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

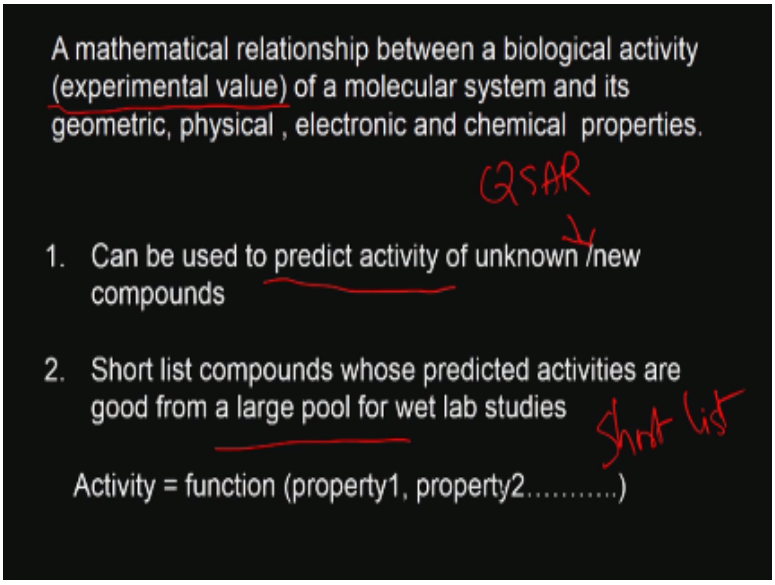
Lecture – 25
Quantitative Structure Activity Relationship(QSAR)

With the structural properties; structural properties, structural descriptor, structural features okay that is why you call it quantitative structure activity relationship. We will talk about what are those structural descriptors, structural features and so on okay later as we go along, so you get a mathematical relation that is why you say, call it a quantitative structure activity relation. Chemists have been using SAR for a long time that is structure activity relationship for a long time.

If electron donating, you may have activity, electron withdrawing less activity or hydrogen bomb bond forming more activity and so on, alpha beta unsaturated may give more activity, so that is structure activity relation. So, they used to synthesize molecules with different features and see how their activity changes from there they develop an understanding that is an understanding that is why it is called structure activity relation.

Whereas in computational work, we bring a quantitative aspect to that that is why we call it a quantitative structure activity relationship here okay, let us look at this in very detail. This is widely used for ligand based drug design especially when we do not have any idea about the target, we know certain molecules give activity; certain do not give, so we find out what are the structural features which give.

(Refer Slide Time: 01:46)



And then try to find out whether we can develop a mathematical relation okay, so it is a mathematical relation between the biological activity that is the experimental value that means that is very important, we need to have experimental data apriori of a molecular system and its geometric, physical, electronic chemical properties, large number of properties okay, so we try to get a mathematical relation.

Can be used to predict activity of unknown new components, so what it means is okay, suppose I developed a QSAR, now I have a unknown new molecule, my medicinal chemist says cannot we synthesize this molecule, a new structure, so I can try to predict activity of this new structure that is the unknown, before going to my lab I can predict what will be the activity that is why QSAR is very important okay.

Another advantage okay, suppose I have a large pool of molecules instead of spending a lot of time going to my wet lab and checking their activity over a long period of time, I can predict the activity of these molecules using the QSAR and then sort of grade them or categorize them as those which will have good activity and those which might not have and then select those, maybe top 10 or top 15, which may have good activity and then take it to my wet lab.

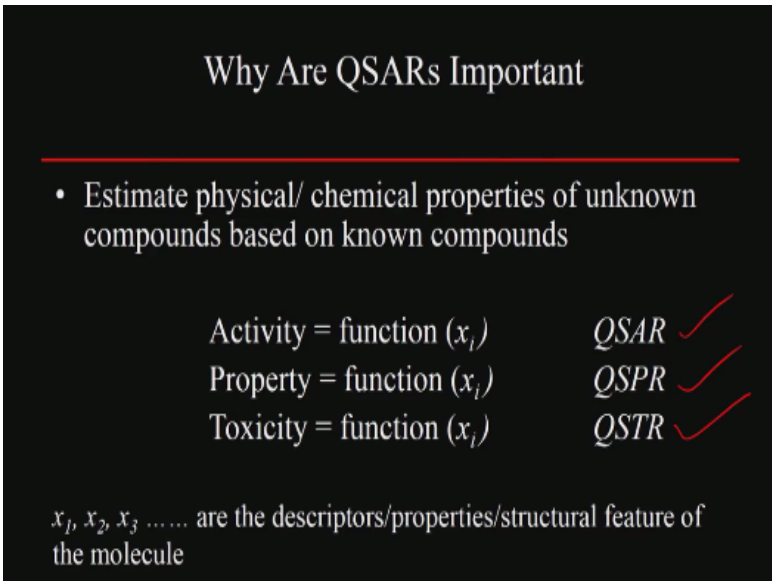
So that way I can shortlist basically, I can shortlist my molecules, if I have a large pool instead of taking all of them into my wet lab and performing experiments, which may be very time consuming, I may take the top 10 or top 15, which I have using the QSAR may have good activity, so this is the advantage of QSAR, so I start with the known structures whose experimental activities are known okay.

Then I develop a QSAR, so with that QSAR I can do 2 things; I can predict activity of a new molecule before going into my wet lab, if a medicinal chemist says why not the structure immediately with the QSAR I can predict of unknown molecules and if I have hundreds of molecules or thousands of molecules instead of taking all of them into my wet lab and then checking out the biological activity which could be very time consuming, I may predict the activity of these hundreds or thousands of them.

