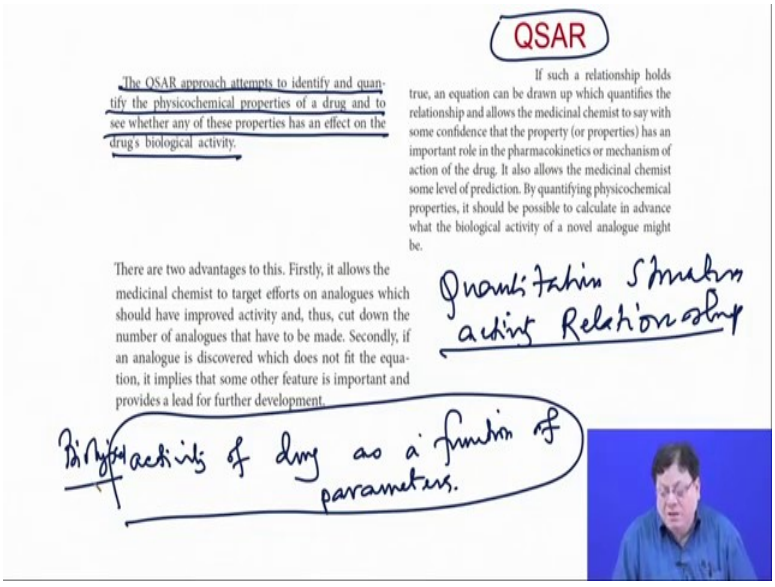
Organic Chemistry in Biology and Drug Development
Prof. Amit Basak
Department of Chemistry
Indian Institute of Technology, Kharagpur

Lecture – 63
QSAR Principles

Welcome back. In the last session, we have discussed the pharmacokinetic pharmacokinetics of the drug discovery process. Now we will discuss the very important aspect of drug discovery process. Although it appears to be little empirical but still these empirical things are useful like Lipinski's rule.

So, many empirical rules are there which are actually derived out of the experimental results. There is no mathematical derivation for those empirical rules but they are useful. One of them in drug discovery process is quantitative structure activity relationship (QSAR).

(Refer Slide Time: 01:31)



Which is quantitative structure activity relationship? This is activity relationship QSAR. QSAR try to formulate the activity of a drug by an equation. One thinks that they are very important for the drug to be active.

We are trying to say that activity of a drug is a function of parameters. Here activity means the biological activity. In case of biological system it is very difficult to have an equation.

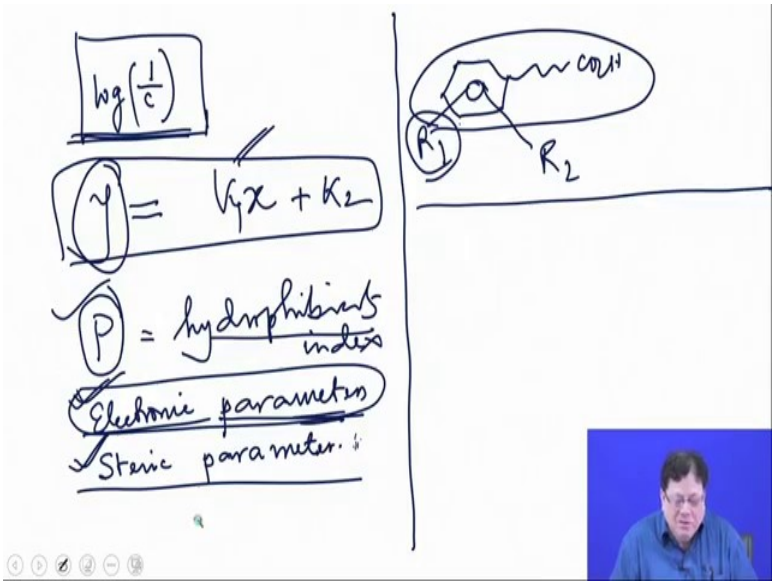
In computation, you have different molecules and you can tell that the pathway followed by a particular molecule in a particular reaction. There are ways to define the chemical activity of the molecule by computational studies. That has progressed rapidly in the last few years because of high power of the computation.

