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

Lecture - 31
3D-QSAR

Hello everyone, welcome to the course on computer-aided drug design. Today, we will talk about 3D-QSAR. We talked about the normal QSAR. Today, we will talk about 3D-QSAR which means the conformations, the molecule stake becomes very important and then we try to superimpose those conformations and then try to get some activity related structural features okay.

So here conformations become very important the molecule that occupies because as you know when the molecule goes and binds into an active site of an enzyme, the conformation plays a very important role okay.

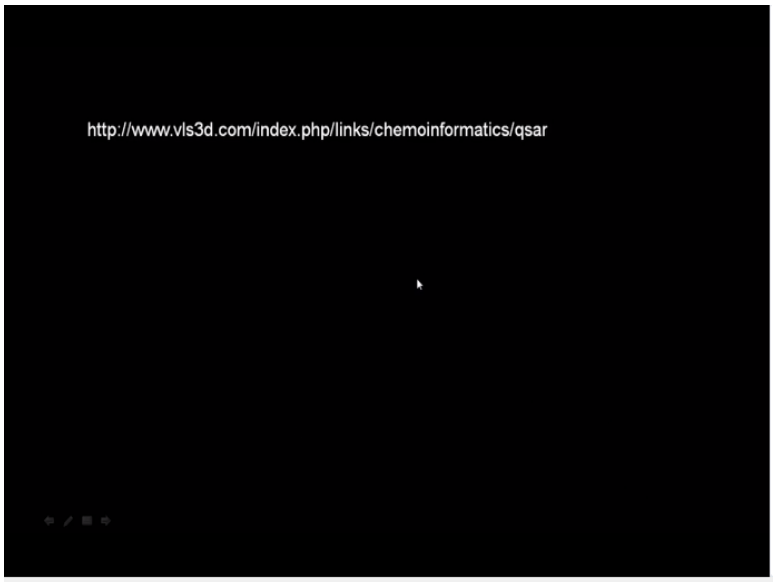
(Refer Slide Time: 00:57)

So before that if you remember in the past few lectures we talked about QSAR and other so we used to lot of these data bases and then we used Dragon, SwissADME, ChemDes, Ochem to get the various descriptors okay. Then, I also talked about BuildQSAR, which almost looks like an excel and then I showed how to build that.

Then, if you are interested in the semi-empirical quantum mechanic descriptors like electron energy or proton and highest occupied molecular orbital energy or lowest unoccupied

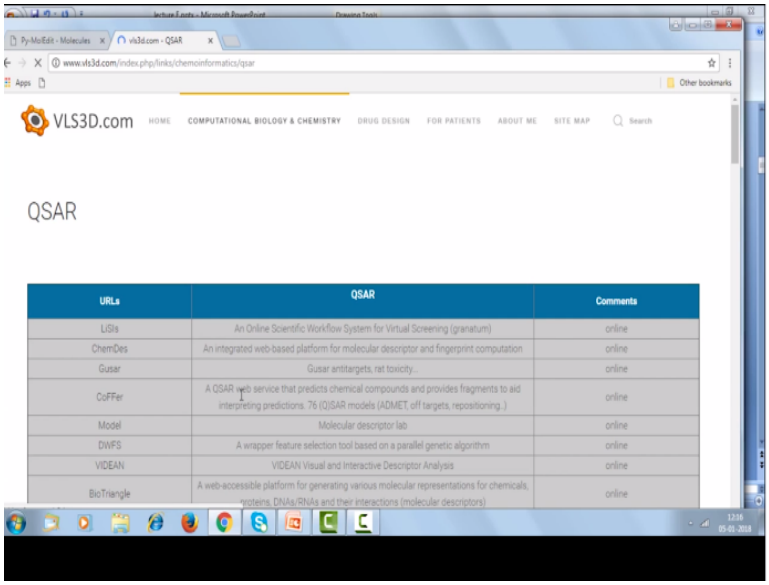
molecular orbital energy and so on you have to use MOPAC. Jmol is for visualization okay. So we looked at all these.

(Refer Slide Time: 01:46)



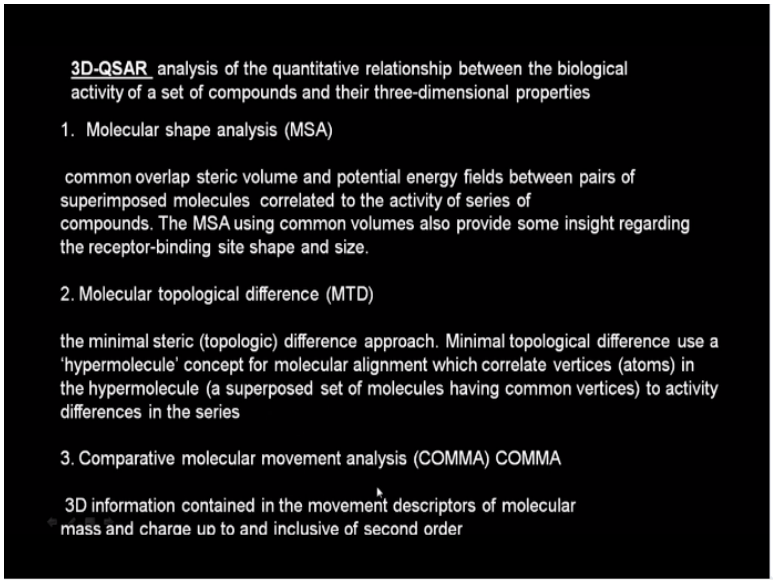
Now there are more softwares on QSAR using this link. For example, I will just show you, we will not have time to spend on so many okay.

(Refer Slide Time: 02:14)



So lot of softwares you can see on QSAR you can start exploring as you can see this okay, a lot of softwares are there in QSAR and some of them are downloadable and some of them are based through the webserver. So we will not as I said have time so please start exploring okay. So what is this 3D-QSAR?

(Refer Slide Time: 02:42)



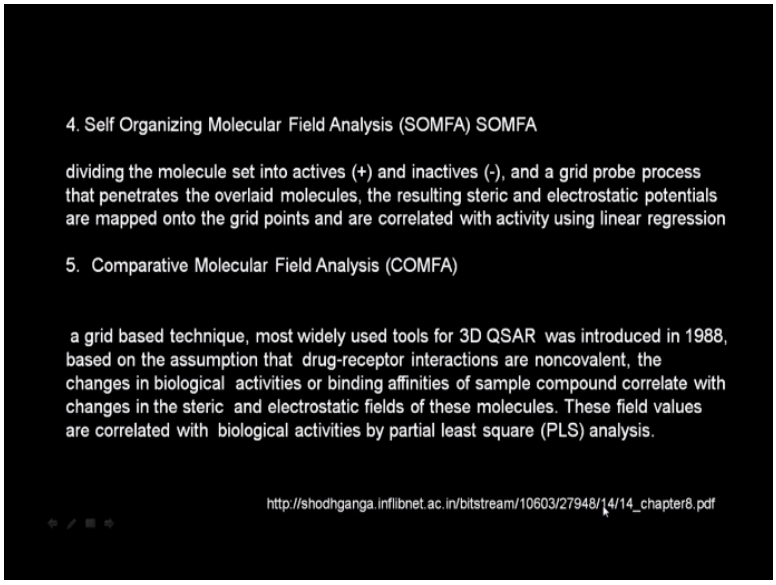
See quantitative relationship between the biological activity of set of compounds and their three-dimensional properties okay. So ultimately like I said the molecule occupies a certain conformation, which is very important for the activity, so not all conformations maybe useful for the activity because as you know it finally goes and binds to an enzyme so the conformation plays a very important role okay.

So there are different approaches one is called the molecular shape analysis approach. So what do we do here is we look at common overlap steric volume and potential energy because its molecular shape between pairs of superimposed molecules, they are correlated to the activity. So it uses common volumes, provide some insight regarding the receptor-binding site shape and size because it is more to do with the volume that is the molecular shape analysis.

Second is called the molecular topological differences okay. This is the another approach, so what it looks like the minimal steric topological difference approach, minimal topological difference uses a hypermolecule concept for molecular alignment. Then, it correlates vertices in the hypermolecule to activity okay that is called molecular topological difference. The third approach is called comparative molecular movement approach.

So basically it looks like a movement descriptors of the molecular mass and charge up to an inclusive of second order. This is called the comparative molecular movement approach.

(Refer Slide Time: 04:25)



The fourth approach is called the self-organizing molecular field analysis, SOMFA so it divides the molecule into actives and inactives, then a grid probe process that penetrates the overlaid molecules so we overlay all the molecules and then you have a probe molecule and then the resulting steric and electrostatic potentials are mapped on the grid points and/or they are correlated with the activity using a linear regression okay that is the fourth approach.

