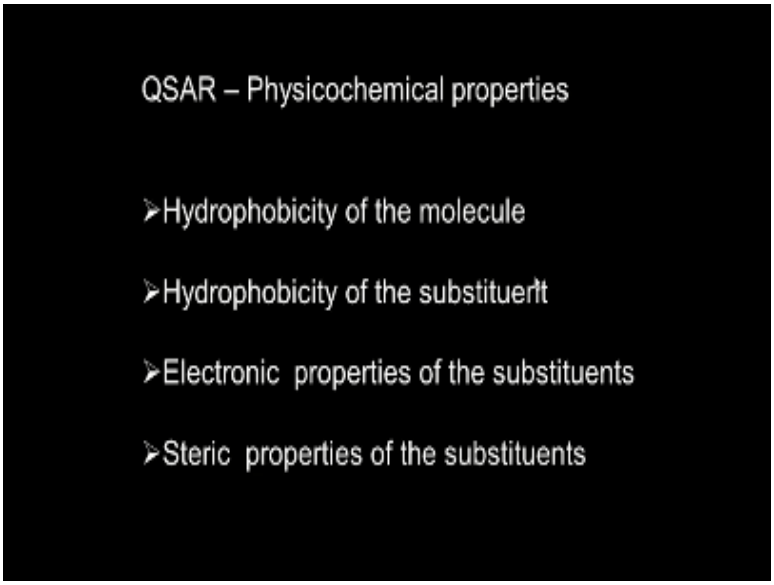
Computer Aided Drug Design
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology - Madras

Lecture - 27
Quantitative Structure Activity Relationship(QSAR)

Hello, everyone, welcome to the course on computer aided drug, we will continue on the topic of quantitative structure activity relation that is QSAR. So, if you look at the physicochemical properties there are four aspects that needs to be looked at.

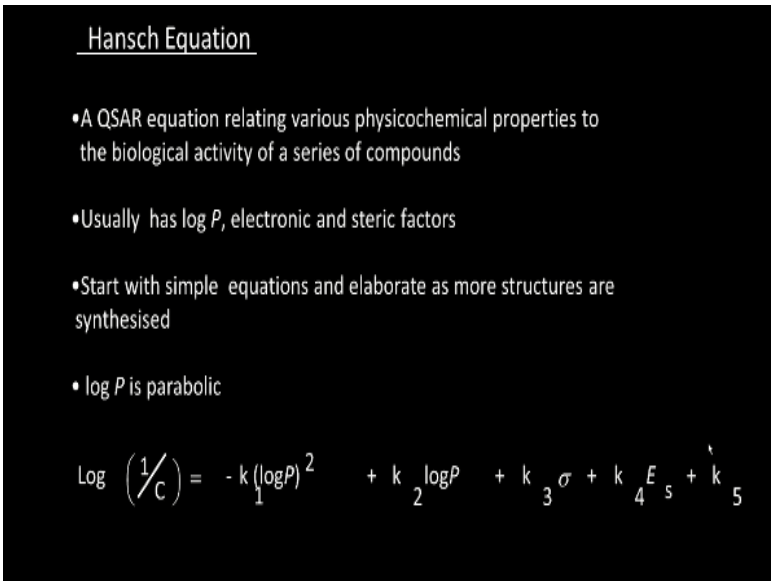
(Refer Slide Time: 00:33)



That is what we are looking at yesterday also the hydrophobicity of the molecule, and then you have different substituents right, so the hydrophobicity of the substituent that you are going to change electronic properties of the substituents, Steric properties of the substituents so all these play very important role in the physicochemical properties of the molecule. Okay so we need to understand these effects of these which in turn will affect the activity.

So, that we can develop a good quantitative structure activity relation that is the regression. So, the substituent can change the hydrophobic, hydrophilic balance the electronic properties, the substituents can change the steric properties of the parent molecule.

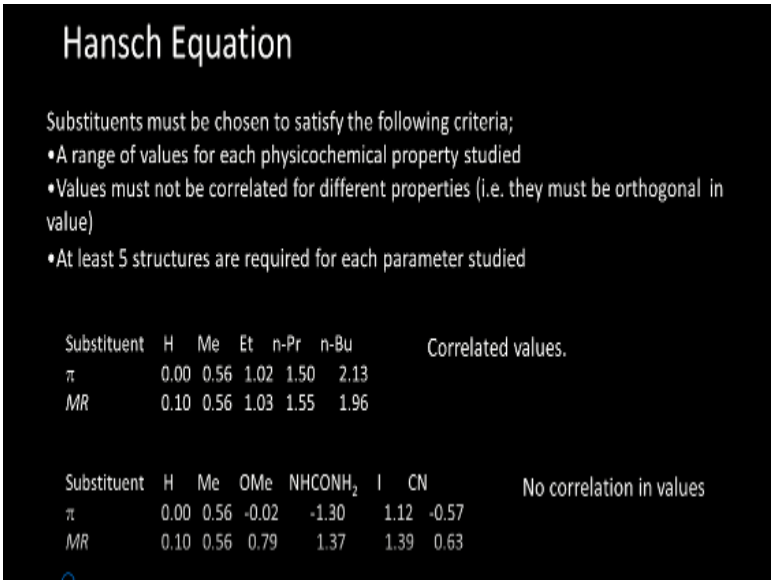
(Refer Slide Time: 01:21)



We looked at Hansch equation this is a equation relating to various Physicochemical properties to the biological activity, okay so we, we have the electronic, steric factors. So, you may start with a simple relationship like activity on the left-hand side and on the right-hand side log p we have a parabolic type of relation. It need not be all the time parabolic, but it can be linear also then we have the electronic, then we have the steric effect.

And this is a constant okay that is the Hansch equation. Generally, they suggest Hansch suggest this.

