PHARMACOGNOSY AND PHYTOCHEMISTRY

Dr. Galvina Pereira

Department of Pharmaceutical Sciences and Technology

Institute of Chemical Technology, Mumbai

Week 12

Lecture 58

Week 12: Lecture 58: Natural Product Databases for Studying Phytochemical Properties

Thank you. Hello everyone, and welcome to the third session of week 12 of our NPTEL course in pharmacognosy and phytochemistry. This week, we are learning a set of computational tools that will help our knowledge or help us gain more knowledge in the field of pharmacognosy and phytochemistry. In the very first session, we saw how we can observe and study which plants we choose for our research. What plants do we see?

Can we compare them with herbarium sheets present universally in these databases and understand the taxonomy of these plants? In the second session, where we discussed the utility of these plants in traditional medicines, we saw how different databases are curated to accommodate the drugs present in traditional medicines, know their properties, and help us translate so that we do not have to go to Sanskrit or other scriptures, and we can get a deep understanding right from the servers themselves. In session 2, we saw how we can use the servers or databases to see the different phytoconstituents that are present. So, if you recollect, In week one, we saw databases such as Plants of the World Online, JSTOR, Tropicos, New York Botanical Gardens, and Herbarium Sheets.

Then we saw some traditional medicinal plant databases. We all recall it was the Traditional Knowledge Digital Library. Then you had your MPPAT, as well as we saw the Indian Medicinal Plants Database. In the last session, we saw a few more servers and databases that helped us in phytochemical searches. These were the Knapsack database,

Dr. Duke's database, PhytoHerb—essentially for food-based ingredients—and Coconut databases, which gave us phytoconstituents along with the structure.

IMP-PAT which also gives us the phytochemical correlations of the plant. IMP-PAT also gave us the therapeutic applications. If you recollect, there were two windows, one pertaining to phytochemical and other pertaining to therapeutics. In addition to that, we also saw the AUSHADI database and AUSHADI database gave us the whole thing like the phytochemicals, the therapeutic application. It also gives us the chemometric database.

Now, in today's session, we will move a step further. Now we got the plants. We got the pharmacological activities, especially with respect to traditional medicine. Now we got the phyto constituents. So today's session will be seeing different different servers or web based applications that are going to help us analyze the phyto constituents in terms of their properties.

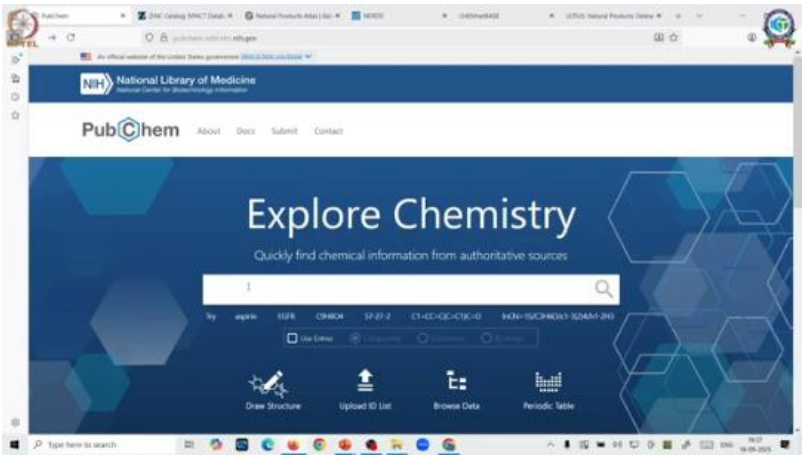
Now, understanding the properties of phyto constituents are very important. important in terms of their application. So I should know what they are soluble in, maybe something about the molecular weight, something about the structure, the orientation. All of this information is vital when I have to go a step ahead and do some chemical reactions or use those phytochemical as pure compounds for analysis. Not only that, this pure phytochemicals have numerous applications in biosensors,

They have been used in therapeutic studies, especially molecular docking studies, where the structural attributes play a vital role. So this session particularly emphasizes understanding the phytochemical properties of the phytoconstituents. So today we are going to use a few databases and web-based searches. This includes PubChem, Lotus, Zinc Natural Product Database, NP Atlas or Natural Product Atlas, Dictionary of Natural Products, as well as NP Scout. So let's see what each of the search engines has to offer in terms of structural information about our compounds.