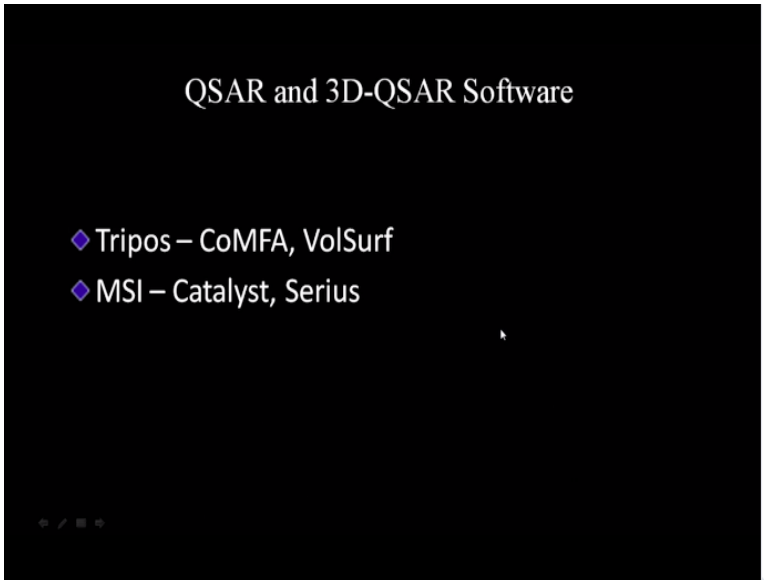
The fifth approach is also a grid based technique okay and is called the comparative molecular field analysis, COMFA. This is very, very common. If you look at 3D-QSAR most of the commercial softwares use this, so this is also a grid based technique and this was introduced in 1988 on the assumption that there will be drug-receptor interactions are noncovalent.

As you know, the drug goes and binds to an active site, it does not form a covalent bond but mostly non bonded interactions like electrostatic, Van der Waals, hydrogen bond and so on. So the changes in biological activities or binding affinities of sample compound correlate with the changes in the steric and electrostatic fields only. So these field values are correlated with the biological activity using partial least square analysis.

This is a statistical tool, statistical approach so we will not go into that but what approach is they have a probe, they first will superimpose all the molecules and then they have a probe and then it looks the probe moves at different locations on the grid and then it determines the steric and electrostatic fields interaction and then that is when correlated with activity using a partial least square technique.

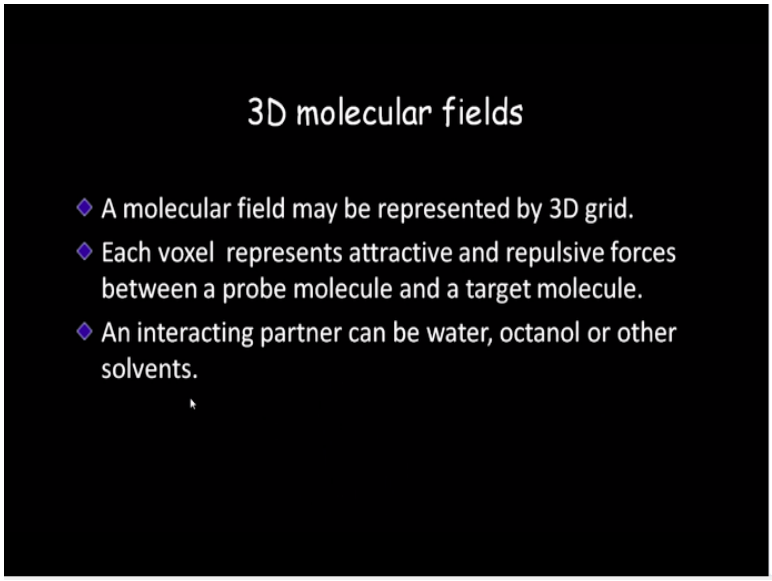
So this is the reference which I would like to acknowledge for all these five techniques, which they describe more in detail. We can have a look at it if you are interested to understand more of that.

(Refer Slide Time: 06:31)



So the commercial softwares also used Tripos, MSI. They all use this type of CoMFA analysis, volume surface, they all used for 3D-QSAR okay.

(Refer Slide Time: 06:44)

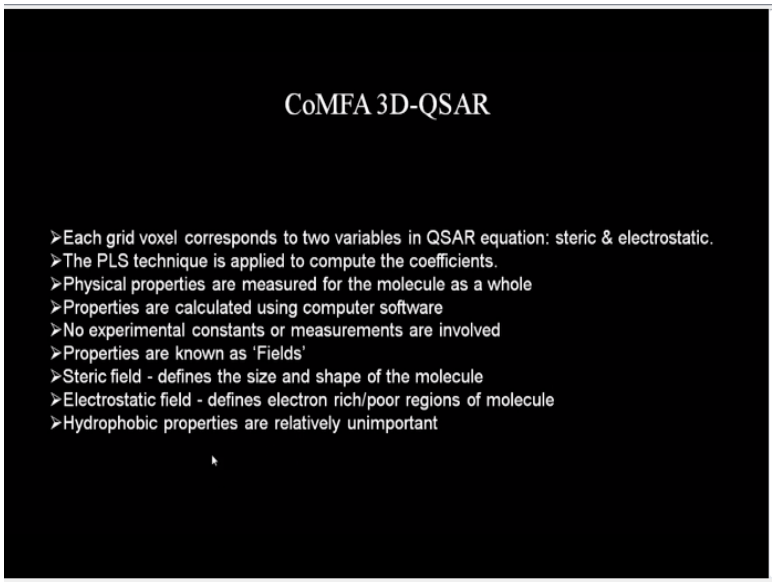


So the molecular field may be represented by a 3D grid okay. So you divide, you create like a big rectangular box and then divide into small, small grids. Each voxel represents attractive and repulsive forces between the probe molecule and the target molecule. So if you have the

target molecule placed and you have a probe molecule placed at each of this grid we look at to the attractive and repulsive forces based on steric interaction, electrostatic interaction okay.

So you calculate that. So it can be water, Octanol, it can be even carbon, simple oxygen, H.

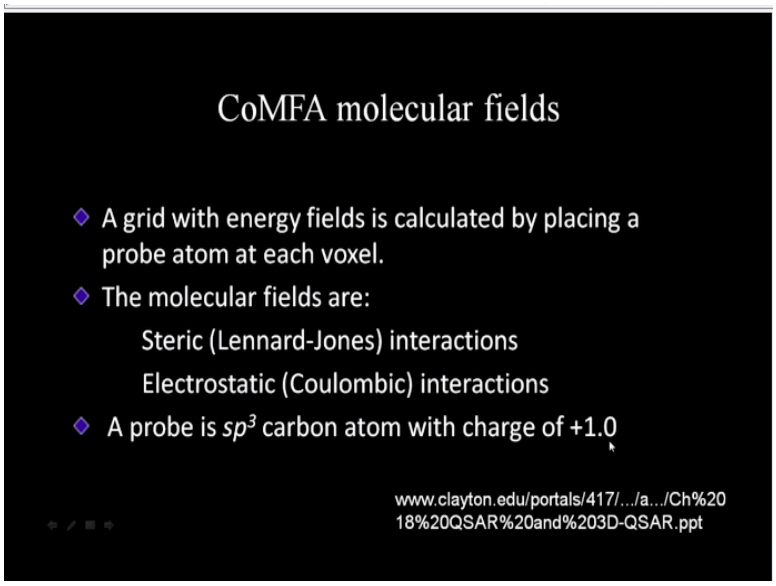
(Refer Slide Time: 07:28)



So each grid voxel corresponds to two variables in QSAR, steric and electrostatic. Then, the partial least square technique is applied to compute the coefficients okay. Once you place a probe and then measure the steric and electrostatic interaction between the molecule and that probe okay, you measure that and then you try to correlate with the activity okay. These properties are known as fields.