And then put them in the descending order and take the top 10%, top 20%, take it to my wet lab and check it out for their biological activity, so that way I can shortlist okay that saves me a lot of time instead of taking in the entire thousands of molecule. So QSAR is useful in that way also okay, so basically you develop a mathematical relation, left hand side you have the activity.

Right hand side; we will have some function of property 1, property 2 like it could be size of the molecule, shape of the molecule, number of hydrogen bond donating groups, number of hydrogen bond forming groups, electrostatic so many things, we will look at them more in detail, these are called the properties of the molecule okay.

(Refer Slide Time: 05:09)



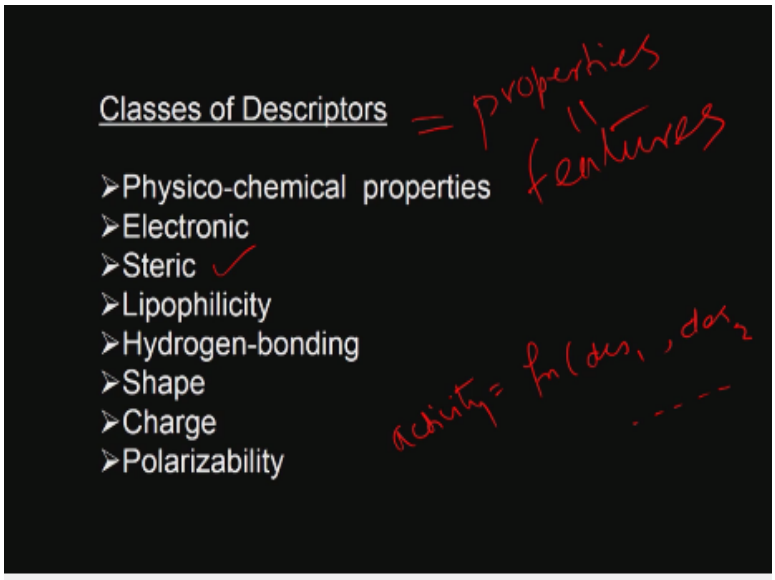
So, we can have QSAR; quantitative structure activity relationship nowadays, quantitative structure property relationships have come that means, I can predict properties like for example, log P, if you want to predict log P for new molecules, how do the softwares do? They have QSPR, so they can predict the property or the log P of a new molecule from known molecules they would have developed a mathematical relationship.

Solubility; there are many QSPR for solubility from known molecules, they predict the; I mean they develop the QSPR and if you input unknown structure for example, in Swiss ADME, it calculates the solubility of that molecule using QSPR. Similarly, QSTR; quantitative structured toxicity relationships, so if you have toxicity data of some molecules you develop a mathematical relation QSTR, so if you input a new molecule, it will predict, what is the toxicity of that new molecule okay.

So, nowadays QSAR, QSPR, QSTR are very common, some of the software's I showed you, there are many software is pretty has this QSPR, QSTR in built in them for predicting properties like solubility, log P and then toxicity values, different types of toxicity values given the structure okay. So, basically so on the right hand side, we have the various properties okay like I said size, shape, number of carbon atoms all these are properties okay.

So, the QSAR is a quantitative relationship, do not forget that it is not a qualitative but a quantitative mathematical relation that is why we say quantitative structure activity relationship okay.

(Refer Slide Time: 07:10)



Now, these properties or structural features they are also called descriptors, these properties or molecules or features, they are also called descriptors, physico-chemical properties okay, like I said shape, size, they are physico-chemical, electronic property okay, so charge of the molecule or largest; highest negative charge, highest positive charge, steric factors okay, the shape gives you the steric factor.

Lipophilicity; the log P value and the those are all lipophilicity, hydrogen bond forming that means, the hydrogen bond forming, hydrogen bond donating, hydrogen bond accepting shape of the molecule, charge distribution and the polarizability, so all these are different types of structural properties or features or descriptors, so they are all used in QSAR and structural descriptors we say, we say structural property, structural features.

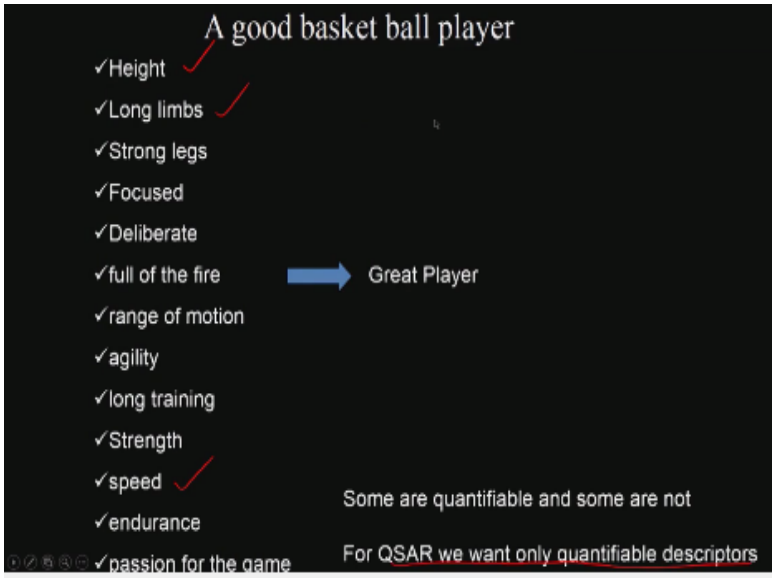
All these are various types of numericals, which we can estimate for a given structure, so if I draw a structure, we can calculate these descriptors. Nowadays, there are software's which can calculate thousands of descriptions, it is almost 2500 descriptors, we will look at some of those software's as we go along and there are many web based softwares, downloadable software which can calculate descriptors for any given structures okay.

