdm which the they have used in tutorial the sequence which they have used in tutorial. So, this 1 bdm.

(Refer Slide Time: 36:46)



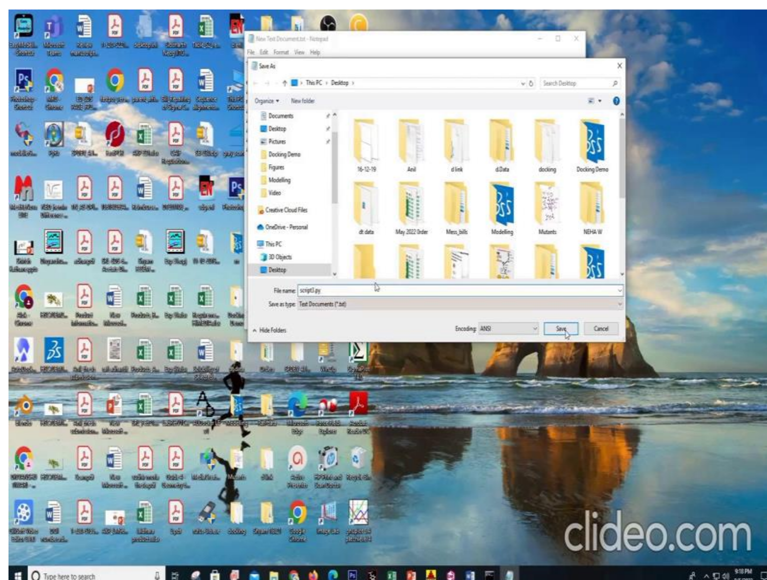
This we will replace with our template sequence that is temp 1.

(Refer Slide Time: 37:02)



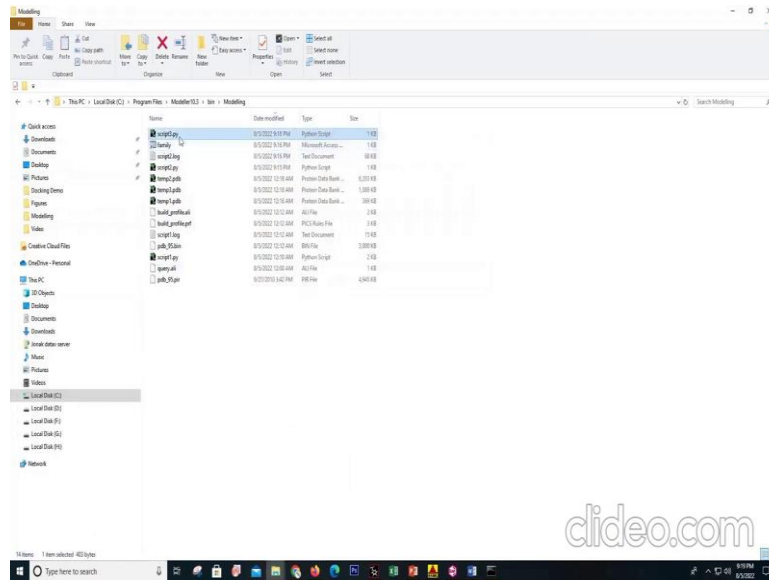
So, I replace all and then here we can see it is written as TvLDH. So, TvLDH we will replace with our query sequence I will replace TvLDH with query. So, we have made the required changes. Now, I will save this script file as in dot py format.

(Refer Slide Time: 37:25)



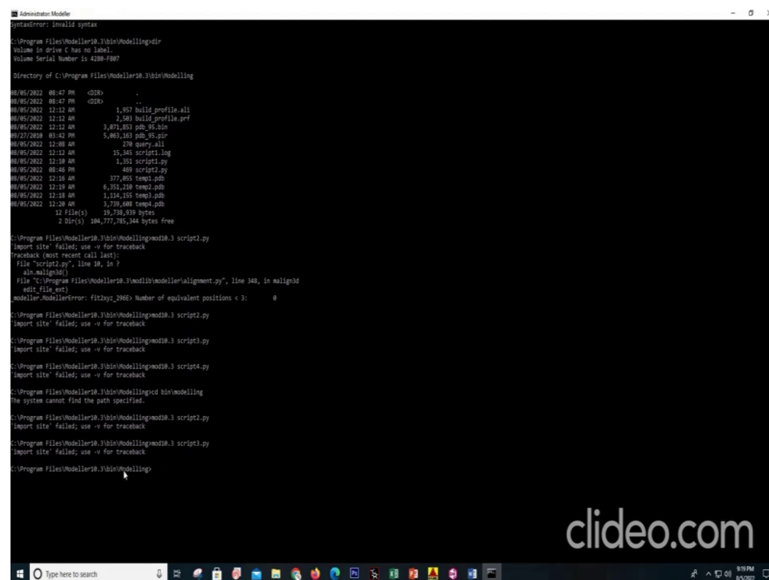
So, I will save it as script 3 dot py. Now, we can see on the desktop the script 3 dot py file is created. So, I will just delete the text document file cut this script file and paste it in the modelling folder.

(Refer Slide Time: 37:51)



Now script 3 dot py is pasted here. So, again we will go to modeller to run the script 3.

(Refer Slide Time: 38:00)



So, here we can type same command mod 10.3 space script 3 dot py. So, now, the script 3 is has also run. So, it will create some new files here we can see these top 2 files dot pap file and dot ali file has been created after running the script 3.

(Refer Slide Time: 38:14)

So, now we will move to our final step again.

(Refer Slide Time: 38:30)

We will go to the tutorial. So, in tutorial the final step is model building. So, this is the script for the building the model based on the selected templates. So, I will copy this script again.

(Refer Slide Time: 38:48)

And paste it on desktop in a new text document file.

(Refer Slide Time: 38:51)

And here again we need to do some changes. So, here we will replace TvLDH with query.

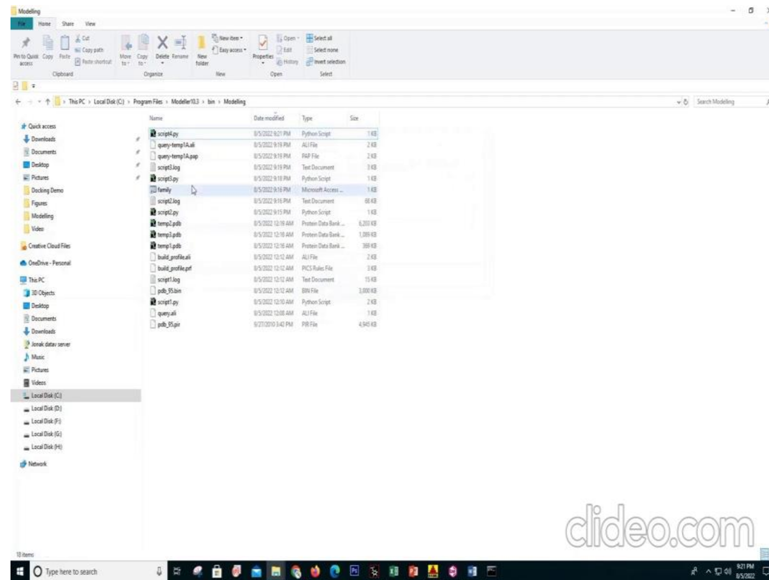
(Refer Slide Time: 39:08)

So, let me written here. So, and this 1 bdm I will replace with temp 1. So, I will type temp 1 and then replace all. So, we can see that it is replaced TvLDH is replaced with temp 1. Now, again we will save this script in.py format on desktop.

(Refer Slide Time: 39:37)

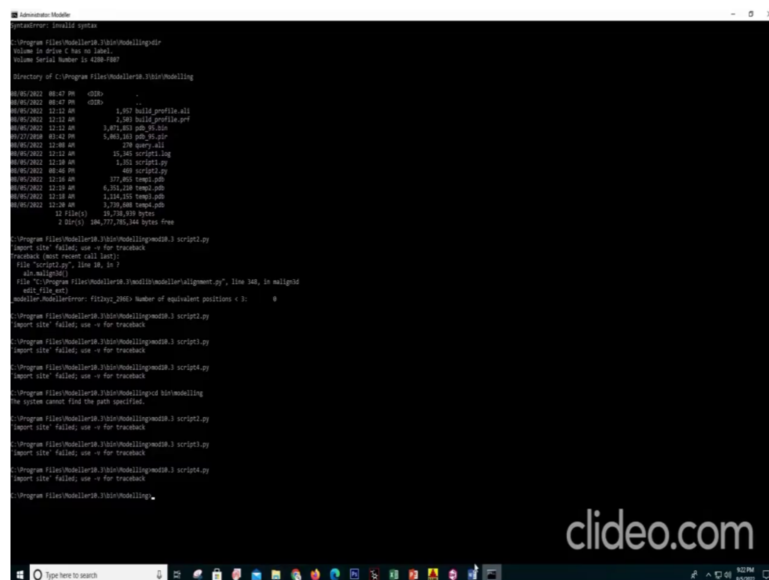
So, this is script 4 dot py. Now, we can close this and we can see on desktop script 4 dot py file is created. So, I will cut this file here and paste it in the modeller modelling folder.

(Refer Slide Time: 40:06)



Now, the script 4 is dot py is present here. So, we will go to modeller command line again.

(Refer Slide Time: 40:14)



And we will type mod 10.3 space script 4 dot py. So, this script will you know create the model.

(Refer Slide Time: 40:30)

So, while this script is running we can see in our folder some new files are being created. So, these are the new files being created and the models for our query structure based on the template are also getting written here. So, this is the first model query B9990001 is the first model and this is the second model it will create few more models and then we will analyze these models.

So, it will take around 5 10 seconds to create all the models and in this we can also see script 4 dot log file this script 4 dot log file this we will use to see the dope score which is assigned to the model by modeller and based on the dope score we will select the best model and then we will evaluate it using certain online servers and tools. So, we will go to modeller command line and here we can see the that the script 4 has run successfully.