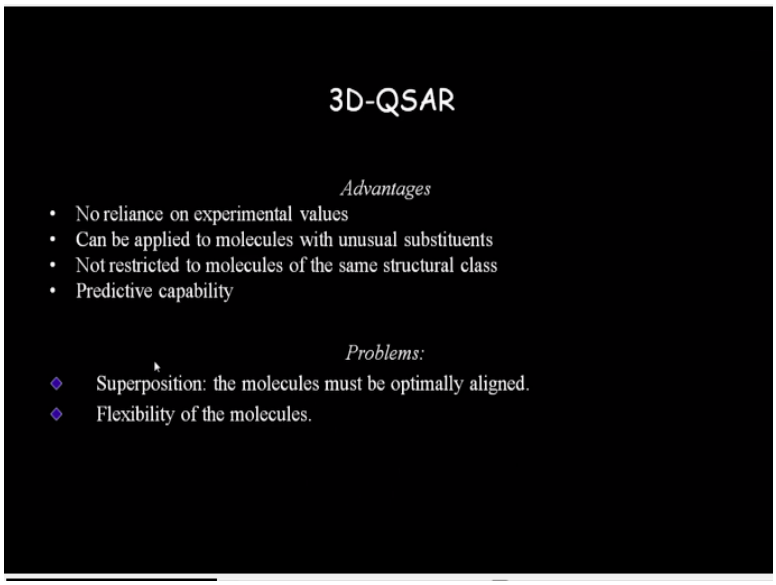
Steric field, it is an indication of the size and shape of the molecule. Electrostatic field, electron rich or poor regions of the molecule okay and then hydrophobic properties are relatively unimportant. It does not bring in hydrophobic property, the COMFA 3D looks at only steric and electrostatic. So it does not bring in the hydrophobic interaction at all.

(Refer Slide Time: 08:28)



A grid with energy fields is calculated by placing a probe atom at each volume. Then, you calculate the steric using Lennar-Jones potential, you calculate the electrostatic Coulombic interactions okay so a probe could be sp3 carbon atom with charge +1 that is the probe okay.

(Refer Slide Time: 08:52)

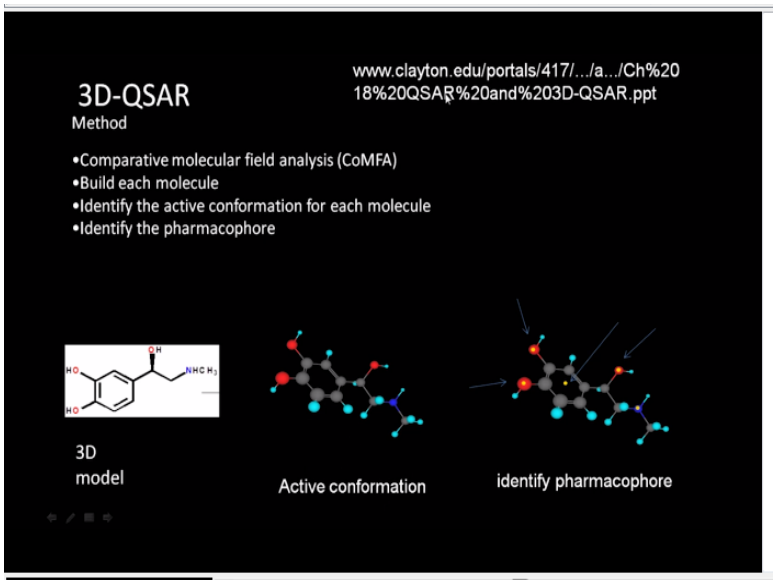


So what are the advantages of this? No reliance on experimental values. We do not need much of experimental accuracies and we apply to molecules with unusual substituents okay, it may be very difficult to do with 2D-QSAR. It is not restricted to molecules with the same structural class because when we talk about in the normal QSAR it is always good to use the same class.

So your QSAR if you are developing for ethers if you extend it to aldehydes, sometimes it may not work okay. It has got predictive capability also. So what are the problems? The

molecules must be optimally aligned that is very important because you are placed one on top of another, they are very, very different, pleasing them may not be really practical. Flexibility of the molecules because the molecules can take different conformations, if you have many rotatable bonds that is the problem okay.

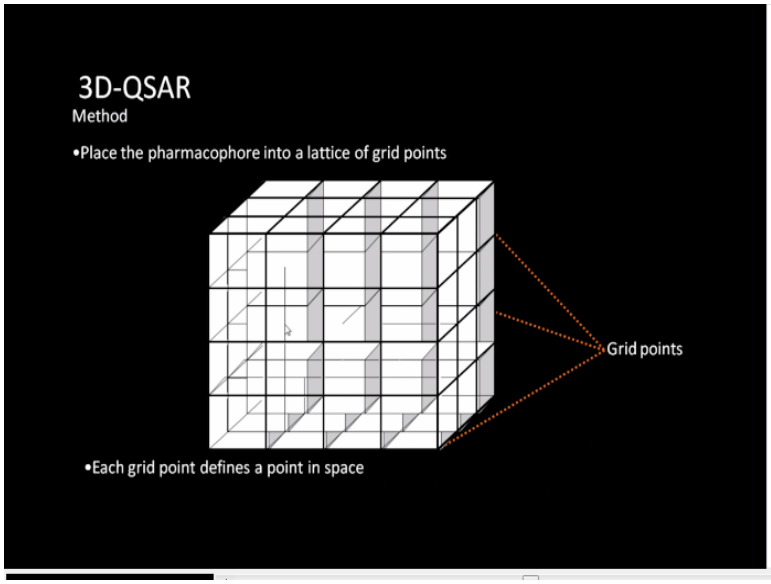
(Refer Slide Time: 09:50)



Let us look at what it is. Say imagine you have a simple molecule like this okay. Then this could be the conformation of the molecule, you identify the pharmacophores of this molecule okay. Pharmacophores could be say hydrogen bond acceptor because this is all oxygen's or it could be a hydrophobic (()) (10:22). These could be the pharmacophore so what do we do? Build each molecule, identify the active conformation for each molecule, identify the pharmacophore.

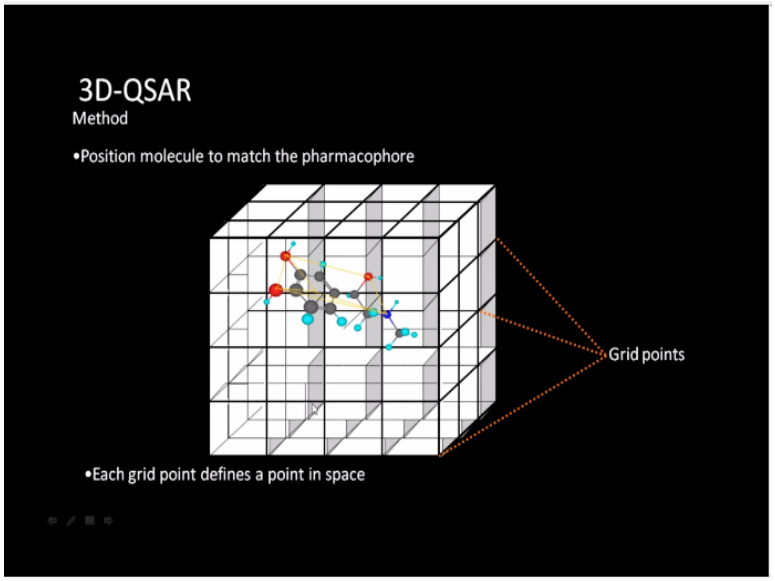
So if you superimpose all of them then it is easy so these knowledge was taken from this reference, you guys can have more look at this to get so I would like to acknowledge that.

(Refer Slide Time: 10:49)



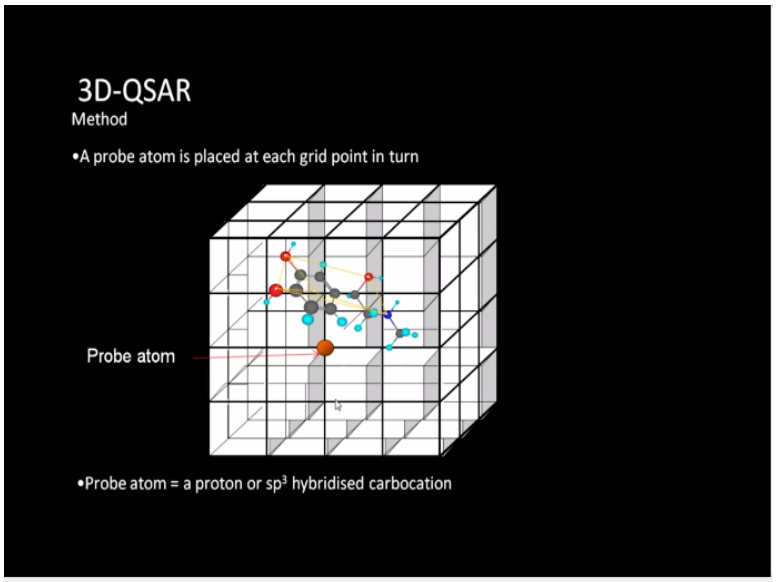
Then, you place the pharmacophores into a lattice of grip points so we may have a rectangular box and if you divide it into lots of grids so you place it. So this could be your rectangular box and then you divide it into small, small grids and then you place your molecule inside okay. So these are the grid points, each grid point defines a point in space okay.

