**AI in Drug Discovery and Development**

**Prof. Rajnish Kumar**
**Dept. of Pharmaceutical Engineering and Technology**
**IIT-(BHU), Varanasi**

**Week-12**
**Lecture-59**

Welcome to the course "AI in Drug Discovery and Development." So, in earlier sessions, we have seen how we can, you know, use RDKit to play with the molecular structures; we can import different types of structures, different types of files like SDF or, you know, CSV. Then import molecules, calculate their fingerprint properties, and visualize them. And then in another session, we saw how we can use, you know, regression modeling to predict the solubility. And then in another session, we saw how we can use classification modeling techniques for predicting the bioactivity. So, in this session, we will talk about pharmacophore modeling based on screening.

So, the pharmacophore is, as we discussed in the theory sessions, a three-dimensional spatial representation or arrangement of features that are responsible for bioactivity. So, these features are, you know, the functional groups, for example, and then those functional groups are converted into features. Like hydrogen bond acceptors, hydrogen bond donors, aromatic regions, and hydrophobic features, you know, all those things we saw. So, in this session, what we will do is use an online server, the PHARMIT server, and then I will show you how we can screen a large number of compounds and large libraries to identify compounds that are similar to our input compounds.

We will take an example of a drug, convert it into a pharmacophore, and then use that pharmacophore to screen a large number of compounds. So, I am going to show you that in today's hands-on session. So for that, what we have to do is visit this link, which is pharmit.csb.pitt.edu. Once you open this link, this will be the landing page, and you will end up here. Then you can see that this is a Pharmit server interactive exploration of chemical space. The best way is to create an account so you can register for a new account, and after creating the new account, you can log in to that account. And then, because if you create your own account, you have multiple possibilities: you can store your libraries, you can create your own database, and you can upload some databases as well.

But for simplicity, we will just quickly go through without registering; here, at least, there are multiple ways. Where you can use this web server for searching pharmacophores for pharmacophore-based screening, actually. So, the simplest way to start is with a PDB. So, you just enter a PDB ID where you know the co-crystal structure of a ligand that is crystallized with your protein of interest. So, for this, we will be using 4EY7.

So, for this, we will be using 4EY7, which we use for docking as well. So, this is the crystal structure of donepezil bound to the acetylcholinesterase inhibitor, and donepezil is a drug that is used for the treatment of Alzheimer's disease. So, and here you can see that these are you know the ligands you can choose you can pick which ligand you wanted to use for making the pharmacophore. So, you have the EDO, NAG, NO3, and then E20. E20 is the donepezil molecule, which we will use to make the pharmacophore model, okay? So, after selecting this, we submit it.

What you will see here is that the protein structure will get loaded, okay? So, the protein structure will be loaded. You can see that this PDBID has two chains, actually. So, both of these chains are similar to each other, and what we can do here is quickly visualize. So, we can quickly remove it, or we can just hide the protein structure. We can do that by reducing the opacity of the protein structure.

So, if we reduce the opacity, it will be gone. And then here you can say, for example, I wanted to see my protein, but you know, you just click on none, actually, so that you will not see the proteins. So there are multiple ways to render your protein structures. You can render it as a stick, a wire, a sphere, a cartoon, and then a cartoon plus wire, a binding site, and none. So you can just say, for example, the binding site, or, you know, I just click here; none.

So now what you see here is that you have your, you know, molecule that is donepezil, and then it has automatically identified these pharmacophoric features. So represented by this spherical mesh, actually, okay, so then we go back to the top. Starting from here, you can see that it has identified one aromatic. So if you click on this, for example, you can see this got highlighted. So this is a hydrogen bond acceptor, right? And then this is an aromatic compound.

And then the third one is the hydrogen bond acceptor. So it has identified these three, and then it had just turned on these three; however, there are, you know, more pharmacophoric features. So you can see, for example, if you wanted to turn this on, you have another aromatic, which is this benzene ring, the terminal benzene ring. And then you have a hydrogen bond acceptor, so you have this methoxy as a hydrogen bond acceptor; you have another methoxy, another hydrogen bond acceptor. And then you have this, so then you have another hydrophobic feature.