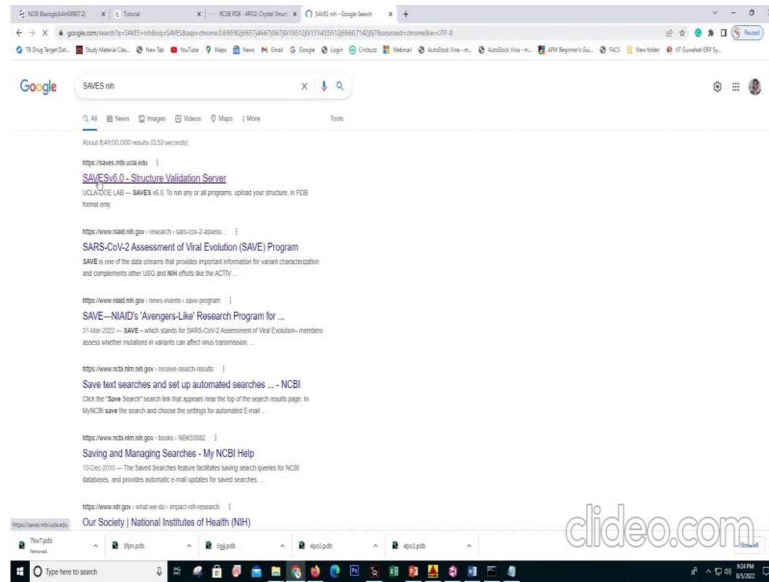
So, now we will go to the modelling folder again and then here we will open script 4 dot log file in notepad.

(Refer Slide Time: 41:50)

(Refer Slide Time: 41:55)

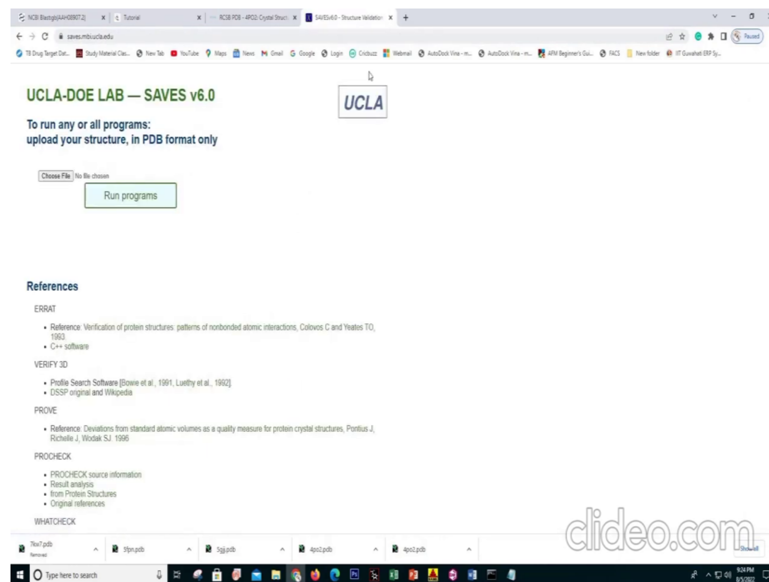
So, here we can see in at the last of this file all the models are given and their small pdf and dope score and ga 341 score is given. So, based on the dope score the model with having the lowest dope score or the most negative dope score we will select. So, we can see this model 4 this is having the lowest dope score. So, we can now evaluate this model using online servers.

(Refer Slide Time: 42:33)



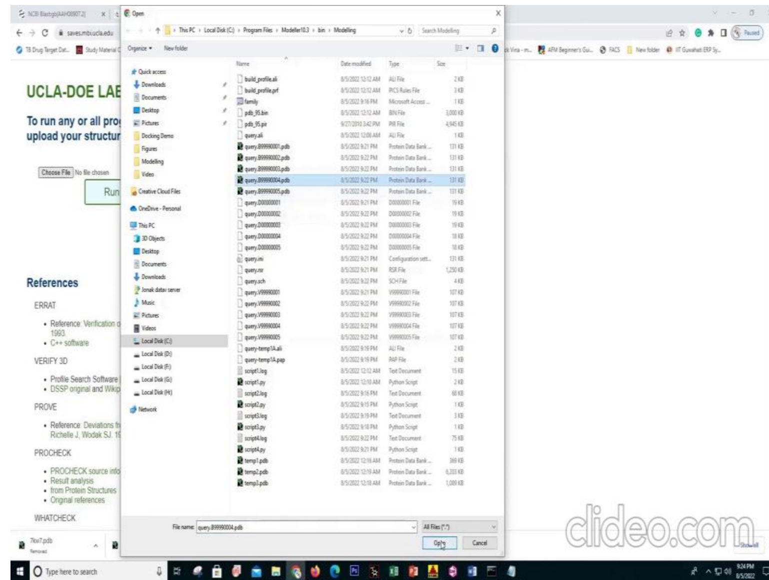
So, we will go to we can use saves server which is created by nih and so I will open this server here.

(Refer Slide Time: 42:44)



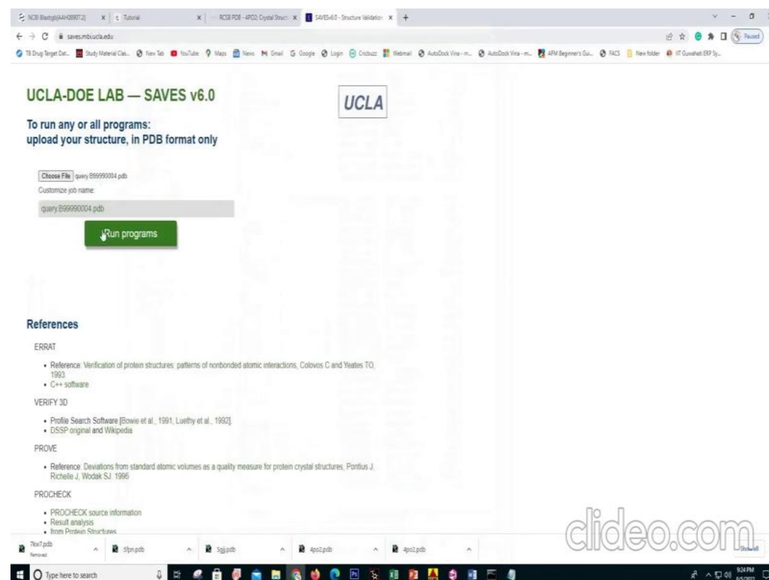
So, here we can choose the model which we want to evaluate.

(Refer Slide Time: 42:52)



So, I will choose we have seen in the script 4 dot log file that is model 4 has the maximum sorry most low lowest dope score. So, we will select model 4 and open it here.

(Refer Slide Time: 43:06)



Now it is selected here and now we will run programs.

(Refer Slide Time: 43:12)

So, here we can evaluate or model it will it is a combination of several evaluation checks. So, we will evaluate for ERRAT, we will evaluate for Verify 3D and we will evaluate for PROCHECK. So, we will start with ERRAT.

(Refer Slide Time: 43:27)

So, now, it will take few seconds. So, overall quality factor for this model is 84.9462 which is a; which is very good and very significant. Now we will verify that with verified 3D. So, it determines the compatibility of the atomic model with its own sequence.

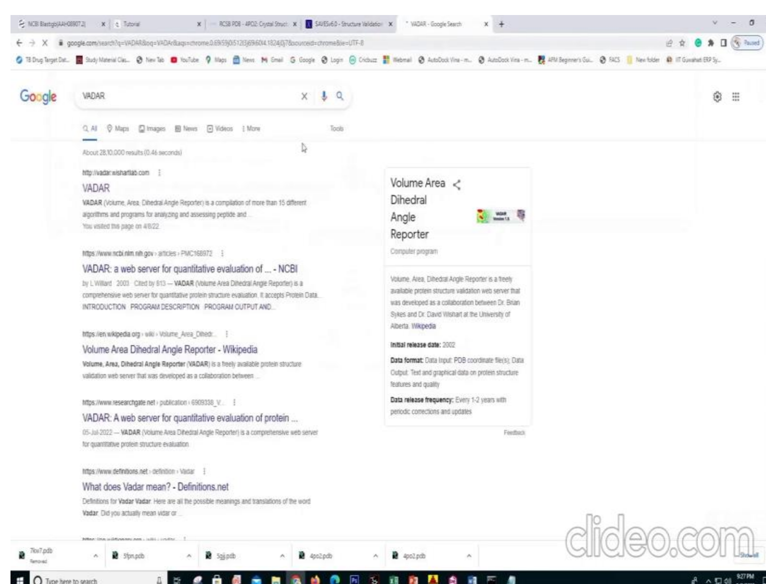
So, we will start this Verify 3D check and here we have to wait for 5 6 seconds and it will give the results. So, and after that we will go for pro check evaluation that pro check evaluation does the stereo. It checks the stereochemical quality of the protein structure by analyzing the geometry residue by residue.

So, this Verify 3D is taking little more time we have to wait see the results of Verify 3D has come now. So, it is showing that 99.54 percent of the residues have averaged 3D 1D score that is greater than or equal to point. So, if 80 percent of the residue shows the score greater than 0.2 that models is passed.

So, we can see our model is passed here. Now we can go to pro check and start the pro check evaluation. So, in it will this pro check will perform few and I think 6 evaluation 6 or 8 evaluation and out of this it will show how much are passed and so here we can see it is showing out of 8 evolution. There are errors in 2 and 6 are pass and there is no warning. So, this is a good result for the model.

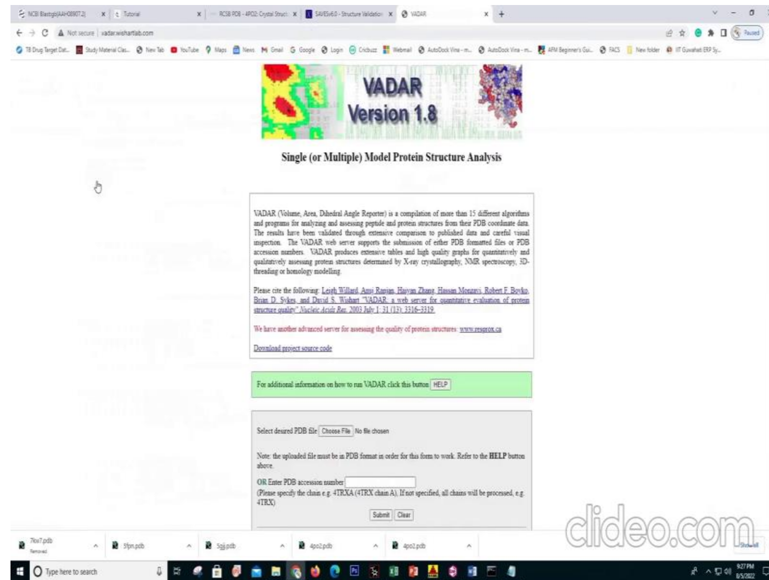
So, we can see the evaluation our model based on these evaluation we can say that our model the our query sequence has been modelled correctly. So, now, we can do some more evaluations with other tool that is VADAR.

