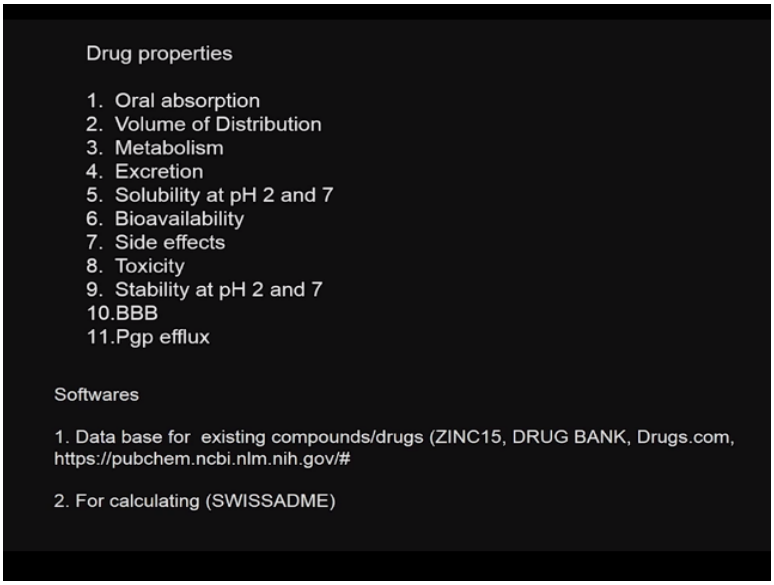
Computer Aided Drug Design
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology – Madras

Lecture – 06
Drug - Properties/SMILES

Hello everyone. Welcome to the course on computer-aided drug design. We will continue on the concept of drug properties and we will also look at something called SMILES.

(Refer Slide Time: 00:26)

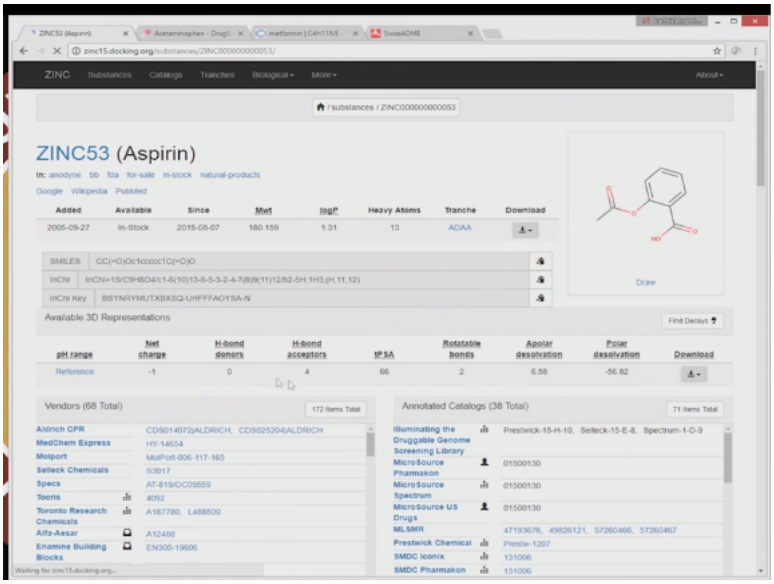


So in the previous class I had talked about properties, important properties, the drug absorption okay, volume of distribution. It is called absorption not adsorption, oral absorption because most of the drugs are taken orally and then volume of distribution it gets distributed into the blood stream as well as into the tissues, the metabolism, how it gets (()) (01:01), how it gets excreted is it have solubility at pH = 2 that is your stomach.

And does it have solubility at pH = 7 that is in the blood region, bioavailability, what is the concentration of the blood available in the target side does it have side effects, toxicity, stability at pH2 and at 7, the blood brain barrier penetration, BBB means blood brain barrier and then Pgp efflux. So many of these properties are given in many of the databases I will show you as examples also now as well as there are some online softwares where we can calculate theoretically some of these properties.

So we will look at some of these databases like ZINC15, DRUG BANK, Drugs.com. PubChem so some of the data are of these drug properties are available and this particular SWISSADME is a software were we can calculate theoretically some of these properties not all of them, some of these properties when we draw a structure okay. Let us look at some of these examples as of now.

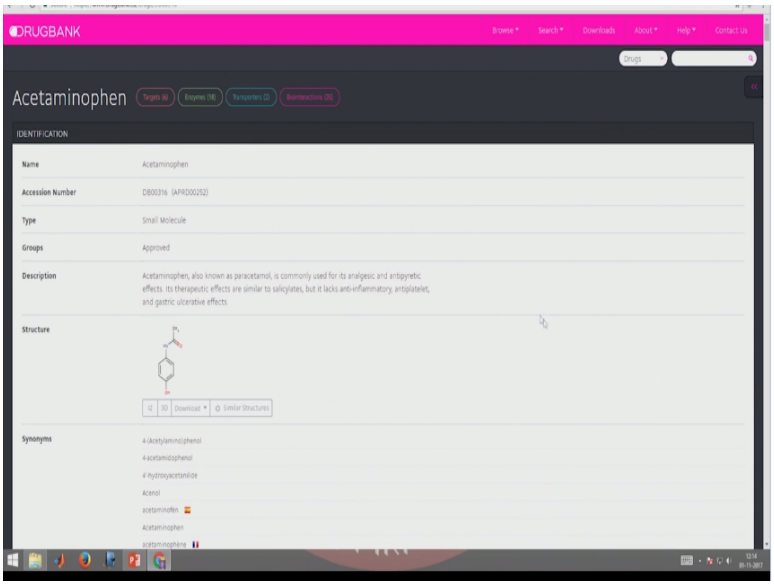
(Refer Slide Time: 02:12)



The first one is you ZINC15 as I said the ZINC15 I just typed in aspirin. Aspirin is a very common drug. It is for antipyretic, pain. It is also used for platelet and de-aggregation and so on so. This is also known as the acetyl and salicylic acid. Now as you can see some properties have given net charge, hydrogen bond donors that means does it have any donors like OH NH, but this OH is not considered as a donor because generally it dissociates into O - and H +.

Hydrogen bond acceptors it has got 1, 2, 3, 4 and a total polar surface area is contributed by the oxygens and nitrogens, rotatable bonds it has got 2 rotatable bonds okay. Basically, these and all there and the apolar desolvation and polar desolvation. So some properties either experimentally determined or properties which are obtained from theoretical modeling are given in addition I think other information is also given about this molecule. Now let us look at another database.