In case of biological activity, the situation is more complicated. We are in the age of this empirical rules because the situation is so complicated. So, ultimately through various experimental results you come up with an equation. I can compare this with Woodward's Diene rule of calculating the λ_{max} of an of a Diene functionality.

There is no basis. He just checked all the uv λ_{max} of various compounds. Finally, he made a rule.

How do we express the biological activity? We are concerned about biological activity. Biological activity is going against different targets. People have decided that better to do this different parameters IC_{50}, ED_{50}.

(Refer Slide Time: 05:36)



So, biological activity is expressed as a logarithmic term log (1/C). C is the concentration needed to have the desired effect up to a certain extent. Why you are taking 1/C, the

reciprocal? Because we know the more concentration is necessary to have the desired effect. It has got a reciprocal relation. Usually these activities are dependent on parameters, they usually follow some logarithm.

Logarithm is not a direct relationship. The activity is proportional to the hydrophobicity. In a linear equation, you have a x part and y part.

$Y= k_1x + k_2$

y here is log(1/C). Why is it (1/C) and why it is logarithmic? There cannot be any direct relationship like that. Now you have to identify the parameters. One has to be P, the hydrophobicity or lipophilicity. If it is extremely hydrophobic or polar it will be extremely soluble in water. As the molecule goes inside the GI tract having water at different pH the molecule will remain in the water and it will never be absorbed. On the other hand, if it is extremely lipid soluble it will remain in the membrane because the membrane is full of lipids.

So, basically the membrane will solubilize that molecule. So, it will not come out or it will not go inside. Those are the problem. This biological activity is dependent on P. It is a hydrophobicity index. Apart from P you can have electronic parameters.

Electronic parameters definitely perturb the chemical reactivity of molecule. For example, benzoic acid versus para nitrobenzoic acid -which one is more acidic? You can tell from your organic chemistry knowledge.

Balance between the ionic form and the neutral form is very important. How much is the percentage of the neutral form? How much is the percentage of the ionic form? Those can perturb the electronic property of the molecule. So, there has to be an electronic parameter on which the biological activity will depend.

We have identified P as the hydrophobicity index or hydrophobicity parameter. There is an electronic parameter and there is a third one called steric parameter. If the molecule is too large then it will not reach the target. If it is sterically highly crowded there is a problem of reaching the target. These are the minimum three parameters.

Other parameters are also possible. But for this course, we will just stick to these three parameters. We will just discuss how these parameters are evaluated. See one thing you

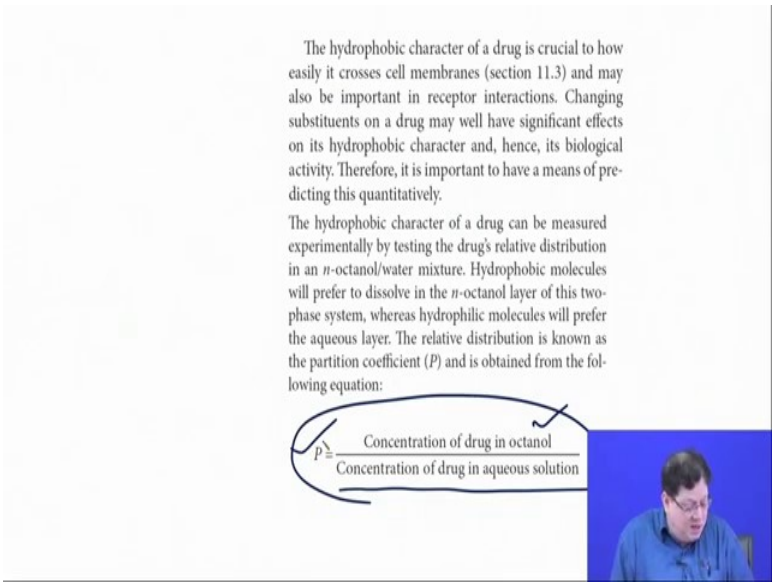
must also know that suppose I have a lead compound which is represented as a benzenoid compound.

So, I will put adequate length of carboxyl. This molecule is having some property against the target. Then you have to optimize the lead. You have to put different substituents in different position of the molecules.