(Refer Slide Time: 45:46)



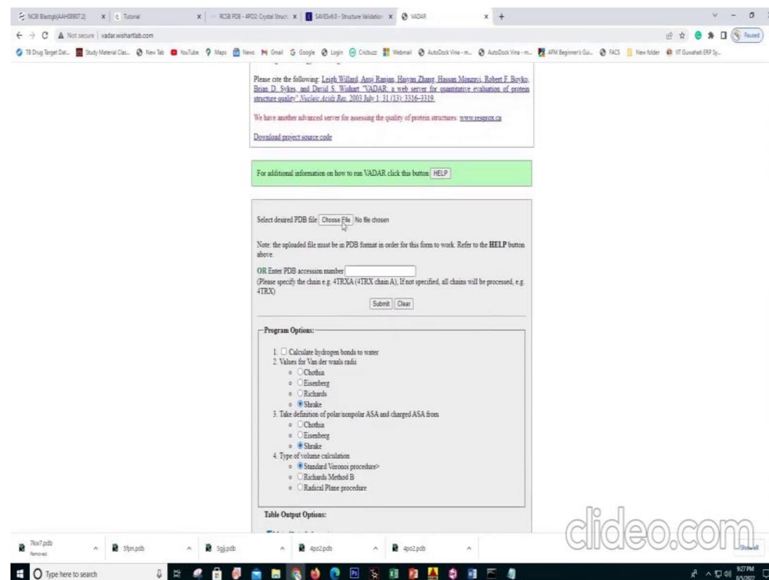
So, this VADAR tool we can generate our Ramachandran plot here.

(Refer Slide Time: 45:52)

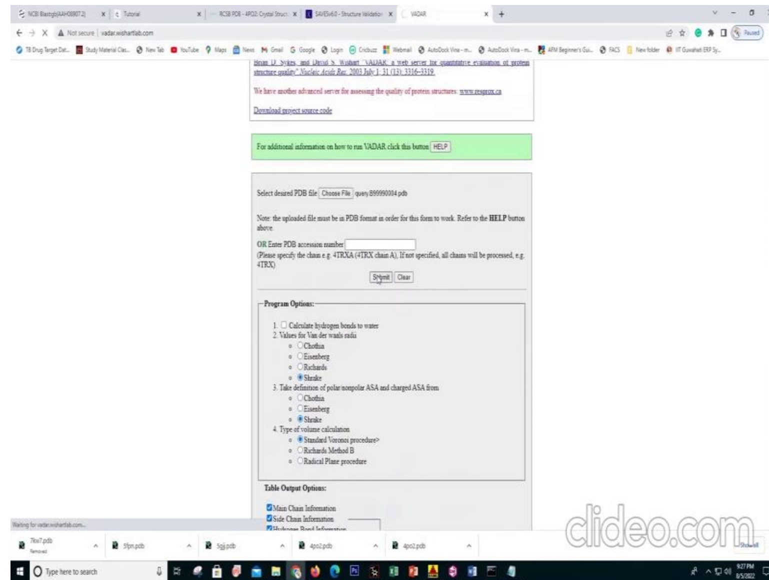


So, I will choose the our model 4 again here.

(Refer Slide Time: 46:55)

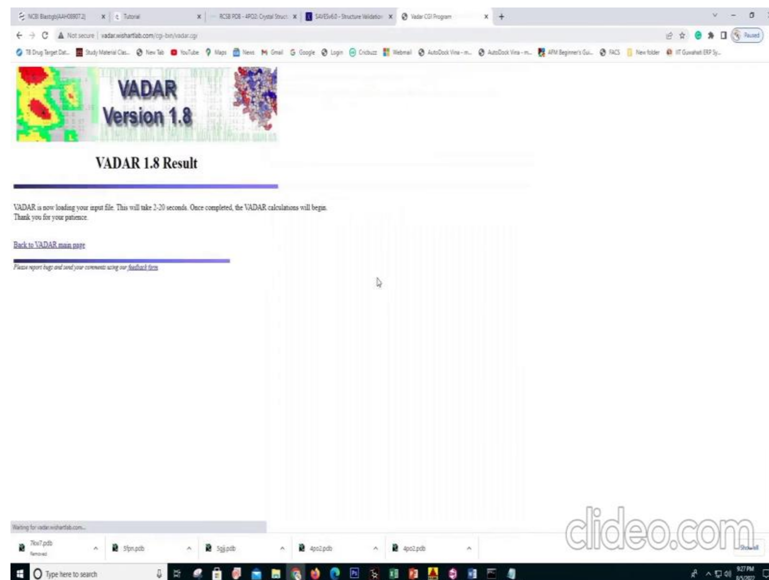


(Refer Slide Time: 46:02)



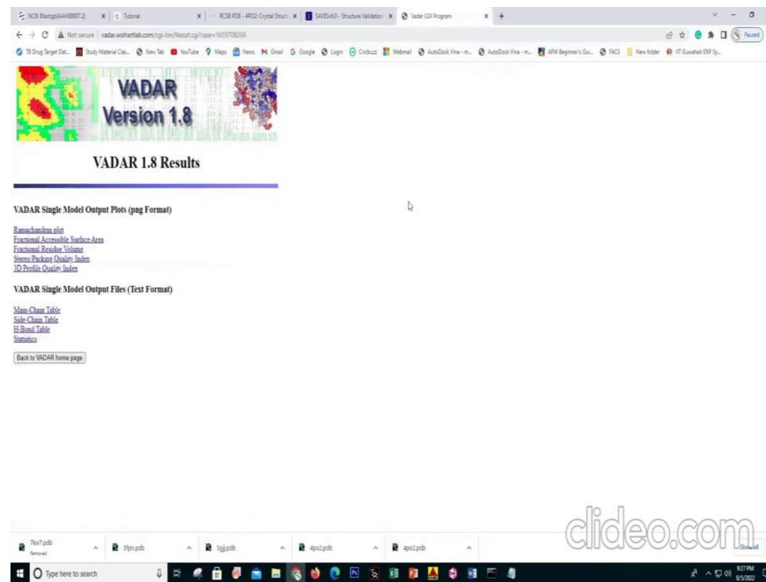
And then I will click submit.

(Refer Slide Time: 46:09)



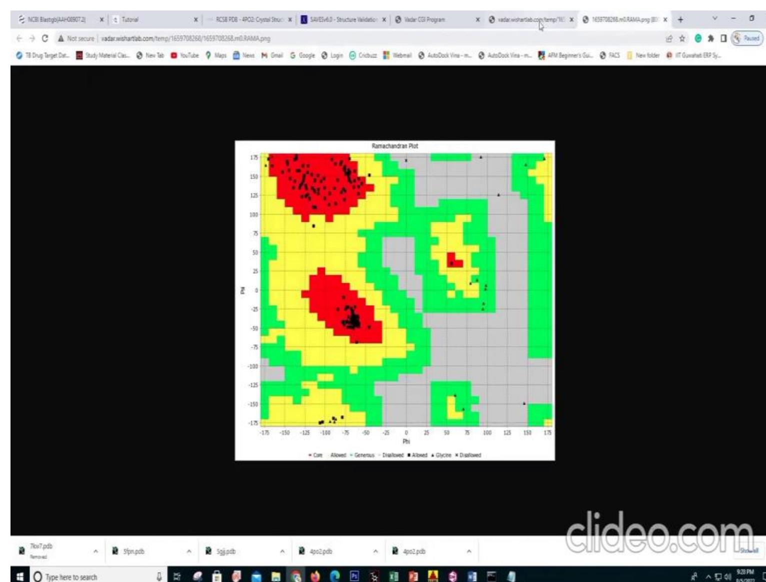
So, it will generate the Ramachandran plot.

(Refer Slide Time: 46:13)



And other statistics for our model here we can see these are the plots it has created and these 4 are the output files it has created. From here we can click on Ramachandran plot.

(Refer Slide Time: 46:28)



And it will give the Ramachandran plot for our model. So, we can see that most of the this Ramachandran plot see this with the most of the residues are in yellowed reasons and then again we can go to the output of the VADAR 1.8 results and then here we can click on statistics.

(Refer Slide Time: 46:54)

(Refer Slide Time: 46:55)

So, this statistics will show the percentage of alpha helix percentage of beta sheets and the coiled structure in our protein.

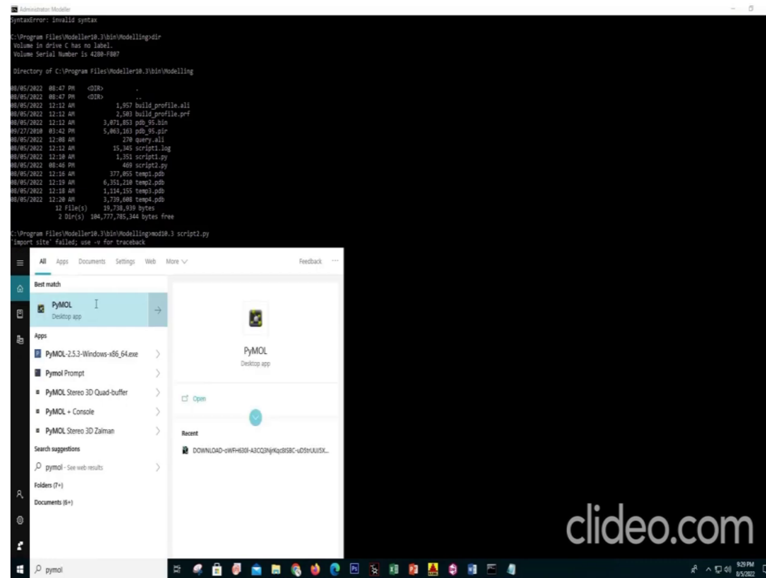
So, alpha helix is 40 percent beta sheets 21 percent and coiled is 38 percent. So, and other statistics of our protein modelled structure we can find out here. So, based on this evolution we can say that our model is significant and correct.