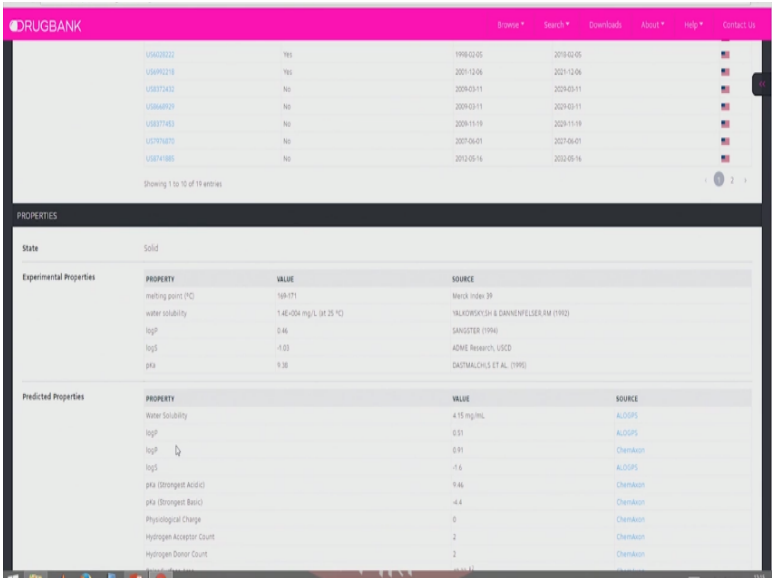
(Refer Slide Time: 03:39)



This is called the DRUGBANK database. As I mentioned drug this DRUGBANK. This molecule is called acetaminophen it is like a paracetamol. It is a paracetamol. So it can be used for analgesic, antipyretic. So okay these are the applications. It does not have anti-inflammatory anti-platelet activity, but it can be used for anti-pyretic. So if you look at some of these properties of this particular database also gives you some properties.

The molecular formula okay and as you can see some of the route of elimination, half life, clearance, some toxicity, metabolism, what type of enzymes metabolize, what is the absorption of this rapid and almost complete. So immediately it gets absorbed. Okay then which organisms get and if there is any interactions? okay. So, lot of information as you can see are given about logP values. Let me talk more about this logP.

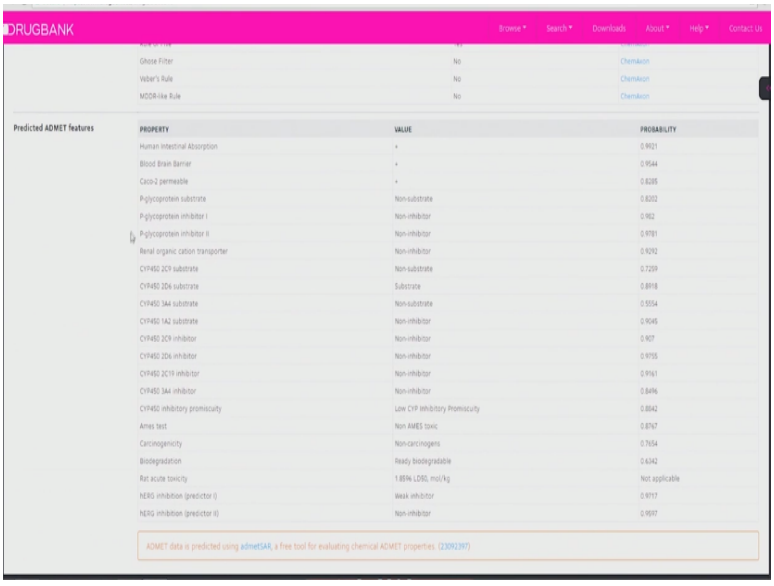
(Refer Slide Time: 04:46)



LogP is nothing but the hydrophobic, hydrophilic balance of the drug and logP is it tells you how much hydrophobic the molecule is so that it can cross the lipid bilayer in the gastrointestinal tract and the hydrophilic determines the solubility of this. So higher the logP more hydrophobic it is, lower the logP more hydrophilic it is. So this is a very hydrophilic molecule.

So these logP are calculated by 2 different softwares then logS is the logarithm of the solubility, PKA dissociation as you can see PKA is strongest basic so many things hydrogen bond acceptor count, hydrogen bond donor count, polar surface area, refractivity, number of rings it has got, bioavailability.

(Refer Slide Time: 05:39)

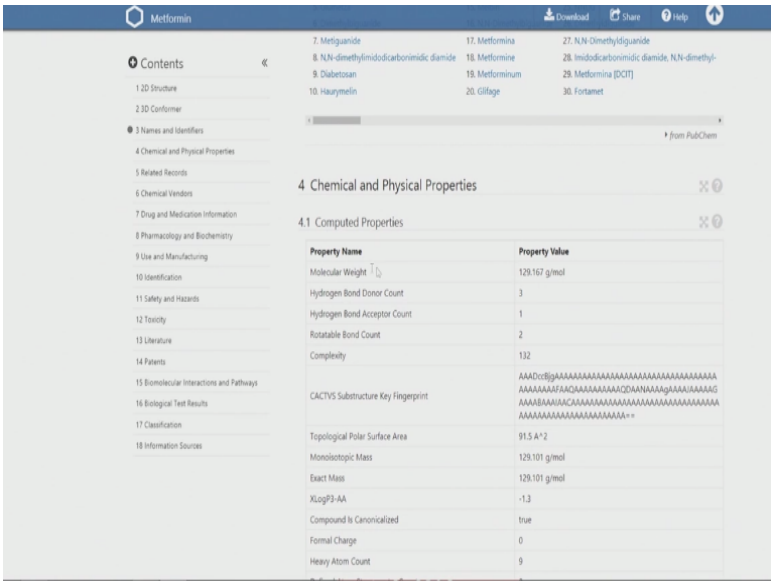


So it is very good available then you can see blood brain barrier, yes it may cross blood brain barrier. Very high probability human intestinal absorption again it has got very high probability. p-glycoprotein substrate no so it does not and then bio-degeneration readily biodegradable, and toxicity details are given so if it is a drug then we can use the DRUGBANK and also other one I talked about drugs.com. So we can use that also and you will get experimental data as well as theoretically estimated data using different types of correlations.