These are, you know, the hydrophobic and aromatic; these are overlapping, actually. So if you see, for example, this is the hydrophobic feature; this is the benzene ring, which is earlier recognized as aromatic. It is also recognized as hydrophobic as well, okay? And

then you have, you know, similarly. So, you have the terminal methoxy; the methyl of this methoxy group is hydrophobic, and then you have another hydrophobic group. So, these are, you know, different features that are calculated, different pharmacophoric features that are calculated from this co-crystallized molecule, which is donepezil, right? So, what we are going to do here is that, when you load this structure, it automatically keeps only the pharmacophoric features that are relevant for this molecule.

Based on the distance of these features from the binding pocket, this is co-crystallizing the binding pocket. So, it takes into account the binding pocket of the receptor as well. So, we have only selected the features that are very relevant, actually. So, we will keep only those; for example, we will just turn off all those, right? And then, as I said, it is a three-dimensional spatial arrangement. So what you see here are the XYZ coordinates, right? So you see, the XYZ coordinates of all these features mean that now you can just draw lines, actually, so you can connect this feature with this, and then this feature with this, and this feature with this.

So you will see that what you will see is just the three-dimensional arrangement of these features, which are responsible for activity. So, by any means, if you have some information, like from the structure-activity relationship or from, you know, the interaction with the binding pocket. Atoms and functional groups like these are essential for activity, as they make contact with the protein structure and protein binding pocket residues. So, you have to keep them to improve the chances of actually getting a better molecule. So, for simplicity, we will just keep these three features.

So, we just keep these three features, right? So, now we have a pharmacophore; the program has converted this molecule into the pharmacophoric features, and now we have these pharmacophoric features. So, now the next step is what we can do: we can screen the libraries using these features, and then here on the top you see. You have multiple libraries, actually. So, you have ChemBL 34, Chemdiv, Chemspace, Enamin, Mcule, Mcule Ultimate, Molport, the NCI Open Chemical Repository, PubChem, Lab Network, and Zinc. So, you can see that they actually have millions of molecules.

And they have, for example, in this one you have around 21 million conformers of about 1 million molecules. Okay, so these can be searched; you can search any of this library. And then you have contributed libraries as well; if you click on contributed libraries, you can see there are hundreds of libraries that people like you and me have contributed to this server. So if you create your own account, you can also contribute libraries to this server, where these libraries can be dedicated libraries, such as SARS-CoV-2 specific libraries or other specific target libraries. So you can use them as well for screening.

Okay, so that is for selecting the library, right? So you just select the library. Here, for example, we will just select the Chemdiv. So, you can select the Chemdiv, and then the next thing here is the pharmacophore search. You can see it here. So, it has both the possibilities of pharmacophore search and the shape filter.

So, the pharmacophore search is like, you know, these functional groups that are responsible for activity based on that, and then the shape filter is, you know, how we can. Can we, because of course this molecule is binding to the pocket, and then this pocket must have some shape actually, So it might be that the binding pocket has a slightly bigger size compared to this ligand, so that there is a possibility that this ligand can grow into some more, you know, some more space. Okay, so in that case, we can use the shape search as well. So, here you can decide whether you are going to do the pharmacophore search first and then the shape filter, or if you want to do the shape filter first and the pharmacophore search later. So, ideally, as suggested by them, we shall go for the pharmacophore search first and then we will do the shape filter.

So, let us see what "shape" means here. You can see here. They have both; you can include as well, and you can exclude as well. Like if you wanted to say, for example, I just want, you know, a benzene ring or at least some shape here in this part. So then it will include that in the inclusion of the shape.

And then, if you wanted to say that, OK, I don't want any part of the molecule to be here in this part. So then you can add an exclusion filter so that part will be empty. And in that part of the pocket, no molecule, no molecule will be, no atom will actually be there. Right. If you see the shape filter, you have none here, and then you can see, for example, a ligand.

So, when you click on the ligand, you can see that this ligand has been converted into this shape. So, now, if you use this shape, then the algorithm will search for molecules that are similar to this shape. And as we know, shape-based screening has, you know, one of the advantages that we can quickly identify similar molecules, like the molecules that are similar to the input query. So, they can be identified, and the hypothesis behind this is that if the molecule has a similar shape, then it has a high possibility of binding or showing activity towards that target because this is like a lock and key, actually. If a key has a similar shape to another key, then that key can also open that lock very easily because it can fit into that pocket nicely.