(Refer Slide Time: 02:01)



So, Substituents must be chosen to satisfy certain criteria, range of values for the

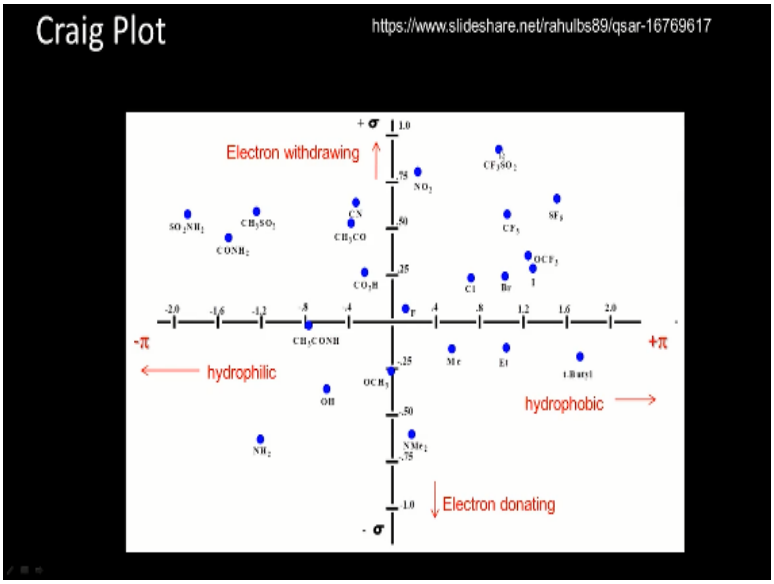
physicochemical property. So, I need to have which substituent to put into the para molecule to achieve this range of changes. The value should not be correlated that is very, very important that is also important and then now of course rule of thumb at least five structures for each parameter which we are studying.

So, for example if you get the pi this is related to the substitution this is Molar Refractivity, okay so Molar Refractivity is related to I mentioned in the previous class the density the molecular weight and so on and so look at this as pi increases Molar Refractivity also increases so obviously they are corelated with each other they are correlated as pi increases molar refractivity also increases okay, what is this pi, pi is related to the hydrophobicity.

Okay when the hydrophobicity increases MR also the volume and density also, so they are correlated. Now look at this set of substituents so a pi increases MR does not increase so you can see this unlike this. So, need to consider using this type of substituent in my para molecule rather than going this way that is what the conclusion is okay. So, look at it there is no correlation between pi which is related to the hydrophobicity these are this MR.

The molecular refractivity okay negative pi means they contribute more towards lipophilicity positive means they contribute towards hydrophobicity.

(Refer Slide Time: 04:01)

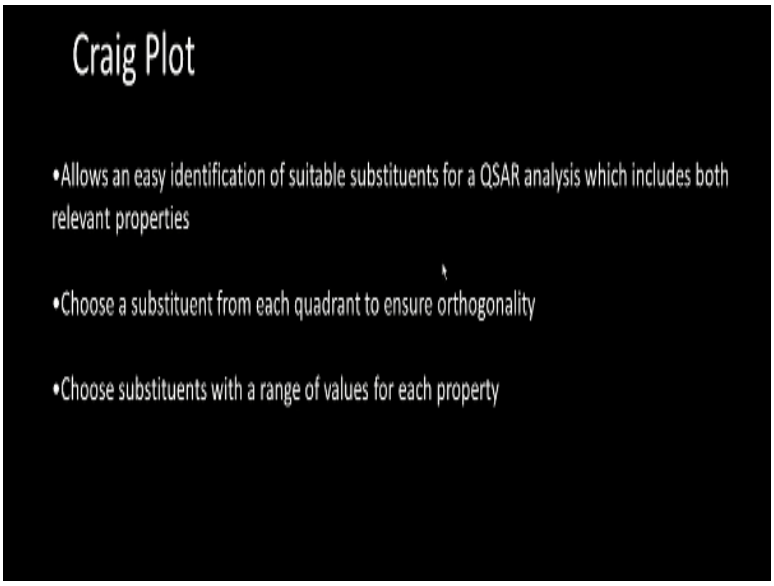


There is something called as Craig plot which I showed you in previous class this is taken from this particular reference. So, you have the electron withdrawing here electron donating here we have the hydrophobic here we have the hydrophilic. So, various substituents like chloro, rho or NO2 they all lead to increase in electron withdrawing increase in hydrophobicity whereas if you look in this quadrant we have electron donating and hydrophobicity.

These types of substituents if you look at this quadrant we have hydrophilicity as well as electron donating OH type of system NH2 OH. So, if you have they are more hydrophilic as well as they donate the electrons. If you look at in this quadrant we have CONH2 CN COOH and so on. So, they are electronic drawing as well as hydrophilic, so this is a very interesting plot, so we can modify our parent compound by having substituents.

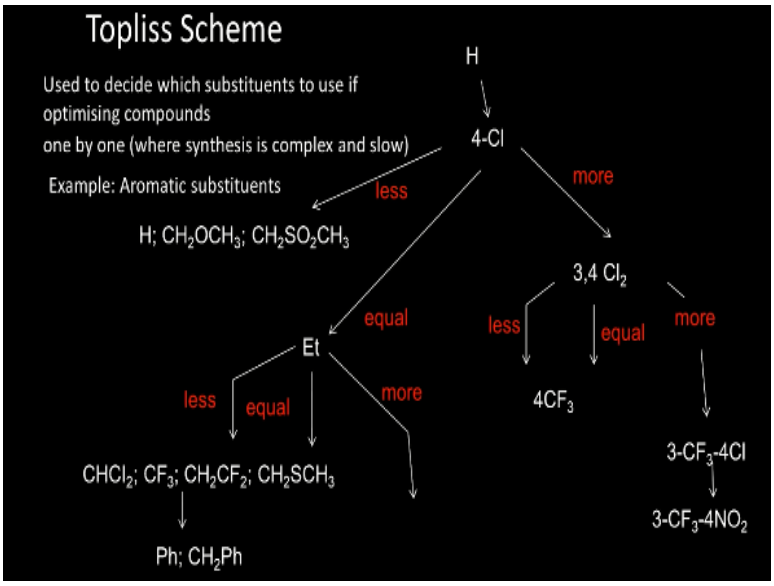
It covers all these four quadrants as well and it covers a range of electron withdrawing, electron donating hydrophobic, hydrophilic that is what this plot is all about okay as I said this is taken from this particular reference.

(Refer Slide Time: 05:29)



So, it allows an easy identification of suitable substituents for QSAR which includes both relevant properties. The substituent from each of the quadrant to ensure orthogonality and range of values. So, they can select lower and higher so that we can cover a big range of value okay.

(Refer Slide Time: 05:48)



Okay there is another approach that is called the Topliss scheme okay imagine we have a parent compound where we place the H in the 4 th row, so the activity may be increasing activity may be decreasing or the activity may be same okay so if it is more then of course you know that chloro in the fourth position has an effect. So, we can have a 2 chloro substituents at the third and the fourth again we can have a activity increasing activity equal and activity can even go down okay.