Suppose I put a R1 here and another substituent R2 is in the same benzene ring. R1 can be methyl /ethyl /propyl /butyl and R 2 can be also different groups. Are you going to make these molecules and then try to find out its activity? If I have a methyl versus methoxy or a nitro what is going to happen?.

This QSAR model is so important. Without making the molecules, you can say the type of molecule (like hydrophobic molecule) from the database. Suppose the molecule is hydrophobic. If you put a methyl group the hydrophobicity will increase. So, you can have a huge database and from there you can calculate the hydrophobicity of a molecule without making it. Without making molecule we can predict electronic parameter/ steric parameter.

(Refer Slide Time: 12:42)



When you connect all these then you get the biological activity. Biological activity will be -

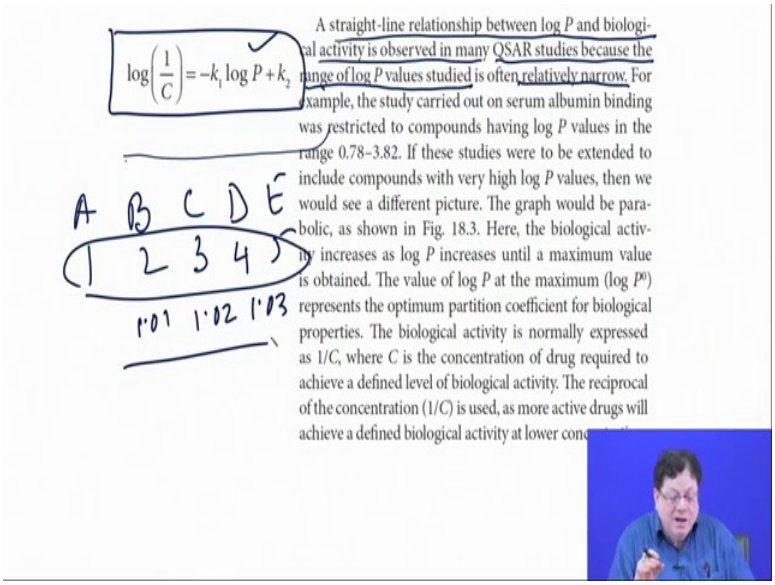
Biological activity = some constant * hydrophobicity + some constant * the electronic parameter + some constant * the steric parameter + the c factor.

So, it is like-

Y = mx + c

What is the hydrophobicity index? This is done in octanol water system. It is very close to our membrane constituents. So, you get the partition coefficient.

(Refer Slide Time: 13:38)



Now, people started taking drugs of belonging to the same class and then determining the partition coefficient. log (1/C) is the biological activity. So, initially they just disregarded the electronic and the steric effect. They only thought that only partition is the actual factor that will decide the biological activity.

So, they had some kind of reaction like –

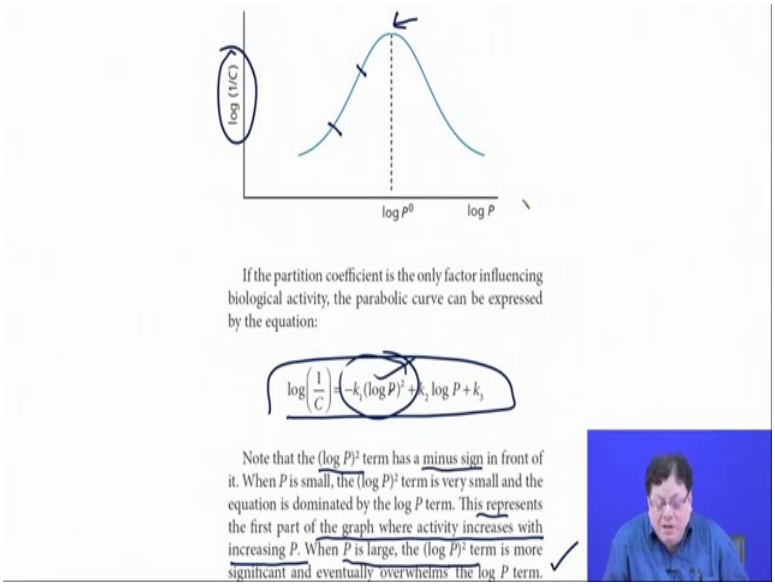
$\text{Log}(1/C) = -k_1 \log P + k_2$

If the molecule is only targeting the membrane there will be question of crossing the lipophilic membrane. So, usually this type of formula holds good for CNS active. Anesthetic drugs actually go to the membrane and then ultimately affect the central nervous system. So, this formula holds good for those type of molecules.