(Refer Slide Time: 06:24)

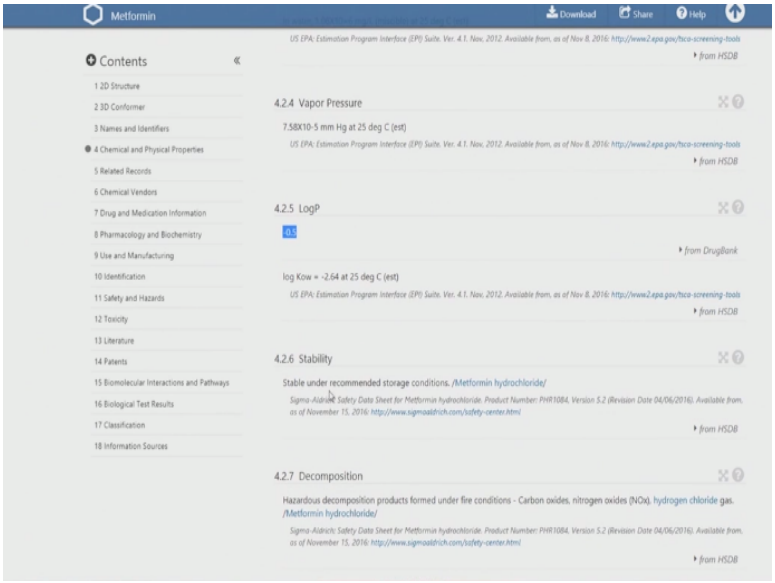
Now another software, which I said is the NCBI database which I mentioned NCBI.NIM.NIH so we can use that for example I typed metformin here so it gives you molecular formula okay. It is biguanide so it gives a lot of chemical details about it. Okay then it gives you toxicity, we can click on this so it can give you toxicity, safety hazards okay then it gives you a 3D confirmation.

(Refer Slide Time: 07:05)



Okay as you can say again molecular weight, hydrogen bond donor count. It has got a 3 hydrogen bond donors okay. So it has got this one, this one and this one 3 hydrogen bond donors then it has got hydrogen bond acceptors okay how many heavy atoms it has got, hydrogen acceptor count 1 okay only 1 acceptor okay.

(Refer Slide Time: 07:45)

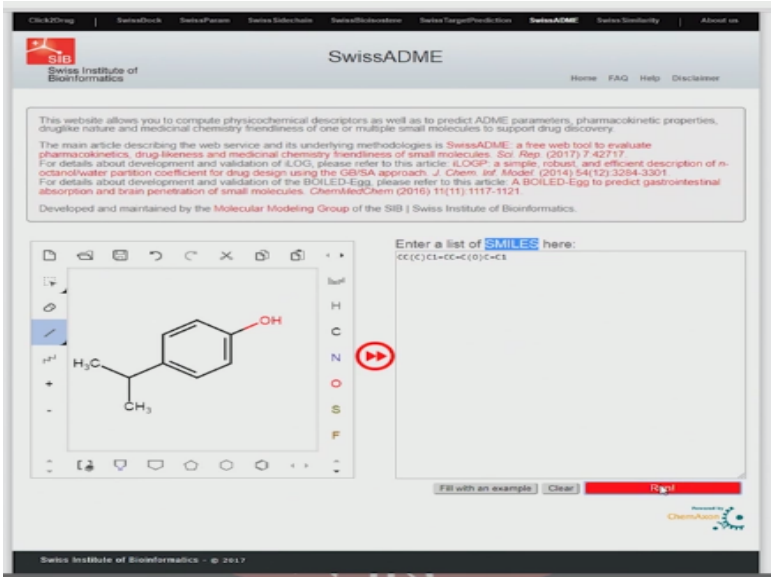


So hydrogen bond acceptors then rotatable bonds 2 bonds, and topological polar surface area, mass, logP, heavy atom count, and so on actually as you can see lot of properties. Then some experimental properties, melting point okay what it is solid or liquid, vapour pressure, logP. LogP as I said more negative it is or lesser the values it means it is highly hydrophilic as you can

see it has got lot of nitrogens right and stability, stable under recommended storage, decomposition, decomposition details, pKa okay.

So that is the dissociation (()) (08:14). So lot of information is given so experimental and theoretical. So we can get many of these properties of molecules which are stored either in ZINC database or if it is a drug they can go to DRUGBANK or drugs.com or even PubChem database. PubChem can also look at new structures for example if I have totally unknown molecule and I want to know some of these properties I can find out using this particular software.

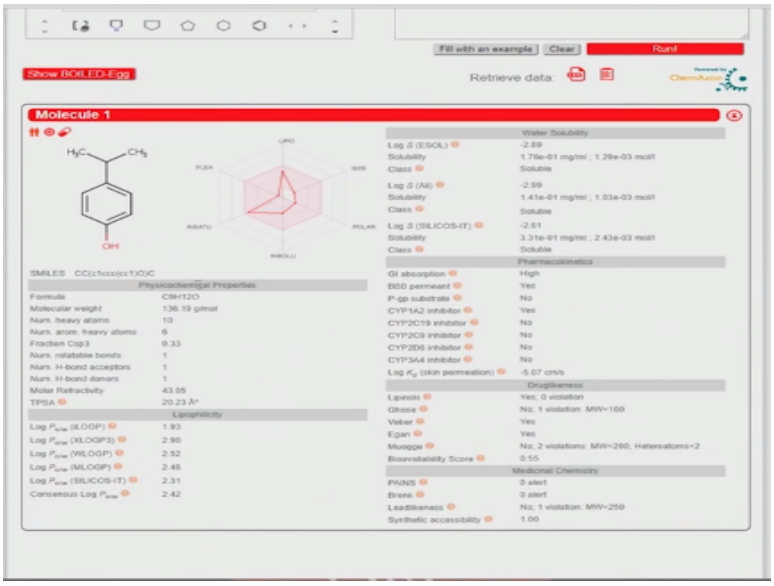
(Refer Slide Time: 08:53)



This is called a SwissADME software. So I can draw a structure here okay. By drawing the structure, I can say or try to predict some property. For example, I can draw the structures so I put in a benzene ring then I put in a this one then I put another one, then I put some oxygen here okay and then I put a territory group here then OH here okay. So this is a structure I have drawn imagine ya this is the structure I have drawn you understand.