So, if it is increasing then you can think about having a fluoro okay and then we can even think of replacing with a nitro like that you no. If it less or equal, we may think of replacing the chloro with the fluoro. Now if you go here so by putting chloro it becomes less so you can think about having a CH2 CH3 CH2 like if its equal then we can think about the other substituents. So, this is very interesting a tree like conclusion we can arrive at.

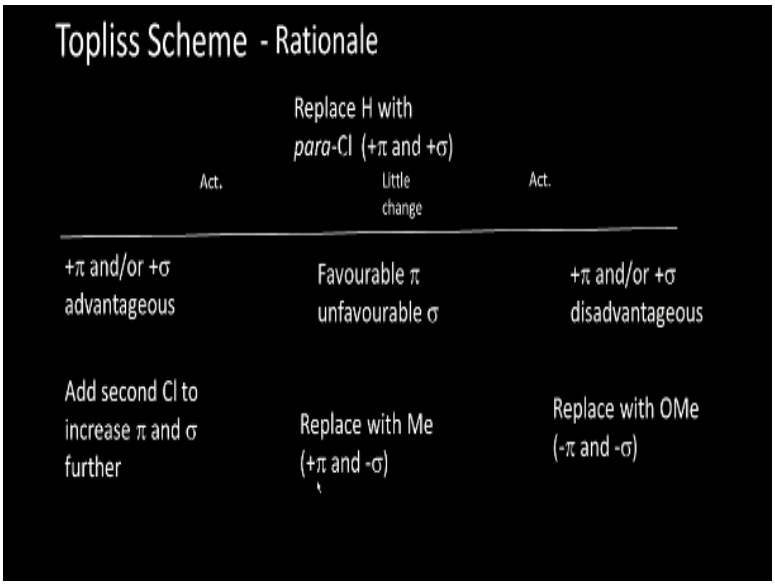
So, at each of the substituents we can find out whether it is increasing the activity or no change in the activity or less. So, by that way by using this tree we can slowly make changes to the parent molecule and achieve a highly active compound. okay so if you know that you are ending up like there is then you how do we change it to this and so on actually. So, that is called a Topliss scheme, so we used to decide which substituents to use if optimizing compounds.

So, we can do one by one synthesis so if the synthesis is very complex and slow this is a very

good approach with the synthesis is very simple you may just blindly substitute different functional groups you can use that Craig plot which I showed you before okay and then synthesis a large number of compounded various electron donating withdrawing hydrophilic, lipophilic whereas if it is a very long slow complex synthesis.

We can use this type of Topliss Scheme. So, after each new molecule synthesis we can go into the branch of the tree depending upon the activity is more or less or equal. So, this is very useful for aromatic type of substituents.

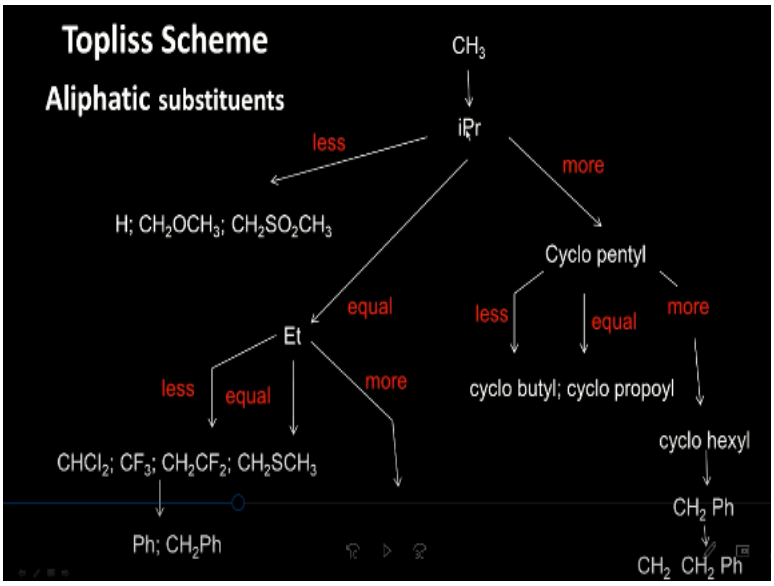
(Refer Slide Time: 08:27)



For an Aliphatic also we can look at we can look at that later so what is the rationale here. So, we are changing + pi and/or + sigma this is related to hydrophobicity this is related to electronic. So, activity increases. So, no change little changes okay +pi and + or disadvantages okay so we are adding second chlorine to increase pi and sigma further replace with Methyl. Okay so Methyl.

When you do you get a -sigma but we maintain the +pi and replace with o Methyl getting both of -pi and – sigma okay so this is the logic by which it goes actually okay. So, further changes suggested based on the arguments of pi and sigma and steric. Okay this is the logic by which or the rational by which changes are made once you start with the para chloro and see what is happening.

(Refer Slide Time: 09:42)



Similar you can do for Aliphatic Substituents also so what you do imagine you have a CH3 replace the CH3 with Iso propoyl so activity can increase activity can be equal or activity can be less so depending upon the situation if it is more so what do you do instead of iso propoyl you may replace with cyclo pentyl so if the activity increases you can go to cyclo hexyl then you put into Ch2 phenyl. Okay then CH2 CH2 phenyl then you end up with a highly active component.

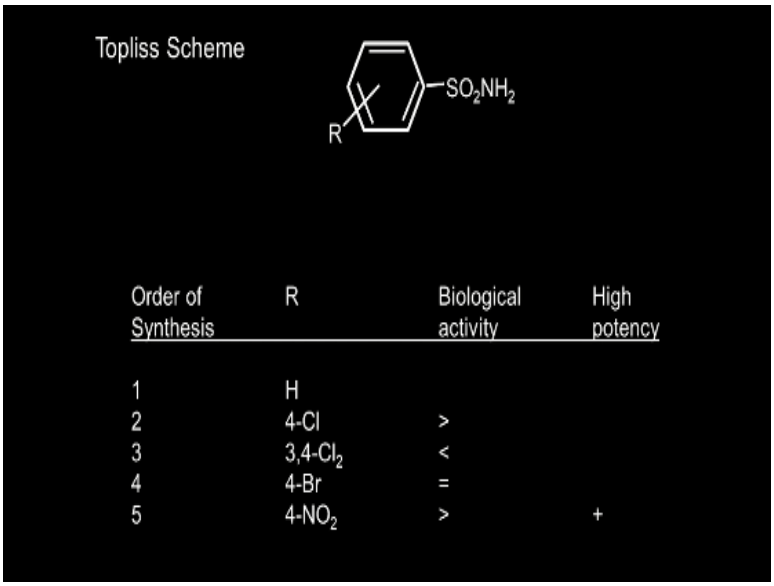
But the activity does not change after cycle pentyl you may go to cyclo propoyl. If the activity decreases so you reduce from pentyl to butyl cyclo butyl okay this is the logic here. Now when you put iso propoyl there is no change in the activity so that is equal so what do we do there we put in the CHCL2 CF3 CH2CF2 CH2SCH3 and so on activity may go down or activity may be equal.