It is written here that a straight line relationship between log P and biological activity is observed in many QSAR studies but only have a limited range of log P values which is relatively narrow. A narrow range holds good. If you have a widespread your log P value you will have different molecules.

It is a huge range and for that this linearity does not hold good. Linearity holds good only where -say A, B has 1.01, this is 1.02, 1.03; then only this type of formula holds otherwise it does not hold.

(Refer Slide Time: 16:18)



If you start plotting log(1/C) with log P it will be a very wide spectrum. So, the curve goes like this. It reaches the maximum. It is a parabolic curve and then you have a the optimum log P value. Now, if you take a very narrow range suppose up to this point it is almost linear.

In that case, that formula will hold good. But actually the formula should be like a square term. Earlier it was-

$\log(1/C) = -k_1 \log P + k_2$

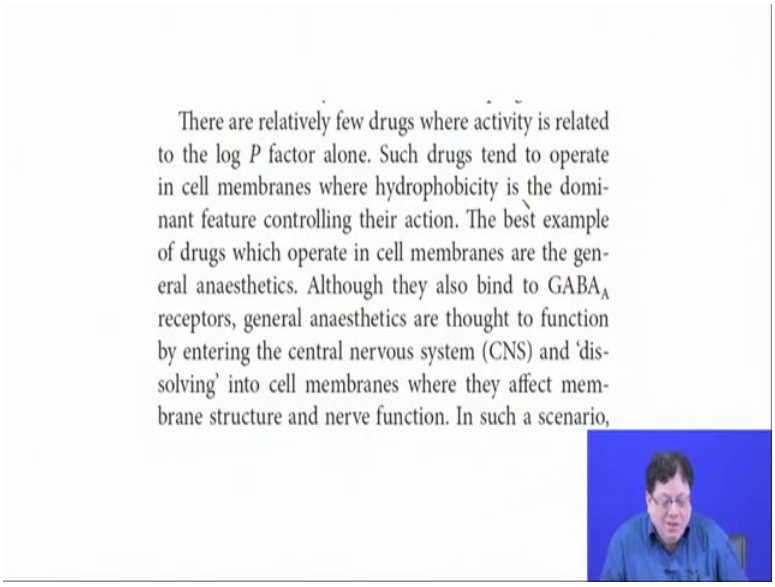
Now it becomes-

$\log(1/C) = -k_1 (\log P)^2 + k_2 \log P + k_3$

When P is small log P square will be very small at that time. So, you can neglect this log P square term. It brings the earlier formula. But as you represent the first part of the graph where activity increases with increasing P, but when P is large log P square term is very significant.

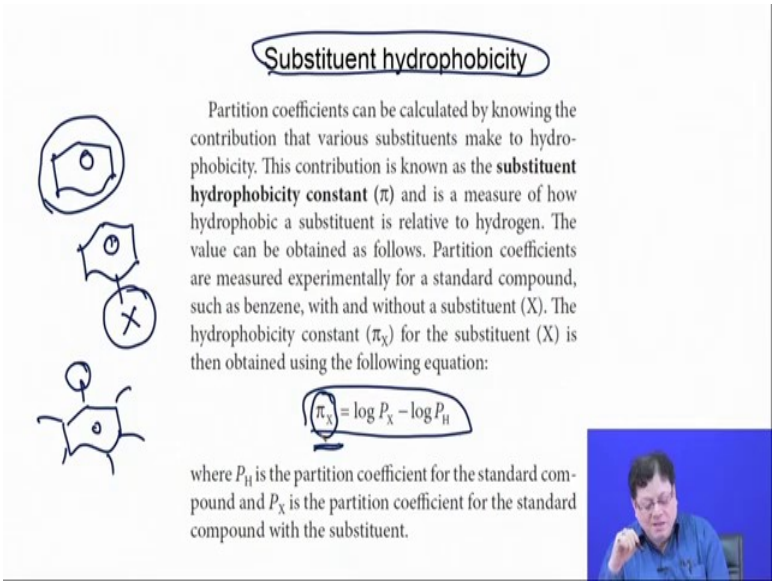
Eventually the log P term is decreasing at that point. When log P is very high this term will predominate at that point. So, the whole thing becomes negative. Negative means it is going down.