So we'll start with the PubChem database. Now, the PubChem database is not exclusively a natural product database. It is a database pertaining to natural as well as allopathic medicines, as well as non-medicinal molecules. So it provides curated information about molecular properties, spectral characteristics, as well as metabolism. So let's first see the

PubChem database, which is non-specific, and what it has to offer for natural products, especially the phytoconstituents.

So when you search the PubChem database, you will land on a page like this. So you have this, which is run by the National Library of Medicine, and you will find a search that is a chemistry-based search. And here you can input your queries either as a name, a structure, a molecular formula, or even what is called a 2D format. We call it a SMILES notation or even an INCI number.



So let's simply put a molecule called quercetin and let's see what happens—it's already started running with the queries. So I am just clicking on quercetin. Now when I click on quercetin, it has come up with quercetin, and it says that this quercetin has 701 compounds. Now what does that mean?

That means this quercetin, which is a native nucleus, is part of the nucleus of many other molecules, and these molecules could be simple quercetin or glycosides. Can you see there? You know what glycosides are. You remember when they are complex or condensed with sugar. Gentio biosides, soforoside.

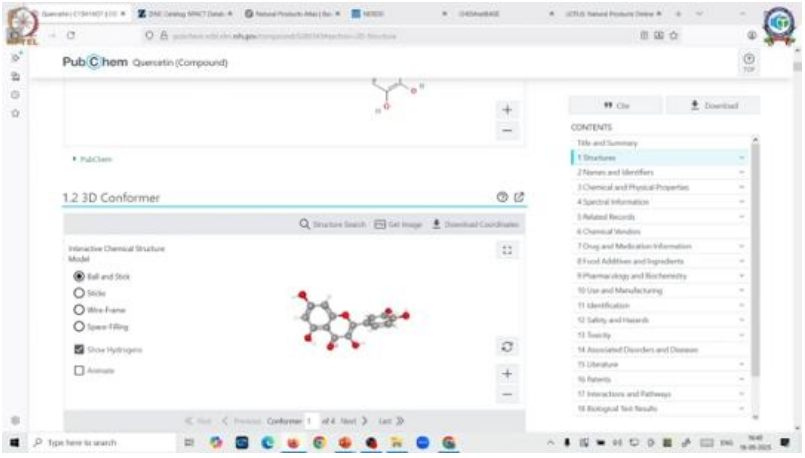
You can see here the list is enormous—almost 701. In many cases, you might even come across dimers of quercetin. So it will take this query as one search and find all very similar structures. It will return to you all such similar structures. So this is a very elaborate list.

So let's simply click on quercetin as a native molecule. Now, when I click on quercetin as a native molecule, you can see on this side whatever the site has to offer. So you will see

PubChem gives you information about structures, names and identifiers, chemical and physical properties, spectral information, chemical vendors, suppliers, etc. Drug and medication, food additives, pharmacology, use and manufacturing, identification, something to do with the MSDS, that is safety and hazards, toxicity, any associated disease, some literature patent. Can you understand here the elaborate literature is not brought in just pertaining to the structural attributes?

Interactions, pathways, biologicals, taxonomy, classification. So you can see it is giving you a whole chemistry in one page. So you can see here if I just see quercetin, I can click here, I get what is called the 2D structure of quercetin wherein the heteromolecules are marked by different colors. Now, if I have to use, like I said many times, we see phytochemical structures. We want the coordinates of it.

You can download the structures. So you can see the different formats are available for this. You can also see quercetin, how it occurs in nature mostly in the 3D format. You just keep your cursor here and you can enjoy beautifully how this molecule occurs in nature and the most what is called as NMT minimized or the most stable conformation in different forms, that is the ball and stick model, just the stick model, wire and frame model, or even you can see it in the space-filling form.



So you can appreciate how the flavonoids are actually there or localized in that. Now we go to the crystal structure. These are all the crystal structure records. Now what happens is when you purify any phytochemical, it will not always settle into one particular crystal

conformation. Now depending upon the solvent, temperature, state of hydration, and the interactions it undergoes, you might come up with different crystal forms.

So all these crystal structure details are brought in here. So these are brought through your crystallography open databases. Then you have your descriptors. These descriptors are the ones used to annotate it. This includes your IUPAC name, INCI, or basically INCHI name.

Then your key. This is just an alphabetical representation of that. Then the 2D format of the structure, which we call the SMILES notation. The molecular formula, CAS ID—that is, the Chemical Abstracts ID. Then you have your EC number, if it has any.