(Refer Slide Time: 47:31)

(Refer Slide Time: 47:33)

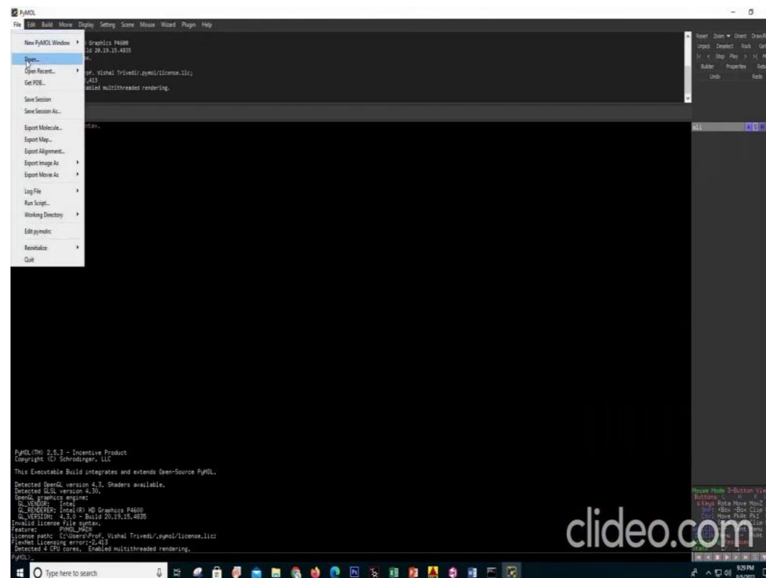
Now, we can visualize our model in many visualization software like I will open this PyMOL or discovery studio.

(Refer Slide Time: 47:41)



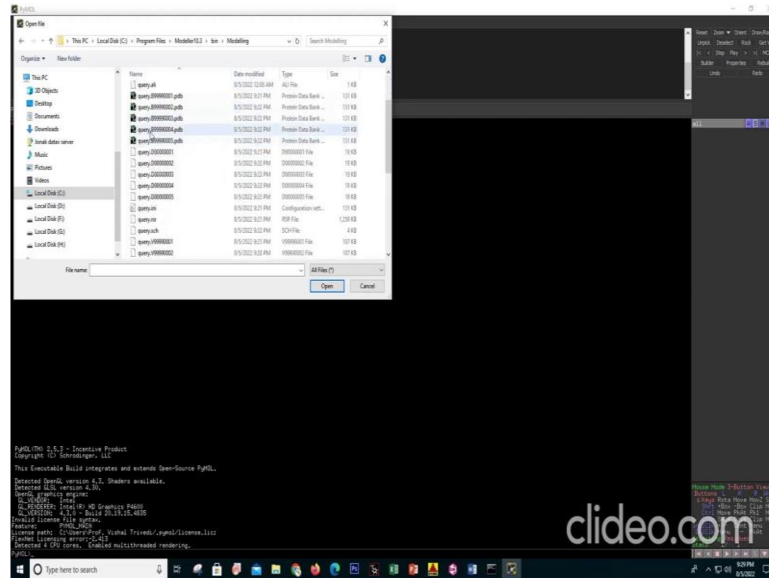
So, I will open PyMOL now and we will visualize our model in the PyMOL. So, save in the PyMOL is opening ok.

(Refer Slide Time: 48:06)



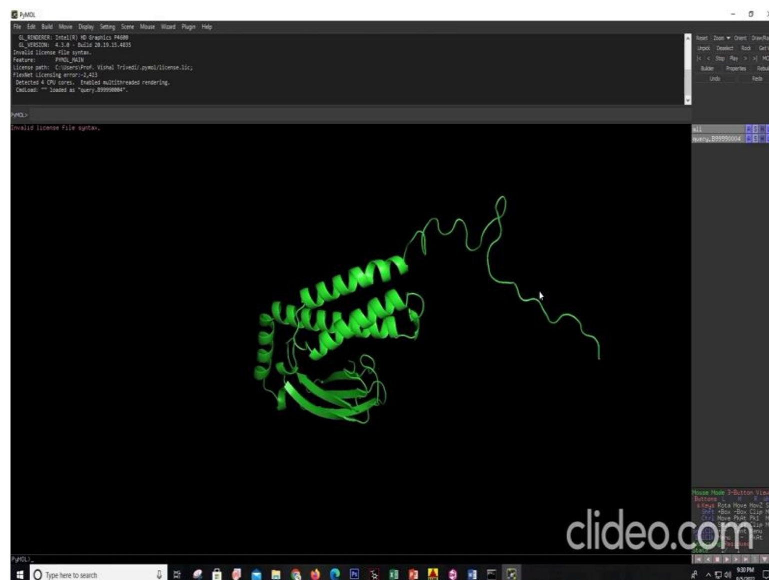
So, in PyMOL we can just go to file open.

(Refer Slide Time: 48:14)



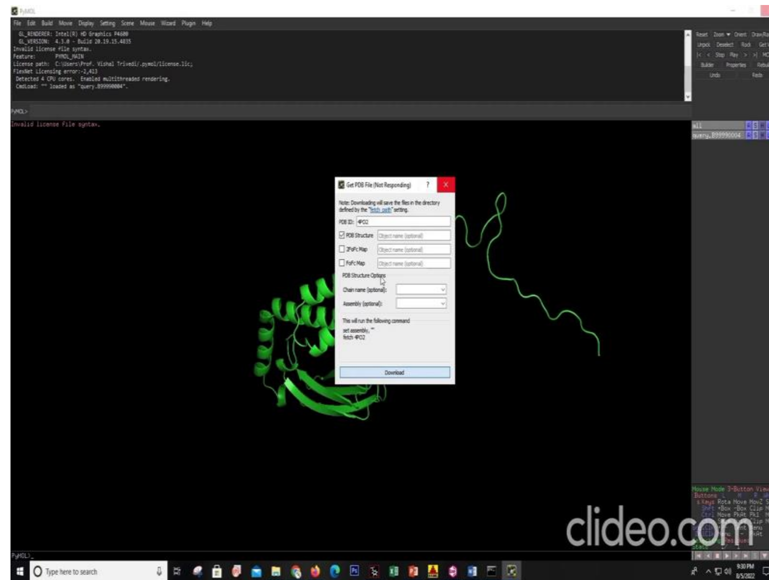
And we can select our model from that fold from the modelling folder. So, this is the model 4 I will open here.

(Refer Slide Time: 48:29)



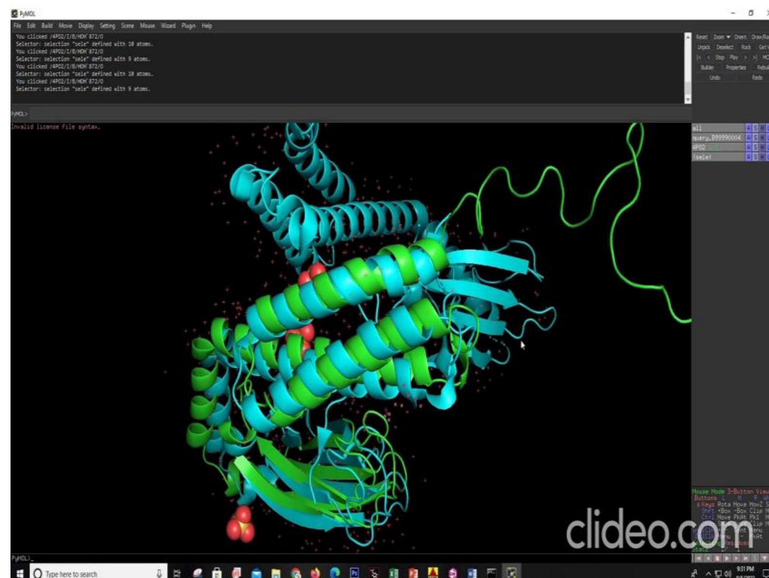
So, here we can see this is the model structure for our query sequence and we can check its alignment with the template sequence. So, we know that our we can go to the NCBI and we know that it is modelled based on the template 1 which is 4 PO 2. So, we can just in in PyMOL we can open this 4 PO2.

(Refer Slide Time: 49:02)



And download it will download the structure of 4 PO2 in PyMOL.

(Refer Slide Time: 49:19)



So, here we can see it has downloaded the structure. So, we can see here that our structure is nicely aligned with a with a segment of the template sequence and the rest of the sequence in the template does not have identity with our protein or with our query sequence our query sequence was 219 amino acids only. So, we can see the those amino acids are aligned nicely with the amino acids of our template sequence.

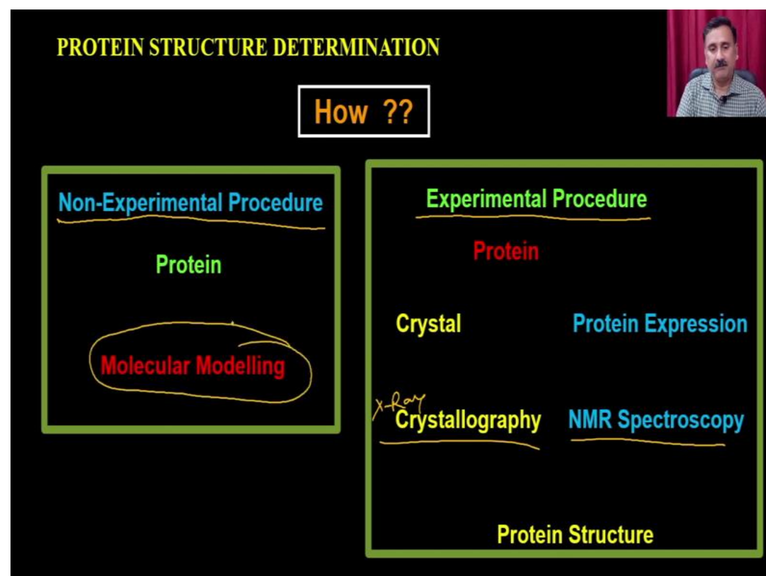
So, [FL] this is all with modeller we have we have performed all the steps like preparing the query sequence then selecting the template then building the model and the model evaluation. So, this is all with this is all with basic modelling and there are certain other things like advanced modelling which will, but in this demo I will be showing only basic modelling. So, that is all with the video thank you.