(Refer Slide Time: 11:14)



Place the pharmacophore into the lattice of this grid point okay. Position molecule to match the pharmacophore okay like this, you have matched so that now these pharmacophores are matched exactly on some grid points. Each grid points now define a point in space.

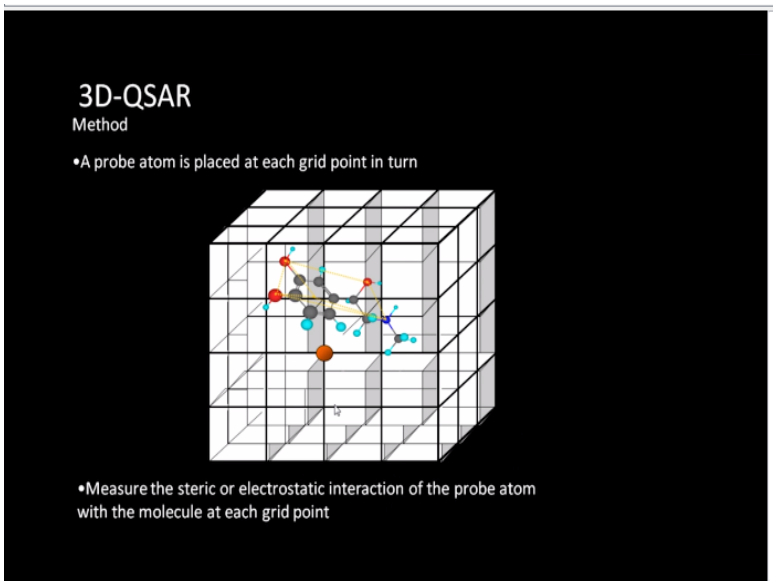
(Refer Slide Time: 11:36)



Now you can have a probe atom like I have mentioned it could be a sp3 carbon so what we do is we move this sp3 carbon at each of this grid points and measure the steric interaction with this molecular repulsion. Let us try interaction of repulsion and calculate it so we will get lots of data and at each of this grid point. So suppose you have 1000 grid points you will get 1000 steric interaction data around and then 1000 electrostatic interaction data okay.

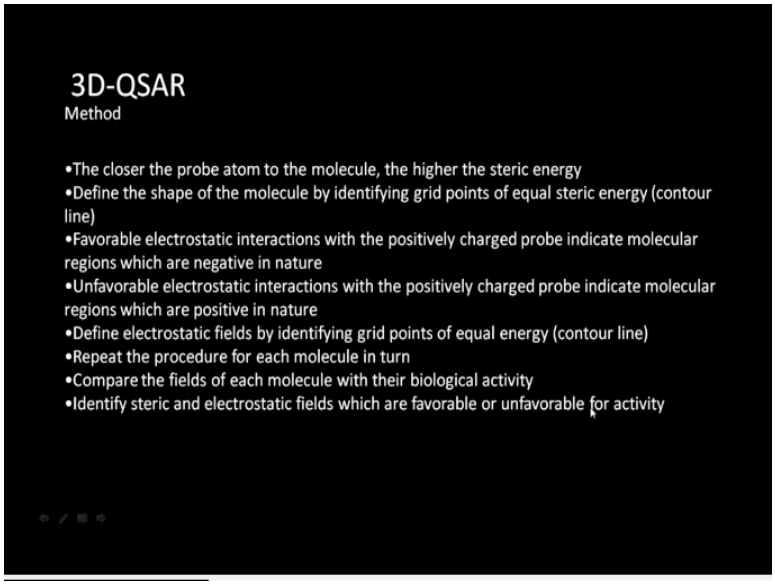
So you will have and then you try to correlate that with the activity using partial least square data okay. So probe atom like I said could be sp3 carbocation, it could be even H or it could be O and so on actually.

(Refer Slide Time: 12:27)



So you place each of these molecule at the grid point okay.

(Refer Slide Time: 12:31)

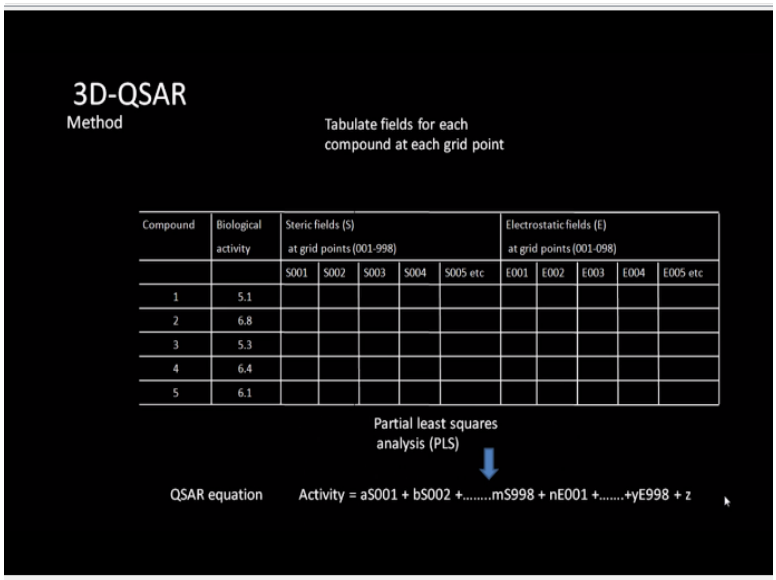


Once we do that so obviously when closer the probe atom to the molecule higher is the steric energy. So it defines the shape of the molecule by identifying grid points of equal steric energy. So we get a contour line on this okay and you can have favorable electrostatic interaction that means with positively charged probe indicate molecular regions which are negative in nature or unfavorable electrostatic interaction that means positively charged probe indicate molecular regions are positive in nature okay.

So it could be favorable or unfavorable. Then we can have electrostatic fields by identifying grid points of equal energy those are all called contour line. So we keep repeating the procedure for each molecule so it places each molecule and then we repeat the same procedure. So for each molecule I will get lots of data with respect to each grid point related to the electrostatic interaction or repulsion, steric interaction or repulsion.

Then, compare the fields of each molecule with the biological activity. So obviously we can do a QSAR which is called the 3D-QSAR. So you identify steric and electrostatic fields, which are favorable or unfavorable for activity.

(Refer Slide Time: 13:53)



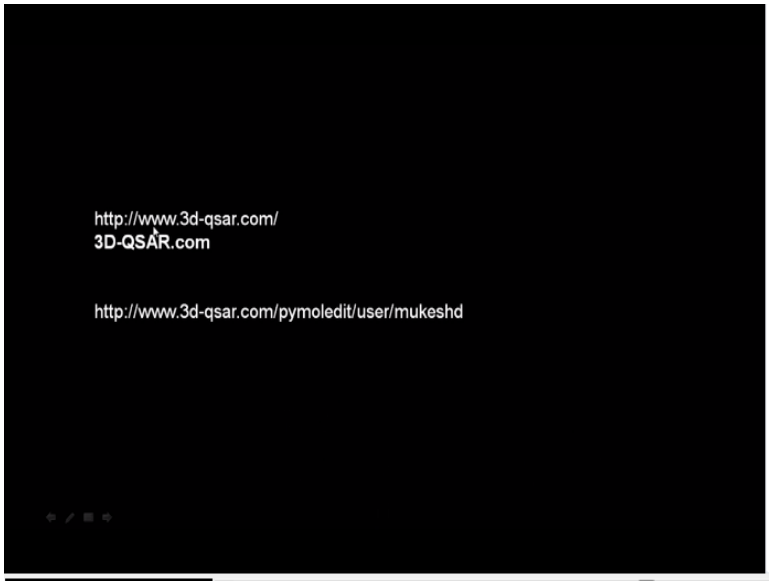
So compound 1, 2, 3, 4, 5, so this is the biological activity so this could be the grid points okay. So at each of the grid points you have the probe and then you get the steric interaction. Then, at many places it is here, you get some number, when you place it here you will get some number, when you place it here you get some number and so on but with respect to steric field.

Again with electrostatic when you place it here, see 001 we get some number and so on. So then you take the second molecule, then you take the third molecule, then you take the fourth molecule, then you take the fifth molecule and so when you keep doing that you get lots of data with respect to molecular weight and you want to correlate with the activity.

And we will get lots of each one sort of could be a descriptor of course like a normal QSAR we cannot just take one and then try to get a regression relation but we make use of all of them and that is how partial least square technique works. You take all of them and create some composite descriptors and then try to correlate with the biological activity that is how we does okay.