Sometimes, some compounds do have. So, all the numbers it is identified with will be seen here. You can see different IDs. DrugBank ID, Toxicity Substance ID, HMDB ID, KEGG ID. So, all the different IDs—whatever molecule is present in those software, by which number it is listed—you can see those numbers here.

Then you have the MeSH entries, which are often used to search literature. The synonyms of it. The physical properties. You remember physical properties and the computed properties, which influence the molecular weight. Log P, which is the lipophilicity.

How many hydrogen bond acceptors the molecule has. How many hydrogen bond donors. Now, this is necessary for interaction studies. How flexible it is. So, it just has that one rotatable bond we saw.

That is the B-ring rotatable bond. The exact mass, isotopic mass. That is, this is the total polar surface area. How polar is the molecule? How many heavy atoms does it have?

You can see the color. Many times, you might come across a brown-colored substance in nature. And you say that you have a purified compound. But when you check PubChem, you see your quercetin. If you are purifying, if it is brown, it's not your purified quercetin.

Your purified quercetin should appear as yellow needles. The boiling point. In this case, it's not boiling. It's actually subliming. Melting point.

Solubility. Drug banks, details, optical rotation, decomposition. So, you can see a lot of properties and other experimental properties are there. Now, we go to the spectral data. So, you can see if this analyte.

Now, this is basically not data that is kind of reproduced. This is data that is calculated by servers. So, it is calculated that if you use a 600 megahertz quercetin, this is the kind of graph you are going to get in terms of ppm. Now, this is particularly for the region between 6 to 8, but definitely, you will see some of the peaks beyond this region as well. Then your C13, remember this.

Then your 2D NMR, especially you see your HSQC in this region where your proton and C13 NMR are brought together. Then the mass spectroscopic details, especially the GC-MS, what all peaks it will show in terms of its fragmentation. This is one of my go-to databases when I search chemical properties of any phyto-constituent. This is a database. Like I said, it's not just pertaining to natural products, but it's a very exhaustive database.

So most of the chemical information that you require about any of these molecules, you will see it here. And not only do you see it here, it's very accurate here. So the fragmentation peaks, the UV spectra, you know typical flavonoids at what nanometers. So you know we said the flavonoids have two peaks: band 1, band 2. So you can see here the band 1 is at 373.

Then you have a small hump here and again you have your band 2 at 256. The IR spectral detail is not there, but you You can see the list is so exorbitant. Who supplies this compound in case you want to purchase it? So your PubChem kind of makes all the details ready for you.

The metabolism, the drug indication, where it is known to be used. Age-related macular degeneration or Alzheimer's, childhood cancer, hepatitis. Almost different entries here. And at what phase it has reached. That also gives you the detail.

The summary. Whether it is toxic, especially to your liver, or not. The drug labels. Yes. So we just reached a 7-point.

So understand the kind of depth this kind of server gives you. Drug warning, biomarker where it is seen in case of a particular disease, food additives, associated food which will also give you this compound. The pharmacology, any reports pertaining to the pharmacological use of this. Metabolism: what happens to those if you consume quercetin? See, that's why I said when you're doing literature, you can get this server's help with the whole thing. So I need not study this again; I already know what my quercetin is going to be metabolized into just by browsing through this software. The biological half-life, what is the mechanism of action? Transformations, uses—mostly antioxidant is what it's used for. Methods of manufacturing: so if you're thinking of synthetically producing it rather than taking it from plants, this kind of server gives you that as well. Formulations, manufacturing information, then you can have your laboratory safety.

Hazard categories, like I said, your MSDS data is also partially here. Then what do you do in case of accidental release when you are exposed to it? Storage conditions, stability: is the molecule stable or not? Transportation, any regulatory information, toxicity information, carcinogenicity information, acute effects on mice. It's seen interactions with any drugs or any other substances.

So, as I said, you can see this software is like a whole package for your literature. That's why it's mostly the go-to software when you search. And many of the phyto-constituents are available here. Environmental deterioration, concentration, soil adsorption. See the depth they've gone to in curating this—where it all occurs.

And at what concentration? So you can just say that this quercetin or its glycoside is present in apples at this concentration, pears at this concentration. Otherwise, it's difficult to search paper by paper. See some literature—like I said, this is just page one of 9,300. So imagine how much they have to offer us in terms of literature.

