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

Lecture - 33
Target Based Drug Design

Hello everyone, welcome to the course on Computer aided drug design. Today, we are going to start a new topic that is called a target based drug design or a docking as it is popularly called. So far we looked at structure-based drug design, that means we looked at the 3 dimensional structure of the ligand, we looked at the physicochemical properties of ligand, we looked at the pharmacophore features of the ligand.

we developed structural activity relationships, so we never knew anything about the target at which this particular drug is going to go and bind, it would be a protein or an enzyme or it could be a channel, so we never knew anything about it. So our focus has been only on the cheminformatics properties of the drug okay. Now let us look at how that drug goes and acts on enzyme or protein and makes the protein or enzyme inactive.

So this protein could be in the pathway of the particular business process okay, so that means I need to know the mechanism of action of the particular disease, and I need to know what are those enzymes involved in that particular disease process, and which enzyme I am going to target. So I need to know more, if you look at 15 or 20 years back there was no knowledge about much understanding about mechanism or target proteins and so on.

Because our analytical tools were not very great, so most of the drugs were discovered based on their activity. For example, if you look at sulphur drug, you look at the penicillin they were tested on bacteria various types of bacteria, and they were found to act very effectively that means they were able to kill the bacteria, hence the drugs came into the market, but nobody knew how they acted.

But later on lot of research done to understand how the drugs acted the mechanism of action was understood. So in the past 15 years with the improvement in analytical tools, with improvement in computational tools or there was a lot of possibilities of finding out how the

drug acts, and what are the targets involved through which the drug acts, that means which proteins get disturbed or get inhibited because of this drug, so one was able to do that.

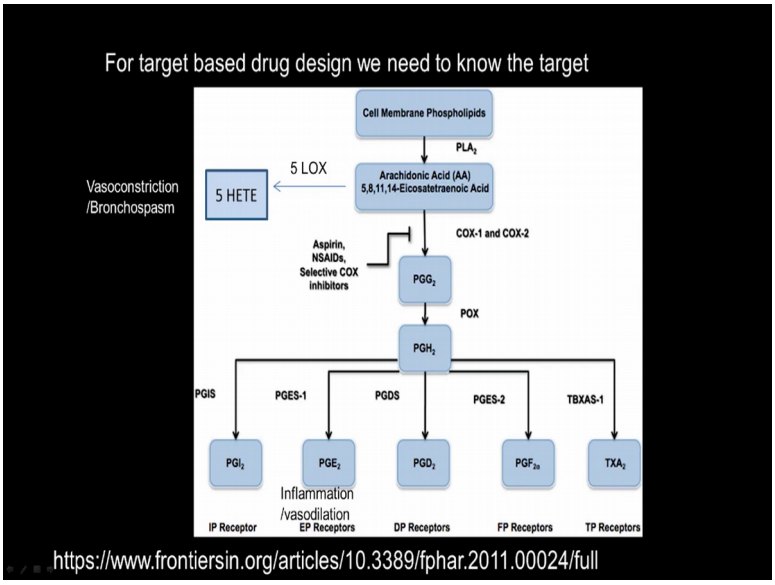
By looking at the target one can design very effectively or very efficiently drugs which may be very selective to the particular target, that means they may not go to and bind to other targets, so you are reducing the side effect profile of the drug, so they can go and bind very effectively to the target, and stop by the disease process. So the target based drug design looks at the specific enzyme structure, the active site of the enzyme.

And design molecules which will go and bind very effectively very selectively and not go and bind to other enzymes or proteins thereby they can become very selective as well as they could have better efficacy. So this is what is called as target based drug design, that means I need to know a lot of about their mechanism, I need to know the 3 dimensional structure of the target protein or enzyme which I want to focus on, I need to know the active site of this enzyme.

And as I said in the past 15 to 20 years' lot of development happened in proteomics, mass spectrometry, x-ray crystallography, NMR, computational tools, protein purification, separation, just like 2D gel electrophoresis. So all these simultaneous developments of computational as well as experimental tools has helped quite a lot in going towards the target based drug design.

And as you see nowadays FDA that is the food and drug administration of US expects all the pharma companies when they file a new chemical entity to have the details about the mechanism or action as well as the target proteins on which the drug acts. They are not just happy about the efficacy of the compound, but they would like to know exactly where the compound goes and where it binds, which enzymes it inhibits and so on actually okay.