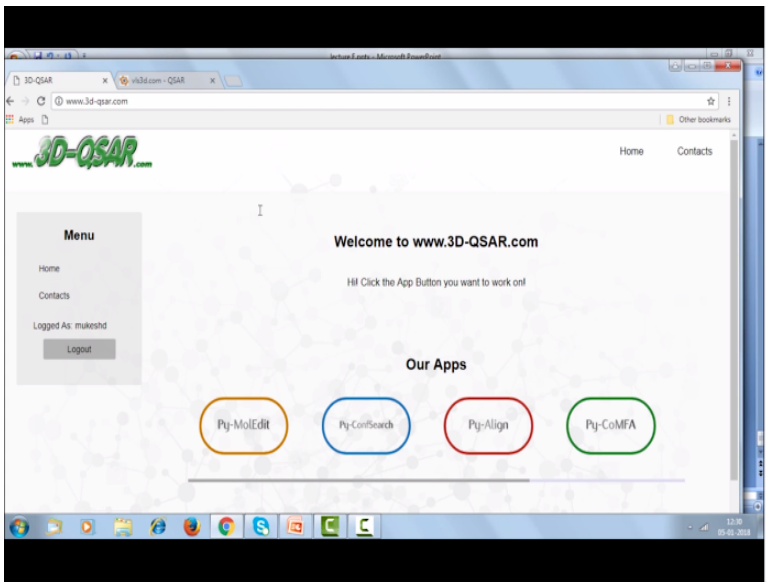
So you use a partial so QSAR equation have to be like this like this like this like this okay. So I will not go into the statistical technique called partial least square analysis so you people have to look into some standard book.

(Refer Slide Time: 15:27)



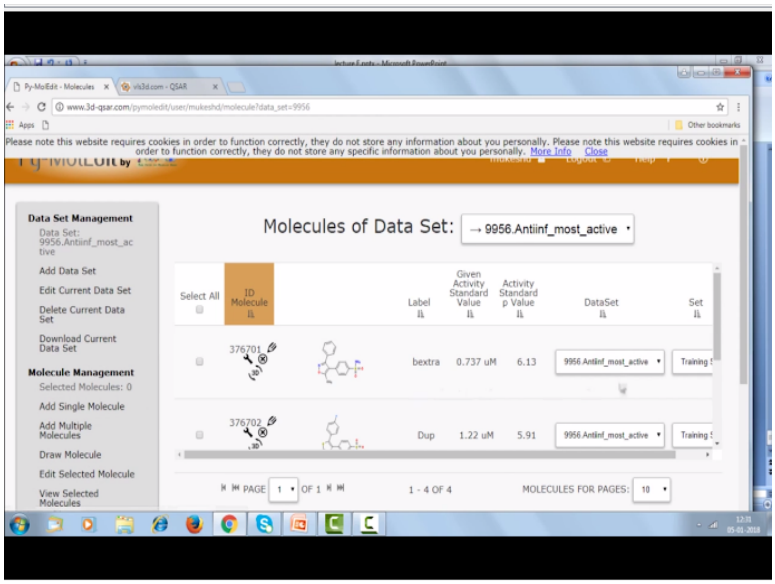
So let us also look at software. This is freely available software. It is called 3D-QSAR.com. So all you need to do is you need to get a username, password so with that username, password you can log in into that okay. Now you can log in to that so this is the QSAR, this is the internal menu.

(Refer Slide Time: 15:56)



So we need to get in username, password so we write to them, they will give you access. So I had already loaded some molecules. These are anti-inflammatory okay. So I have loaded 4 molecules, I have loaded their activities also.

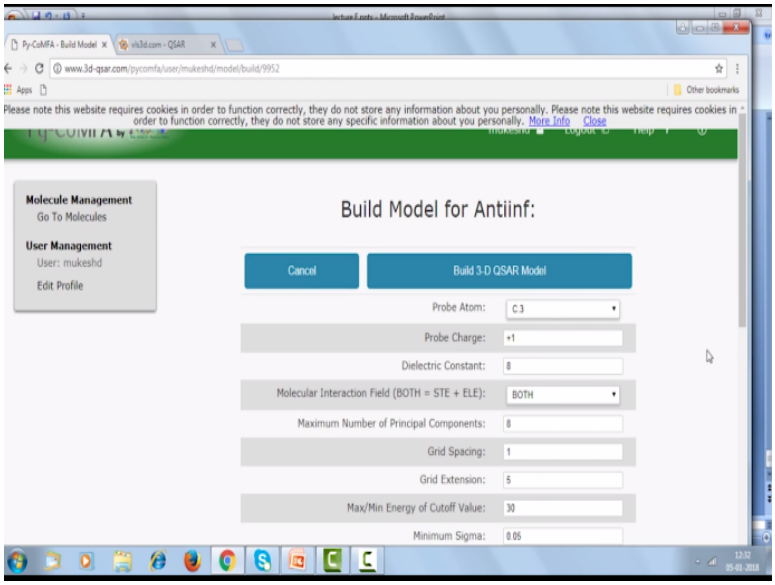
(Refer Slide Time: 16:20)



These are selective Cox-2 inhibitors as you can see okay this is Bextra, this is DuP, this is Rofecoxib and this is called Celecoxib okay. So 4 molecules I have loaded and I have also loaded their activity values also here as you can see here okay these are the activity values. As you load DuP has the highest activity, so these 4 molecules are selective Cox-2 inhibitors. You can load it through SDF file.

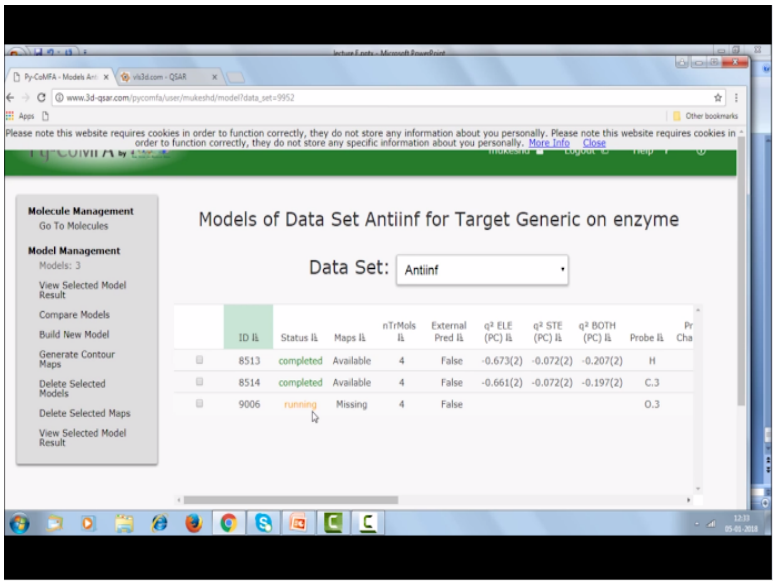
I can load more molecules, add another molecule or even draw molecules okay. So these are all anti-inflammatory drugs okay. So I have this I can go to the models. I can say build a model for selected data set.

(Refer Slide Time: 17:30)



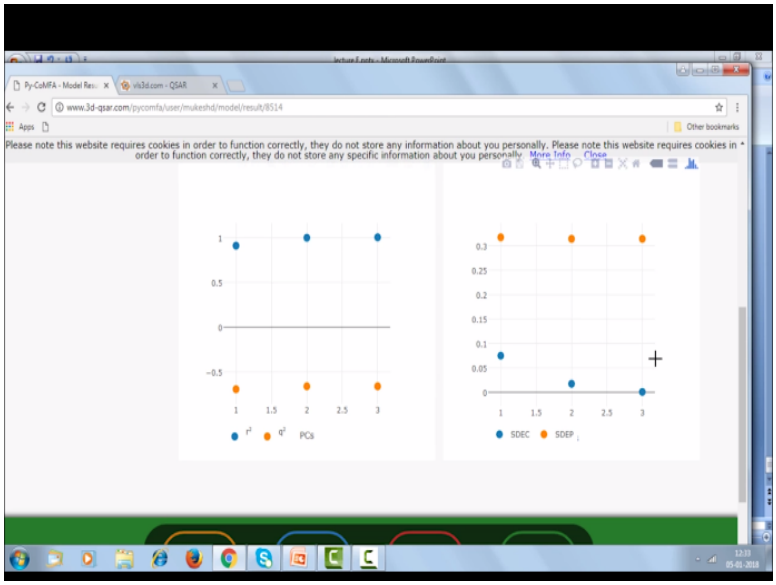
So the probe atom could be C3 that is the sp3 carbon like I said it can be O3 or H or C2 or N3 okay. So I can say O for example, probe charge I could change it to -1 so both steric and electrostatic okay.