Then I want to know some of the properties which it can calculate. So it converts into something called SMILES. We will talk about this more in detail. SMILES is a linear representation of the molecule. We will talk more in detail after a few minutes. So it converts that into SMILES. Then we can say run. So run say it does a lot of calculations okay.

(Refer Slide Time: 10:05)



As you can see formula, molecular weight, this is the molecule which we have drawn, number of heavy atoms, okay number of heavy atoms that is carbon, nitrogen, oxygen like that. So you have 6 carbons here, 7, 8, 9, and 1 oxygen that is why 10 okay. Then rotatable bonds okay this bond is rotatable that is why 1 number of hydrogen bond acceptors that is this oxygen, hydrogen bond donor, the OH, then total polar surface area is given here.

LogP there are different software which can give you logP so as you can see each software gives you some different values and average logP then solubility, log solubility okay and as you can see when you say solubility class if it is < 4 as you can see here it is soluble and then different types of solubility. Then Ga absorption is very good high, BBB penetration again penetrates through the blood brain barrier, because it is a small molecule.

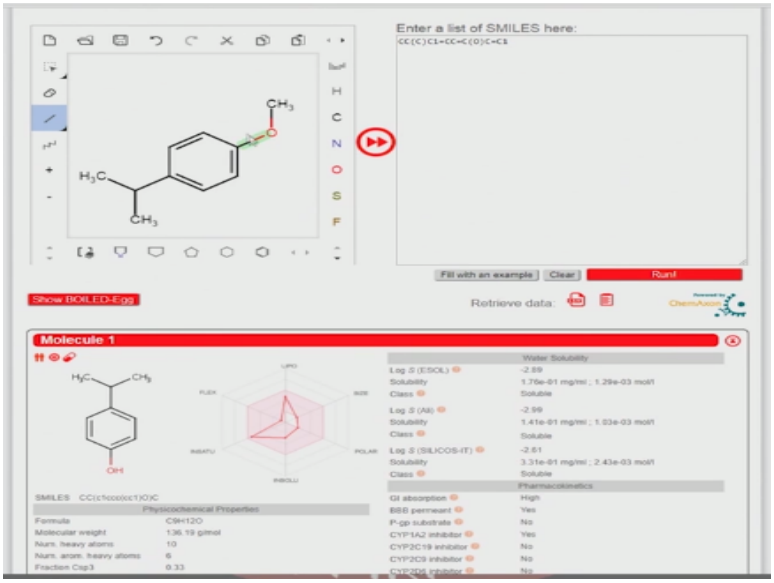
P-gp substrate it says it is not a substrate and then does it go and bind to these cytochromes okay yes, no, no, no. then we will talk about this some of the drug likeness property. They are called there are different rules like Lipinski's rule, Ghose rule, Veber rule, and Muegge rule, we will talk about all these rules and then later on and then bioavailability so 0.55 is good actually so it is bioavailable lead likeness, does it have a lead likeness property yes, it has got a lead likeness.

Synthetic disease for a chemist is it easy to make it or very difficult to make it okay. So gives lot of information for new molecules it is very useful so I feel it is got too much hydrophilic, if I

want to reduce the hydrophilicity, I can draw another modify this structure and so on actually so I can do many things and check it out and see whether I can synthesize this molecule and take it to my lab and start synthesizing. So this SwissADME is a good freeze online tool.

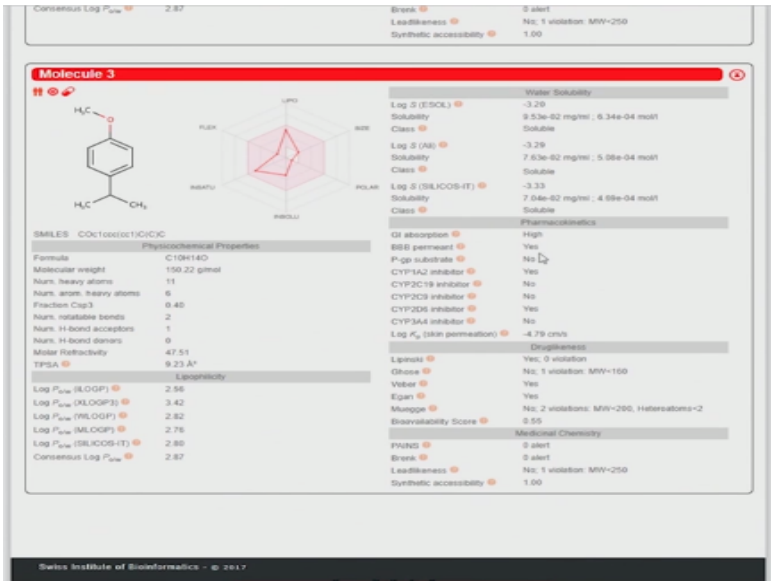
So we can draw any structure here we can even import structures from different softwares and as we can see import so we can import from different softwares also different okay softwares the files with the different extensions. We can import okay and then we can do lot of calculations as I showed you below. So this molecule which I have drawn appears to be good and it has got good lead-likeness, it has got good bioavailability, but it can penetrate the blood brain barrier okay so that is one of the problems of this particular molecule.

(Refer Slide Time: 13:27)



So I can say modify a little bit for example I then I have put something like this and then I have put this, then I put this, then I put this, then maybe I put this, okay so I have instead of OH I made it into a OCH3 so it will become more hydrophobic. Let us see what happens to it okay. So around this molecule so molecule 2.

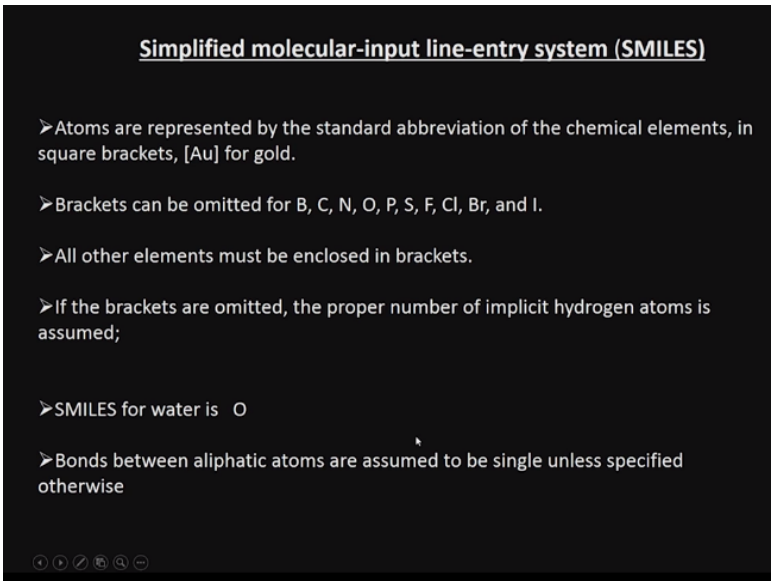
(Refer Slide Time: 14:22)



So this is the molecule I have drawn so BBB still it is penetrating as you can see logP has gone up to 2.87, bioavailability has not changed much okay number of heavy items has gone up to 11 because we have added 1 CH3 total polar surface area has come down to 9.23 okay solubility is still good still soluble. It is BBB penetrating Ga absorption is still good so we can draw different types of structures okay.