And then with those descriptors, we try to select the best descriptors and then try to develop a mathematical relation between the activity on one side that is the left hand side was as the function of descriptors okay, 1 descriptor, 2, to so on. So, obviously we need to have a large number of data points if you want to have many descriptors in the model okay that is statistics okay.

If you have done a course on statistics, we will know how many descriptors we should have for activity okay, let us look at some simple system okay. Imagine, I want to have a good basketball team, so I am selecting basketball players for my team from the college, one approaches to make all of them play and then select the team that is going to be very time consuming. Imagine a college with 8000 students that will be very time consuming.

So, I may look at some features of the people right, based on these features, I may shortlist a group and then that group I may test it out to see whether they play well okay that is the best way to do instead of making entire 2000 people to play a basketball which is going to be very time consuming okay.

(Refer Slide Time: 10:01)



So, what are these features height for example, ideally you would like to have tall guys, guys with long limbs; long hands, long legs, strong legs, they should have strong legs because they should be able to run around and all that focused, okay deliberate, full of fire, range of motion, agility, long training, strength, speed, endurance, passion all these make a good basketball player.

See but then some of them are quantifiable, some of them are not quantifiable okay, say height means I can tell he is 200 centimeters square okay, long limbs I can tell the hands are so many centimeters, legs are so many centimeters, whereas things like deliberate full of fire, these cannot be quantifiable. For QSAR, we need to have only quantifiable descriptor that is very, very important.

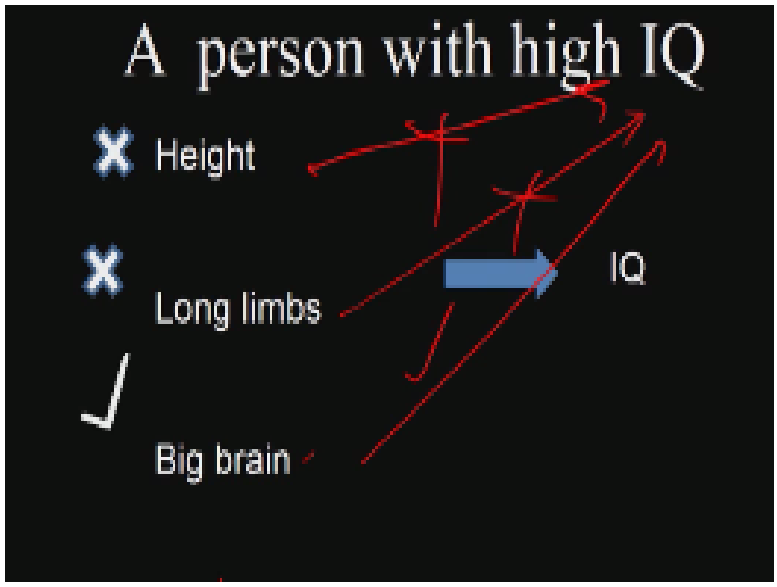
So, we need to look at only quantifiable descriptor; speed maybe I can make them run 100 meters and then find out how long it; how many seconds they take and I may say okay people who take < 11 seconds in 100 meters can be considered as the team, people around 180 centimeters can be taken height and so on, so I can select some of these as my descriptors. So, from 2000 students, I may select maybe 100 students who have certain cut off on height, limbs, speed and so on.

And then I may see whether they play take them to the court and make them play and see whether they are able to play okay, so these could be the descriptors or features, structural features for selecting a good basketball team and that is what we do in your drug discovery, so

we know a lot of descriptors for a molecule, electronic descriptor, shape descriptors, size descriptors, thermodynamic descriptors and so on.

And then I develop a mathematical relation between some of these descriptors with their activity, so that I allow a QSAR okay. There are some problems in this okay, for a basketball team or if you are talking about yes, an athletic team, sports team I can look at height, limbs, speed all those things right.

(Refer Slide Time: 12:19)



But then look at it, I want to select there are 2000 students, I want to select students who might think have high IQ, one thing is I can make them do an exam test can conduct an IQ test for couple of hours or 3 hours and then take those who have 120 hour instead of that from their structural features can I shortlist? It is not so easy; can height be a feature representing IQ? No, it might not be right, it need not be that tall people will have IQ or vice versa.

Can limbs be; so there are many descriptors which did not be a representation of the IQ, may be big brain; may be size of the brain may be may be okay, maybe the weight of the brain or size of the brain may be a representation of the IQ but look at this, there are problems right, we do not know which descriptors will predict their IQ. So, in QSAR that type of problem can also happen, we do not know which descriptors to select to predict the activity okay.

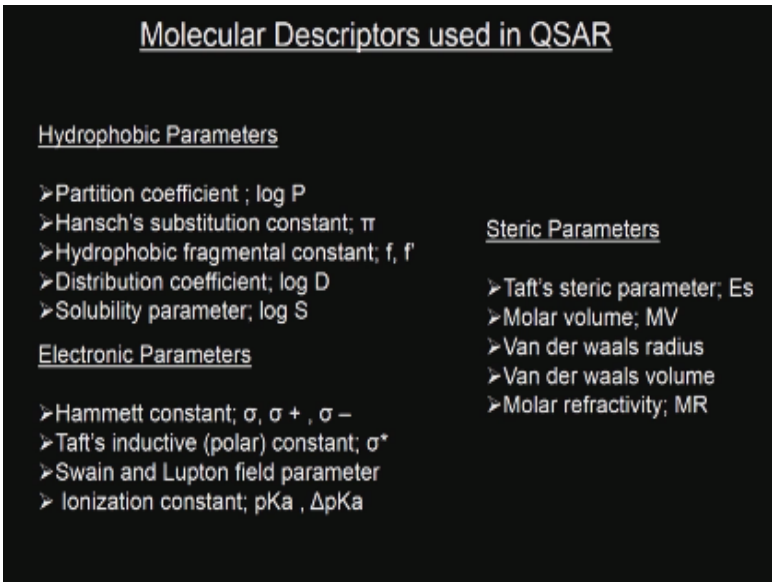
Sometimes, we can predict for example, if you are looking at antioxidants we may say a phenolic group or voyage type of group mainly to good antioxidant property similarly, antibacterial we may look at molecules, which may have alpha, beta, unsaturated or which may

form quickly radicals and so on but it is not so easy to say which descriptors are needed to predict a particular activity, so that problem can happen like this you know.