So, more than 5,000 pages you can search, where you'll get the literature pertaining to that. Wiley references, Nature journal references, chemical co-occurrences. If there are any other flavonoids which are very structurally reactive. Then you have gene co-occurrences, chemical-disease co-occurrences. After the literature, you have patents.

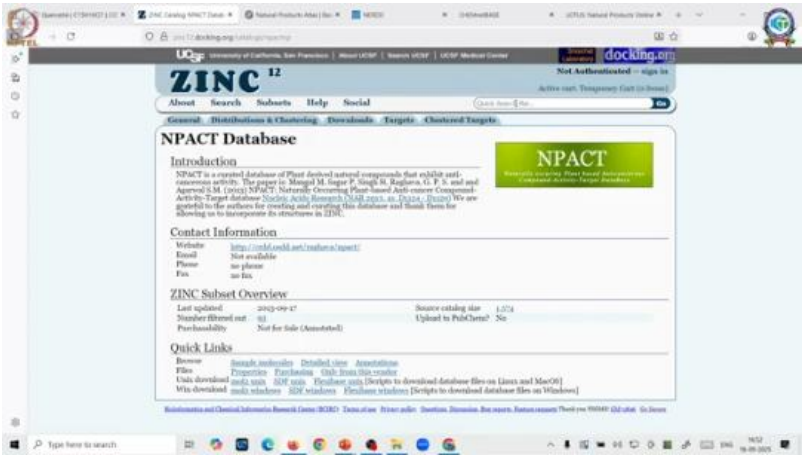
What patents have been filed for quercetin? Which mention quercetin. So all that patent data will be there where the quercetin name is mentioned. It will come here. And also interaction data.

Is quercetin studied in terms of its binding ability with certain receptors or proteins? You can even download those from here. So again, if you see, there are 7 pages with different receptors wherein your protein receptors are complexed with quercetin. The ligands, the interactions with different genes. Again, you can see the amount of data that comes on your PubChem.

It's almost a thousand five five one kind of searches or queries it answers to. And then Biosynthesis. So if you want to know how it is biosynthesized, that detail is also there on this database. Bioassays.

The taxonomy where it occurs and what the taxonomy of that is. The MESH terms which help you in literature. The lipid KEGG, phytochemical compounds KEGG. So all in all, if you go to see, you know, this is kind of a very elaborate database. Just search the compound and you will get a complete history about it.

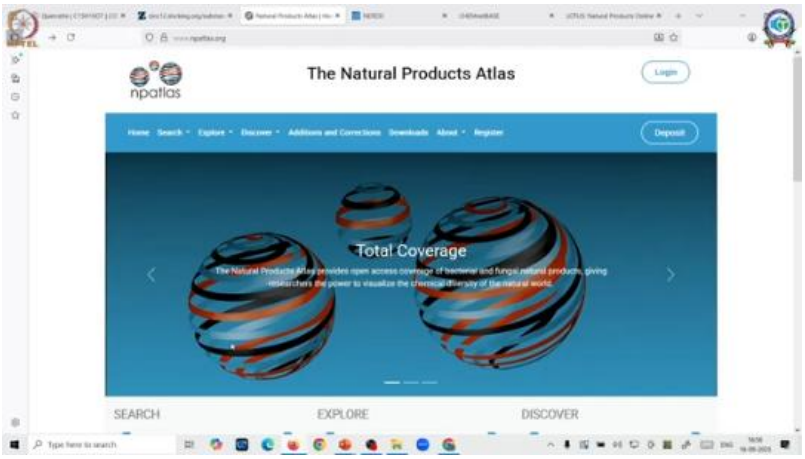
But we are not stopping here. We'll move to the next database, which also has some details about it. And let's go to the next database. And that's your Zinc database. So again, in this, I'll raise the same query and let's see what Zinc has to offer.



Zinc and PAT is the Zinc Natural Product Database. So let's see which are the structures. Now you can see it has come back with three queries. The slight difference between these

is the first structure is racemic, and you can see the other ones are the optically active ones. So let's just click on the racemic one to see what they have to offer us.

The NP Atlas database is another curated database for phytochemicals. It contains properties of most phytochemicals with respect to their structures as well as their pharmacological activities. So let's search or input the query for our compound and then check whether it offers something additional compared to the previous software. In this case, I'll just enter my query in the form of a structure. Let's draw our quercetin.



