PHARMACOGNOSY AND PHYTOCHEMISTRY

Dr. Galvina Pereira

Department of Pharmaceutical Sciences and Technology

Institute of Chemical Technology, Mumbai

Week 12

Lecture 59

Week 12: Lecture 59: Databases for Studying Spectral Properties

Thank you. Hello everyone, and welcome to session 4 of week 12 of the NPTEL course in pharmacognosy and phytochemistry. This week, we are learning a slightly different topic from what we usually study in our syllabus. These include using computational-based techniques, computational databases, as well as certain servers that help us understand the plants, their applications, and their physical and chemical properties. So far in this session,

you have learned about different databases that can help you identify plants. These include Plants of the World Online, Gestor, Tropicos, and the New York Botanical Garden. There are many more paid and unpaid software options, but since these are freeware and well-documented, we are very familiar with them right now. In the next session, we saw the different software, databases, or web-based servers that help us understand traditional medicinal preparations. We saw TKDL, the Traditional Knowledge Digital Library, which contains all the traditional formulations with their ingredients, procedures, and preparations.

We saw IMPPAT, where even the traditional medicinal plants, their geography, and phytochemical searches were obtained. We also saw the Indian Medicinal Plant Database, which provided more details about the plants, their occurrence, their information, and various applications, including the shlokas and Ayurvedic texts that mentioned them. Thereafter, in our previous session, if you recall, we discussed phytochemical searches. Through these, once you know the plant—whether it's a herbal medicine or a traditional medicine—you can easily search for all the compounds present using these databases. The few databases we studied included Dr. Duke's, Knapsack, Phytoherb, Coconut, IMPPAT,

as well as the Osteo database. Now, in today's session, we are going to take a step forward. When you purify these compounds from your plant sources, what happens is you have a compound at hand.

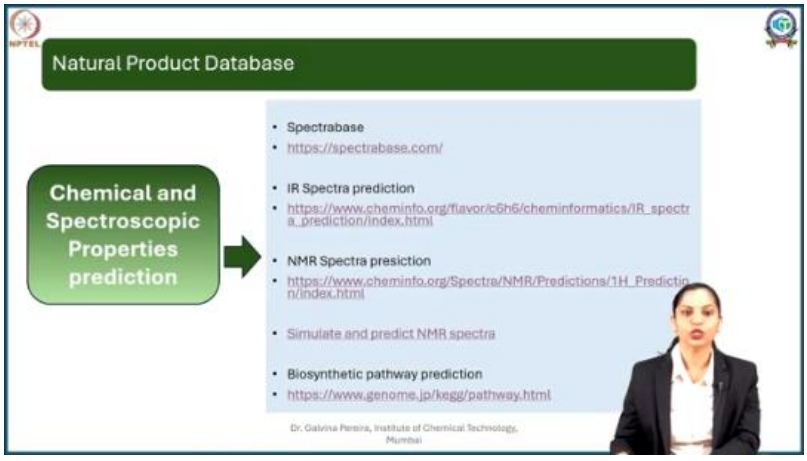
And then the question is, I have to analyze it to ascertain its purity. So far, using the quality control methods, we saw that definitely my HPLC and my GC can help me ascertain the purity. But for identity, that is for the establishment of identity, you need to resort to not only chromatographic but also spectroscopic methods. Now, many times your LC, that is HPLC, or GC, that is your gas chromatography, is often coupled with mass spectroscopy. So your HPLC as well as GC is coupled with a mass spectroscopic detector, which gives us MS-MS fragments, and thanks to the advances in machine learning techniques, we now have a library search.

Wherein these mass fragments are analyzed by computers, and the computers come up with the possibility that probably this is the plant you have and these are the compounds that are present with it. Now, given a case where due to some incidents you purified a compound and maybe you want to see whether it matches or not, or you have an intention to purify one compound. Say, for example, we have been dealing with our search for quercetin. So let's put a quercetin example here. So tomorrow I say from all the sources that my previous software gave me, I tried to purify quercetin, and to a certain extent, I have a molecule in hand.