I do not know which descriptors to select to predict the IQ, so I am; if I take a height has a predictor of IQ that is wrong, a long limbs that also can be wrong, whereas, so we see in this particular case, you may be able to select descriptors to predict whether he will be a good basketball player or a good athlete or a good sportsman but here it may be very difficult to find descriptors.

So, the same thing can happen when you are doing QSAR in drug discovery, so we may not have the correct descriptors to select, maybe selecting the wrong descriptors also, we are not sure what should be the descriptor for this IQ, so we may be selecting the wrong ones, so that is a danger that can happen in QSAR also, so that is a shortcoming in QSAR, we need to think of it okay.

(Refer Slide Time: 15:08)

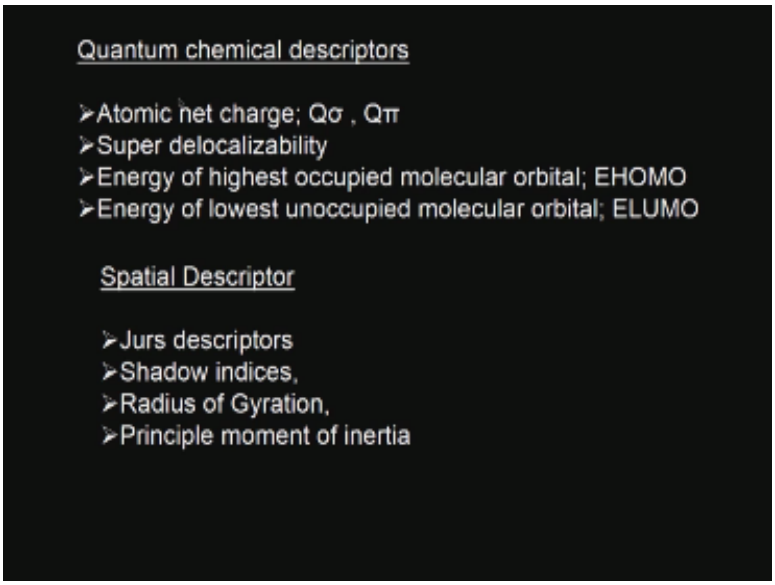


There are a lot of descriptors like I said molecular descriptors used in QSAR, for example hydrophobic, electronic, steric. Hydrophobic; you can have log P, Hansch's substitution constant, so if you have CH3 groups, OCH3 group, OH groups, some of them lead to liphophilicity, some of them lead to hydrophilicity right, hydrophobic fragmental constant, distribution constant, solubility they all come under the hydrophobic.

Electronic; Hammett constants, Tafts inductive polar constants, Swain and Lupton field parameters, ionization constant. Steric; Taft steric parameter, molar volume and Van der Waals

radius, Vander Waal volume, molar refractivity, so you can have descriptors, you can have sub descriptors under each category also, so you could have lot of descriptors like I said we can have 2000+ descriptors for given structure.

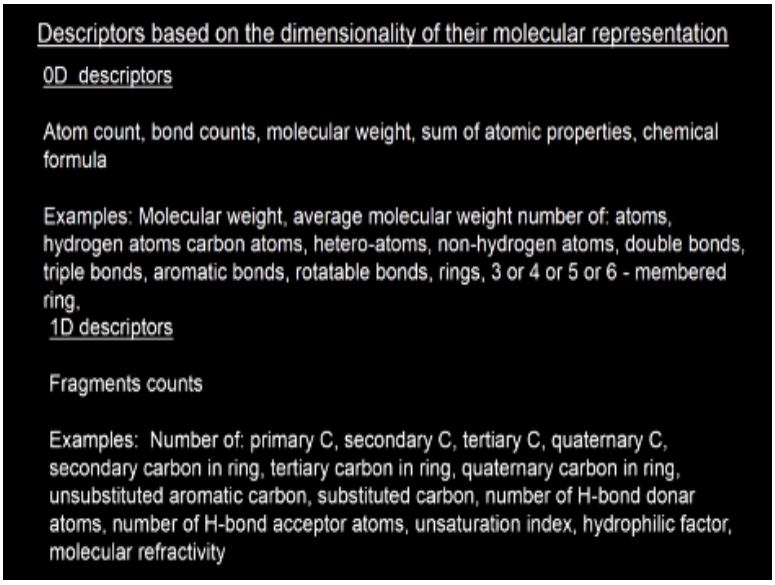
(Refer Slide Time: 15:58)



There are software's which can calculate them, quantum chemical atomic net charge, super delocalizability, energy of highest occupied molecular orbital, energy of the lowest unoccupied molecular orbital then spatial descriptors, Jurs descriptors, shadow indices, radius of gyration, principle moment of inertia, so a large number of descriptors, see hydrophobic related, electronic related, steric related, quantum chemical related, spatial related.

Of course, we also have thermodynamics and all that also like heat of formation, delta G that sort of descriptors, so large number of descriptors are possible.