Then let's try to search. Now, you can input it as a full structure or as a substructure. When you say a full structure, I want precisely this. But when I say substructure, it would include any molecule that has at least this in its structure and may have additional functional groups. Then I can check for similarities in them.

The threshold is that the substructures or structures that appear should at least be 80% matching to my query. That means 80 percent matching to my query. In addition, I can write the name or provide the molecular formula, chemical formula, or INCHI keys. In this case, I'm just entering all the properties and submitting my query. Here you can see the quercetin structure is definitely present.

But in addition to that, you can see there are some additional groups there. And here in this case, it's quercetin, but you have added a methoxy here and a glycoside here. So these are a few add-on products that have come out back with. So this is the name of the compound. I can just go there.

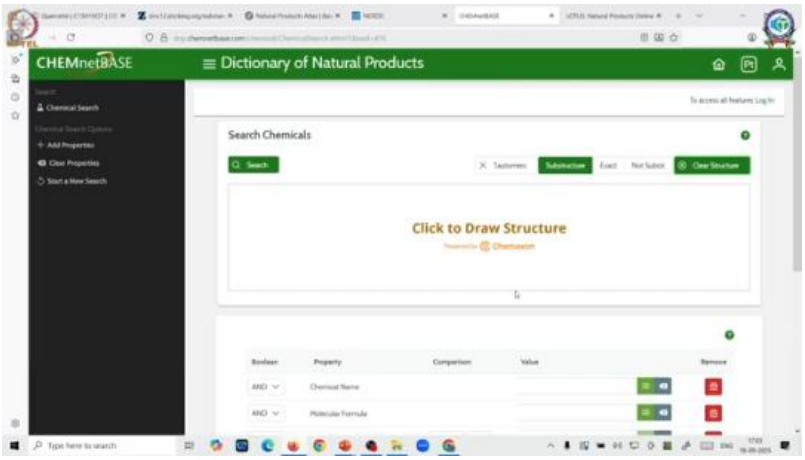
I can download the structural properties from here as SDF options, which will give me the 3D structure as well as the mass and other properties. Now, what I can do is, if I want more of this, like more similar flavonoid kind of structures, I can give a slightly incomplete structure. Say, for example, I'm not very keen on quercetin, but I want some flavonoid. So I'll just give it the flavonoid nucleus rather than giving an entire flavonoid. Let's see then what happens.

So this is just the part of the nucleus that I'm putting in, and then I tell it to submit the query. So this time it has come up with 17 compounds which have it. The reason is, this time I was not very specific with the hydroxy substitution. So there is a good chance that if I have this kind of nucleus, any of these compounds can also be there. So NP atlas kind of gives you an idea that if this particular characteristic pharmacophore or structure is there, what other possibilities or what possible structures are there that you can abbreviate.

Now moving on to the next one here. This is a software called NP Scout, which helps you predict what is known as drug likeness. Now, many times after purifying the molecules, you might come across or derivatize certain structures based on their spectral properties. Now, after that, is that molecule naturally feasible in plants or not? We can check it here. So what we need to do is, for example, based on the spectral characteristics, I found out that this could be my molecule. So, is it a natural product-like molecule or not?

I can just find out by doing this. So again, I raise my query as quercetin and let's see what happens. So it says NP class probability. Once means full probability that it's a natural product. It occurs in nature, and the majority of the areas where the hetero groups match are marked with green coloration. So anytime after spectral derivation, you come across a structure and want to see whether it is

actually occurring in nature or you have drawn something wrong. Whether it's naturally possible or not, I can understand by doing this. A good score is an NP class probability of 0.9 and above. Then we move on to the next database. Here also, we can see the dictionary of natural products.