So, you plan accordingly if the activity is less by putting iso propoyl then what do you do you put H because obviously when you make it more hydrophobic it is not working so you go to H just remove it or you can make CH2 OCH3 try to make it a little bit hydrophilic as you can see in his and hopefully you may get a more active compound than the original parent compound Okay this is how Topliss scheme works actually this is ideal if you have very complex.

And this is slow and very time consuming synthetic protocol. Whereas if the synthesis is very fast I would go and look at the Craig approach and synthesis large number of molecule

substituent molecules and test their activity.

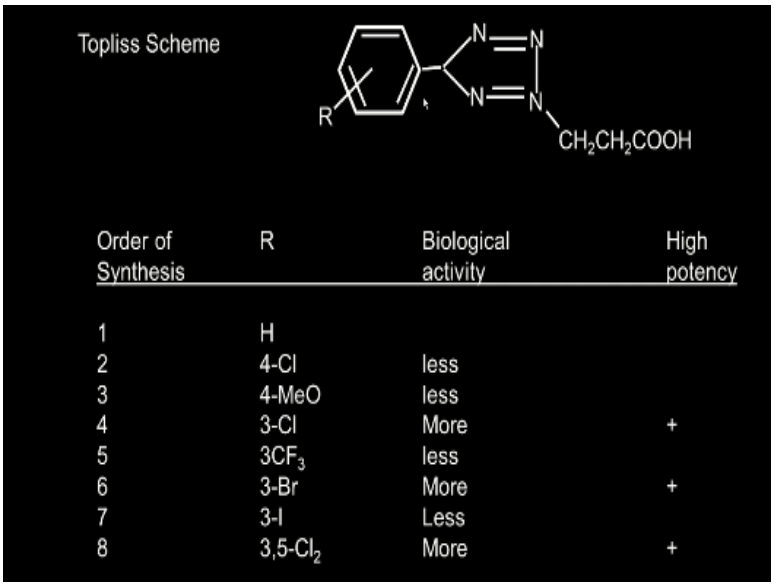
(Refer Slide Time: 11:37)



Order of Synthesis	R	Biological activity	High potency
1	H		
2	4-Cl	>	
3	3,4-Cl_2	<	
4	4-Br	=	
5	4-NO_2	>	+

So, the logic as I said if you look here you put R ads H the activity that is your starting then you put 4 chloro activity increases then you put 34 chloro okay 3 and 4 chloro then the activity decreases then you try 4 Bromo activity same put 4 nitro biological activity is very good. So, you would have achieved a molecule and by following this particular logic.

(Refer Slide Time: 12:09)



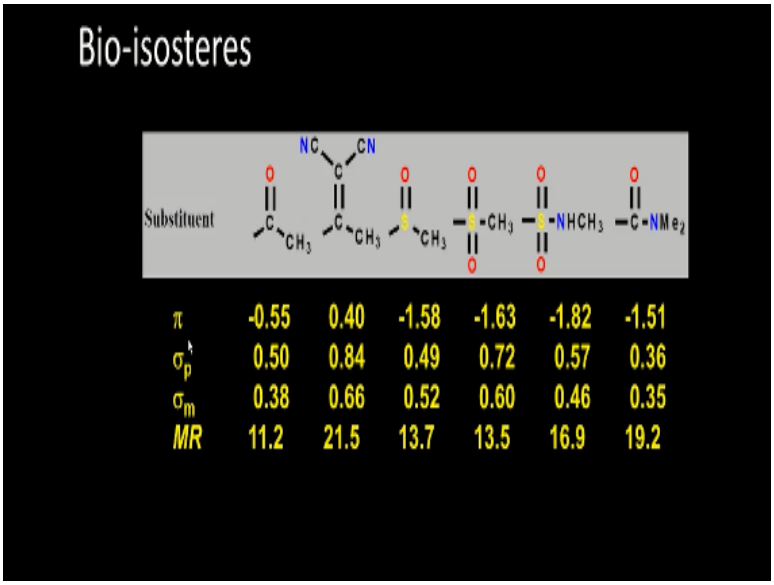
Order of Synthesis	R	Biological activity	High potency
1	H		
2	4-Cl	less	
3	4-MeO	less	
4	3-Cl	More	+
5	$3CF_3$	less	
6	3-Br	More	+
7	3-I	Less	
8	3,5-Cl_2	More	+

Okay again look at his you have a R here imagine this is your molecule R starting with H then you put 4 chloro activity less. So, what do you do you try 4 Methoxine again the activity is less you tried more hydrophilic by putting oxygen, but it does not. So, you try 3 chloro here activity

has increased so high potency. So, okay so this position is better than this position so here you do 3CF3 activity is going down.

So, instead of chloro you put Bromo again the activity is high 3 IDO activity is less. So, chloro is okay Bromo is okay so what do you do 3. 5 chloro activity s good so you found some active compounds. Okay by following this type of logic here.

(Refer Slide Time: 12:59)



You can also replace with Bio isosteres okay we talked about Bio isosteres long time back. So, o looks at these groups of molecules it is got a pi that is related to hydrophobic hydrophilic balance it is negative here because it has a O here. We have a lot of carbons that is why it is positive this is becoming more negative here okay so sigma for this look at the sigma is like this. MR for the molecule Molecular Refractivity is like this.

Sigma for para sigma for Methyl sorry not Methyl Meta substituents. So, so you can check it out whether there is any correlation cross correlation between and then we can make the substituents accordingly. We can select either I want to have Hydrophobic system or a Hydrophilic system it is quite hydrophilic here. So, Bio isosteres understanding the Bio isosteres is very, very important.