(Refer Slide Time: 16:40)



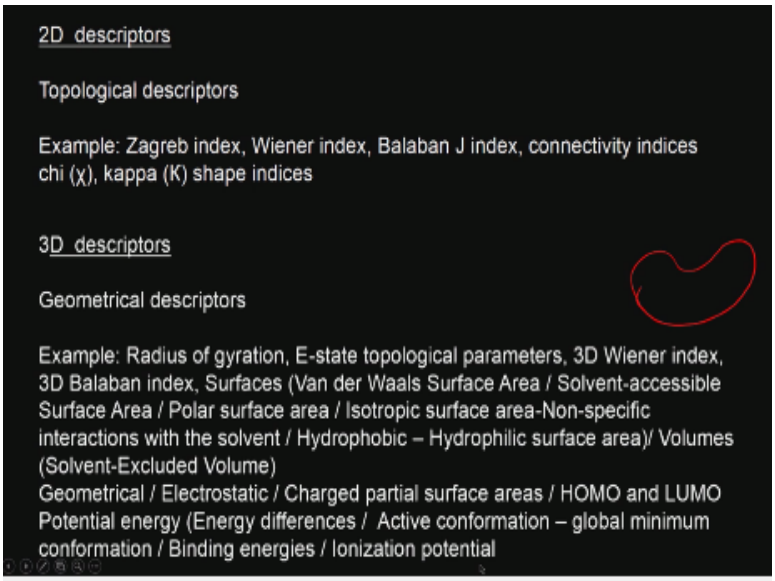
There are some D Rho dimensional descriptors, one dimensional descriptors, 2 dimensional descriptors, 3 dimensional descriptors okay, let us look at it. Atom count; number of atoms the molecule has, bond counts; number of bonds the molecule has, molecular weight; what is a molecular weight? Because the molecular weight may have an effect for example, as you know large molecular weight diffusion may be a problem small molecular weight diffusion is easy.

Some of atomic properties, chemical formula; so molecular weight, average molecular weight, number of atoms, hydrogen atoms, carbon atoms, hetero atoms these are all non-carbon and non-hydrogen atoms, number of double bonds as you know double bonds, it is easy to oxidize or they are easily reactable, triple bonds, aromatic bonds because they form resonance, rotatable bonds because as you know a rotatable bond can change the conformation of the molecule.

So that it can go and fit into active sites, rings; 3 membered rings, 4 membered, 5 membered, 6 membered as you know highly strained 3 membered, 6 member is very good, so all these are possible structural features of a molecule, they are also called a descriptors of the molecule, structural features, structural descriptor and so they are called zero dimensional descriptors okay; zero dimensional descriptors okay.

Fragment count okay, these are a one dimensional descriptors; fragment count, number of primary carbon, secondary carbon, tertiary carbon, quaternary carbon, secondary carbon in ring, tertiary carbon in ring, quaternary carbon in ring and substituted aromatic carbon, so many variations on carbon itself, the number of hydrogen bond acceptor, unsaturation they are all called one dimensional descriptor okay; one dimensional descriptor.

(Refer Slide Time: 18:45)



These are zero dimensions, these are one dimensional, 2 dimensional; these are all 2 dimension as you know is topology okay. There are indices like Zagreb index, Wiener index, Balaban J index, connectivity indices, chi, kappa shape indices, so they are all topology related, they are all 2 dimensional descriptors, so when I draw a structure and so I can get all these okay of course, then we have 3 dimensional.

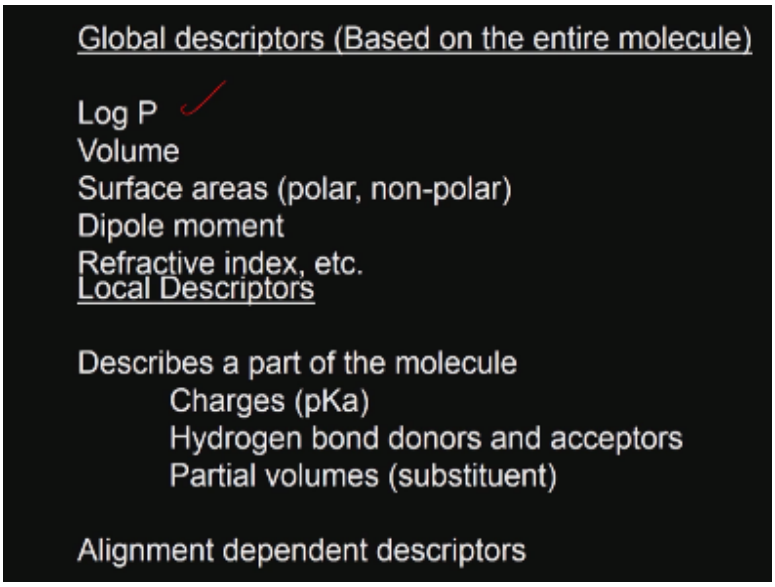
When it takes a conformation, the geometric descriptors radius of gyration because that depends on the shape of the molecule okay, 3 dimensional electronic state, topological parameters, 3 dimensional Wiener index, 3 dimensional Balaban index, surfaces, Vander waals, surface area, solvent, accessible surface area, polar surface area, isotropic surface area, nonspecific interaction with the solvent, geometric, electrostatic charged partial surface area, highest occupied molecular orbital energy, lowest unoccupied molecular orbital energies.

Energy difference between this homo and lumo, active confirmations, global minimum confirmations, binding energies, ionization potentials, so these are all 3 dimensional descriptors. So, as you can see once I get the minimum energy conformation, I can calculate all this, if I draw a structure using a structured draw program, I can calculate all these, without drawing from the molecular formula, I can calculate these.