Right, so this is if you wanted to use the shape of the ligand itself, or you can just add some spheres as well. Like, for example, you can add a sphere here, as I wanted to have some atoms in this part, or you can add this one as well; I wanted to have part of the molecule in this area as well, okay? So, that is, you know, the inclusion, but for simplicity,

we will not use "ok." And then you can similarly have the exclusion criteria as well; you can say that, based on the receptor and the residues of the receptors, we are considering the binding pockets. So, you can identify. So, now you can see that this has, you know, converted the receptor into shape, okay.

So, you can now see, for example, if you just zoom it a little bit. So, you can see that there is a lot of, you know, pocket actually which can be explored beside the co-crystallized ligand, right? So, you can just say, for example, "sphere," where you can add the information about, "Okay, I just wanted to have, I just do not want to have anything here, here, here around that ligand." So, this one, also for simplicity, we do not do that. So, that was, you know, a shape filter. So, if we choose this first, it will perform the pharmacophore search based on the features we have, and then it will, you know, apply the shape filter.

But since we did not put in any shape filter, in this case we will just go for the pharmacophore search. Yeah, so then after the shape, the next thing is the filters. For example, we can reduce the filters as well. So we can just, you know, fix the maximum hits per confirmation, the maximum hits per molecule, or the maximum total number of hits. So we can say, for example, the maximum total number of hits we wanted to have is 100; actually, that will restrict only up to 100 molecules.

And then you can have, you know, hit screening as well. So, this is again a very useful filter where you can filter your molecule based on these parameters. So, these are just like Lipinski's rule of five filters or physical chemical properties filters, and these are sometimes very useful. Because we know that, for example, for CNS permeability, the logP should be between 1 and 3.5, and the polar surface area should be less than 90 angstroms.

So, for all those features, we can just put some restrictions, like we can put a filter for molecular weight; we want to have molecules with a minimum of 250 molecular weight and a maximum of 450. And then we can put restrictions on the rotatable bones, like we can set it from, you know, 1 to, you know, 10, and then we can set log P from 1 to, you know, 4, and then for PSA we can set it from 20 to 90. And then for a number of aromatic groups or aromatic rings, we can put, for example, 1 to 5. And then for hydrogen bond acceptors, we can put like 1 to 10, and then for donors, we can put like 1 to 5, right? So, now once we have put in this detail, what it will do is whenever we press this button, search ChemDiv. So what it will do is filter; it will first use the pharmacophore filter, and then it will filter those molecules based on these parameters.

So, we will only see the molecules that fit this criteria, such as the molecular weight, rotatable bonds, log P, PSA, aromaticity, and all those things. Okay, so that can be done, right? And then in the end, there is, you know, just the visualization. Like, you can have

black, or you can have white, or you just wanted to turn on the auto zoom function. And then you can always save the session by clicking on "Save Session," so all the things we have done will be saved, and you can reuse them later. And then, for loading the session, once you have saved it, you can load the session and reuse the settings, which we have done, and then you can open the project.

So now we have fixed all these things, let us screen the library. We go to the top, and then you can just click on the search ChemDiv. So once you click on the search cam div, on the right-hand side, it will show you that it is searching around 1 million molecules and 21 million conformers. And then here you will see the RMSD (root mean square deviation), the mass, and a rotatable bond; so these properties you will see here. And then it will show us the compounds, so it will take some time, and then the compounds will start popping up here in this table, actually.

Right, so meanwhile it is working. Just give you some other details. So here you can see. For example, you have the possibility to load the receptor and load the features. If you remember, we directly started. Okay, so now we can see that we have got compounds here.

So you can see that we have this compound ChemD134, the RMSD is 0.023, the mass is 367, and the number of rotatable bonds is 9. So, now that you can see what you can see here is that we used these physicochemical filters. So, it has only given us the compounds which are satisfying that criteria otherwise if we are not using that filter. So, it would have given us, for example, molecules that are much, much bigger and much, much larger than this compound.