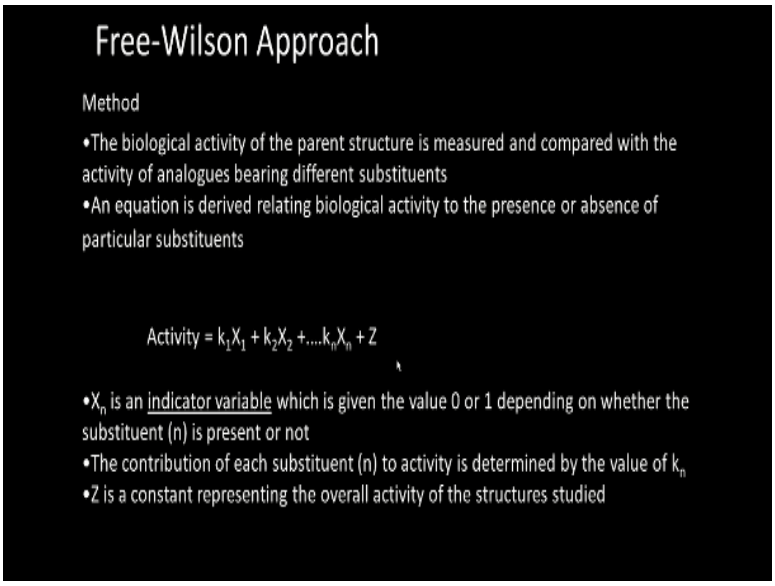
If you want to substitute new functional groups okay to improve the activity. Choose substituents

with e similar physicochemical properties okay look at Cn NO2 and COMe they could be Bio isosteres. Choose Bio isosteres based on most important physicochemical property. Okay what is very, very important is the electronic feature very important or the hydrophobic is very important or molecular refractivity.

Molecular Refractivity as we said is volume and density that means the sizes so depending upon which you will realize it is important you can choose by your system based on most important physicochemical. For example, COMe SOMe. COMe SOMe are similar look at this sigma p when it is substituted in the parabolic position. SOMe and SO2Me are similar in pi as you can see here hydrophobicity not electronic.

This is electronic this is hydrophobicity. Okay depending upon which physicochemical property you think is important we can select so BIO isosteres can be very, very useful when you are synthesizing new molecules with different substituents.

(Refer Slide Time: 15:37)

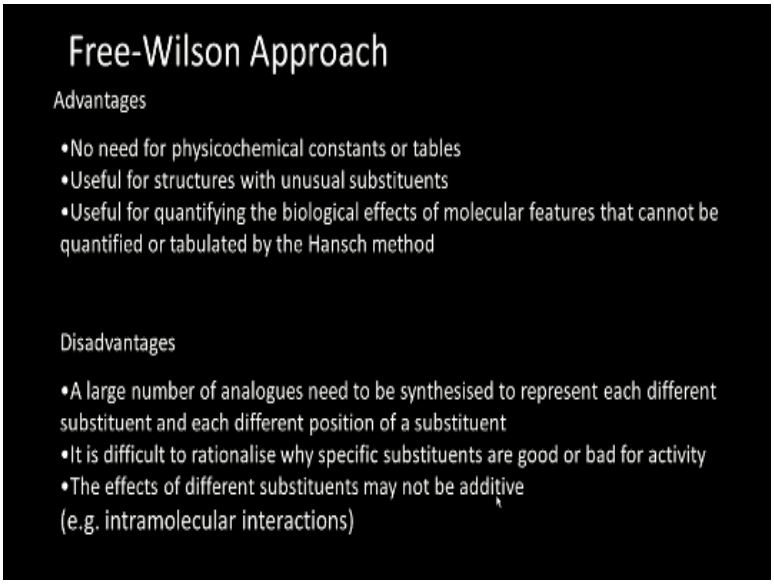


Then there is something called free Wilson approach okay free Wilson approach. The biological activity of the parent structure is measured compared to the activity of analogues bearing different substituents okay that is free Wilson. An equation is derived relating biological activity to the presence or absence of particular substituents. Activity=K1x1+k2x2+ knxn and so on knxn+z okay z is the parent molecule and then when I put different substituents.

How the activity changes Xn is an indicator variable which is given the value 0 or 1 depending on whether the substituent is present or not. So, if it is not present we will make K10 otherwise if it is present k1 will be 1. The contribution of each substituent n to activity is determined by the value of kn. Okay so, this kn tells you whether It is contribution. Z is a constant representing the overall activity of the structures studied.

Okay this is called a free Wilson approach. Okay so what we do we have say different substituents coming into the picture so if it is present k1 or k2 can have a value otherwise it will become 0. The contribution of each substituent n to activity is determined by the value of kn the contribution. Z is a constant representing the overall activity of the structures studied.

(Refer Slide Time: 17:17)



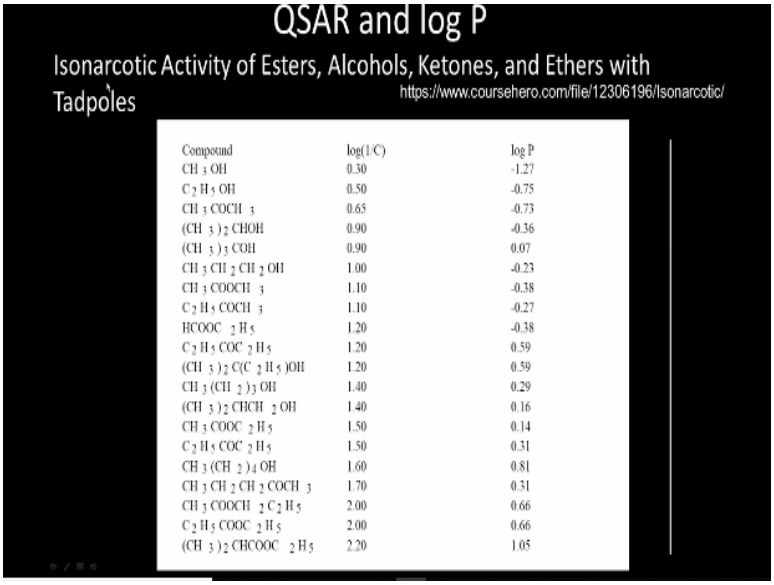
So, what are the advantages no need for physicochemical constants or tables. Useful for structures with unusual substituents. Useful for quantifying the biological effects of molecular features that cannot be quantified or tabulated by the Hansch method. So, for each of these complicated substituent we can find out what is its contributions okay. So, here we need to have a large number of analogues to be synthesized to represent each different substituent.