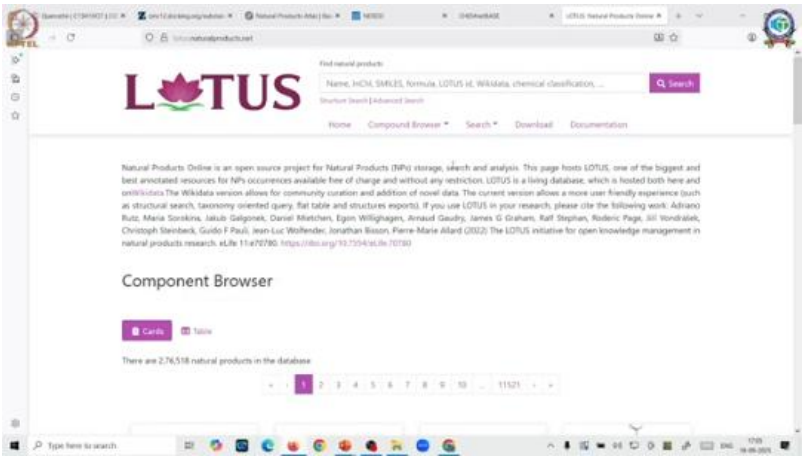


So I just have to draw maybe a pharmacophore and then just see whether it is possible or how many such natural products are possible. Again, I go back with a little raw query. I'll not give it a full structure, and then let's see what this software comes back to me with. I'll just give it one hydroxy for convenience, and then let's see. So what I am putting here is in the dictionary: how many natural products have this kind of structure.

So let's see and search. So you can see the total hits. In nature, there are 7,212 molecules which have that particular structure, and here is the list of all those molecules. So the dictionary of natural products, based on the structure that you have curated, will come up with all the possibilities and all the structures. So this is very important in terms of pharmacological activity or whenever you are doing docking studies.

I want a particular molecule with this pharmacophore. So how many molecules are there in nature which possess this particular pharmacophore? So just put that pharmacophore in your dictionary of natural products. It will come across this many compounds. So you can use these structures in docking and then you can optimize which natural product will react well.

Now, in addition to that, you also have your lotus. Here, you can directly put the name as a query. So I'm just putting it, and then I click search. So there are so many molecules—about 32 natural products—which have quercetin in their structure, plus some add-ons. So lotus also gives me all the molecules which will have that in a very similar manner to your atlas.



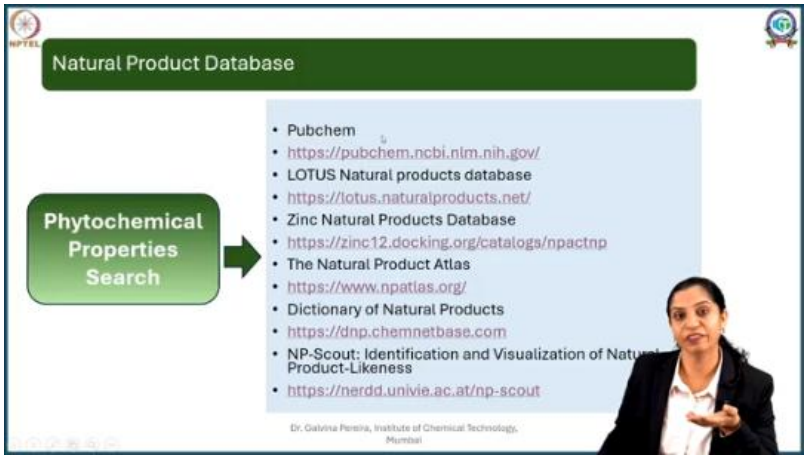
And as a result, that will be very useful for me in terms of predicting properties. If I just put my native quercetin here, let's see. So I start getting properties of it, not as much in-depth as your PubChem. But yes, I get something which is called the link to taxonomy. So if I say it is found in Vitis vinifera.

So which all plants in that particular taxa have it is how they see it. So you can just click on one. Let's click on the Vitis vinifera, which is nothing but your grapes. So you can see in the Vitis family, it occurs in all these families. Places or all these families.

So your quercetin per se here also, you know how many plants contain this—not only that, but in that family, how many plants contain this can be obtained via your Lotus database. So this is, you can see, a very exorbitant list of so many plants containing your quercetin as a molecule. So just huge, and that's where you say, if I want to see quercetin, which all plants contain when I'm trying to find the source, Lotus is a go-to software which can help you out. So quickly summarizing it back. So today we learned how to use natural product databases.

To see the properties, both the physical as well as chemical properties of the drug. The software includes or the web-based servers include your PubChem, Lotus Natural Product Database, your Zinc Databases, NP Atlas, Dictionary of Natural Products, and NP Scout, which will help you predict the natural product likeness. So thank you everyone for listening. But I want you all to check by putting your phytochemicals in the queries and seeing if you are actually able to assess all the properties through all the software. When

you try to do it yourself, maybe you will discover a lot of other things you didn't know about your compound.



So happy learning, all of you. Thank you.