So, I hope you have enjoyed the demo clip and in the demo clip my student has shown you the different steps how you can be able to find the query sequence, how you can be able to use that query sequence to determine or select the templates, what are the different parameters you should use and while he was showing you the demo he has used the program which is called as modeller 10th version right.

And then he has shown you how to do alignment of query with the template. So, that you can be able to screen out which template you should use and which template you should avoid and then he has shown you the molecular modelling of building the you know 3D models and then ultimately the validation of the program with the saves servers.

So, now you got the 3D models at right. So, what we have discussed?

(Refer Slide Time: 51:33)



So, far is we have discussed about the tertiary structures, how to determine the protein tertiary structure of the proteins and we have discussed about the computational method and we have also discussed about the experimental wet lab experiments.

So, in the wet lab experiment we discussed about the X-Ray crystallography and the NMR spectroscopy whereas, in the case of the non-experimental computational method, we have discussed about the homology modelling.

(Refer Slide Time: 52:03)

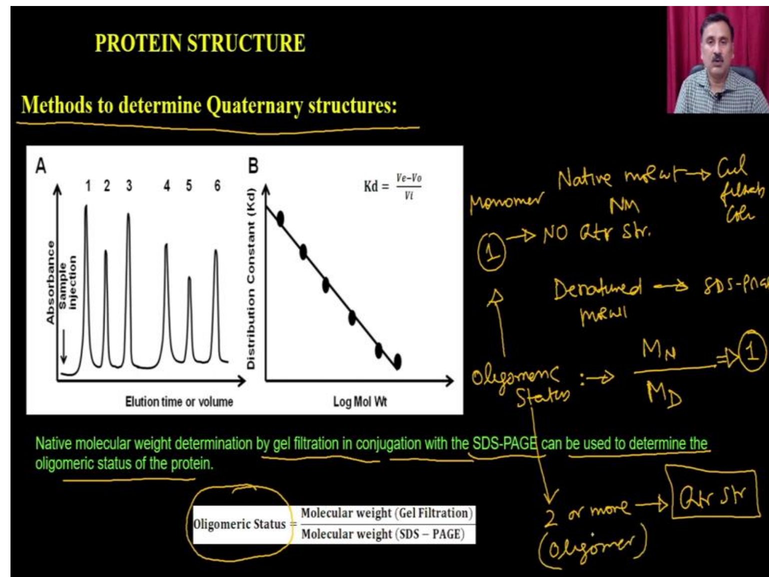
Now, once we have done with the tertiary structures you can also have the quaternary structure. So, if the multiple polypeptides are involved in the constitution of the protein the tertiary structure of these different polypeptide chain come together to form the quaternary structure. Now the question comes under what conditions you can have the quaternary structures? So and how you will know that the protein has the quaternary structures?

So, if so the question so answer in the of this question is if the multiple polypeptides are involved in the constitution of the protein the tertiary structure of the different polypeptide chains are going to come together and that is how it is actually going to give you the quaternary structure. Classical example is hemoglobin where you are going to have the 2 alpha chain and 2 beta chains right.

So, it is actually a hetero tetramer right where you have the 2 alpha chains and 2 beta chain. So, it has a four chain and that is how it is actually going to give you a quaternary structure, but the question is how experimentally when I give you a protein sequence or when I give you a protein how you can be able to determine whether it also has the quaternary structure or not.

So, there is a simple experiment what you can actually do to determine whether the quaternary structure is present in this particular protein or not.

(Refer Slide Time: 53:34)



What you are going to do is methods to determine the quaternary structure is what you are going to do is you are going to determine the oligomeric status and this methods anyway we are going to discuss in detail when we are going to discuss about the gel filtration chromatography and as well as the electron electrophoresis right.

So, what you are going to do is you are going to take the protein and resolve that protein onto a gel filtration column in conjugation with the SDS page and what will happen is that when you do that it is actually going to give you the two molecular weight. It is going to give you a native molecular weight which you are going to get from the gel filtration column and it is also going to give you the denatured molecular weight which you are going to get from the SDS page.

Now, what you are going to do is you are going to determine the oligomeric status right and when you do the oligomeric status what you can do is you can determine the NM right. So, native molecular weight you can actually be able to divide by the denatured molecular weight.

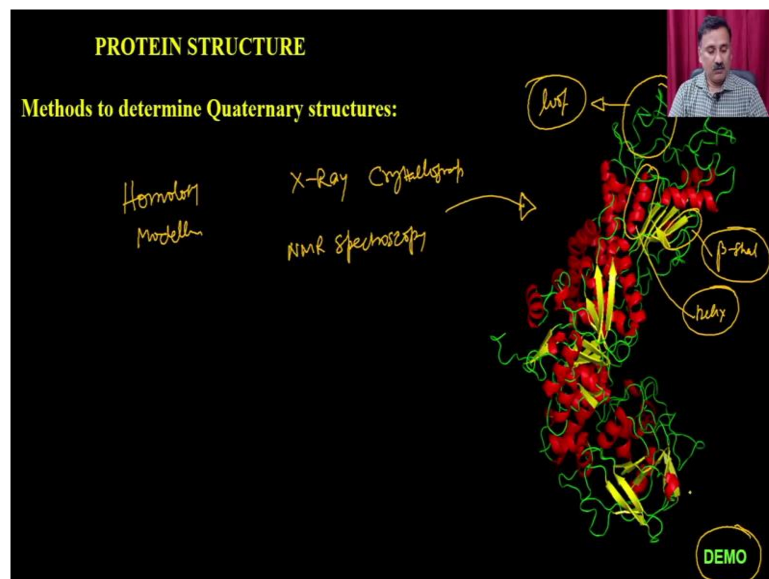
So, if this number so you will get a number right. So, if this number is 1 it is actually going to have. So, if this number is going to be 1 right you will not going to have. So, no

quaternary structure right. Because its only has 1 polypeptide chain. If this number is actually going to be 2 or more right then it is actually going to have the quaternary structure because it is going to have the multiple polypeptides.

So, if so I think you will not be you will you would be more queries curious that how we have actually so I am sure you will not be able to understand it very clearly because so far we have not discussed about the gel filtration chromatography even the SDS page electrophoresis.

But the idea is that you should determine the oligomeric status and then it is actually going to give you the it is going to give you the idea whether the quaternary structure is present or not. If the oligomeric status is 1, then it is actually going to be a monomer right and if it is quaternary structure is 2 or more then it is actually going to be oligomer and oligomers are going to show you the quaternary structure.

(Refer Slide Time: 56:30)



Now, irrespective of whether we use the molecular modelling which means whether you use the homology modelling or whether you use the X-Ray crystallographyc, you are going to or you are going to use the NMR spectroscopy you are going to get a 3D structure of a protein. So, this is a typical 3D structure of a protein where you will see that these are the helix; these are the beta sheets and they are arranged together and what you see is these are the unstructured loops.

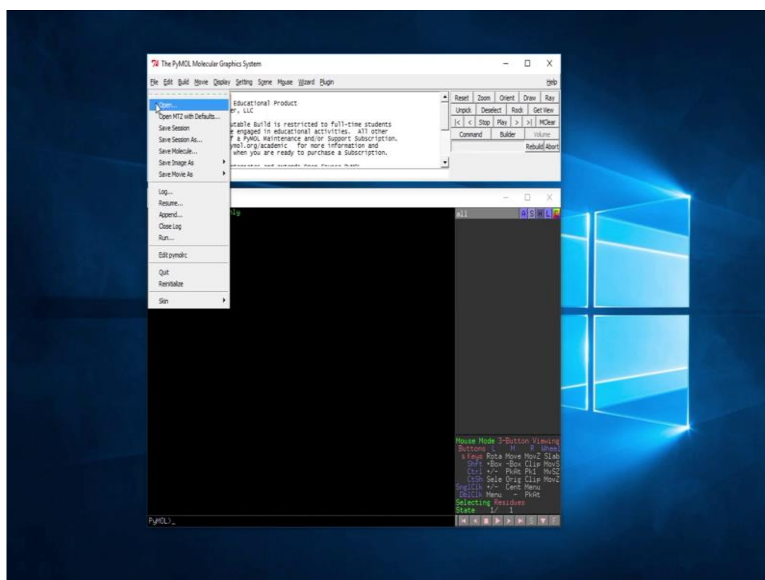
So, these are the loops, these are the beta sheets and these are the helix right. Now, when you have a protein structure it actually requires a very extensive study with the help of the different types of software's. What you can use to determine many properties of this enzyme or many properties of this particular structure right.

So, we have prepared a very small demo clip to show you what are the different properties you can actually be able to study from the protein structures and how you can be able to exploit that information's for developing the substrate or the inhibitors and how you can use that for studying the different types of interactions. So, this demo clip is actually going to be very useful for learning the how to analyze the protein structures.

(Refer Slide Time: 58:11)

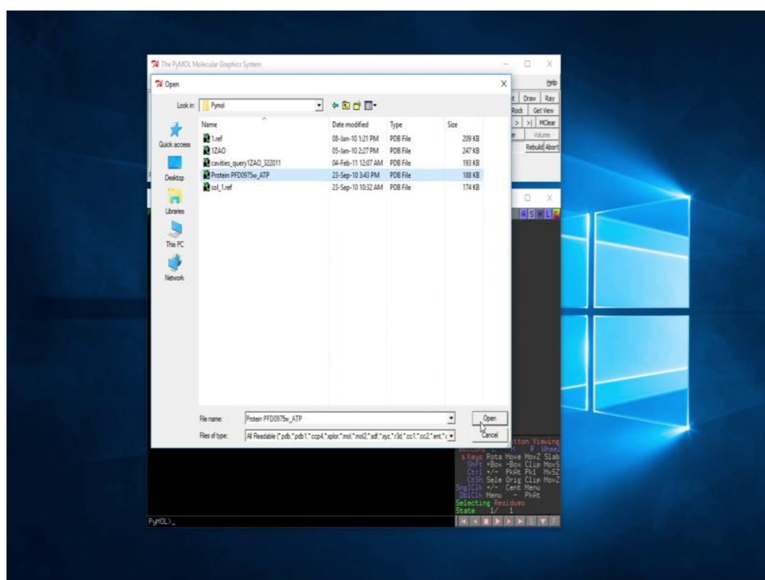
Hello everyone. So, in this particular demo we are going to see how you can be able to analyze the protein structure and so for demo purposes we have used software which is called as PyMOL and PyMOL educational version can be easily be downloaded from the PyMOLs website and then what you have to do is first.