(Refer Slide Time: 18:12)



So, there cannot be a very simple formula like that. If you are only tackling a small subset of molecules the log P values are within a very narrow range.

(Refer Slide Time: 18:28)

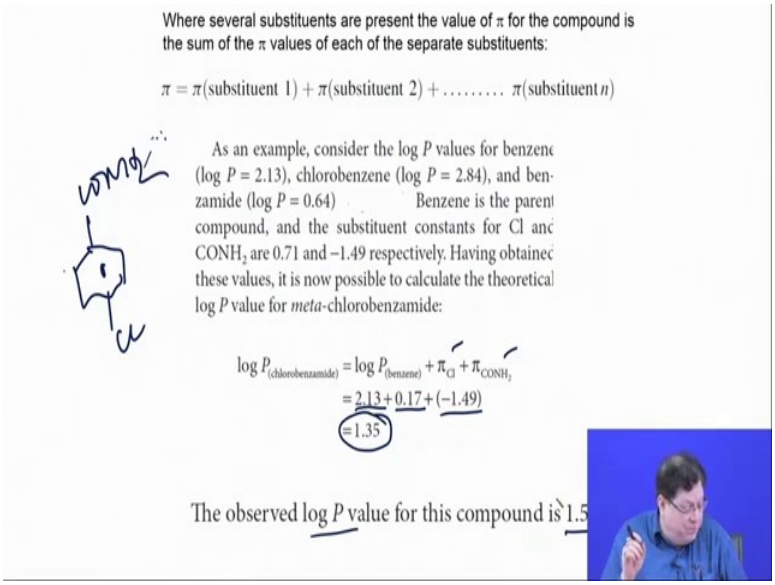


Substituent hydrophobicity is basically that you have a benzene molecule and you have a substituent X. Now if you want to assign this is very useful in the sense. You can measure the hydrophobicity of benzene.

The hydrophobicity constant(π_x) of the substituent (X) is obtained by the following equation-

$\pi_x = \log P_x - \log P_H$

(Refer Slide Time: 20:09)



When several substituents are present in a system what will happen the overall π.

For p-chloro benzamide, get a database which is the hydrophobicity index of methyl, nitro, carboxy are now known. Suppose I want to know the hydrophobicity of chlorobenzamide.

From the equation, we can get it-

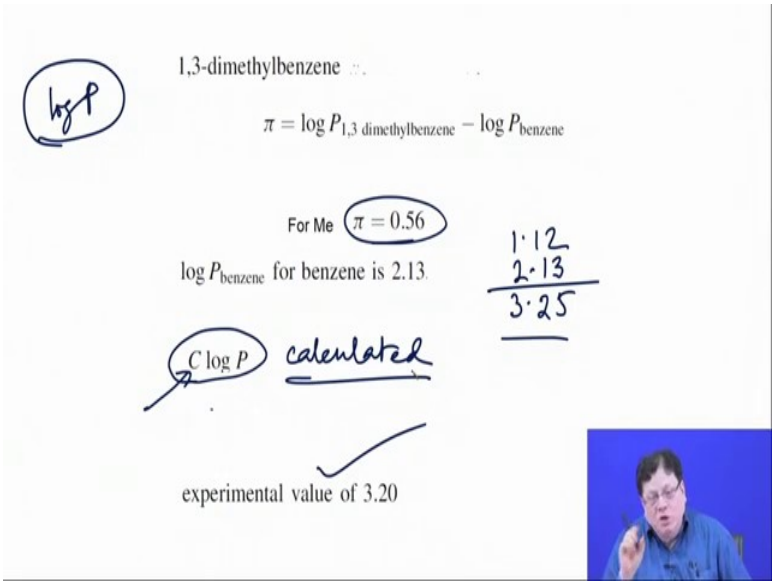
$\log P_{(\text{chlorobenzamide})} = \log P_{(\text{benzene})} + \pi_{Cl} + \pi_{CONH2}$