And each different position of a substituent that is also important. Okay that is the big problem it is difficult to rationalize why specific substituent are good or bad for activity. Because if you

look at the other approaches the hansch approach we know sigma the electronic fact or okay electron donating or withdrawing effect hydrophilic or hydrophobic effect. So, we can have some rational MR positive or low and so on.

So, we can have different logic for why a particular substituent is giving certain activity and also the effects of different substituents may not be additive that is very, very important. I put in say for example have 4 chloro then I put in 3 4 di chloro the effects may be different, so it may not be an additive effect. So, intramolecular interactions are always possible.

(Refer Slide Time: 18:41)

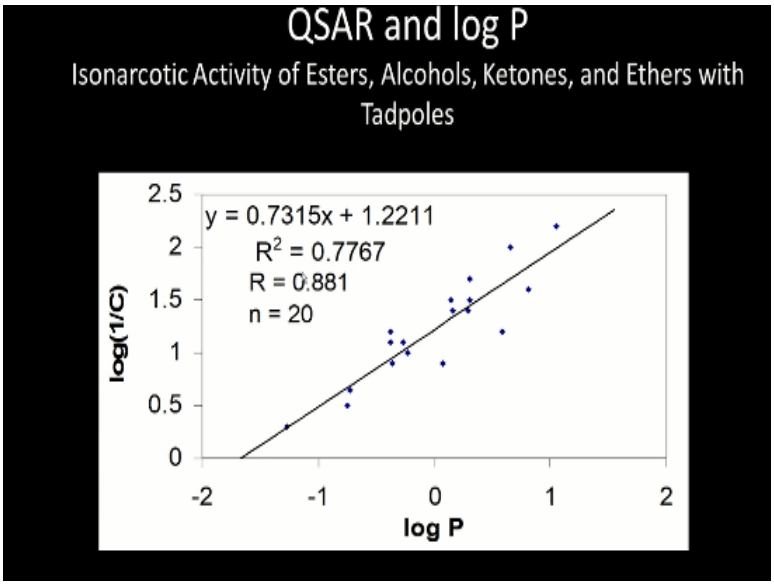


Compound	log(1/C)	log P
CH_3OH	0.30	-1.27
C_2H_5OH	0.50	-0.75
CH_3COCH_3	0.65	-0.73
$(CH_3)_2CHOH$	0.90	-0.36
$(CH_3)_3COH$	0.90	0.07
$CH_3CH_2CH_2OH$	1.00	-0.23
CH_3COOCH_3	1.10	-0.38
$C_2H_5COCH_3$	1.10	-0.27
$HCOOC_2H_5$	1.20	-0.38
$C_2H_5COC_2H_5$	1.20	0.59
$(CH_3)_2C(C_2H_5)OH$	1.20	0.59
$CH_3(CH_2)_3OH$	1.40	0.29
$(CH_3)_2CHCH_2OH$	1.40	0.16
$CH_3COOC_2H_5$	1.50	0.14
$C_2H_5COC_2H_5$	1.50	0.31
$CH_3(CH_2)_4OH$	1.60	0.81
$CH_3CH_2CH_2COCH_3$	1.70	0.31
$CH_3COOCH_2C_2H_5$	2.00	0.66
$C_2H_5COOC_2H_5$	2.00	0.66
$(CH_3)_2CHCOOC_2H_5$	2.20	1.05

So, these are the disadvantages of this approach QSAR and log p log p is very important ISonarcotic activity of Esters. Alcohols, Ketones and Ethers against Tadpoles okay this is taken from this reference. As you can see the log p keeps changing the activity also keeps changing these are the various substituents groups that comes into the picture CH3 OH starting from that we keep on adding CH3, so it is more hydrophobic.

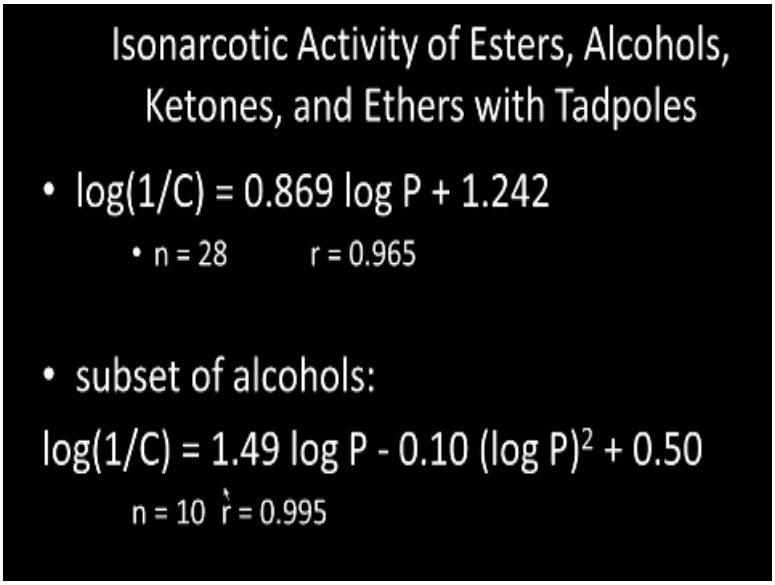
Okay activity has also become 1.05. So, log p plays a very important role.

(Refer Slide Time: 19:24)



So, we can plot same reference we can plot log p versus activity so as it becomes more hydrophobic the activity also increases the R square which is a measure of the fit of the data. So, this is a straight-line equation x is the log p here y is the activity here. So, the regression coefficient is 0.88 and R square which is a measure of the fit is 0.7767 which indicates a good fit okay.