(Refer Slide Time: 58:39)



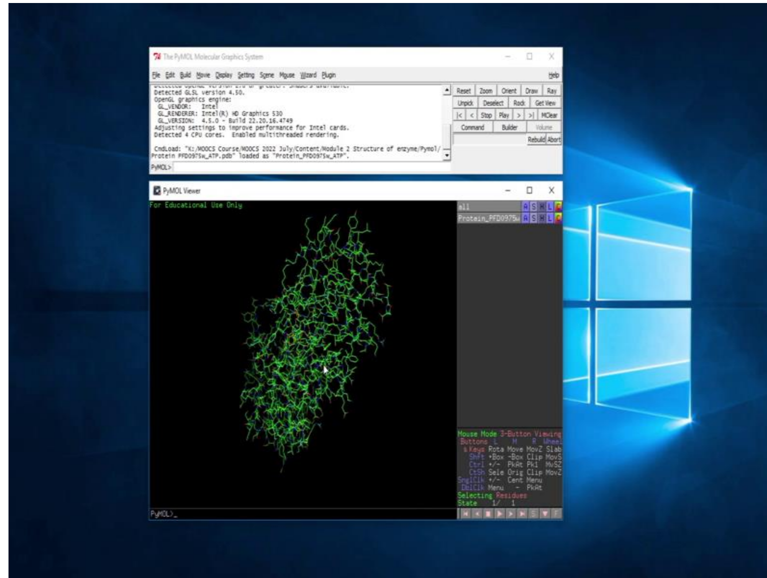
We are you are going to load the molecule. So, for loading the molecule you just first click the file.

(Refer Slide Time: 58:45)



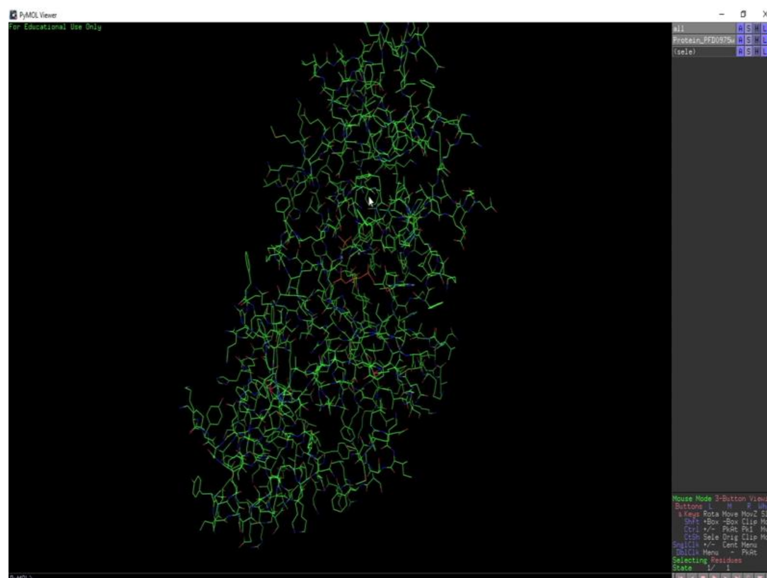
Open and then you go to the respective directory. So, we will go to ok. So, I will load the molecule. So, I am loading a molecule which is called as protein PFD 0975 w it is a malarial protein.

(Refer Slide Time: 59:08)



So, once you load the protein it is going to show you like this.

(Refer Slide Time: 59:13)



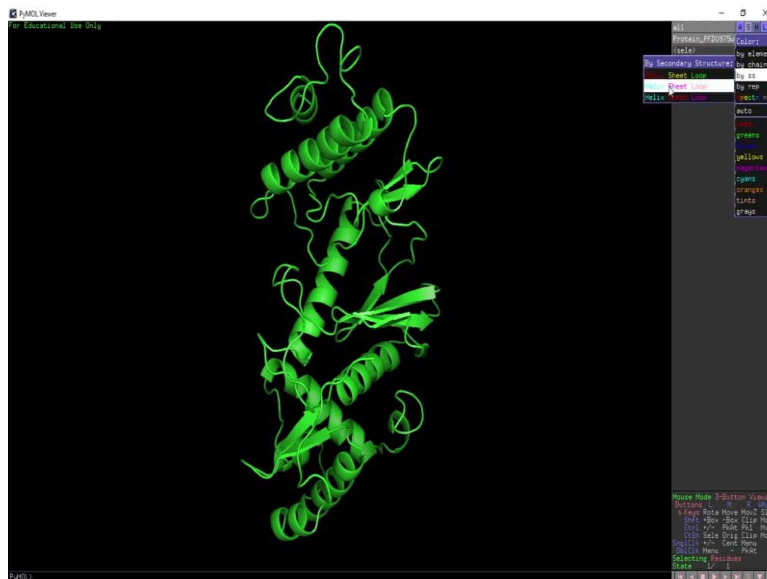
And what you are going to do is so what you see in this protein is that there is a ligand what is actually been present right and now first objective would be that we would like to do the active side analysis ok. So, for doing the active side analysis first I will do is what I will do is I will make it little beautiful.

So, what I am going to do is I am going to convert this into a cartoon model that is the model what you are going what you normally see and then we are going to see the

primary structures, we are going to see the secondary structures, we are going to see the tertiary structures and then ultimately we are I am also going to show you how you can be able to do the analysis for the active site.

So, for making a cartoon model of this particular protein structure what you are going to do is you click the hide.

(Refer Slide Time: 60:09)



So, you hide first the protein and then you are going to put the cartoon ok. So, moment you do that it is actually going to give you the cartoon. Now, if you are interested and you want to make it like secondary structures or you want to make little coloring to this structure so that you it is easy for you to identify what you can do is you go to this color tab right and then you can be able to color everything by secondary structure elements or you can do the molecule color by its choice the color of your choice right.

So, what I am doing is I am putting the by secondary structure. So, I am choosing this right. So, where the helix is going to be colored as cyan sheet is going to be colored as pink and loop is going to be colored as orange ok.

(Refer Slide Time: 60:57)

So, now, this is what you are going to see ok now if I want to study the primary structure. So, if you recall primary structure means the amino acid sequence.

So, if you want to see the amino acid sequence of this particular protein what you can do is you are going to click this s you see there is there are. So, many buttons here. So, if you click the S there will be a window which is going to be pop up on the top and that window is actually going to give you the idea about the amino acid sequence ok and it actually is a very handy tool because it is actually going to help you to select a portion of the protein which you are interested to zoom or something ok.

So, in this if you see the second primary structure what you see is that different region is colored with a different color. So, for example, the beta sheets are colored in a pink right. So, these are this is the beta sheet residues ok. So, these are the residues which are responsible for making the beta sheets. So, you see all these beta sheets are being selected and similarly for helix for example, this is the amino acids which are making the helix. So, this is the helix what they are making actually right and so on.

In some cases you might have to select for example, if you want to use this protein for or you want to use this enzyme for designing the drugs right. So, in that cases you might have to study some region of the protein and then you have to select. So, in that is also very handy right if you just like select for example, I have selected the amino acids from 150 right to 200 for example. So, I will select up to 200 right and then what I will do is I

will I want to color that region in a separate color ok. So, I have selected it right. So, this is two button right one is for protein the other one is select.

So, in the select I will go and I will say ok this is the color I want to choose. So, I will choose like magenta or I will choose like some other color which is not there ok. So, I will use say yellow.

(Refer Slide Time: 63:09)

So, this is the portion you see this portion is got labelled as yellow color. So, that actually is going to help you to identify ok this is the portion which is responsible for catalyzing the reactions or interacting with the substrate and so on.

Now, as far as the so this is all about the primary structures. If you scroll this at the end it will actually going to show you. So, this is actually is showing the pdb. So, it is also going to show you that there are manganese two molecules of manganese which are present in the protein structure then it is having the protein bound phosphate and it also has the ATP at the end ok. So, it also has the protein bound ATP because this protein is a kinase. So, it is actually going to have the active side bound ATP as well.

Now, once we have gone through with the primary structure we can just click this and then it is actually going to disappear the tap and then you can also study the secondary structure. So, for example, this is the helix right. So, what you see here beautiful cyan

colored helix and then these you also have the beta sheets. So, pink color beta sheets what you see and see all the two strands are.

So, this is actually anti parallel beta sheet. So, one sheet is going in this direction the other one is going in this direction. So, this is a anti parallel beta sheets. Whereas, what you see this is also anti parallel beta sheets one sheet is going in this direction the other sheet is going in this direction.

Whereas if you see very carefully the this sheet this sheet or this sheet and this sheet is actually in a parallel mode ok. So, they are arranged in the parallel mode for example, these also are anti parallel beta sheets and what you see right all these secondary structures right the alpha helix or beta sheets are connected by the loop structure and you see how the loops are unstructured right.

And because this is a modelled structure what you see is basically having the lot of unstructured region. Because this is the structure this is the region where the protein does not have the any kind of homology with the existing structure. So, if I want to improve this I probably could have to do ab initio molecular modelling.

So, this is all about the this ok. Now, if I want to do the active side analysis what I will do is I will put first put these substrate right.

(Refer Slide Time: 65:49)

So, I will put the molecule and so I will put the ligand and I will show that as sticks right.

(Refer Slide Time: 65:55)

So, this is I have selected. Now, you can see right this is the ATP which is protein bound ATP. So, this is the adenine ring this is the sugar and this is the phosphate. If I want I can actually be able to color this as per the elements. So, I can color it as per the element. So, in that case it is actually going to show me the phosphate and the you know so wherever it has the negative charge it is actually going to show me the red color and wherever it has the polar groups or positively charged it is actually going to give me the yellow blue color actually.