(Refer Slide Time: 05:07)



So for example, let us look at this particular picture is taken from this reference okay, this deals with the inflammatory pathway it is called arachidonic acid pathway, arachidonic acid pathway inflammation okay so there is a cell membrane, phospholipids that are produced, there is an phospholipid enzyme which converts them at arachidonic acid okay. So this is the site of inflammation once AA is found there are some enzymes like cyclooxygenase 1, cyclooxygenase 2 which convert the arachidonic acid okay into PGH2 that is PGH2 to here sorry as PGH2 here okay.

And so PGH2 here and this PGH2 is acted upon large number of enzymes, prostaglandin E synthase that is this one which converts the PGH2 into PGE2 okay, so this is how the inflammation progresses okay inflammation vasodilation and so on. This PGE2 goes to EP receptors okay, then there is a PGDS enzyme which converts that into PGD2, then there is a PGFS enzyme which converts into PGF.

And then there are thromboxane, then there is PGIS enzyme which converts into PGI2, so the PGH2 is being converted by large number of enzymes to various products okay. Simultaneously, we also have another enzyme called lipoxygenase okay, lipoxygenase which converts the arachidonic acid into 5 HETE which is involved in bronchospasm vasoconstriction okay. So you have large number of enzymes in the arachidonic acid pathway.

Large number of enzymes in the arachidonic acid pathway which are involved in the inflammation vasoconstriction and vasodilation and so on actually. So if you look at original

inflammatory drug aspirin is quite a good drug okay, which is called a nonsteroidal anti-inflammatory drugs, because originally steroids are given for inflammation especially steroids at here.

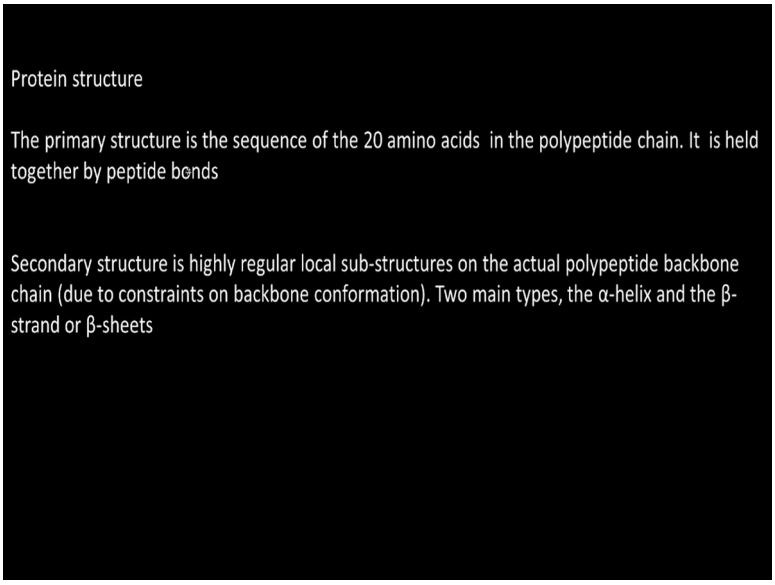
And so aspirin the nonsteroidal anti-inflammatory drug, so it goes and acts on many of these enzymes cox-1 and cox-2 so it is being not selective actually. Then came a selective cox-2 inhibitors, I have been talking about it like (()) (08:17) which will go and bind only to this particular enzyme called cox-2 okay, but not to other enzymes. And then there are some drugs which may be targeting, this there are some drugs which maybe targeting this.

So if you take a mechanistic pathway like inflammation, one could target different enzymes to achieve different okay scenarios okay, one could target this enzyme alone, one could target this enzyme alone or this enzyme or combination of this enzyme and this enzyme depending upon what is decide. So then the target based drug design comes into picture, so I need to know something about the target protein, the size of the protein, the active site of the protein and so on actually.

Okay, this is where the target based design comes into picture okay, so that means I need to know something about the target, this mechanism I need to know, I need to know the details about the target, the structure, the shape, the 3-dimensional conformation the protein takes, all these things I need to know that is very, very important. And another important interesting thing is you maybe discovering drugs for this enzyme cox-2 there may be some companies which may be discovering drugs for lipoxygenase.