$= 2.13 + 0.17 + (-1.49)$

$= 1.35$

If you can make this molecule or you can buy this you just do it in partitioning between octanol and water. You see this is 1.51 which is not a bad matching.

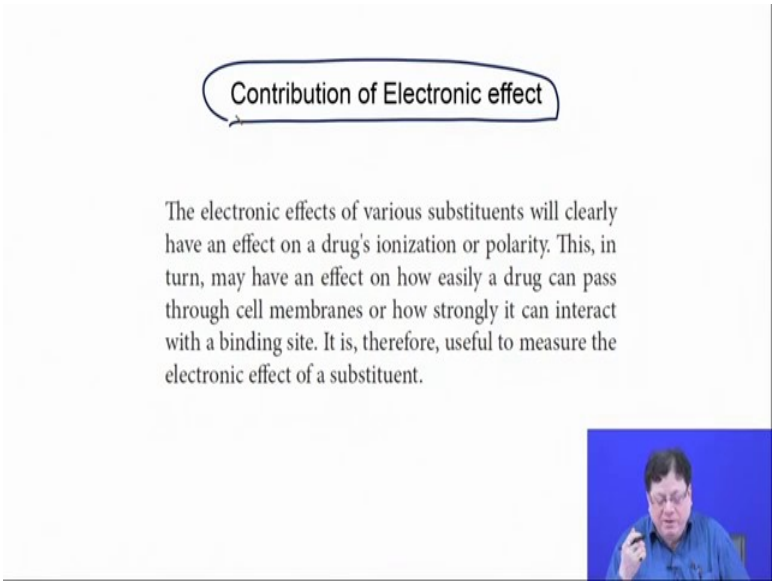
(Refer Slide Time: 22:12)



This is another example 1, 3-dimethylbenzene. So, that will be what benzene plus the hydrophobicity of 2 methyls. π for 1 methyl is 0.56, so for 2 methyls that will be 1.12. So, that will be 1.12 plus 2.13. So, that will be 3.25. The experimental value is 3.20 which is very close.

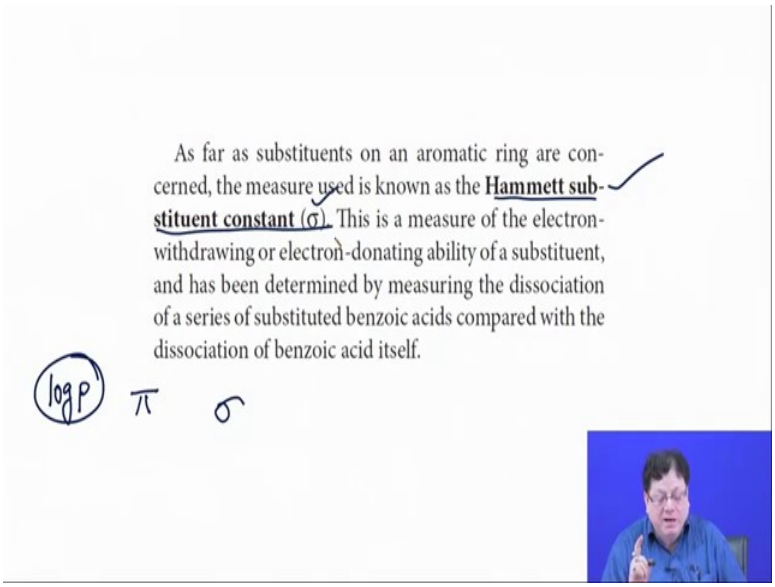
This is called C log P. They have put a term C. log P means experimental value of the hydrophobicity. C log P is the calculated. Without doing any experiment that is why you have this C; so C log P is the calculated one.