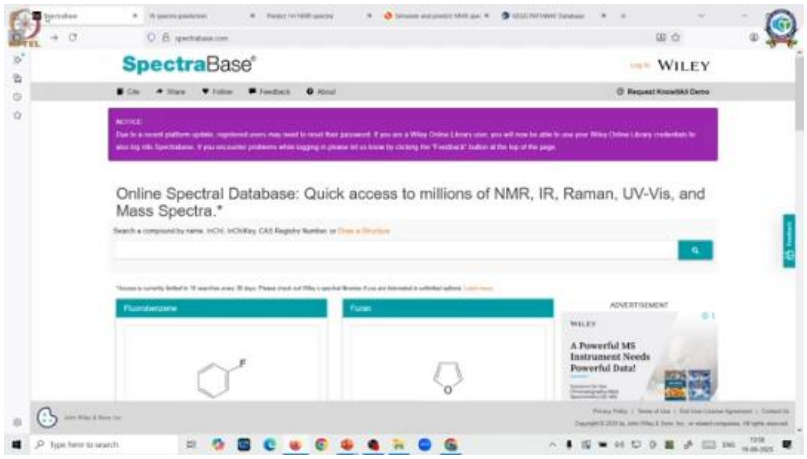
So now what I have is a yellow-colored powder. Now I give it or subject it to my mass spectroscopic technique. Now, many times we as students or academicians even struggle with spectroscopic data. So in order to facilitate this, there are also databases which help us by giving us spectral information. So even if we don't know the spectral data, you can put in the structure of the molecules in these spectroscopic databases, and what you will find is because you put your structure in these databases, now you are getting well-resolved data, and the peaks also have meaning.

So let's see what all databases there are. Few databases that we are going to study today include here your spectra base, IR spectra prediction, you also have your NMR spectra prediction, then you have your biosynthetic pathway. Now biosynthetic pathway prediction

is not clearly prediction, but I wanted to show you or I brought in this also just to tell you that it is also possible for me to put my metabolite in a particular pathway and trace back where it came in from.



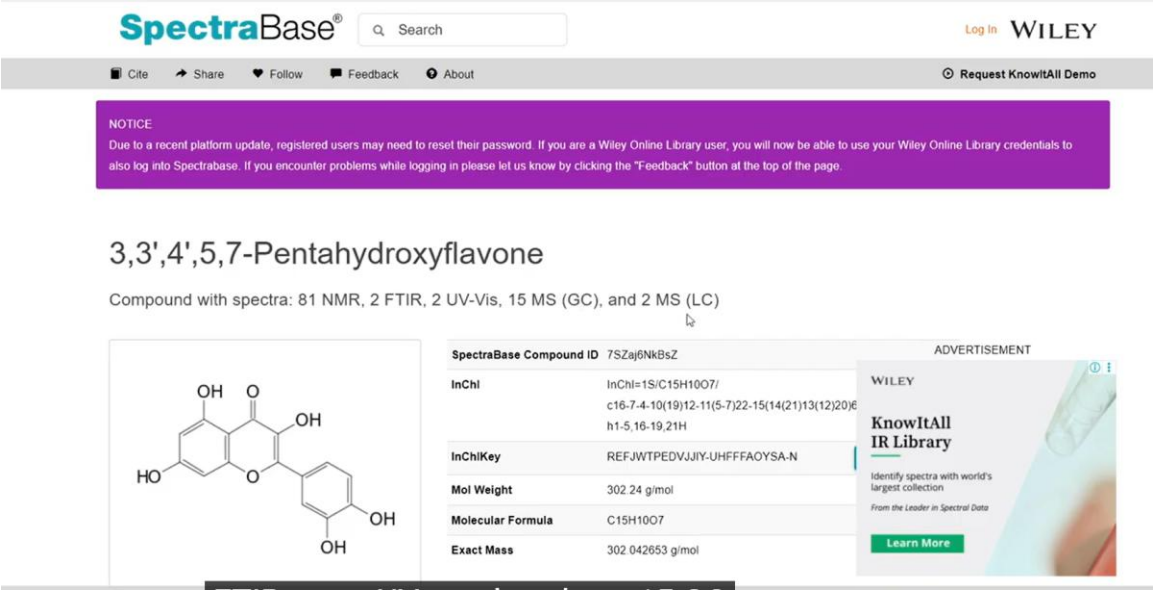
So initially we will see different structural properties but at the end of this session we will also see that there are softwares which will tell me from where or which pathway is my compound biosynthesized. So let's study this pathways one by one. So we go to the first pathway and our first pathway is a spectra-based pathway. Now this spectra-based pathway gives you a quick access to numerous molecules. This is a Wiley-based search engine which has a stored data of numerous compounds and it not only gives you an idea about it but also gives you some references pertaining to where this data is taken from.