And then see how these some of these properties change and then we can select molecules which appear to be satisfactory so we can start with 1 structure and then slightly modify using the software and then select a structure which seem to satisfy most of these rules. Some of these are most of them are theoretical in nature so there is going to be lot of errors, but still it gives you a first and idea about what type of structures to select and which are the issues that particular molecule may face. Okay now let us look this SMILES. What does this SMILES means?

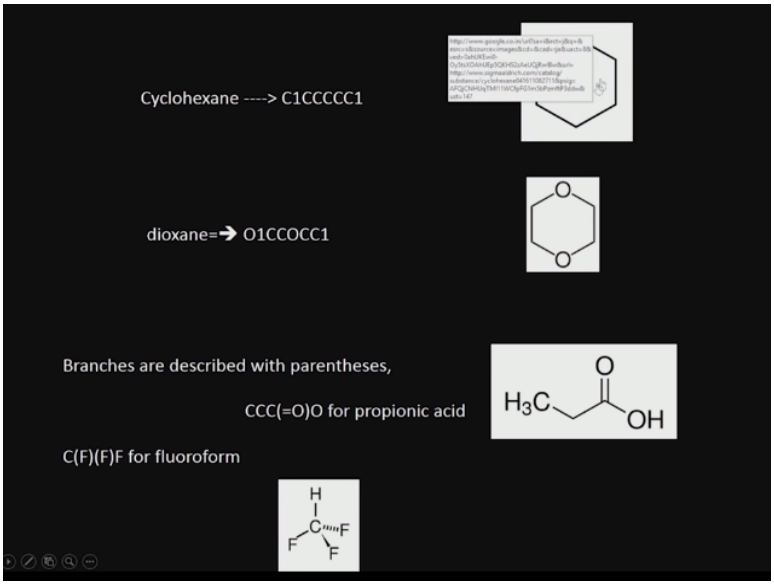
(Refer Slide Time: 15:40)



SMILES expansion of there is Simplified molecular-input line-entry system is SMILES simplified molecular input line entry system. So in single line we can describe the entire molecule. Atoms are represented by the standard abbreviation of the chemical element okay and so square brackets especially if it is gold because it has got 2 alphabets Au whereas C is 1 alphabet. So you do not need like this.

So it can be omitted for Boron, carbon, nitrogen, oxygen, phosphorus, sulfur, all those things whereas other elements if it has got 2 characters we need to enclose with a square bracket. If the brackets are omitted, the proper number of implicit hydrogens are assumed okay. So hydrogens are automatically assumed. You do not have to put. SMILES for water is O because it assumes that O has 2 valency so 2 H are assumed. Bonds between aliphatic atoms are assumed to be single unless it is specified.

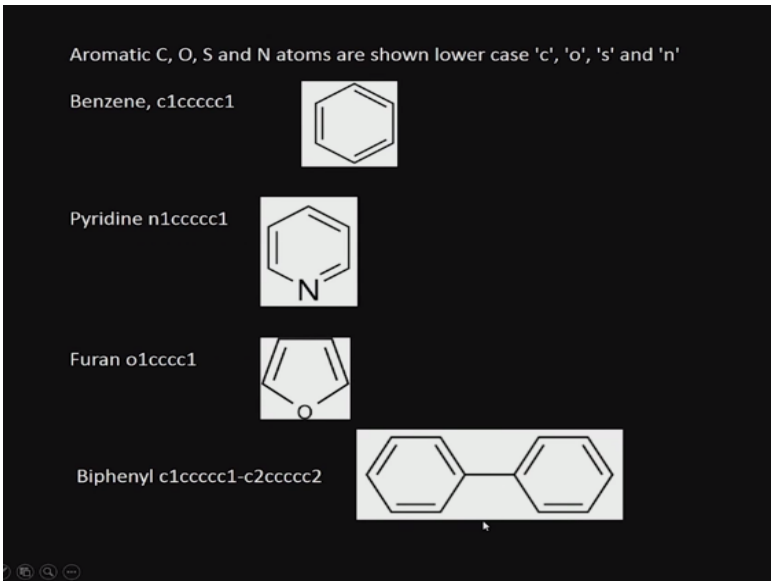
(Refer Slide Time: 16:49)



For example, if I take Cyclohexane I will say C1CCCCC1 there are 6 carbons okay and then when I put this 1 and 1 that means this and this are bonded so it gives you Cyclohexane whereas if I want hexane I will remove this 1 and 1 so I will get 6 carbons so it is a hexane. Dioxane, so I have O1 then CCO1CC and then again OCC okay 1. So this one and this one are connected okay. This O comes in the middle okay so it has got 4 carbons that is why you have 1, 2, 3, 4, and 2 oxygen chains and branches are described in parentheses.

So I want a ketone so I have 3 carbons and there is a double bond O on this carbon as well as there is a single bond O with this carbon. So as you can see I have an acid here okay this is a propionic acid so I have carbon, carbon, carbon. This carbon has a double bond O and this carbon also has a single bond O. For fluoroform so we have carbon there is a F with this carbon. There is another F with this carbon and this carbon and F are connected so fluoroform okay.