(Refer Slide Time: 18:06)



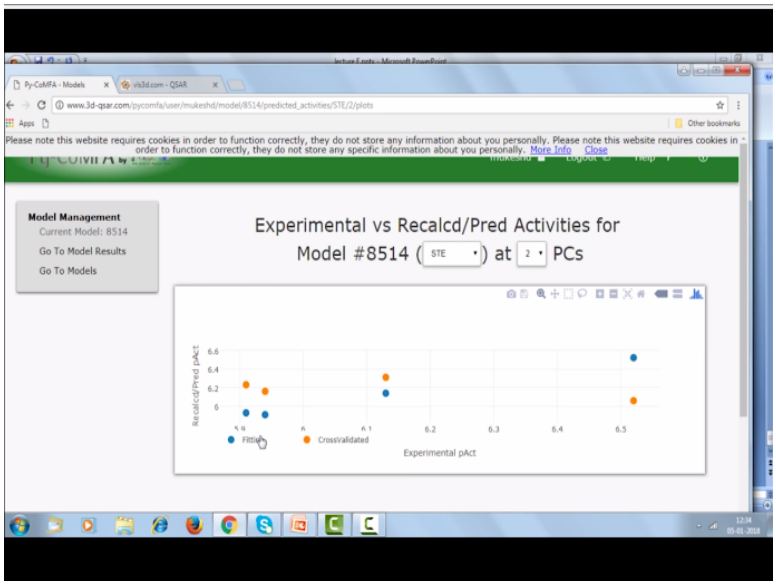
Build 3D-QSAR model, running so I had done before couple of them okay. Now this is oxygen, I have a C3 here so we can have a look at this if you want, it gives you the Q square, Q square is not very good okay. You selected model results okay.

(Refer Slide Time: 18:40)



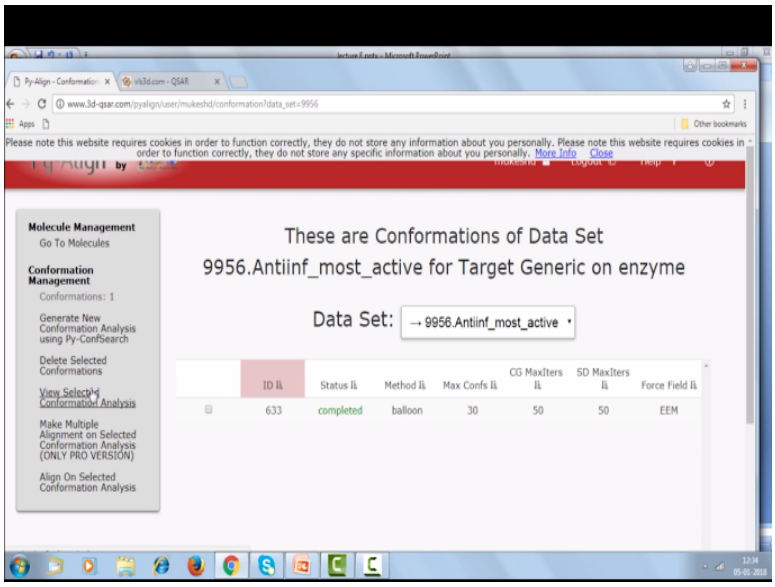
So this gives you some idea about the model results. As you can see, this is the R square and Q square for this and this R square is good. Q square is not good for this model. So we can view experimental recalculated.

(Refer Slide Time: 19:00)



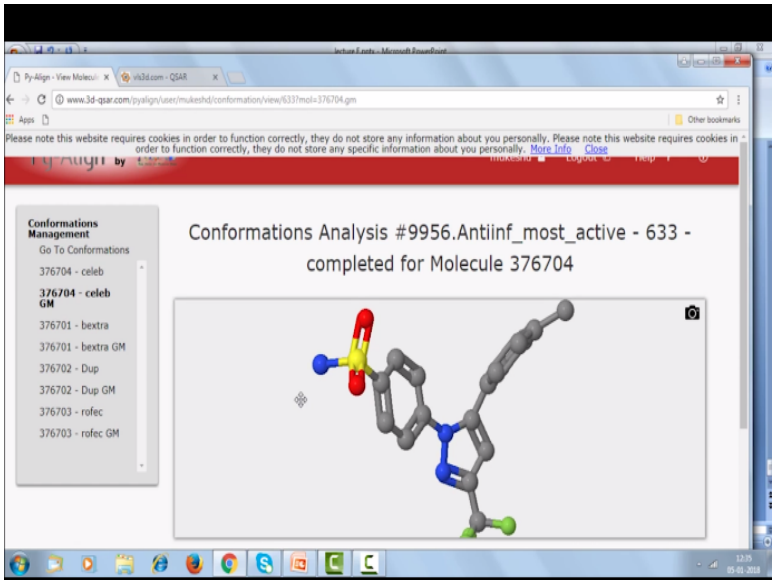
So it gives you experimental value fitting and then it gives you the cross-validated R square okay here and so on. We can also look at alignment. If you look at the alignment of these molecules okay, view aligned molecule, go to conformations okay, view selected conformation analysis.

(Refer Slide Time: 19:34)



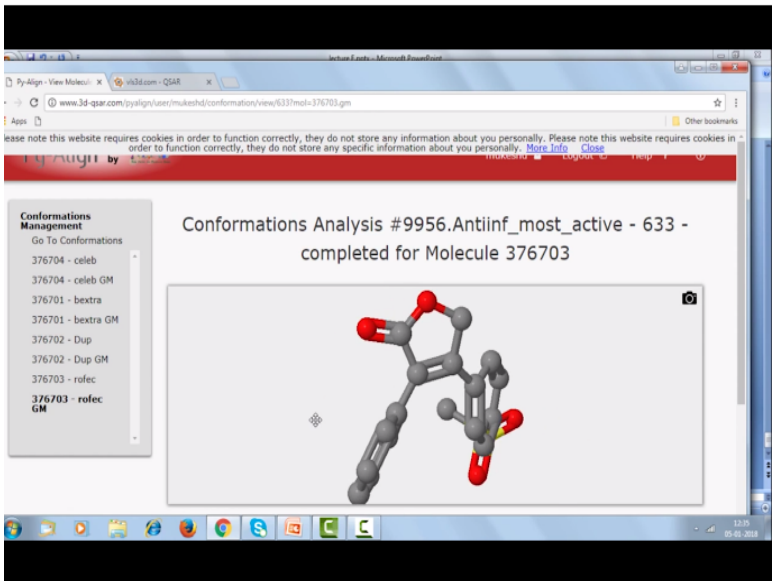
So these are the 4 molecules, Celebrex, Bextra, DuP (()) (20:00). So as you can see this is the Celebrex various conformations aligned. Yeah this is Celebrex various conformations aligned and this is Celebrex alone simple molecule which had given okay.

(Refer Slide Time: 20:14)



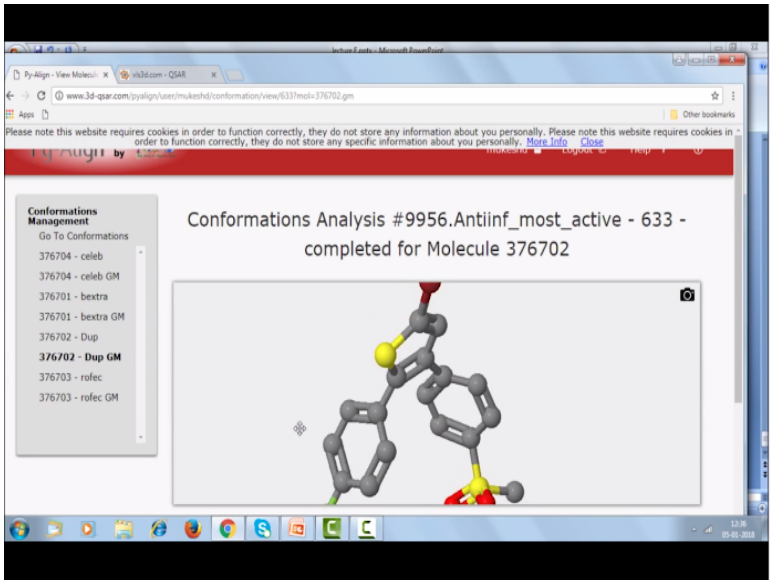
So this is simple which had given here okay, this is Celebrex as you know this sulphur double bond O double bond O. This is various conformations of Celebrex superimposed on each other as you can see here, this is Rofecoxib.