So you can get your NMR, IR, Raman, UV-Vis or in certain cases even mass spectra. But since this is a server base for compounds whose spectroscopic data is not available, you might not find the query coming back with all the answers. So in some cases, a complete

database that is NMR, mass, IR may be ready. In some cases, you might just get up one or two. So let's quickly try to search our same molecule and that is quercetin.

So when I put my query as quercetin, the spectra base is suggesting a molecule. It is also trying to give me analogs. So these are all the compounds that it has come up with. So not only you will get quercetin, you will get all over searches or hits wherever this word is quercetin. Even if it is a sub part of the name or the nomenclature.

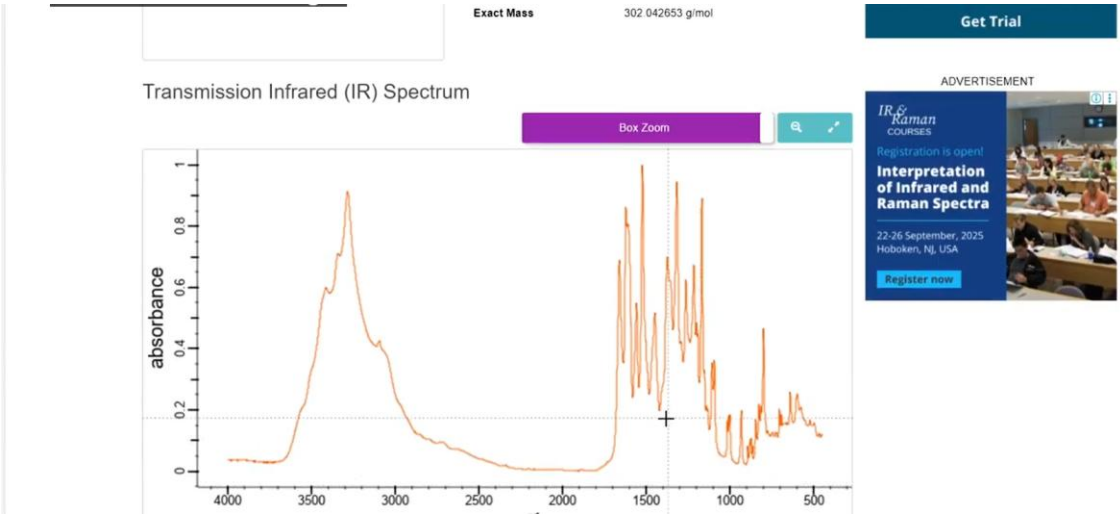


So let's see our molecule quercetin. So, this is quercetin dihydrate and this is your quercetin. I am just clicking on this. So, you can see here it has a good databases. It has about 81 NMRs, 2 FTIRs, 2 UVS graphs, about 15 GC searches and 2 MS searches which you can access through this.

So, this is the IDs. The molecular weight is about 302 which is going to View full spectrum for free. Click on it. OK, it will ask you to log in maybe.

So after entering your registration or after registering, as you can see, you are able to access the spectral data. You can zoom in. You can zoom out. These are mostly your hydroxy protons out here. So now when we go back, let's try to see if you are able to visualize the C13 NMR as well.

So just click on it, and you can see there is some information about it. Let's view the entire spectrum, and after logging in again, you will be able to see the entire NMR spectrum of quercetin. This is the C13 NMR spectrum. Let's move to the IR. Let's click on the IR.