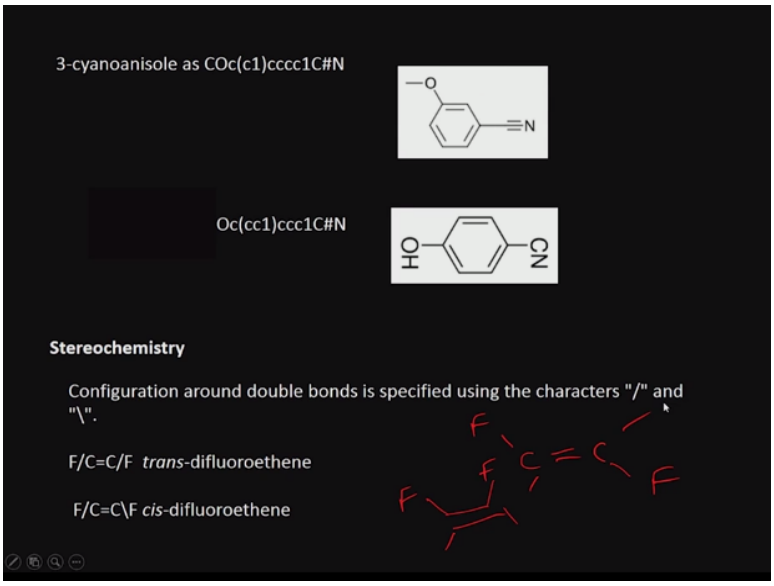
(Refer Slide Time: 18:14)



Aromatic are shown with lower case. So aliphatics are shown with higher upper case, aromatic so Benzene I use just like Cyclohexane, but only thing is instead of a capital C I use small c. you understand. So Pyridine I start with nitrogen 1 then I put 5 carbons and then the last one I put 1. So this carbon and this N are connected and because they are all in lower case it is an aromatic system.

Furan okay it is a 5-membered oxygen okay so I start with O1. I put 4 carbons and then I put 1 so this carbon and this O are connected. Biphenyl okay 2 benzene rings connected like this so I will close this benzene ring then I will have another benzene ring and then I connect them understand this one.

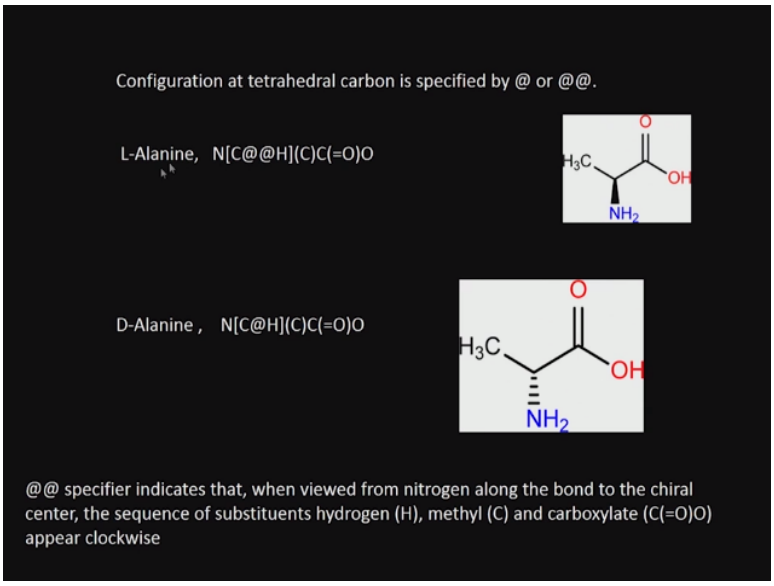
(Refer Slide Time: 19:31)



And 3-cyanoanisole this is the structure so I have a benzene ring okay and then there is a O CH3, connected and there is a carbon nitrile C triple bond N connected. Triple bond is put like this. Double bond you can put, but triple bond is put like this so you can see CO that it is starting from there from that corner, this corner, COC that means you are coming here and then this C1 is like this and then you have CCCC1C triple bond N okay.

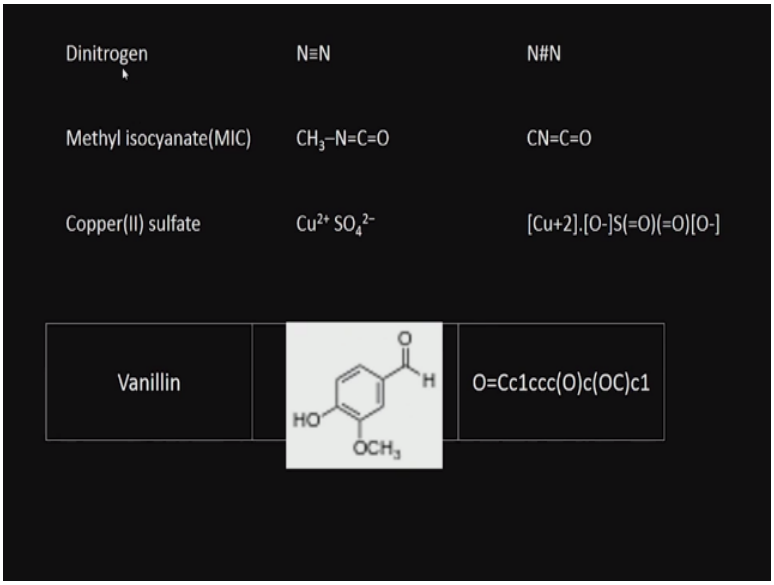
So this C1 means here and then CCCCC triple bond N they are all connected. If you have cyanide and OH here CO here this correspond to this and then CCC1 okay, then CCC1 then C triple bond N. So if you have stereochemistry configuration around double bond is specified using this or this. So if we put like this it means trans-difluoroethene that means these F and F are in the trans form that means okay it is in the trans form okay if I put like this that means F and F are this cis form okay.

(Refer Slide Time: 21:03)



Tetrahedral carbon is specified by like this or like this okay so if you have L-Alanine so you have this bond coming out so you have 2 of them if the bond is going N you have 1 of them. So when you have 2 specifier indicates that viewed from nitrogen along the bond to the chiral center okay so you can see this very clearly D-Alanine so we have the carbon at the hydrogen inside.