And then once I do it by hand, I can calculate all these, so these are some descriptors and of course, some I have; I might have omitted like I said there are software's like E dragon which

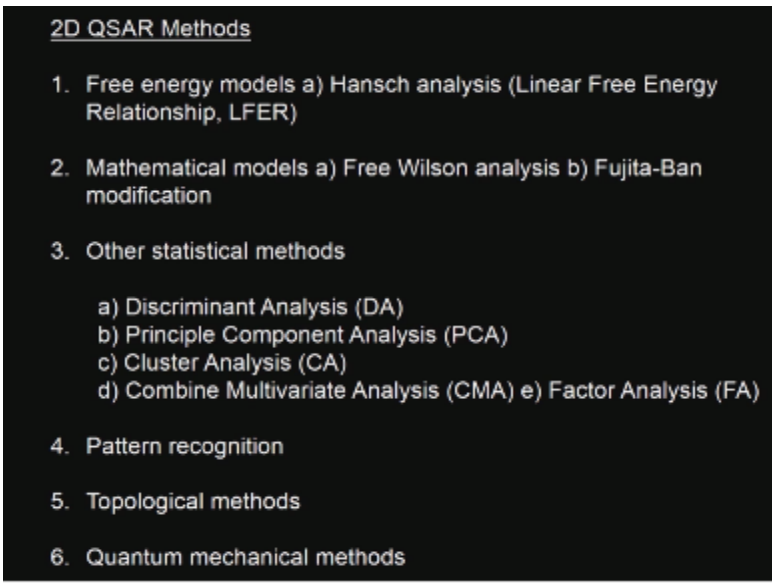
can calculate 2000+ descriptors for performing QSAR. So, large number of descriptors can be estimated given a structure okay.

(Refer Slide Time: 20:40)



Global descriptors; log P okay, log P once I know the molecular formula, I can calculate the log P using aqueous PR, volume surface area, polar surface area, nonpolar surface area, dipole moment, refractive index then local descriptors describe a part of the molecule okay, charge in one portion, hydrogen bond donors and acceptors, partial volumes, alignment then you have alignment dependent descriptors okay.

(Refer Slide Time: 21:13)

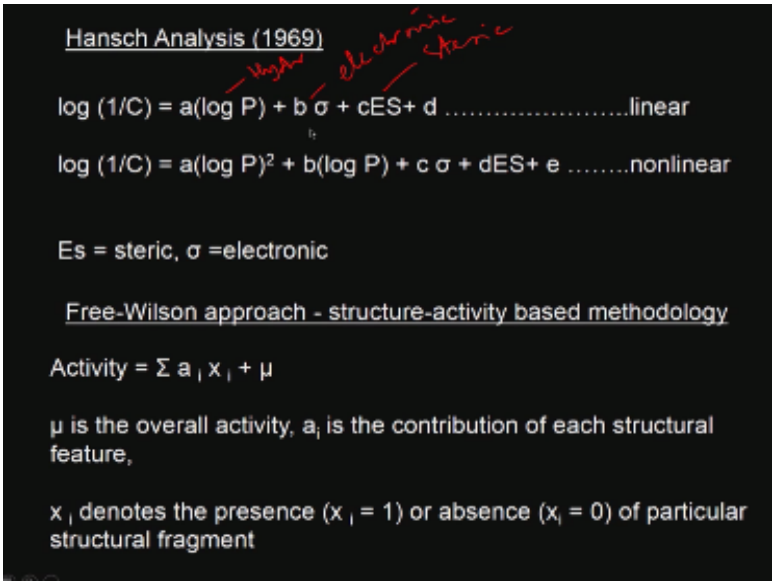


When it is aligned those descriptors, these are called global descriptors okay, so there are many methods to develop the QSAR relationship that is quantitative structure activity relationship, free energy models, mathematical models, statistical methods, pattern recognition, topological

methods, quantum mechanical methods and so on, so large different types of approaches, we are going to look at many of them in the course of this topic QSAR.

So, many different approaches are there, most and as I said they are all mathematical relationship because the quantitative structure activity relationship, so they are all mathematical relationship okay.

(Refer Slide Time: 21:57)



Hansch's analysis; this came in 1969, this called Hansch's analysis, so log 1/ C that is we have the logarithm that is the P IC 50 for example, is equal to you can have A constant log P + B Sigma could be electronic, c ES steric + d; d could be a constant or you could have square and so on actually, so it is a nonlinear. So, you could have a linear relation, so it basically has 3 terms; one related to the hydrophobic, hydrophilic nature.

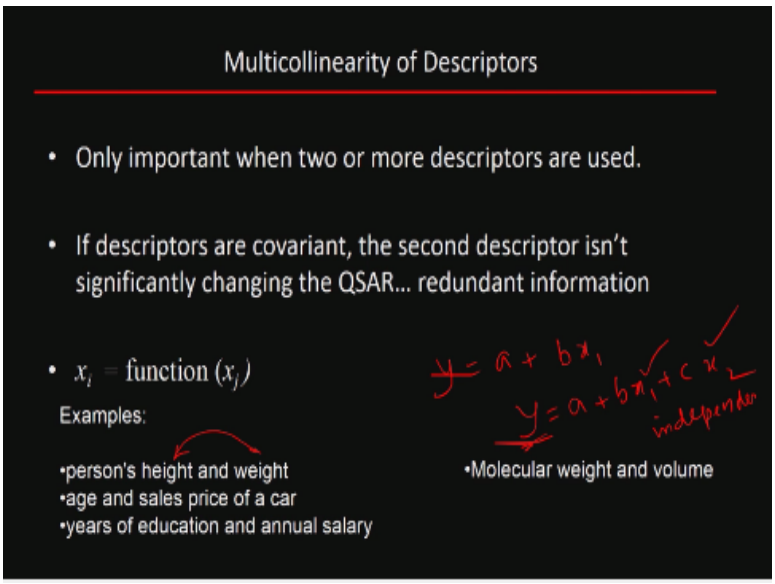
One relates to electronic, one relates to steric, so according to Hansch's, any molecule QSAR could be defined using 3 basic things okay. The activity will be a function of hydrophobic, hydrophilic nature log P electronic, static okay that is what Hansch's said actually. Then we have free Wilson approach, activity = summation of ai xi + mu; mu is the overall activity, ai is the contribution of each structural feature okay.