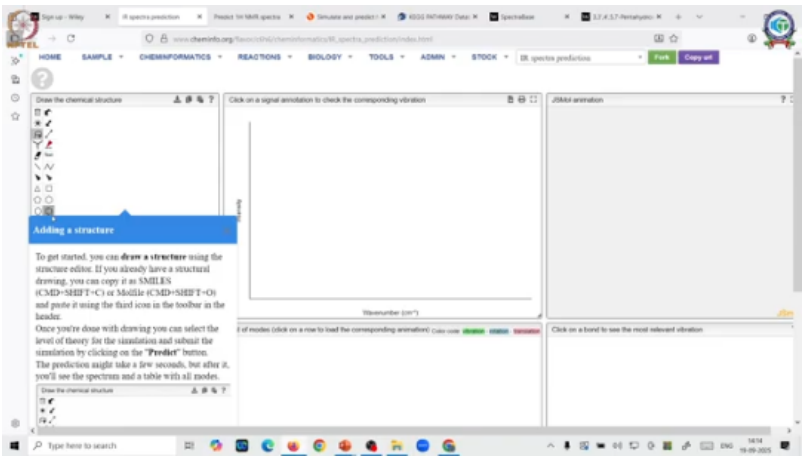


You can see an inverted IR spectrum of this with the fingerprint region as well as a very sharp hydroxyl region for your quercetin. So the point is, through this one software, when you log in, you'll be able to get most of the data associated with it. So you have your NMR data, and you have your IR data. You can also see the stored or saved data. Since this is a server, it is definitely collecting it.

You can see the UV vis. You know it's a typical flavonoid, so you should see two major peaks. One represents your band 1 if you recollect, and this one represents your band 2. So these are the saved peaks. And it also gives the solvent in which they have obtained this.

So all in all, based on this data, it will generate values for you. It will generate a graph for you. So you will get whatever readings you want in terms of that. The mass spectra show its mass as 302 if you recollect. So my point is, if you are not conversant in spectroscopic techniques, this database will give you some reference values to compare with. Based on those reference values, when you compare your spectra, you will know if the peaks are overlapping or not, and based on that, you can determine whether the compound you have isolated is actually quercetin or not.

Similarly, there are other databases or other interesting databases for IR prediction, and this is another one from ChemInfo. Now here you will see that in ChemInfo, again, we can draw the structure. We will create a flavonoid, quercetin again, and let's see. So I am just adding the oxygen groups now. So I submit my query, and this is a simulation-based software.



The ChemInfo is going to predict based on its structure and attributes and let's see how it predicts. So definitely it has come up with a spectrum. So you can see again. fingerprint region as well as the hydroxies region but rather than a proper band which appeared in the previous software which gave you as much as a natural spectra as possible instead of a bands here you will observe discrete lines that's the difference here. You can see every time I move this cursor, a molecule gets highlighted.

Now, the interesting part of this software is now in addition to giving you values, you remember IR spectra principally happens due to stretching, bending. tailing, wagging kind of vibrations between the molecule and if you are not conversant with it this software is very interesting way to learn. Say for example I want to see here that the vibration or the band which is appearing at 3 5 is due to what and you can see carefully here this is what is moving You can see here this is what is moving. So it is telling it is basically the wagging action of this hydroxyl group which is giving me this band.

This is the stretching. So I can take my cursor on every value and what or why exactly that reading is coming. So you can get it basis this structure or you can get basis this molecule. You can see in both manners. So I am just putting it here and you can see when I click here

it is due to which regions, which interactions that is also seen and what exactly is happening that is also seen.

So in the fingerprint region you will see mostly the whole molecule is involved but beyond fingerprint region if you go to specific wavelengths or little higher wavelengths you can see this is due to hydroxy CH. You remember your carbonyl which occurs somewhere around your 15 to 17 centimeter inverse. So let's try and search if that's the case. So these are mostly the double bond ones. And see you can see here.

This is the carbonyl interactions. And these are the hydrogen interaction. I just click on it. You will see the interactions on the top molecule as well. So if I see this, you can see what exactly is due to interaction and the hydrogens involved.

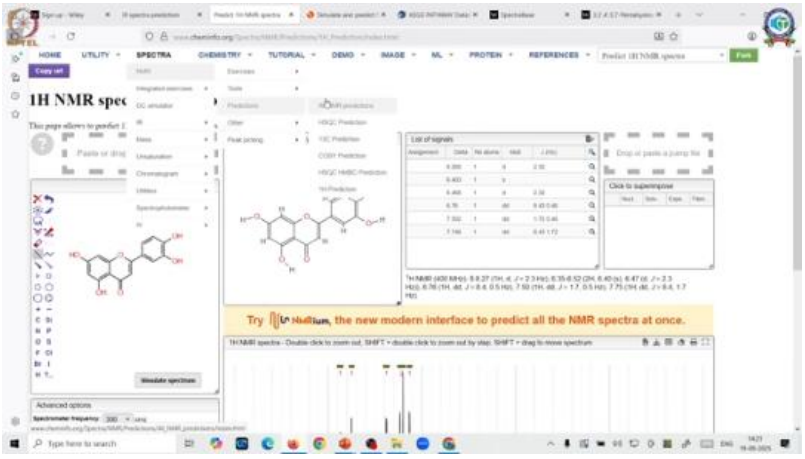
I just need to keep my cursor here. This is the typically carbonyl interaction with the adjacent hydrogen. That is what is happening. So this database that is the cheminfo.org IR database will give you a very nice structural representation of IR spectra and you will be able to understand the peaks in a much much nicer manner. Similarly if you go to the cheminfo NMR spectra you can also get much much more depth as compared to the Wiley one.