Some companies may be trying to target both cyclooxygenase 2 and lipoxygenase and so on actually. So for same inflammation there could be many companies working towards different enzymes okay, so this picture was taken from this reference okay. So I need to know the protein or enzyme or whatever the target, I need to know that that is very, very important okay.

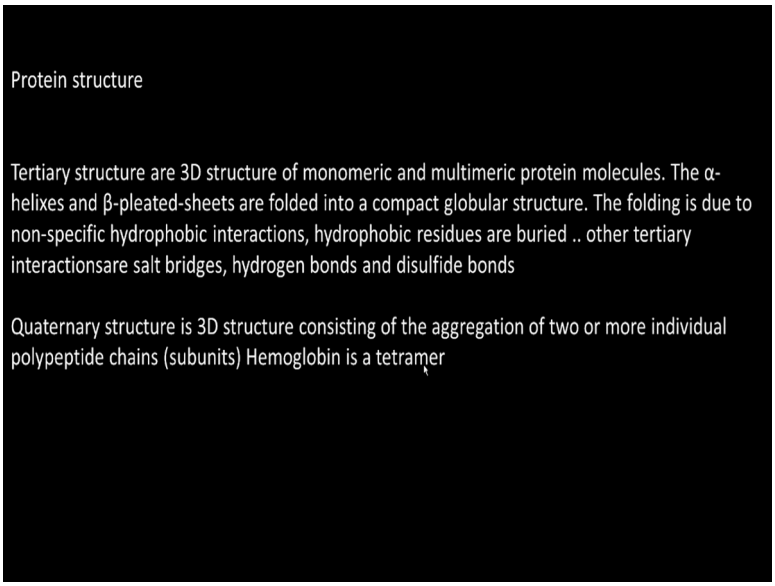
(Refer Slide Time: 10:21)



So the primary structure, so the proteins have 4 different structures: the primary, secondary, tertiary and quaternary. So the primary structure is sequence of the 20 amino acids in the polypeptide chain okay, it is held together by peptide bonds, what is the peptide bond? C double bond N right, so that is the peptide bond that is formed actually. Because you have the acid there be NH2, so they form a peptide bond.

Secondary structure this is very regular local substructure on the actual polypeptide backbone due to constraints on the backbone, because it is not possible for the backbone to form all types of structures okay. 2 main types, we have that is the alpha helix and the beta strand or beta sheets as you can call, alpha helix or beta strands or beta sheets okay. Then we have the tertiary structure, these are real 3D structures of this monomeric protein.

(Refer Slide Time: 11:33)



The alpha helixes and beta sheets are folded into compact globular structures, so thermodynamically it forms a very compact structure, the folding is due to non-specific hydrophobic interactions okay, so we have the hydrophobic interactions which are generally buried inside, hydrophobic interactions are buried inside because as you know protein is generally found more in the water.

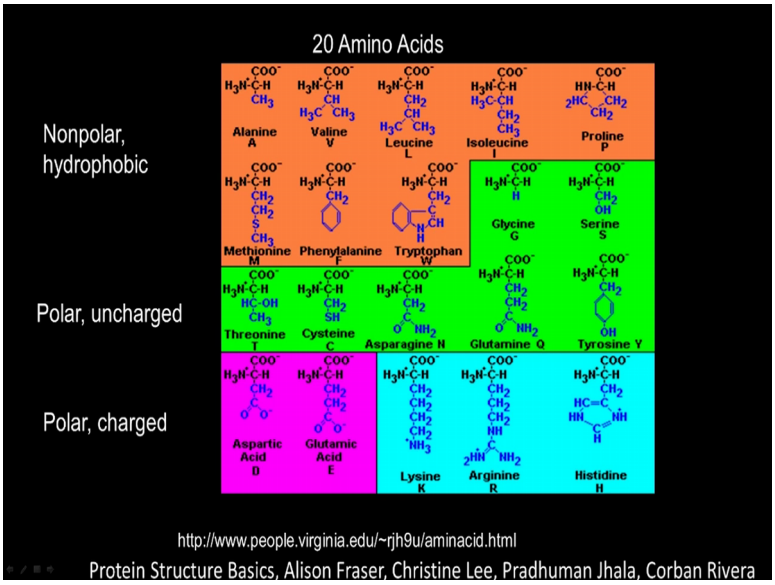
So generally we have all the hydrophilic portions of the protein coming out, and the hydrophobic portion is always buried inside. Then we have the other tertiary interactions are formed salt bridges, hydrogen bonds and disulfide bonds, all these are formed that is why the protein takes a 3-dimensional globular compact structure okay. So we have the primary structure when you say it is only the 20 amino acids arranged in the polypeptide chain okay, by the peptide bond.