Now, if I want to study more about this. So, there are many ways in which I can be able to study the active side. For example, if I want to know what are the residues which are interacting with ATP. So, what I will do is I will select the ATP. So, you have to do nothing you have to just click this molecule either you click here or you click in this tab and then I what I will do is

(Refer Slide Time: 66:55)

I will select and I will they route I will do a right click. When I do a right click I will say I want to see what are the different molecules are interacting with this molecule?

So, what I will do is I will go to the action and then I will say I want to see what are the atoms are present within the 4 angstrom or 5 angstrom. So, when I do that.

(Refer Slide Time: 66:24)

It is actually going to select all the atoms of the proteins which are going to be present in the 5 angstrom radius. Because ideally what happen is that when you are actually looking

for the possible interactors you always have to you know select like that and then what I will do is I will go to the label ok.

(Refer Slide Time: 67:51)

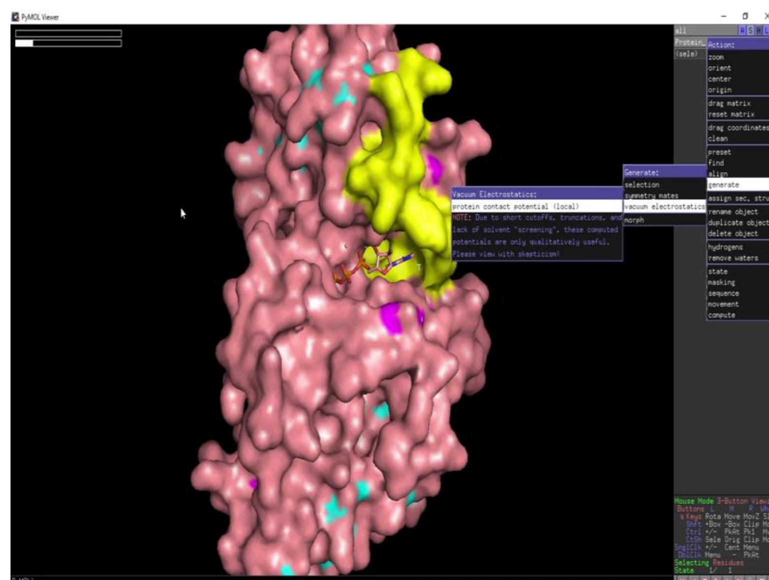
And then I will say I will label the residues. So, what you see here is it has labelled all the residues which are within the 5 of this particular molecule and that is how it is going to give you that ok these are the molecule which are probably be interacting.

Now, ultimately what you can do is you can do distance matrix and you can actually be able to open this molecule in other software's to measure the distances. Apart from this you can also be able to check.

(Refer Slide Time: 68:20)

The you can also be able to check the you can also be able to make the cavities ok. So, you can also study how the grooves are present in this particular protein. So, for that what you can do is you can just go to here and you say I want to see the surface ok.

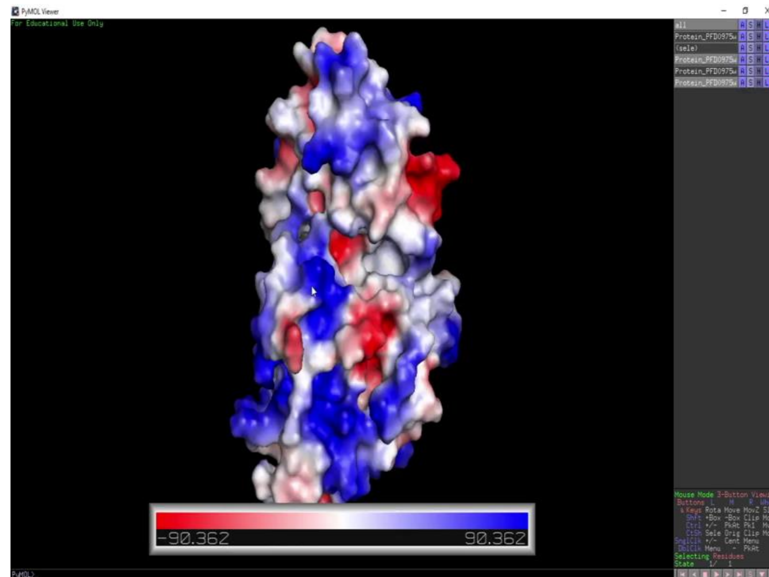
(Refer Slide Time: 68:38)



So, what you are going to do if you want to see the charge distribution on the protein structure is you are going to see this you are going to go to this action button ok. So, click the action button and then you say I want to generate the electrostatic charge distribution under the vacuum ok and then you select this vacuum and then you are going

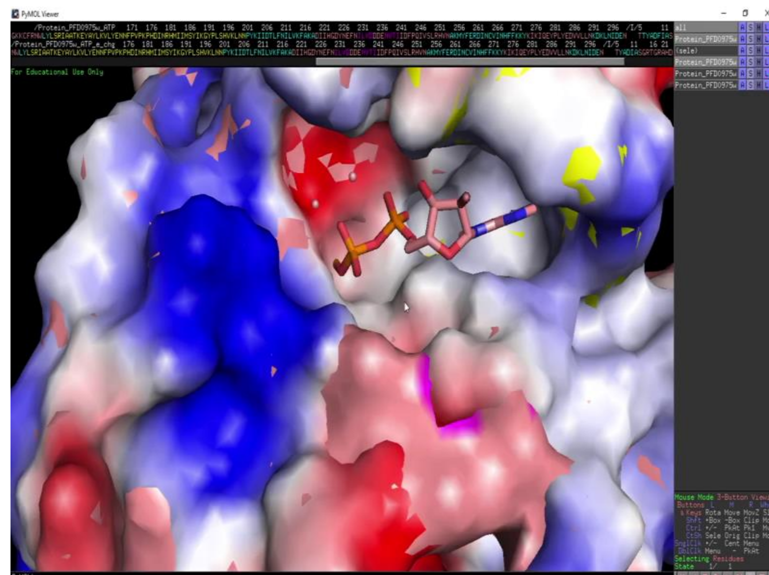
to select this. So, it is actually going to calculate and very soon it is actually going to show you the color.

(Refer Slide Time: 69:03)



So, this is the you know charge distribution what you see and wherever you have you would see the red color that is the actually the negative side and wherever you see the blue it is actually going to be positive side ok. Now, if I will show you again the molecule right. So, if I show you the molecule again right. So, this is the ATP sitting here right.

(Refer Slide Time: 69:45)



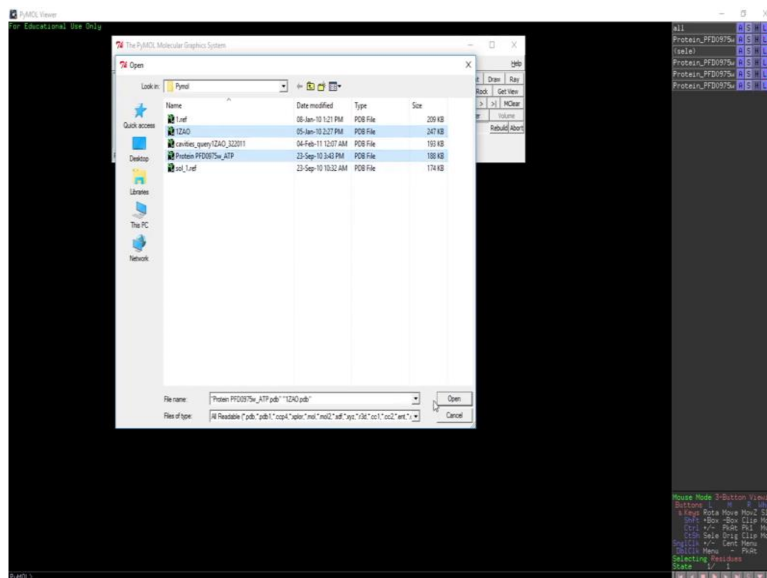
And if I show you that ATP again you will see that how nicely the ATP is being placed and ok. So, this is what you see right. So, ATP is placed and wherever you have the blue actually that is the positive and wherever you have the red that is the negative and wherever you see this white that is either the neutral or the hydrophobic regions of the proteins.

Now, this is all about what you can actually be able to analyze on the protein structure one more thing which you can also do is you can actually be able to you know superimpose the two structures ok.

(Refer Slide Time: 70:29)

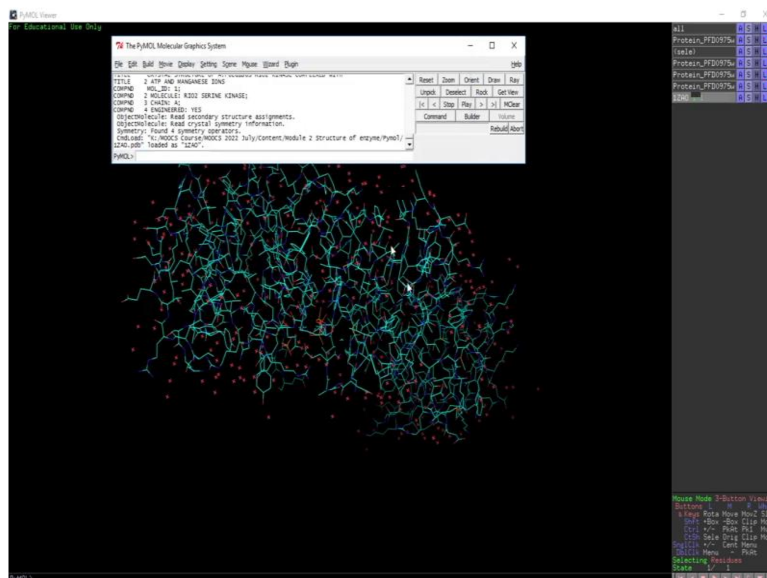
So, how to do that? What first thing what you have to do is you have to load the molecules.