So again, I am putting my query as quercetin. And let's see what ChemInfo has to offer in terms of NMR. So I run these queries and ask ChemInfo to simulate my NMR spectra. So you can see here, as compared to the previous ones, you get the coupling values. So

That one was very predictive. This one is more about analyzing. So you know which proton is coming where and how. So you can nicely move this and see this is the proton of quercetin which is at the sixth position. This is the proton which is at the third position.

So if you want to convert it into hydroxy for quercetin, you might do that. I kind of missed it there. So again, you can see here these particular ones. Now flavonoids have most of the molecules which are aromatic, so you will see all the proton NMR peaks come in between 6 to 8 ppm. Now, what if I want to increase it? You can even increase the resolution of this.

Now you can also add here, or you can change the moles of this. So under the Spectra tab, you just go into the prediction. And then you select all NMR predictions. So it is going to give you all the different NMR predictions. This includes your proton, C13, COSY, even your 2D, such as HSQC, HMBC, and so on.

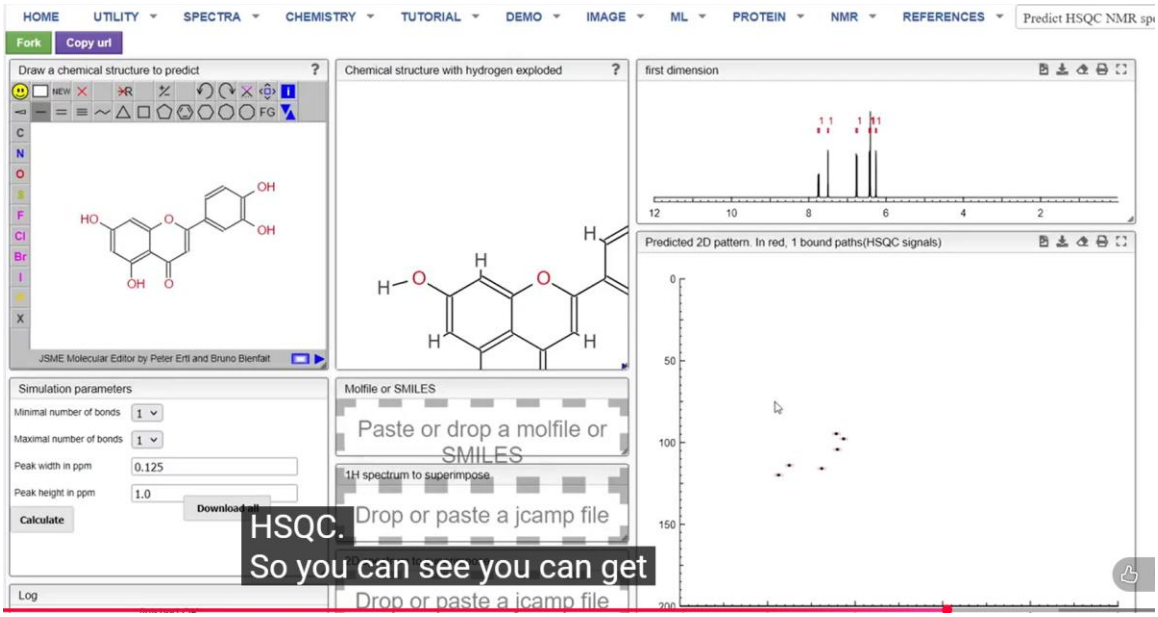


Still calculating. So let's try predicting them one by one. Let's say C13 prediction. That will make it a little faster. So you can see here all the carbons are accounted for.