But when you have the secondary structure, these regular local substructures because the polypeptide bonds have some constraints on the backbone, so it forms alpha helix or beta strands or beta sheets. Then you have the tertiary which is because of the non-specific hydrophobic interactions which are buried inside, and another interaction such as salt bridges, hydrogen bonds, disulfide bonds all these are formed.

Then we have the quaternary structures, we may have a 2 or more individual these tertiary structures combined together okay, to form because as a whole really they start acting okay. So for example, hemoglobin is a tetramer, if you take that 5 lipoxygenase which I showed you which is involved in bronchoconstriction okay, and also which is which takes arachidonic acid as a substrate it is dimeric okay.

So the quaternary structure in the aggregation of 2 or more individual polypeptides, that is 2 or more tertiary structures connected together actually okay, so they act accordingly. So the proteins have 4 different levels I would say of structures okay. 5 LOX is a dimer okay, this was taken from these 2 references.

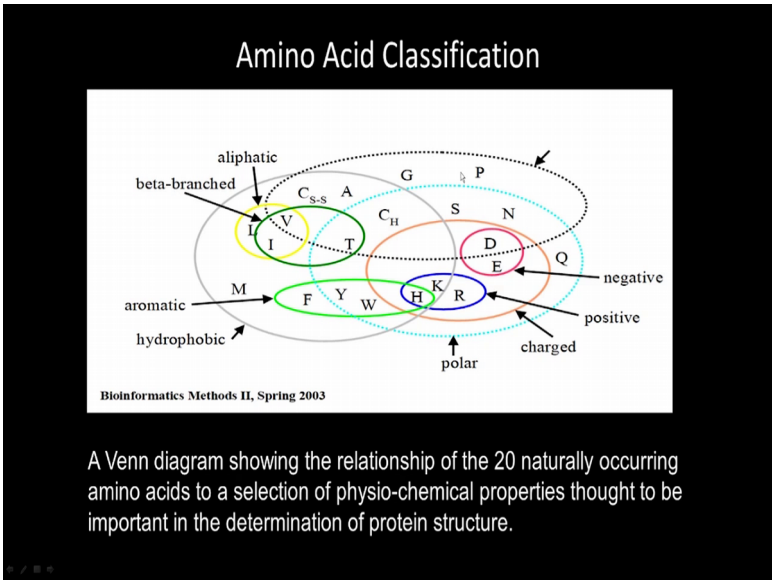
(Refer Slide Time: 14:11)



So it is a very interesting picture if you look at this some of these amino acids like alanine, valine, leucine, isoleucine okay methionine, tryptophan, they are nonpolar hydrophobic okay, they are all hydrophobic. So generally and they will be buried inside in aqueous medium, you generally have only the hydrophilic amino acids oxide. These are polar uncharged okay, cysteine, glycine, serine, tyrosine, glutamine look at that polar uncharged okay.

Because okay charger gets neutralized here, as you can see. Polar charged: aspartic acid okay, so you can see glutamic acid, lysine, arginine, histidine. So we have the polar charged, we have the polar uncharged, we have the hydrophobic protein. So the proteins are divided can be divided into these types of groups, so generally if you have protein 3 dimensional structure, the hydrophobic groups may be buried inside, so the polar uncharged or polar charged maybe sticking out.