(Refer Slide Time: 70:32)



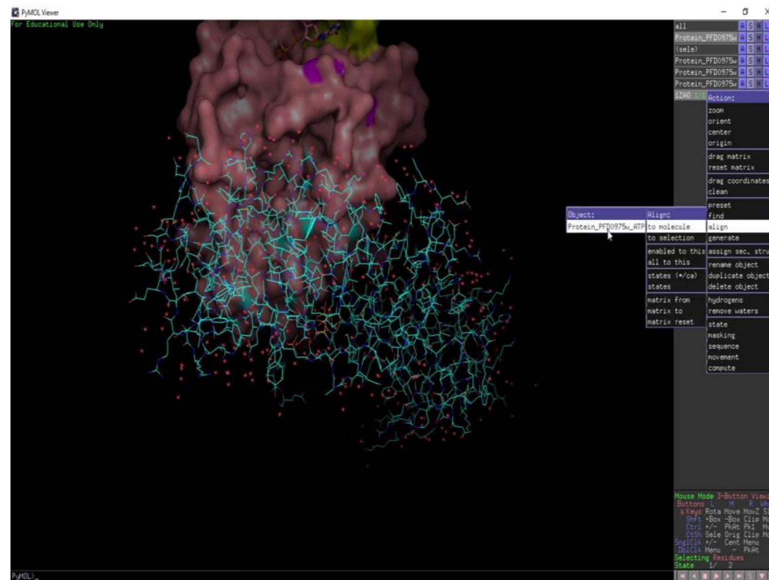
So, for example, I have loaded two molecules like one ZAO and this molecule right.

(Refer Slide Time: 70:41)



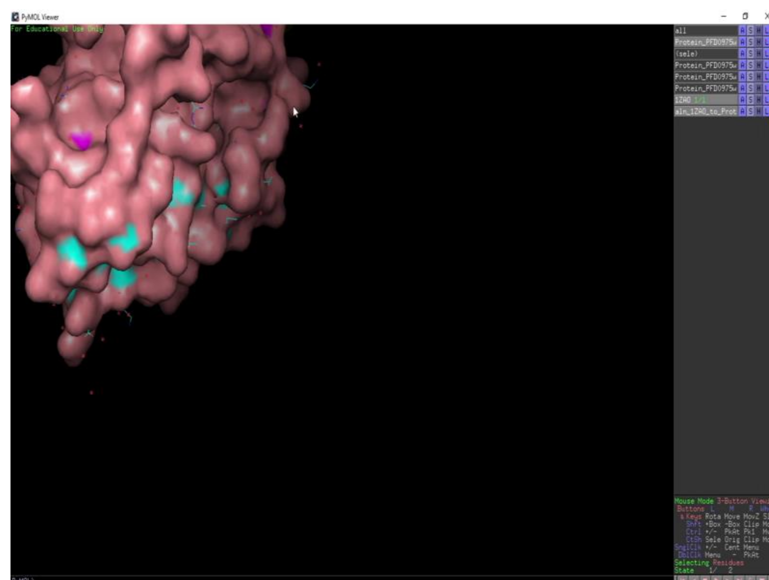
So, if I load the one ZAO and this molecule I am going to see the two structures right 1 ZAO and this one right. So, this is the other one right.

(Refer Slide Time: 70:51)



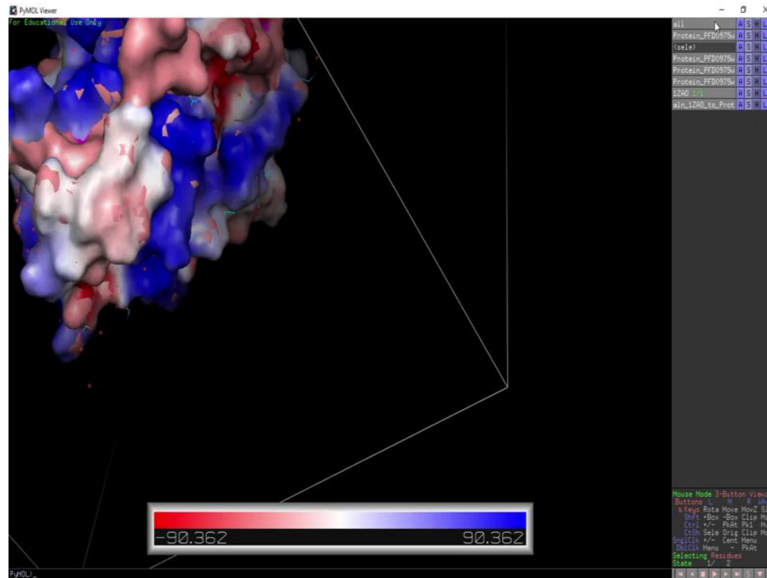
And so, what I can do is, I can just do the alignment. So, I can do is I will click the action button then I will do the alignment and then I will say two molecule and then you select ok.

(Refer Slide Time: 71:05)



The moment you do that ok it is actually going to align on to the other molecule to make it very simple.

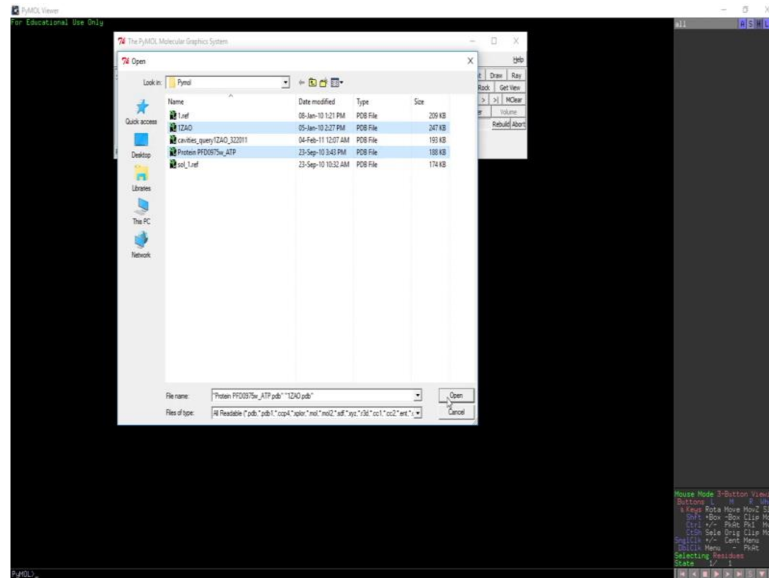
(Refer Slide Time: 71:15)



Let me just remove all these molecules. So, that you know so that it will be easy actually.

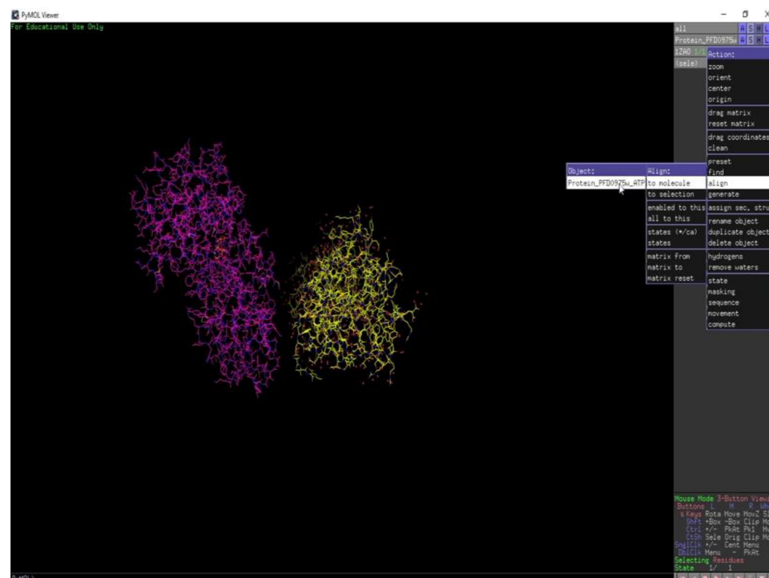
(Refer Slide Time: 71:25)

(Refer Slide Time: 71:28)



So, if I open ZO 1 ZAO and protein right.

(Refer Slide Time: 71:34)



You are going to see that ok. These are the two protein structures right. So, one is the 1 ZAO the other one is the PFD 0975w and if I want to superimpose this to this what I will do is I will go here I will go to the align, I will go to the two molecule and then I will select the other one ok and it is actually going to show me the alignment ok.

(Refer Slide Time: 71:52)

And what is the advantage of making an alignment and the advantage of making an alignment is that it is actually going to show you the contrasting features. So, this is all about the demo on which so that you can be able to do the different types of analysis on to the protein structure thank you.

So, I hope you might have liked the demo clip and you will understand and understand the potential of the protein structures and how the protein structure is actually very very complicated and it requires a very detailed study to understand the function of each and every portion of the proteins.

So, in that demo clip we have shown you how to determine the charges this charge distribution on the protein, how you can be able to you know explore the different cavities, what are present in the protein structures and so on.

(Refer Slide Time: 72:26)

So, what we have discussed in this particular lecture, we have discussed about the primary structures, we have discussed about the secondary structures, we have discussed about the tertiary structures and we have also discussed about the quaternary structure.

While we were discussing about the primary structures we have discussed about how you can be able to determine the primary structures and what are the different types of amino acids are present, amino acids are actually having a general structures where you have the amino groups, carboxyl groups attached to a central carbon which is called as C alpha and then it is attached to the different types of R groups.

And depending on the R groups you can be having the amino acids of different types; whether it is the polar amino acids, non polar amino acids, hydrophobic amino acids charged amino acids like the negatively charged amino acids, positively charged amino acid and so on. And then we have also discussed about how you can be able to determine the primary structures with the help of the different types of protein sequencing methods.

Now, once you have determined the primary structures you can actually be able to understand the secondary structure. So, because these primary structures are only going to direct how the protein is actually going to fold and that is how it is actually going to give you the secondary structures.

So, within the secondary structure we have discussed about the alpha helices, beta sheets and we have also discussed about the loops right I already suggested you when we were discussing about the secondary structure that you should actually go through with some of the standard biochemistry books like the Lehninger white and white or Stryer.

So, this is all about the different details of the protein structures and how you can be able to use the different techniques to determine the protein structure. So, with this I would like to conclude my lecture here.

Thank you.