(Refer Slide Time: 19:56)

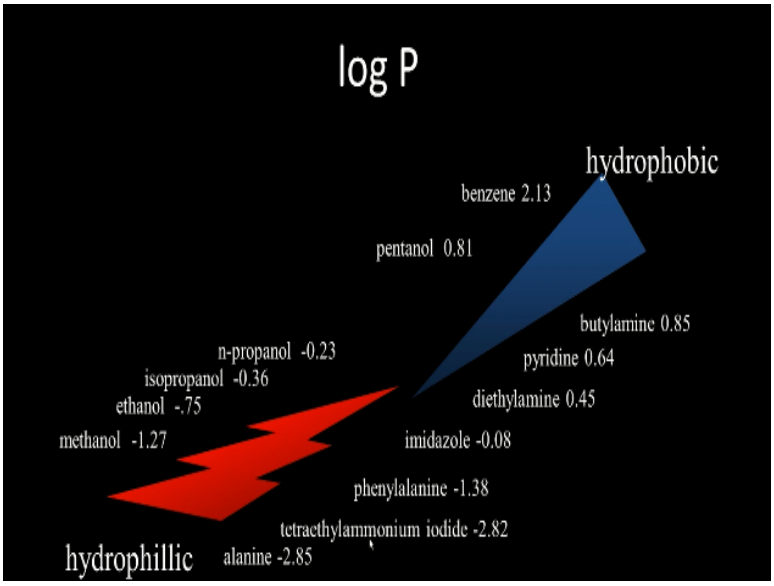


So, yeah this is for ketone so for a subset only alcohols, if you take only alcohols it may have a second order or a quadratic relation. So, you can see R going up much further. So, as I mentioned long time back QSAR is very good if you have a set of molecules when you extend it to other molecules sometimes it might not work actually. So, alcohols alone you can see a very good

relationship.

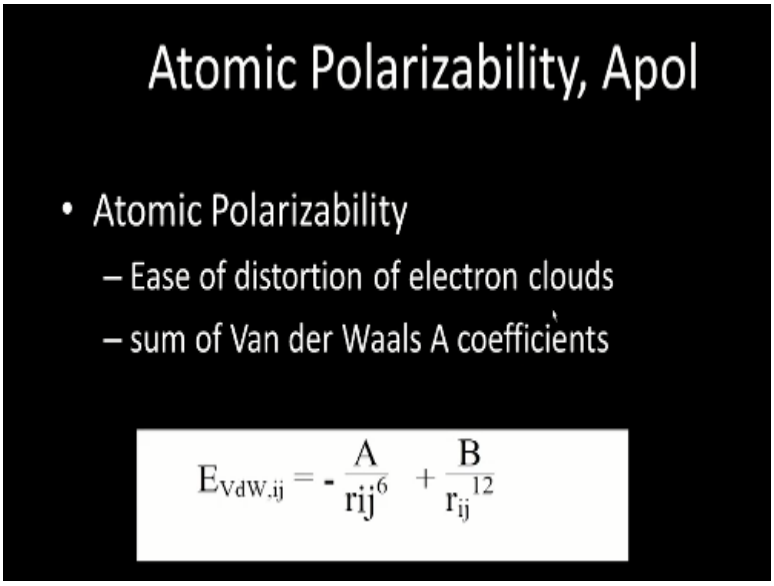
Of course, this is also reasonably good we have taken Ketone, Ethers with tadpoles, alcohol esters also. But if you have taken more hydrophobic substituents then the regression relation could be very poor.

(Refer Slide Time: 20:46)



So, , look at this log p hydrophobic this is very hydrophilic methanol, ethanol, isoporpa ethanol okay phenylalanine imidazole now this is becoming more hydrophilic more positive benzene quite high methanol quite high okay.

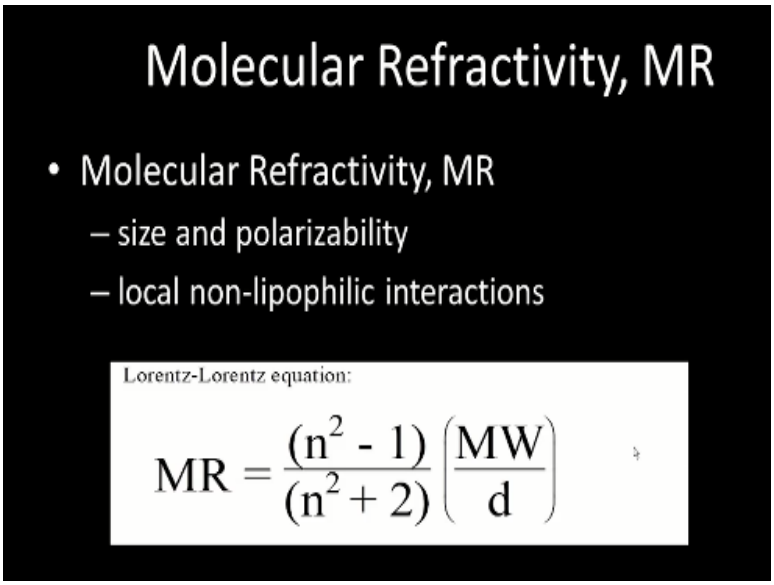
(Refer Slide Time: 21:10)



Log p is a very important thing because log p we have been talked long time back it tells you the solubility it tells you the GI absorption it tells you the HERJ HRG it tells you the plasma binding it tells you the membrane penetration. So, log p is dependent on so many log p effects so many different parameters. Polarizability is called atomic polarizability Ease of distortion of electron clouds okay sum of van der Waals coefficients.

If you remember Van der Waal equations/rij raised to the power 6 +B/rij raised to the power 12 okay I hope you remember that.

(Refer Slide Time: 22:04)



Molecular Refractivity of course is defined before also. Molecular refractivity talks about the size and also the polarizability so molecular weight/ the density. Okay that is called the molecular refractivity this is another important descriptor this is another important descriptor they will come quite a lot in many of your QSAR equation as you go along in your activity.

(Refer Slide Time: 22:29)

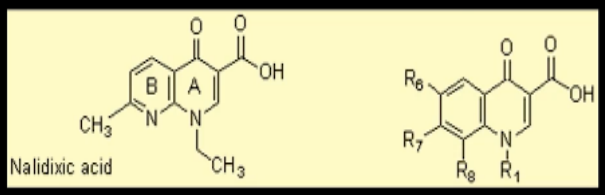
Group Additive Properties, GAPs

Substituent	Volume (SA)	MR	π	Rot Bonds
-H	1.48	0.10	0 (reference)	0
$-CH_3$	18.78	0.57	0.56	0
$-CH_2CH_3$	35.35	1.03	1.02	1
$-CH_2CH_2CH_3$	51.99	1.5	1.55	2
$-CH(CH_3)_2$	51.33	1.5	1.53	1
$-CH_2CH_2CH_2CH_3$	68.63	1.96	2.13	3
$-C(CH_3)_3$	86.99	1.96	1.98	1
$-C_6H_5$	72.20	2.54	1.96	1
-F	7.05	0.10	0.14	0
-Cl	15.85	0.60	0.71	0

So different substituents okay volume, Molecular refractivity okay electronic feature okay sorry this is not electronic feature pi represents the hydrophobicity and then this is number of rotative bonds. So, all these can have effect on your QSAR relationship okay.