Which carbon and what is the value? So remember, typically your carbonyl comes at 180. So that is how you try to match it. So depending on that, all your C13 values. So now that you have your spectra, you can correlate.

You can see whatever I am highlighting here is also simultaneously highlighted as a red line in the below spectra. So you will see the peaks of each of these carbon compounds. It's occurring in sequence. So you need not worry which is your adjacent carbon. Now going to the B ring.

So all the adjacent carbons it will account for. I will give you a reading. So you can just pick up this NMR value and compare with those reported in your literature articles. You will also see the two dimensional if it is able to predict. So let's try.



So you can see you can get your two-dimensional proton and C13 NMR in a HSQC manner. So it's still loading. That's the reason it's taking time. You can see the peaks out here. So it not only helps you with one-dimensional, it also gives you a good in-depth spectroscopic details about your two-dimensional NMR spectra.

Let's try a COSY. You can see a correlation is obtained. So you can know which is adjacent to what. So your correlation spectrum is also obtained with a proton-proton 2D map superimposed. So this is a software that can tell you about your data—ChemInfo gives you much more detail about your mass, NMR, as well as IR.

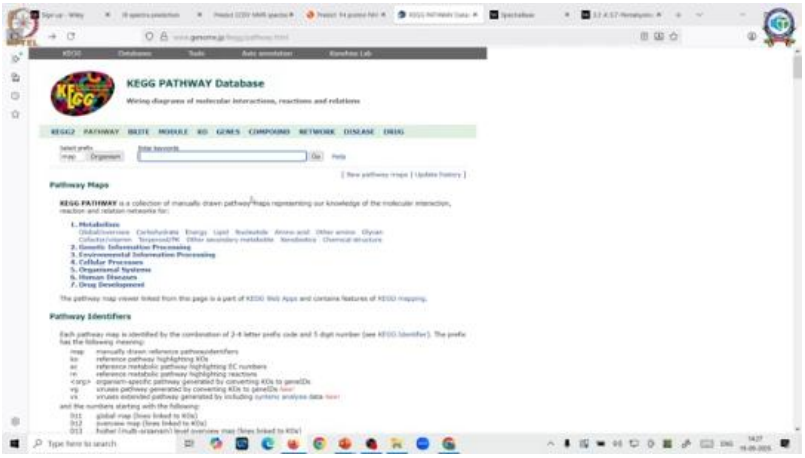
The good part is, with the mass, it gives you the fragmentation pattern; with the IR, it will tell you the peaks; and with NMR, it will tell you the ranges as well as the coupling

constants. Another way to predict NMR is what is called the NMR database, which gives you almost similar results. So here again, we just have to draw our molecule. You can see the interface is very similar. So I will just draw a slightly different molecule now.

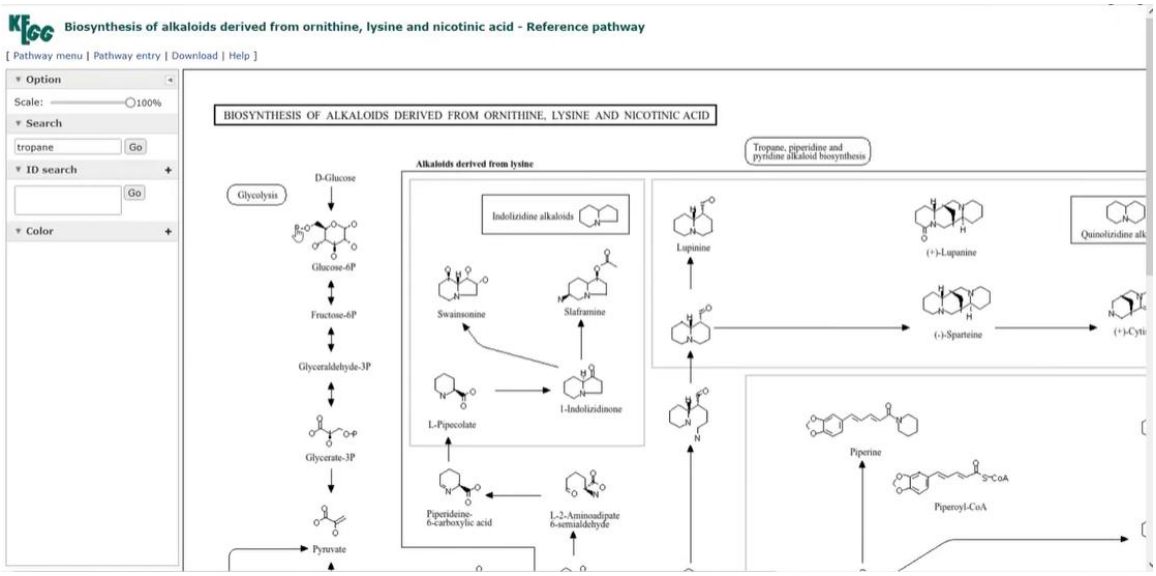
Let's see. I am drawing coumarin. So it is the very same database it is connecting us to. So the NMRDB is also obtained through this. And it's the very same interface.