(Refer Slide Time: 15:54)

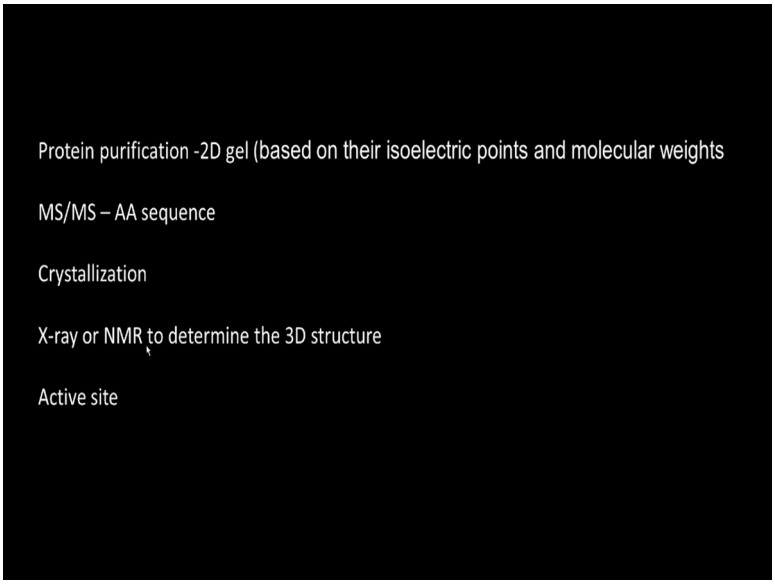


A Venn diagram showing the relationship of the 20 naturally occurring amino acids to a selection of physio-chemical properties thought to be important in the determination of protein structure.

So this interesting picture was taken from these 2 references okay, so the amino acids again this is very diagram okay showing how the proteins could be grouped, we can call it aliphatic protein, beta branched, aromatic, hydrophobic, polar, charged, positive, negative, showing the relationship of the 20 naturally occurring amino acids to a selection of physio-chemical properties.

These physio-chemical properties determine the protein structure. So we could have a polar, we could have a charged positive, polar charged negative, hydrophobic, aromatic, beta branched, aliphatic. So based on their amino acids, we can guess what type of amino acid is going to be, and generally whether it will be pointing outside or inside okay.

(Refer Slide Time: 16:48)



So how do we get the protein 3 dimensional structure about 15 to 20 years back there were no tools like proteomics, so protein structure was not known but with lot of development in the area of 2 dimensional gel electrophoresis, then mass spec tandem mass spec systems, x-ray crystallography or NMR, lot of proteins are being crystallized and their structures are being determined. But still there is lot of scope for doing research in protein crystallization.

Because it is not such a simple job. So if you look at a thousands and thousands of proteins that are involved in human diseases, we have only structures 3 dimensional structures proteins which are one tenth of that number okay, because it may be very difficult to purify to a high degree the protein of our interest, and then crystallize it and get the 3 dimensional structure okay.

So the first step you have in the protein purification is the 2D gel it is called that 2D, 2 dimension 1 dimension is the molecular weight difference separates the protein other dimension is the isoelectric point, you all know what is isoelectric point, this is the pH at which protein is neutral. So if you have pH gradient protein start moving, because of the charge and when it reaches the pH at which the charge is 0 it stops.

And then protein moves based on molecular weight and so the separation happens, so in 2D gel we will get thousands of proteins separated, we may pick up the protein of our interest, and then you put that into your mass spec tandem mass spec system mass spec mass (()) (18:43) and so on. So you get the molecular weight, we can even find out the sequence of amino acids that means how the amino acids are placed next to each other.

So we can get the primary structure of the protein by this approach which is quite simple, if you want to know the 3 dimensional structure of course we need to crystallize the protein okay, crystallization is not so simple. Once you have sufficient amount of crystal protein, you passing through x-ray or a nuclear magnetic resonance there is a lot of interest that is happening here, because this would be in the liquid form that means proteins could be in the solution okay.

Whereas this has to be crystallized, so this has more advantages. So we can go down to very high degree of accuracy, and we can get that 3 dimensional structure of the protein. Then you need to know the active site of the protein okay, so many times when the crystallized protein