(Refer Slide Time: 20:38)



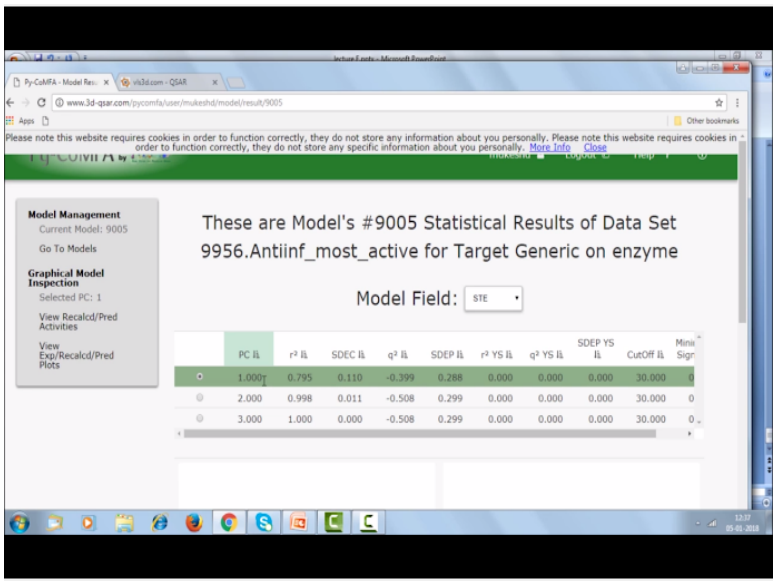
Again, this is called a sulphur and double bond O double bond O here. These are the various conformations of Rofecoxib aligned as you can see. So from this we can get the pharmacophore importance okay. This is the other molecule DuP okay.

(Refer Slide Time: 21:00)



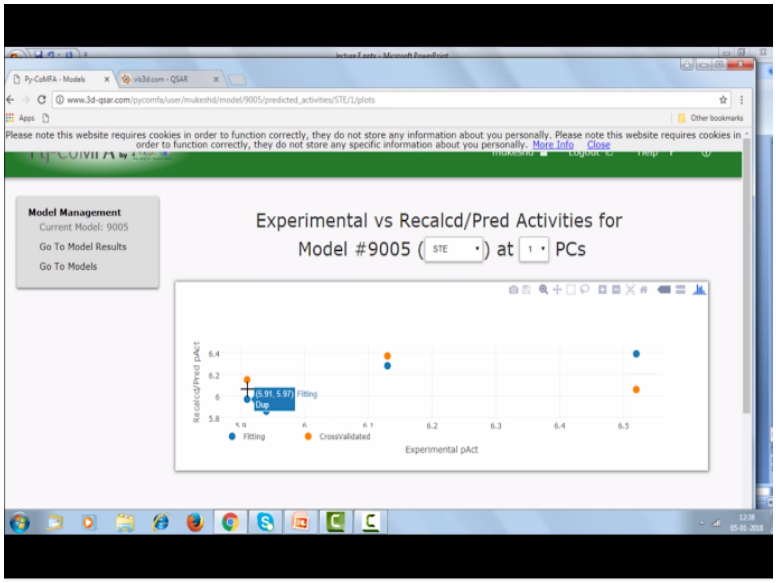
This also has got a sulphur-O2-O2 and this has various conformations of tube as you can see here okay. So this one has been completed just now, yeah this one has been completed just now. We have selected model results okay.

(Refer Slide Time: 21:31)



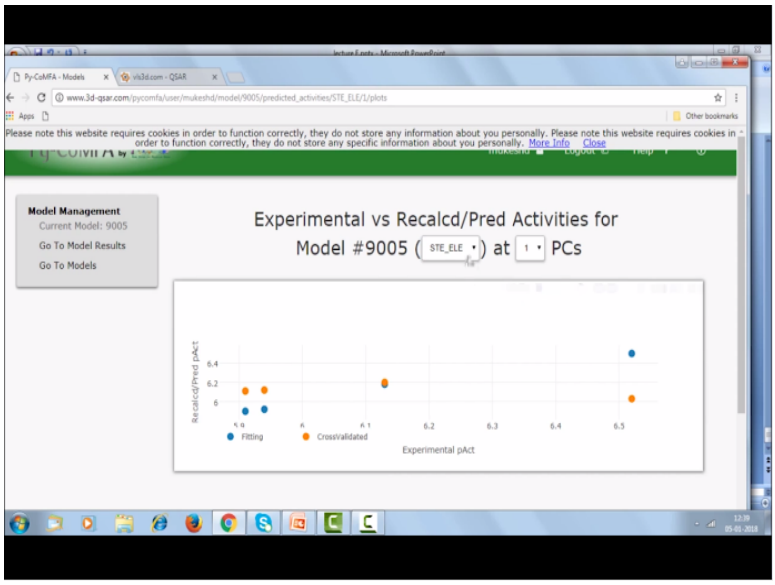
So this has been different partial least squares methods have been used. So we can see this R square spreading good. Q square is not good then these are different statistics. I think these are different statistics, which has been used okay as you can see here for each one of these models how the R square varies okay. So this is the fitting here okay. This is recalculated predicted activity was its fitting here.

(Refer Slide Time: 22:18)



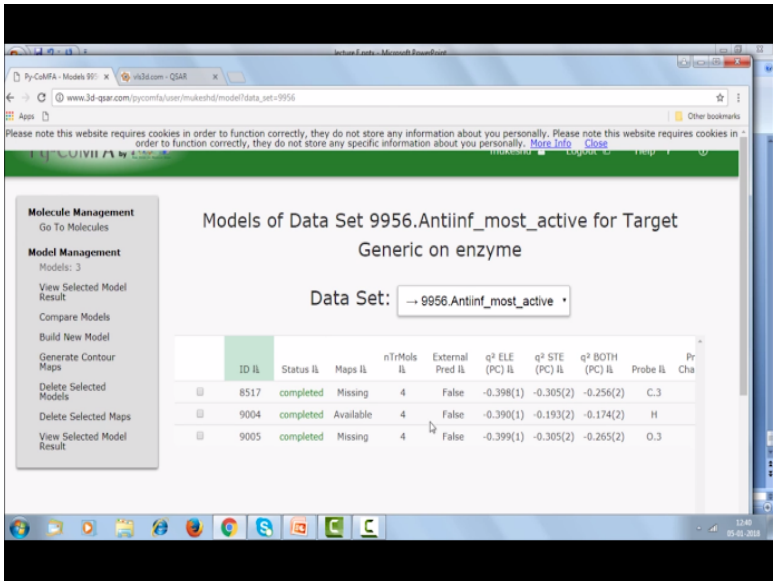
This is cross validated as always as you can see here. So Bextra, this is for Rofecoxib, this is for Bextra, this for DuP and this is for Rofecoxib okay. So this is steric, so we can look at steric and electrostatic here. So this gives you steric and electrostatic.

(Refer Slide Time: 22:52)



So this is the fitting, cross validated this is the fitting okay this is the predicted versus okay as you can see here the fit is good for Bextra, fit is not good for Rofecoxib, fit is reasonably good for the DuP okay. So as you can see go to models okay.

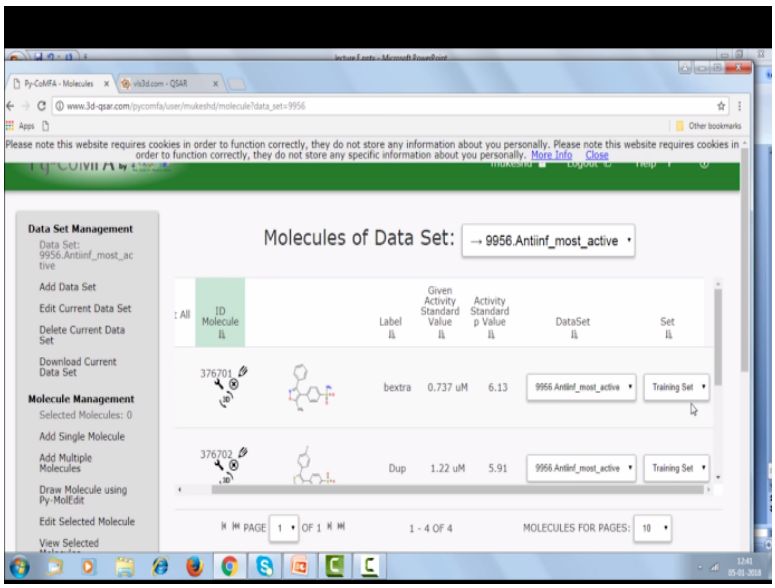
(Refer Slide Time: 23:53)



As you can see I have done anti-inflammatory mostly Cox-2 inhibitors like Bextra, Rofecoxib and DuP-6 okay and then I used probe C3 that is sp3 carbon, oxygen then I also used hydrogen, it is not hydrogen H. You got different values okay for both electrostatic and steric. So you can play around with that okay. So I can have more molecules here in this example I have just taken 4.