OK, so it will take some time. It is still searching. Once it is done, we will see options here, for example, minimize and then save. So let it finish first and then come back to the load receptor. So what I was saying is that when we started this, we started directly from the PDBID. But in many cases, you do not have a PDB structure actually co-crystallized with a ligand.

So, you can start with the docked ligands as well. In that case, you can use this load receptor function as well. And then the other beauty of this is that you can even load the features from other tools as well. It is not necessary for you to use this for making a pharmacophore model; you can even upload the pharmacophoric features or pharmacophore models from other tools like MOE, LigBuilder, LigandScout, Pharmagist, and Pharmer. So from any other tool, you can quickly make the models and then pharmacophore models, and you can use those pharmacophore models to screen the libraries here in this tool. Okay, so now this is done, right? So now this is done, and you can see here we have the minimize and save options.

Since we asked it to generate only 100 hits... So you can see here that it has generated 1 to 7; it is showing 1 to 17 of 100 hits, and we can download them. We can save them if we click on save. So, if you click on save, then it will get downloaded, okay. So, you can see the query result.

sdf, and it is downloaded in the SDF file. And now, since you have downloaded these 100 pharmacophore molecules that are similar to your input query. So, then you can further, you know, go for, you know, structure-based screening; like, you can do them, and then you can rank them. This tool has already provided that facility. So, and that is here in this minimize, actually. So when you click on minimize, right? So when you click on minimize, what it will be doing is, for example, yeah.

So what this tool is doing is using the autodock vina scoring function, and it is scoring those molecules into the binding pocket. So, if you remember, we just hit the receptor structure as it is there, right? So, you can see, for example, we have that structure here; we have that receptor structure here, right? So, this is what it looks like. So, now what this minimize is doing is when we click on the minimize what it does is. It calculates the docking score of these molecules, which we have identified in the binding pocket, and for that, you always need a receptor structure. So, if you do not have a receptor structure, we cannot, you know, use this function, and that is a little bit of a prerequisite for it.

So, if we have a receptor structure, we can use this function, and we can directly, you know, calculate the score, which represents the AutoDock Vina score, indicating how strongly it binds to the receptor. Okay, and since we had only 100 molecules, we limited it to 100. So what you can do is you can just put, you know, the maximum score; for example, you can put 11. And then, if you apply, so now what you can see is that you can just have this, and then you can even save this as well. So, where can you save the docking score of these molecules along with the RMSD values? This is kind of a minimization; it's not like.

.. It is not exploring all the possible confirmations; it is just minimizing the confirmation that it has identified as similar to the pharmacophoric query. So, it is just minimizing that and calculating the docking score. So, it is, this docking score is, we cannot say docking score actually because it did not generate all those conformations, all those poses actually. But it has just calculated the score using the Vina scoring function. So, that can be one of the filtration criteria where you can filter the molecules using this technique, and then you can further use these molecules.

So, now we have what we have done here is we have used, you know, a co-crystal structure of a ligand, and then based on that, we screened a library of ligands for similar compounds.

and we also use some filter criteria based on physicochemical properties like molecular weight, log P, polar surface area, rotatable bonds, etc. Okay, so that was, you know, like that. Another thing I was saying is, for example, even if you don't have a crystal structure, if you don't want to start with a crystal structure, even if you have just a ligand, actually. So even that you can use, and for that you can just click on the Enter Pharmit search, okay? Enter Pharmat search.

And then you can just, you know, load the features, so that is all where I have shown you how you can use, you know, this nice server to screen molecules for your, you know, for your target. Even if you don't have a target, you can start with the ligand as well, and then you can search these large libraries. You can see that many of those libraries have millions of compounds, like this zinc library, which has around 13 million compounds; some 13 million molecules have been uploaded to the server. And then you can screen them, and then you can identify those molecules, but there is no alternative to wet lab testing. So, in the end, what you have to do is always validate your compounds to see if they are giving the results you expect from them.

So, once you identify those compounds, you have to procure them, and many times, most of the time, those are purchasable. So, you can just simulate; you can quickly purchase them, and then you have to test them in your assay. And then you have to make sure that, okay, you can only say that, okay, they work or they do not work.

Okay. So, I think that is all. I hope you enjoyed this session. So, it will be highly useful for some of you, as you can use it to screen large libraries by using pharmacophore modeling. And with that, thank you.