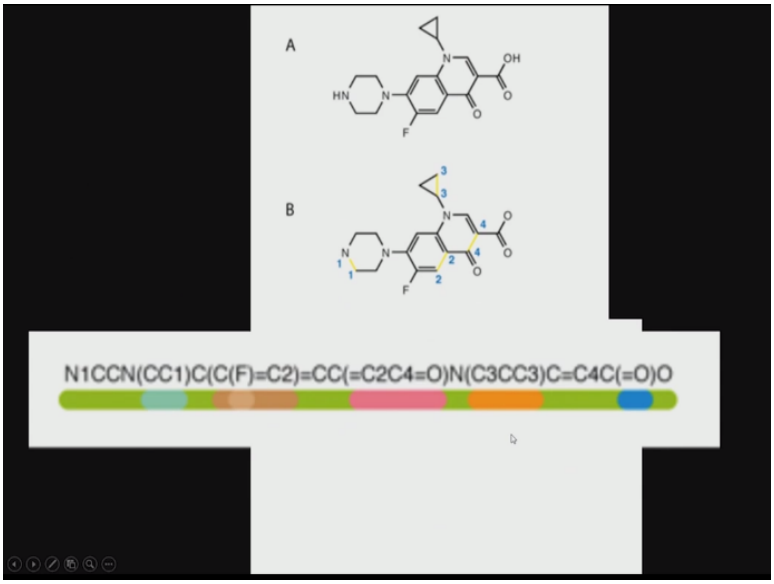
(Refer Slide Time: 21:34)



Dinitrogen and triple bone N is shown like this. Methyl isocyanate okay methyl isocyanate as you can see here CH3-N double bond CO so it shows CN double bond C double bond O it is very simple. Okay what about inorganic copper sulfate CU2 + SO4 so you put like this CU + 2 your square bracket. There is 1 bond O with the - and then we have S then S is connected with a double bond O and again double bond O.

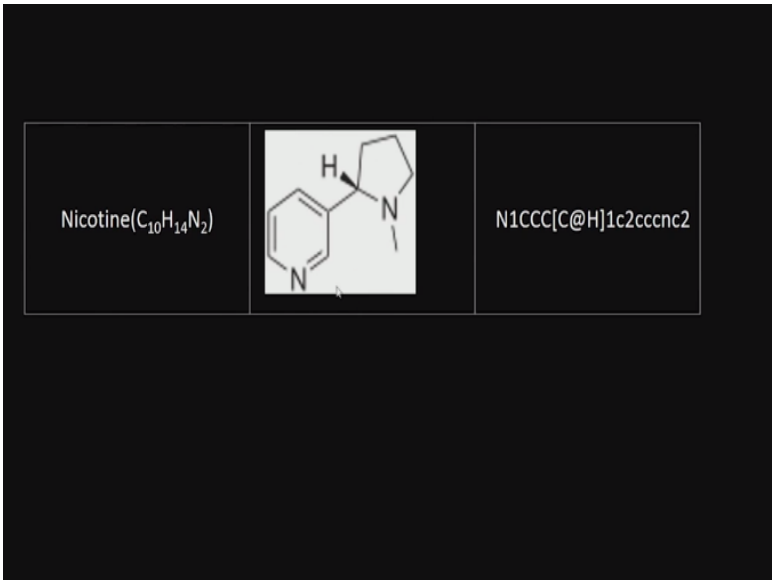
And then there is a square bracket for O - you understand. Vanillin. So we have this O okay double bond C that is this connection and then we have the C1ccc and there is a O connected here and then there is a C again O C that is this connection and this is closing here understand. Okay this is shown here and this is shown here.

(Refer Slide Time: 22:36)



So look at this you can spend your own time and read so you open up somewhere and then you start from here so nitrogen 1 from 1 corner you are starting and drawing it okay. So it is beauty of this SMILES.

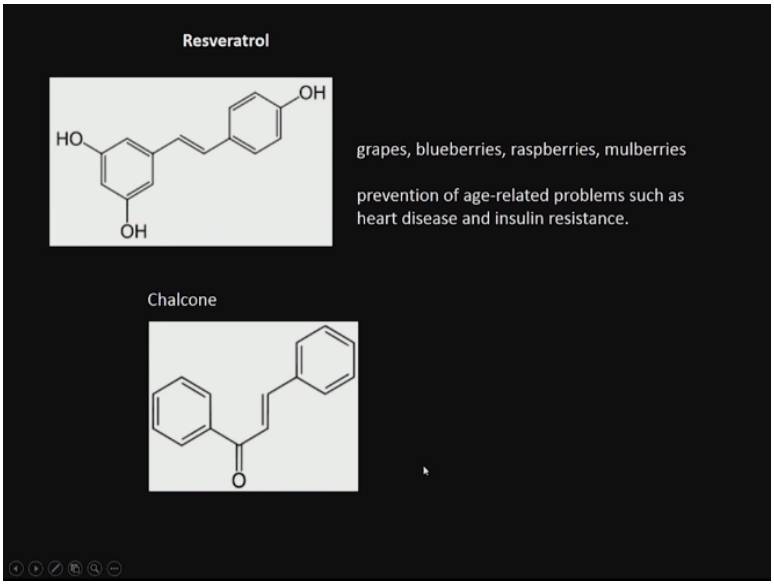
(Refer Slide Time: 22:49)



Nicotine. So we have the nitrogen. Here this nitrogen okay nitrogen 1 and then we have lots of carbon. This carbon, these carbons, carbons and nitrogen here and then there is a connection here so we have 2 rings here so this is related to the ring 1 and this is an aliphatic system so the carbons are capital CCC and then with that at H okay and then there is a closure here and then this is the aromatic system okay they are connected.

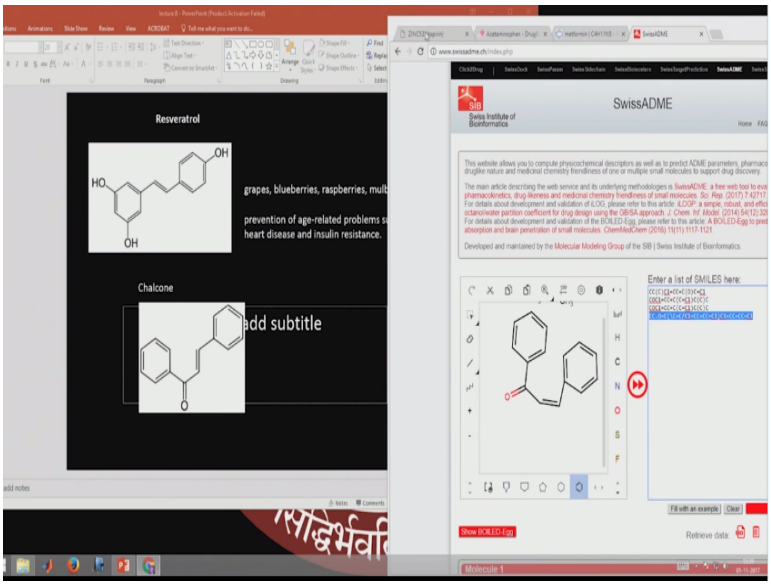
We understand this represents the aromatic system, this represents the aliphatic system and so we have 2 rings here and then C @ H represents this.