So, if I have a structural feature like OH, then that will contribute to the overall activity, if I do not have then the; that will become 0 okay, if I have that will become 1, so each one will contribute to the activity, either some of them may be enhancing activity or some of them may

be pulling down the activity or reducing the activity that is called the free Wilson approach, okay.

So, remember Hansch's analysis free Wilson approach; Hansch's analysis, they assumed hydrophobic hydrophilic contribution, electronic contribution that is here, this is the steric contribution that is here, so linear relation will be of this form okay.

(Refer Slide Time: 23:58)



Some important points we need to consider, so when we use 2 or more descriptors in my equation and if 2 descriptors are highly related with each other, covariant then the second descriptor is not significantly changing the QSAR, so we need to remove that okay. So, if 2 descriptors or a; for example, molecular weight and volume, we cannot take this as descriptor 1, descriptor 2.

Because there is; there could be a very good relationship between molecular weight and volume, the molecular weight increases, volume increases, so there is a relationship, they are okay correlated; person's height and weight, so mostly a tall person will be heavier than a short person will be lighter, so they may be interrelated. So, do not take a equation which has both height and weight as 2 independent variable, okay.

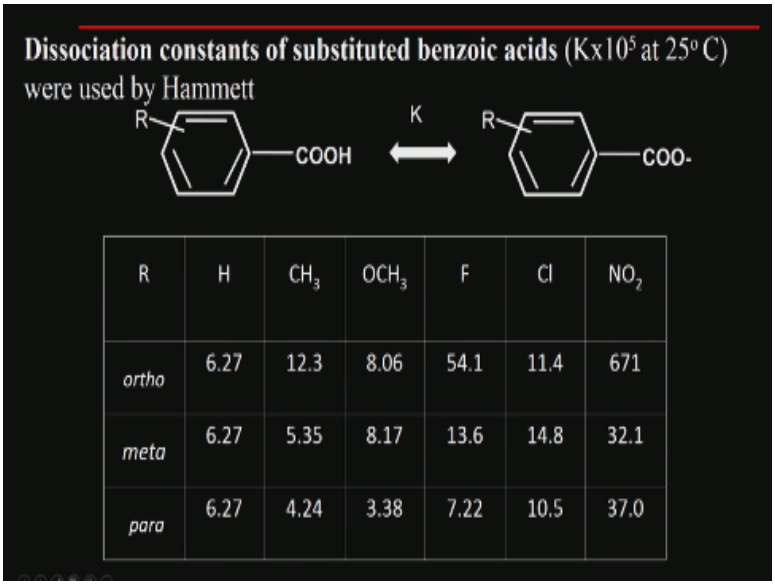
What is independent variable? When we say y is = a + b x1, so when we write y is = a + b x1 + c x2, these are called independent variables and y is called the dependent variable okay. So, basically all these x's that means the independent variables should be independent of each other,

they should not be correlated with each other that is very, very important that is what is called a multicollinearity.

So, person's height and weight for example, age and sale price of a car, so older the car, sale price will be less, younger the car sale price will be high, so these 2 descriptors or parameters are interrelated, so do not have a model which includes both as independent variable, so year of education and annual salary, for example if you have studied longer, your salary may be more, so these 2 variables could be interrelated with each other okay.

So, like molecular weight and volume, so you need to keep in mind that when you have x1, x2, x3, x4 as independent variable descriptors, they should not be interrelated or correlated with each other that is very, very important okay.

(Refer Slide Time: 26:29)



R	H	CH_3	OCH_3	F	Cl	NO_2
ortho	6.27	12.3	8.06	54.1	11.4	671
meta	6.27	5.35	8.17	13.6	14.8	32.1
para	6.27	4.24	3.38	7.22	10.5	37.0

Let us look at some simple structure activity relation, look at this molecule okay; these are called substituted benzoic acid. The benzene ring COOH, R; R could be in any group like this in ortho or meta, para, so they are benzoic acid substitute, now they dissociate okay to form COO – okay. The dissociation constant is given in this table at 25 degree centigrade depending upon the substitution and depending upon the position you can see the dissociation constants.