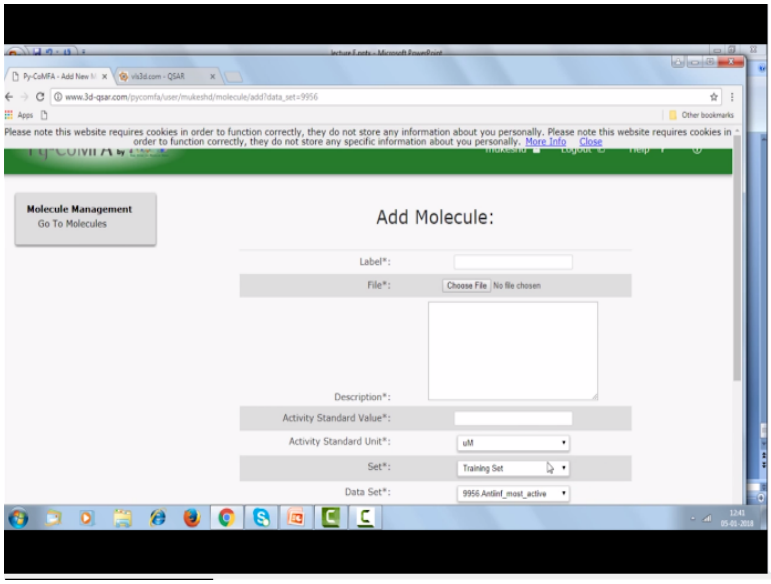
So we can add more than 4, we can also have training set and test set also as you can see here.

(Refer Slide Time: 24:45)



Training set and test set we can do okay. So we can use draw molecules using Py-MolEdit or we can add multiple molecules, add single molecule we can add okay.

(Refer Slide Time: 24:58)

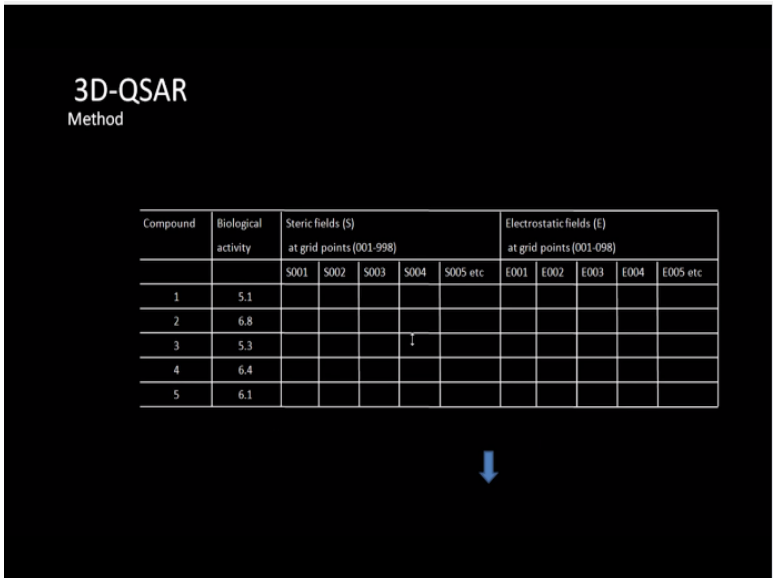


We can give the activity and then we can give title for that, choose file, so we can choose the file from where we want to. So it is quite user-friendly okay you can say done then it can be training set molecule or it can be test set molecule, so it is quite simple. I suggest that you keep playing around with this particular software and because it is free and you just have to get a username, password.

And once you go to that you can start playing with this. So the 3D-QSAR principle is it superimposes the molecules. So first you look at the molecules, you get different conformations, find out the pharmacophore and then you place the molecule in a grid box so that the pharmacophores are properly kept and then you take a probe molecule like atom, sp3 carbon or it could be a O or it could H and then you place it at each grid point and calculate the steric and electrostatic information.

So how you will calculate the steric and electrostatic information? Use the Lennard-Jones interaction, electrostatically look at the Coulombic interaction right and then you place it at each grid point when you do that you get some numbers, once you get these numbers what do you do?

(Refer Slide Time: 27:00)



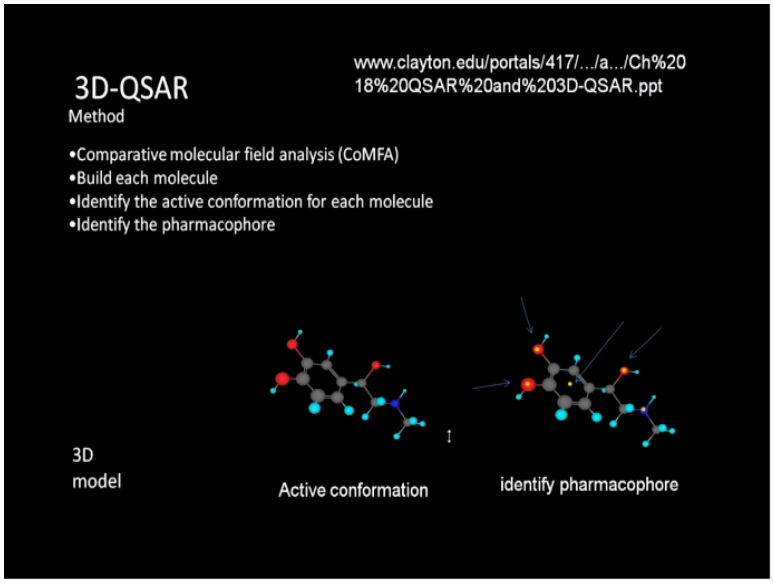
You make a table as I mentioned make a table so at each grid point you for a steric Lennard-Jones we will get some number and then electrostatic okay. Some electrostatic numbers Coulombic then you will keep the second molecule, align it properly and then you do the same thing here, then third molecule, fourth molecule, fifth molecule, you also have the activity here okay.

So then you have actually you have too many descriptors right. You want to make use of all the descriptors unlike your normal QSAR we used to develop a model, a linear regression with one descriptor. Here we want you make use of all the descriptors and that is why we use something called partial least square method okay. So it creates some composite knowledge based on all the descriptors.

Basically, these are all descriptors like okay and then try to develop an equation for biological activity okay. So that is how it generates the QSAR and once it generates the QSAR we can calculate R square, we can calculate Q square, so many things. So we can also have just like normal QSAR we can also have a training set like I showed you in my example of software.

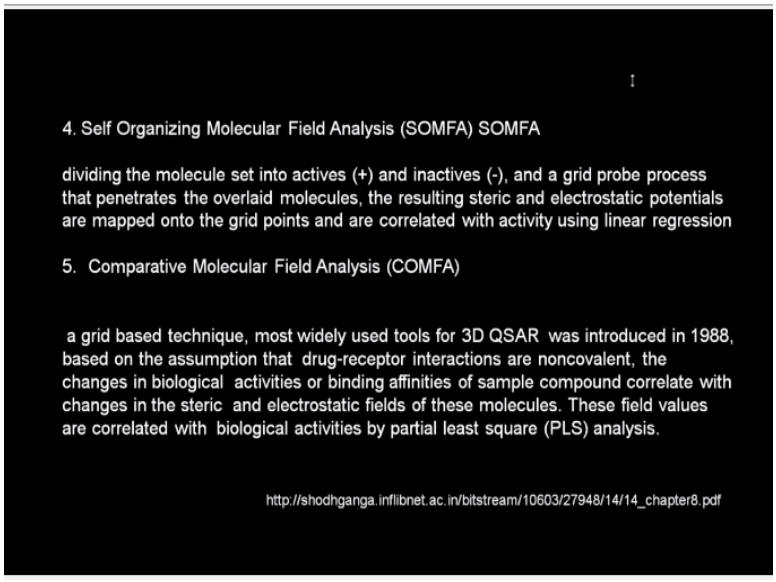
Then, you can have a test set and then you can try to predict the activity of the test set compounds with the model which has been 3D-QSAR model which has been generated using the PLS techniques okay. So 3D-QSAR is very powerful because ultimately molecules take a conformation and they go and bind to the active site the enzyme so that is the very, very important parameter.

(Refer Slide Time: 29:06)



So some of the information which I have shown in the slides taken from this reference as well as from this reference okay.

(Refer Slide Time: 29:15)



So I like to acknowledge both. So there are many softwares for QSAR, which I just briefly showed and I also showed you some examples of softwares, which can be run for both 2D-QSAR as well as for 3D QSAR. So I expect that you can start exploring these okay. We will continue on a new topic in the next class. Thank you very much for your time.