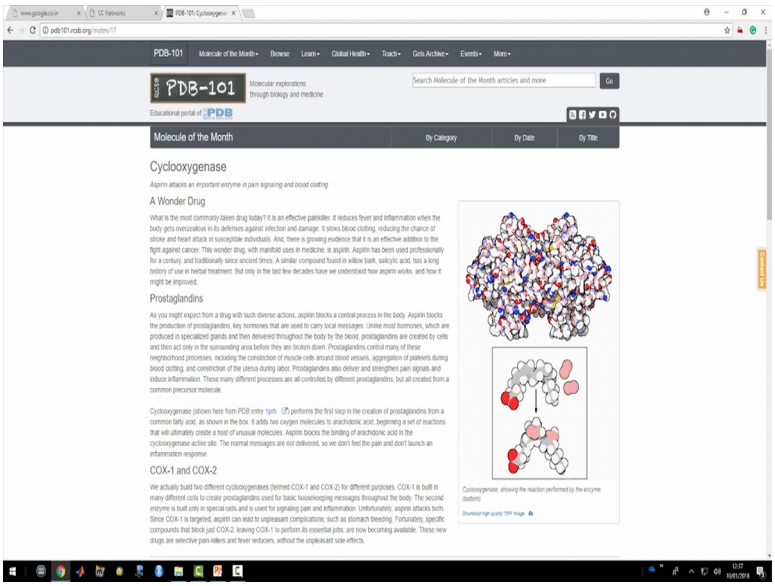
they put in in an inhibitor or the substrate and crystallized, so when you do the crystal structure determination, we get the structure of the protein as well as the ligand or substrate which is bound to the protein.

So that is a very good advantage to have a ligand, because such protein ligand complex is more stable than the protein alone, because the active site may collapse if there is no ligand or if there is no substrate bound to the protein active site. So it is a good idea to generally crystallize with the ligand or with the substrate bound to the protein. It is also good when we are doing docking studies or when we know where the substrate or the ligand has gone and bound, we can take that has the active site.

When we are doing docking we can remove that ligand and try to put a new molecule inside and see whether they bind better and so on actually. So these are the various steps by which protein 3 dimensional structure is determined, and as I said in the past 15 or 20 years with the improvement in the experimental tools more proteins are getting crystallized with substrates and ligand and drugs and the crystal structures are being determined okay.

So if you go to PDB protein databank, you will be able to see lot of proteins and you will be able to see the structures okay, and so it is very interesting to have a look at this okay.

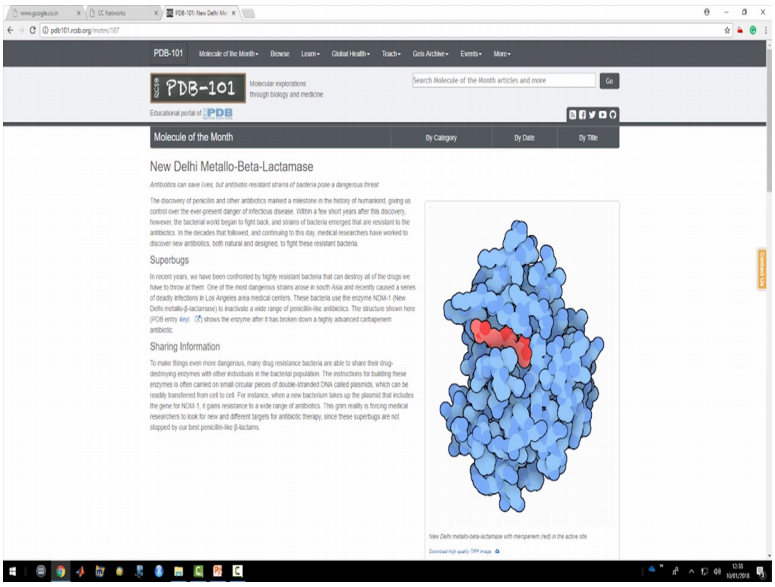
(Refer Slide Time: 21:55)



If we have internet here, we will be able to see, this is a cyclooxygenase 2 it gives you a lot of information in addition plus the references from which protein is downloaded okay. So you can get the 3D structure of this protein which we can download for our further studies, if you

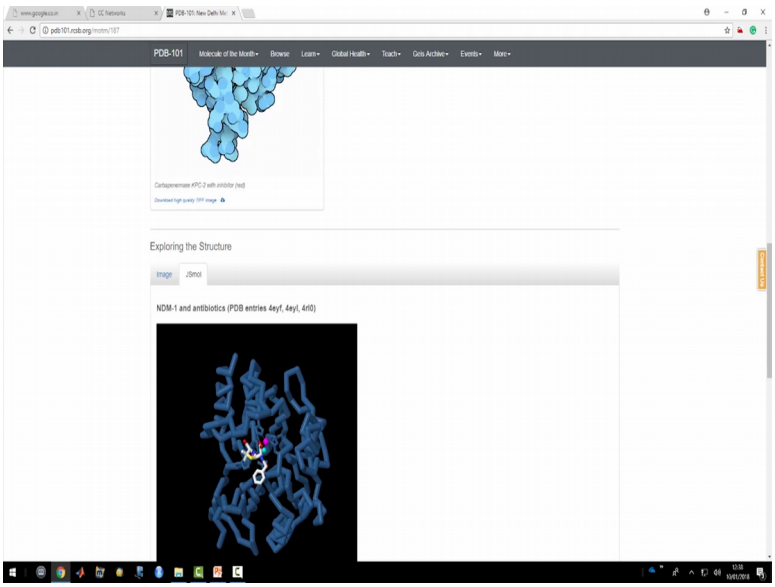
want to perform docking for example, then we need to have the PDB structure. Okay, these are some beta lactamase type of systems okay.