(Refer Slide Time: 23:00)

Nalidixic acid analogues - Nalidixic acid (Negram) an antibacterial agent effective against gram-negative bacteria and used primarily for urinary tract infections, ~weak activity

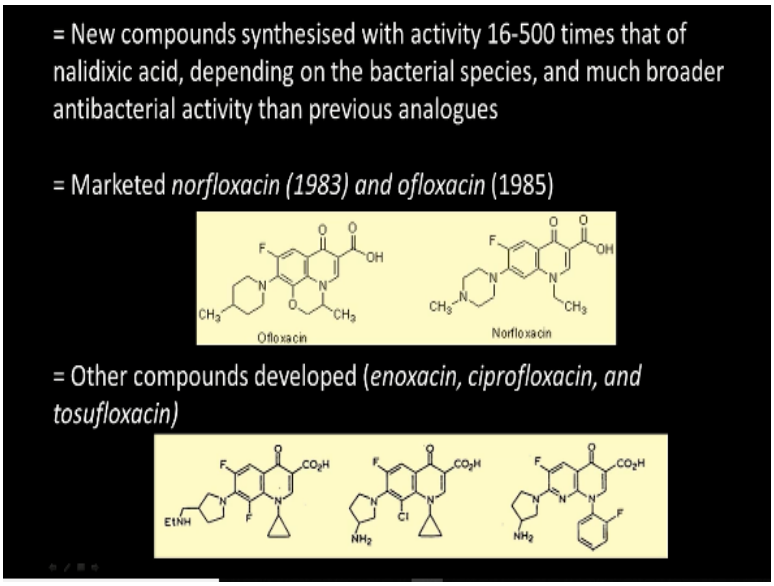


- pyridone ring A is essential for activity
- but ring B could be varied

Let us look at some simple systems also if you know this is called Nalidixic acid Okay this is a commercial anti bio ring this was discovered few decades back this is given for gram negative bacteria and used mainly for urinary tract infections it is visibly weak activity. The structure activity studies are shown the pyridine ring A this is the pyridine ring A is essential for activity and this ring can be changed.

So, we can have kept constant and we can have these things changed quite in order okay. So, we can have a large number of compounds synthesized following this approach.

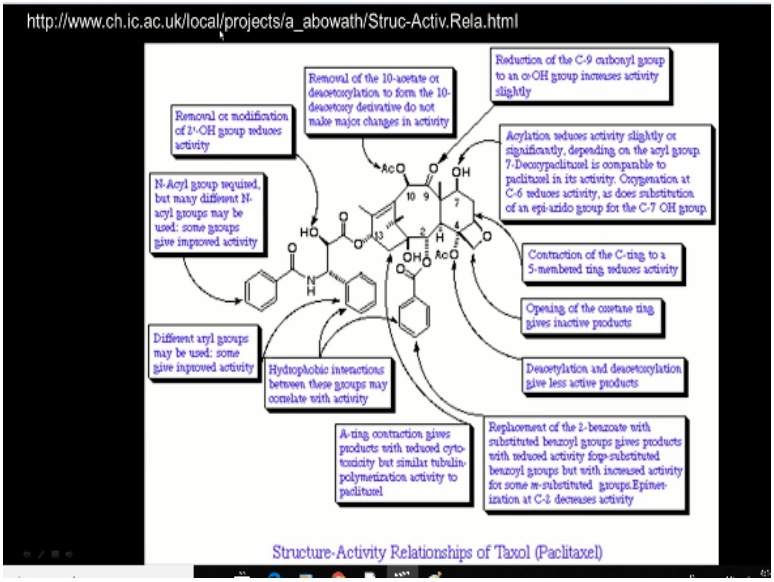
(Refer Slide Time: 24:00)



Many compounds were synthesized by modifying the ring b they were found to be almost 500 times active than the original nalidixic acid which was originally discovered may be about 15 to 20 years back and the normal compounds are much broader antibacterial activity than the original okay. So, in 1983 norfloxacin we can see this the ring A is kept constant and which change on this ofloxacin, ciprofloxacin tosufloxacin, enoxacin.

So, all these are new molecules and that were synthesized starting from the original Nalidixic acid here you can see here, and you can see that that this portion is constant, and you get a lot of changes on the other B ring and you end up with very highly active molecules okay and more potent and more wider range of activity. Okay that is the beauty of understanding the structure activity relationship here that is very, very important actually.

(Refer Slide Time: 25:20)



So, this is taken from this particular reference it is a very interesting picture okay Taxol or paclitaxel this an anti-cancer drug originally it was found from a natural product called Yew trees it is found in some European and also in Himalayan region this tree is found. The original molecule like anti-cancer property molecule had poor bio availability and it had very poor water solubility.

So, the second generation a lot of changes were done to the parent molecule to improve the water solubility and then later on salts of this was also made to improve its water solubility. So, the current Taxol has better bio availability and water solubility unlike the original and so how this was performed because of the understanding and the importance of each one of the functional groups and the positions.

So, look at this removal or modification of OH group reduces activity okay. Hydrophobic interactions between these groups may correlate with activity that means you can have seen two Benzene rings so there could be a Benzene Benzene stacking. A ring contraction gives products with reduced cytotoxicity but similar tubulin polymerization activity to paclitaxel. Opening of this ring gives inactive product contraction of the c ring to a 5 membered ring reduces activity.

A acetylsation that means remove acetylsation gives less active product. SO, by knowing what are the important groups that are necessary for a activity what are the groups that might not

really have a play in activity we can develop a new structures by modifying those places. Reduction of the C9 carbonyl group increase to an okay Alpha OH group increases activity slightly.

Removal of the 10 acetate or deacetylation to form 10 deacetoxy derivative do not make major change in activity okay so these positions do not have much effect. Acylation reduces activity slightly or significantly depending on the Acyl group. So, you see some locations have a major impact on activity some locations have minor impact on activity. So, one with the medicinal chemistry synthetic chemistry knowledge can come up with large number of new derivatives.

And check out their activity and because as I said original N paclitaxel has poor bio availability and poor water solubility and although changes have been made to come up with new generations of paclitaxel sill there is a lot of research that could be done in coming up with new changes. So, this slide was taken from this particular reference. So, we will continue further on this topic of QSAR in the next class as well. Thank you very much for your time.