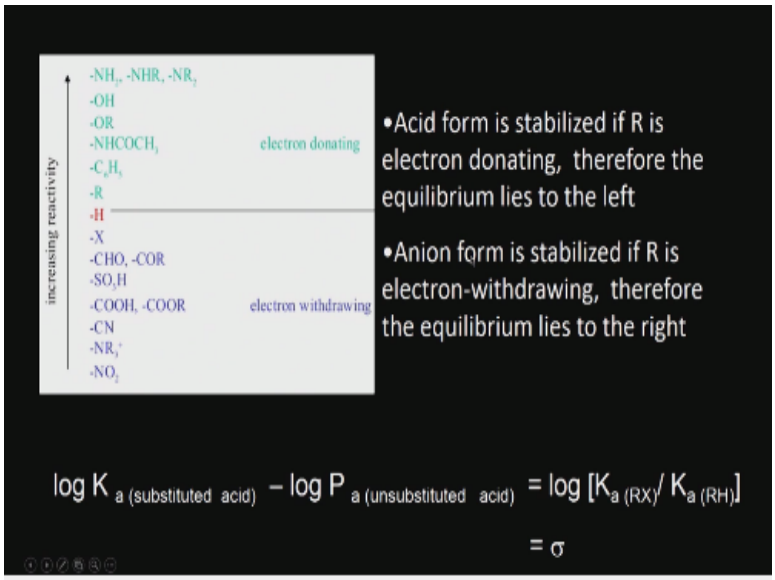
You can see the numbers vary dramatically quite a lot, it can be as small as say 4 or even 3.3 and it can be as large as 54.1, so a fluorine in the ortho position can have a very large dissociation constant whereas, OCH3 in a para position here could lead to a very small

dissociation constant, so fluorine here does not makes it dissociate like this, OCH 3 here pulls the electron, so it does not allow the dissociation taking place okay.

So, this is nice structure activity behaviour you can see depending upon the functional groups like nitro, chloro, fluoro, OCH3, CH3, H, so depending upon whether it is electron donating or it withdrawing and in the position that is the ease with the electron donating or withdrawing can take place, the dissociation constant can change okay. So, acid form is stabilized or is electron donating, therefore the equilibrium lies to the left okay.

Acid form stabilized; acid form is stabilized, if it is like this instead of going like this, so if it is a; if R is electron donating okay sorry, R is electron donating okay, if R donates electron, acid form is stabilized, anion form is stabilized if R is electron withdrawing therefore, the equilibrium lies to the right that means, anion form that is on the right if sorry, R is electron withdrawing okay.

(Refer Slide Time: 29:10)

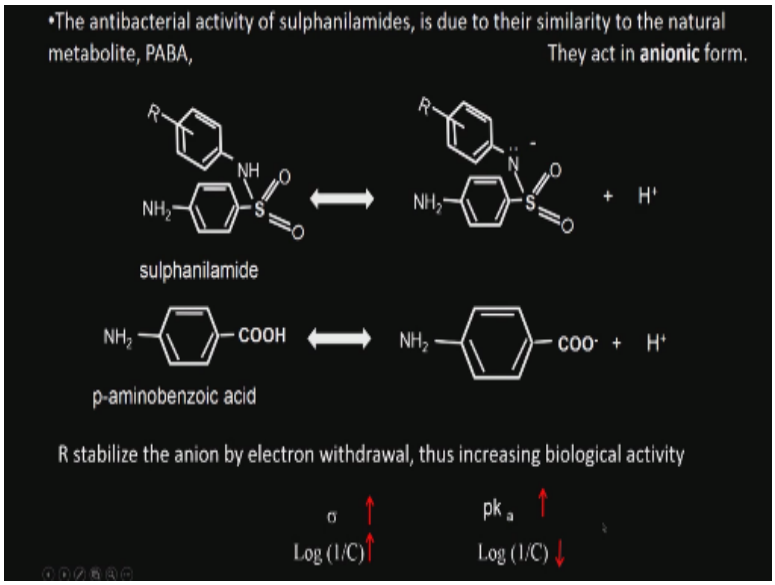


Look at this electron withdrawing, these are all electron withdrawing, electron donating OCH3 here okay, so this is called the acid form, this is called the anion form. Acid form is stabilized if R is electron donating, anion form is stabilized, if his R is electron withdrawing okay that means, this form is stabilized, if its electron withdrawing, acid form is stabilized, if its electron donating.

So, there are increasing reactivity of this, electron withdrawing okay, these are all electron donating, these are all electron donating OR that is OCH 3 OH okay increasing activity, so you

can see down here okay and so on, so log Ka of the substitute acid - log Pa of the unsubstituted that means, if you have H is = log Ka RX/; log of Ka RX/ Ka this is called the Sigma okay, so larger the Sigma, so there is a dissociation taking place.

(Refer Slide Time: 30:18)



Why are we talking about it actually okay? Let us talk about it, why? Look at the antibacterial activity of sulphanilamides, these are called sulphanilamides, they are antibacterial, they are called sulpha drugs discovered after First World War, we can see the structure here, so there is a benzene ring has double bond O, double bond ONH, there is benzene ring with R okay, antibacterial of sulphanilamides is due to the similarity to the natural metabolite paramino benzoic acid.

They act in their anionic form only when they are anionic, they act when they are neutral they do not act okay, why do they act? They look like the para amino benzoic acid, this is the natural substrate for the bacteria, so this sulpha drug goes and binds in a competitive fashion and it does not permit the substrate food for the bacteria to bind, so the bacteria dies okay. So, the anionic form is, what is decide.

Whereas, this form is not desired okay, so R stabilizes the anion by electron withdrawal that is increasing biological activity okay. So, if sigma is larger, activity is also larger, if PKa is larger, the activity goes down, so ideally we should have compounds, the R substitution which makes it the anion form rather than the neutral form. So, what is that? So, we would like to have more of this type of molecules okay.

So, we will have more of these type of molecules okay, anion form is stabilized, if this R is electron withdrawing therefore, the equilibrium shifted the light okay, so Sigma is larger here goes down, so that is a very important point we need to understand how whether it is an electron withdrawing or electron donating, we end up either with the stable anion form or with the stable neutral form.

And as I said this is what gives the activity as an anti-bacterial drug okay, so depending upon whether it is electron withdrawing or donating, we have the proper anion form getting stabilized okay and hence implement in the activity okay. So, we will continue further in the next class on QSAR, thank you very much for your time.