So you can see the values. So you can either go by what is called the ChemInfo database or you can go by NMRDB. The interacting interface is one and the same. Now we go to the last part where I told you that, in addition to spectroscopic data, your databases interestingly also tell you about the biosynthetic pathway of your metabolites. And this interesting pathway finder is what is called your KEGG database.

So in that, I can just insert my molecules and see what biosynthetic pathway they originate from. So let's bring in a variety now. Let's try to put in maybe quercetin or tropane. Let's try tropane. So I'm just seeing it's coming up with a huge number of biosynthetic pathways—different biosynthetic pathways—and where all your tropane will feature in that.

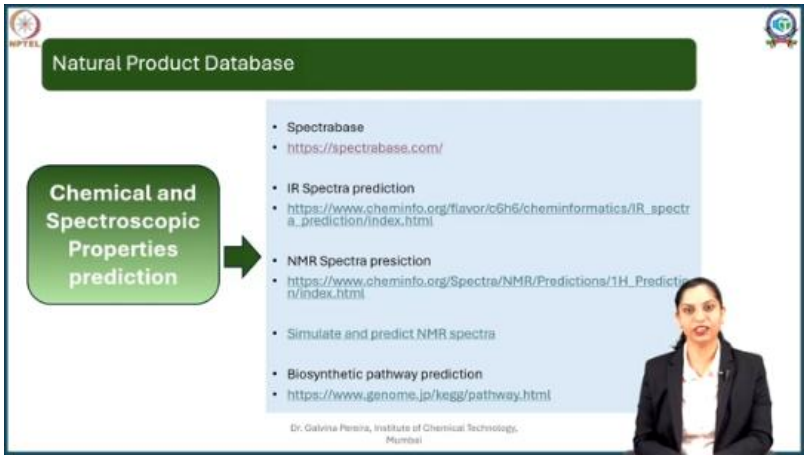


So this is one. You can see I'm just clicking them one by one. You can see from your glucose, that is your glycolysis, slowly moving to your Krebs cycle. You can have most of your glycosides being synthesized. Just moving on this side. You can see it has gone to cuscohygrin, hygrin, tropinone, propene.



And from that, it came into your propane alkaloids. Here also, this includes your hyosamine, hyacine, cocaine, and propane, and so on. This is one. So here, it's not containing structure, but if you see here, it's kind of a pathway that is depicting the biosynthesis of most of your alkaloidal compounds. And here, you have in your biosynthesis. If you go from this biosynthesis or if you come in here, you can see slowly, slowly you get your tropinone as well as tropate.

So tropinone and tropate will give you tropin. So this is where it is. You will see your other alkaloids also coming in this pathway. So KEGG software kind of gives you an idea about the biogenesis of your metabolites. So what we saw right now is there are techniques that will help you analyze the spectra, or there are techniques which will help you analyze the compound by using spectroscopic techniques.



So once you have your compounds, now no matter if you are a little weak in spectroscopic techniques, what you can do now is that the databases we have learned, you can search for your compounds, you can obtain the mass, NMR, IR values, and then compare with the ones that you obtain. If the values are exactly overlapping, plus or minus a little error, because these are mostly computational values, there is a good chance that your compound is pure. But if you are getting additional peaks, there are chances that it might be impure or it might be some other compound. So with this, we end our today's session.

Thank you, everyone, but don't forget to try these spectroscopic databases for the analysis of the different metabolites that we studied in the previous week. Thank you.