(Refer Slide Time: 23:25)



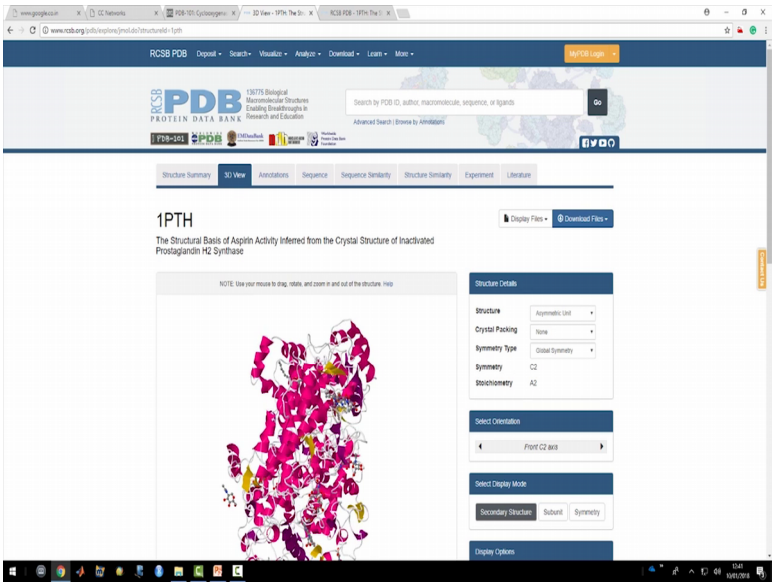
So as you can see they are called, so we can see beta-lactamase with meropenem in the red which is bound to that, so we can get the information and then we will also get structure we can even explore the structure, and there are a lot of references.

(Refer Slide Time: 23:55)



Here we can even rotate the protein as you can see here, so we can rotate this is a software JSmol, this is antibiotics okay, so a lot of information is given about this protein and what it interacts with, as I said there are 2 enzymes one is called cyclooxygenase 1 and other one is called cyclooxygenase 2 okay. This is the 1PTH, that is the prostaglandin H2 synthase or the cox as it is called okay.

(Refer Slide Time: 25:12)



You can see I think aspirin bound or what let us see what is it, this is the prostaglandin H2 synthase, these are the ligands that are bound to this particular okay ligand in the pocket ligands are in the pocket surrounding that. So a lot of information it can be got this is the space filling model, we can see the ligands or and yellow colour there. So this PDB is a very powerful protein databank or a protein database, it contains a large number of 3-dimensional structures of proteins.

We can if we know the area which we are working on, and if we know which proteins we want it, we can download it is all based on for example, this prostaglandin H2 synthase, this is called 1PTH. But you may have a different prostaglandin H2 synthase also found here, there could be depending upon how many researchers have got the crystal structures from maybe from urine or from other species and they might have crystallized with the other ligands okay other substrates.

So you may have many references to this prostaglandin H2 synthase, and depending upon your area of interest you may take up that 3 dimensional structure of that particular protein of your interest okay. So that is the beauty of this PDB, so it is a very useful and we are going to use that quite a lot as we go along, because the 3 dimensional crystal structure of the protein and active site of the protein are very, very essential for us to use in docking software.

And we are going to use docking software such as AutoDock or SwissDock, where we get the 3D structure of this protein of interest, and then look at the different drugs and start placing

them in the active site, and see how the docking takes place okay. So that way this PDB database is protein databank or protein database is very, very essential for us okay. So we will continue more in the later classes, on the concept of target based design or docking okay, thank you very much for your time.