(Refer Slide Time: 23:35)



Resveratrol it is found in grapes, blueberries, raspberries. It is supposed to have antioxidant activity it has got very good prevention related to age related problems, heart disease, insulin so lot of things are there. So I want you guys to practice it or we can even check it out here and look at the SMILES notation of this. You can look at this SMILES notation of Chalcone also okay. So let us try that. We can just try this SMILES notation here. So we can use this particular software to do that and I want to show you guys how it is done and.

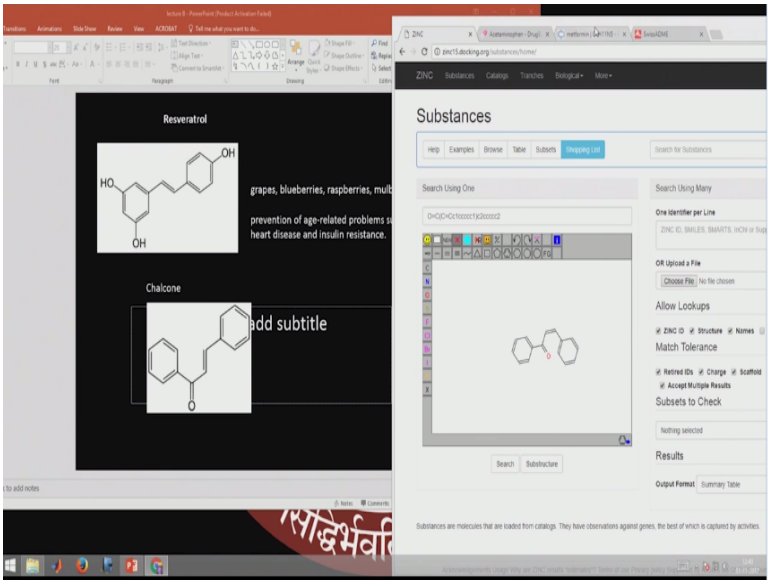
(Refer Slide Time: 24:37)



So Chalcone we can draw so we have benzene here. We have single bond. Then we have another single bond, then we have a double bond made out of this so we have this portion here then we have this portion here okay. Sorry little bit delays on that so we have this, this, this, this done, this done so we need to add a benzene ring okay.

So this is Chalcone so I want to see this new structure. So you can see this. So this is the SMILES notation for Chalcone here okay clearly. It is given here. So like that we can draw structures. We can create. Then the other advantage is we can copy SMILES notation from 1 software to another software, for example.

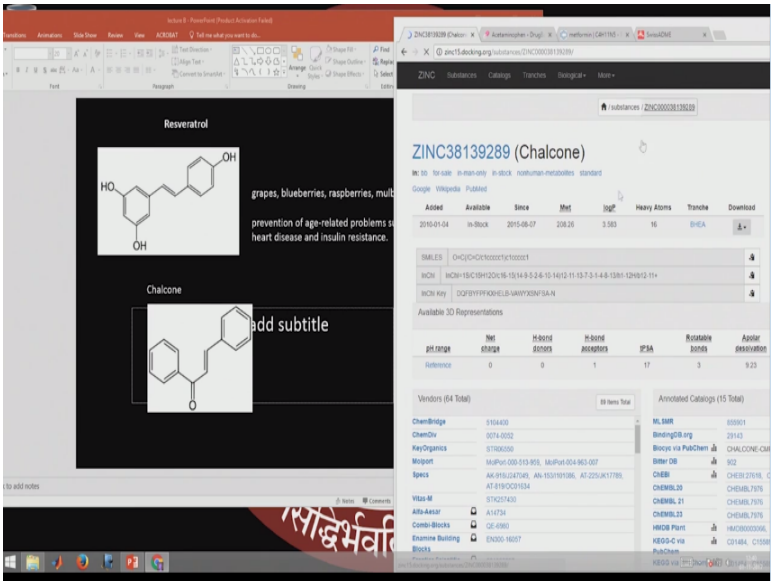
(Refer Slide Time: 25:48)



I can go to ZINC database and as you can see substances and we can put in the SMILES notation. So we can get the SMILES notation okay ignore this part okay. Okay so we can get the SMILES notation copy it from 1 and then we can get this all. Then we can try to get properties of various molecules okay. Once we draw the ZINC database we can use other database and find out if there are any properties existing. So SMILES notations we can copy and get the structure of molecules okay.

So that is very, very simple to do rather than actually drawing structures. So we can move molecules from 1 database to another using this SMILES notation okay so you can see this SMILES notation for this particular molecule. Okay so we can move this SMILES notation from 1 molecule to another as you can see here. So lot of these type of details is given okay so if we can move SMILES notation from 1 software to another and get properties if they are available or we can search for similar structures if they are available okay.

(Refer Slide Time: 27:18)



Okay so we have C 4 hydroxy Chalcone. So those structures are there Chalcone okay. So we can look at this Chalcone. We may get more properties from this particular database. So this type of SMILES notations so you can get some details about this particular molecule okay. Properties, hydrogen-bond acceptors, polar surface rotatable bond and so SMILES notation that way is very, very useful.

It is just a string of characters which describes the structures and we can move this SMILES notation from 1 database to another to get if any information is available in another database. So instead of drawing the structures we can use this SMILES notation. So SMILES notation is extremely useful once more as you said atoms are represented by stranded abbreviation so for bromine, carbon, nitrogen, oxygen, phosphorus, sulfur, fluorine, boron.

We do not need to have brackets okay and hydrogens are implicitly assumed if it is an aromatic we use a small lower case it is an aliphatic we use upper case. So if you have C1 and C1 coming in the beginning and end that means there is a bond. This is a ring okay and then if you are enclosing with a parenthesis that means this is a side chain so this is typically there is a ketone connected to the third carbon that is what it means?

And we can show stereochemistry we can show the trans forms by putting slash in this form you can show the cis form by putting slash in this form. We can also have okay we can also show the (()) (29:18) by putting either single symbol or double of the symbol okay so SMILES notation can represent many type of molecule and I showed you here if I use SwissADME and draw a structure automatically it gives you the SMILES notation.

And then we can use that SMILES notation and that it too different databases and try to get properties of that particular molecule. Okay so we will continue more on these drug properties in the next class as well. Thank you very much for your time.