(Refer Slide Time: 23:21)



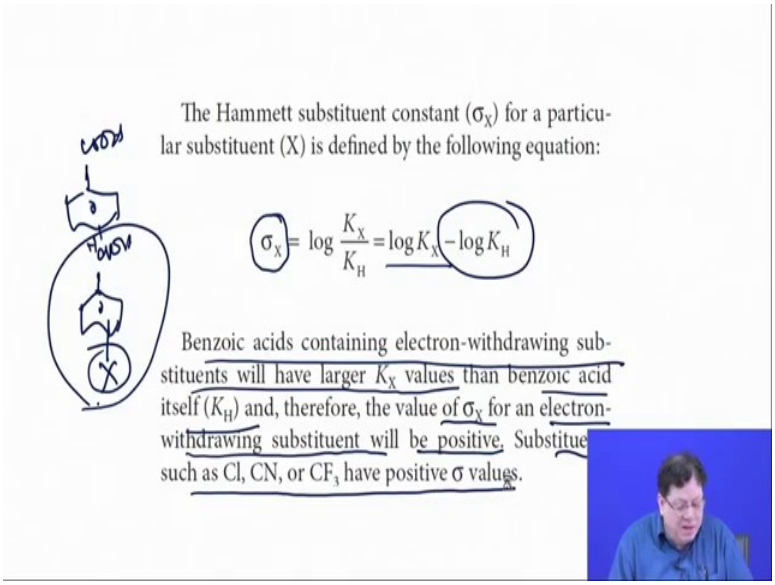
What is the contribution from the electronic effect? We realize that biological activity has to have some contribution from electronic effect that is operating in the molecule. There will be lot of weak interactions which are controlled by this electronic effect.

(Refer Slide Time: 23:47)



Now the electronic effect has been measured by Hammett substituent constant.

(Refer Slide Time: 24:41)

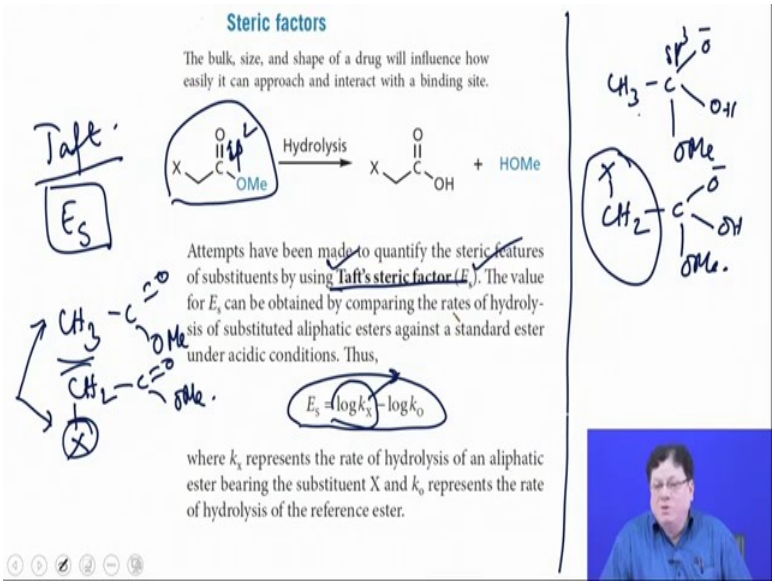


The Hammett substituent constant (σ_x) for a particular substituent is identified by the following equation-

$\sigma_x = \log K_X/K_H = \log K_X - \log K_H$

Benzoic acids contaiing electron-withdrawing substituents will have larger K_X values than bezoic acid itself (K_H) and, therefore, the value of σ_X for an electron withdrawing substituent will be positive. Substituents such as Cl, CN, or CF_3 have positive σ values.

(Refer Slide Time: 26:18)



The steric factor was given by Taft and steric factor is denoted by E_s.

He took say CH_3COOMe, a methyl ester of ethyl acetate and then substituted this with different substituents. Then check the rate of hydrolysis of these esters. Now you know the hydrolysis means saponification *i.e.* esters can be hydrolyzed by acid or by base.

If you do base catalyzed saponification the formation of the tetrahedral intermediate is the rate determining step. So, these are the two intermediates. Now these intermediates will be statically more crowded because you started with sp^2 carbon and now this is sp^3.

The more is the steric bulk less will be the rate of hydrolysis. So, that is the way you can quantify the steric factor. So, basically it is nothing but the difference between the two rates of hydrolysis.

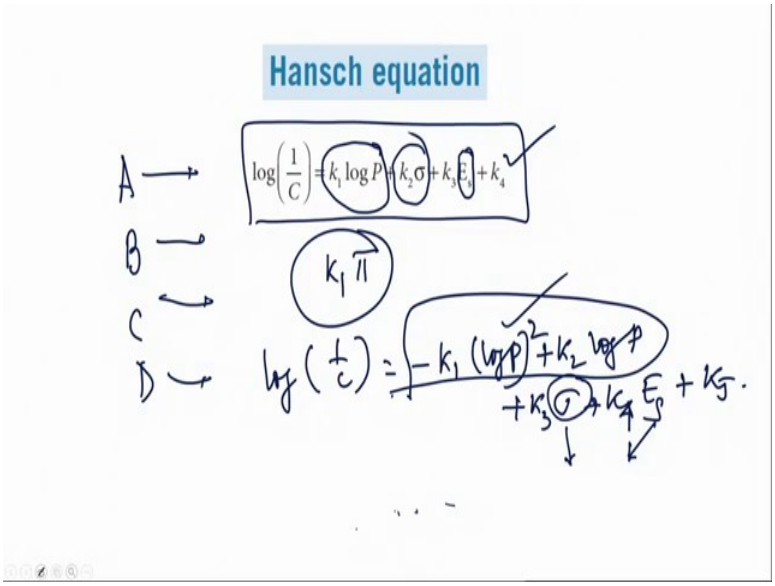
Attempts has been made to quantify the steric features of the substituents by using Taft's Steric factor (E_s). The value for E_s can be obtained by comparing the rates of hydrolysis of substituted aliphatic esters against a standard ester under acidic conditions. Thus,

$E_s = \log k_x - \log k_o$

Where k_x represents the rate of hydrolysis of an aliphatic ester bearing the substituent X and k_x represents the rate of hydrolysis of an aliphatic ester bearing the substituent X and k_s represents the rate of hydrolysis of the reference ester.

So, now we have idea about these three parameters. The hydrophobicity that is given by C log P because that is a calculated one. Then you have this σ and you have this you have the steric factor E_s. What will be the final equation then?

(Refer Slide Time: 29:35)



The final equation will be-

$\log (1/C) = k_1 \log P + k_2 \sigma + k_3 E_s + k_4$

This is called Hansch equation.

So, there are different variations of this Hansch equation. You can make it more accurate by doing –

$\log (1/C) = -k_1 \log P + k_2 \log P + k_3 \sigma + k_4 E_s + k_4$

So, that is the actual one. For all practical purpose you can remove this. If you want to quantify the biological activity rigorously you can use this formula. But remember this is empirical formula. But pharmaceutical companies use this QSAR equation. Then try to figure out before making any molecule. Try to figure out the prospect of the biological activity whether it will increase or it will decrease.

It not only depends on the log P but also depends on sigma. So, if you increase the lipophilicity you have to take care of the steric effect. So, all these combinations will

ultimately tell you that this is not accurate. It will guide you whether you are going in the right direction or you are going in the wrong direction to make the molecule.

After testing these you have to synthesize the molecule and that cost energy, money and human resource. So, that involves human resource. Pharmaceutical companies are extremely careful. They do not want to spend money on something which is bound to fail. So, they check this Lipinski's rule, then they check the Hansch equation. There is a prospect of having good activity of the molecule by combining using this equation.

The different parameters in Hansch equation can be calculated computationally. There could be other parameters like surface electrostatic potential. Computation can give you the surface potential *i.e.* the charge density in the contour of the molecule. So, you can have different variations of this type of equation but this was the original equation that was given by Hansch. There may be modern versions of this. Our purpose of this session is to just to know that.

QSAR principles can quantify the biological activity of a drug.