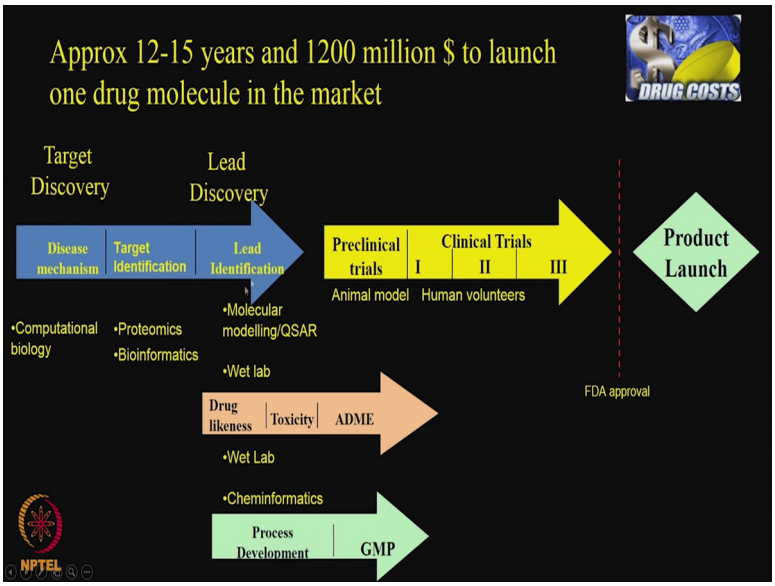
Computer Aided Drug Design
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology - Madras

Lecture - 03
Target and Lead Identification

Hello everyone, welcome to the course on Computer aided drug design. Today, we are going to talk about target identification and lead identification. So where does target and lead identification come in our overall process.

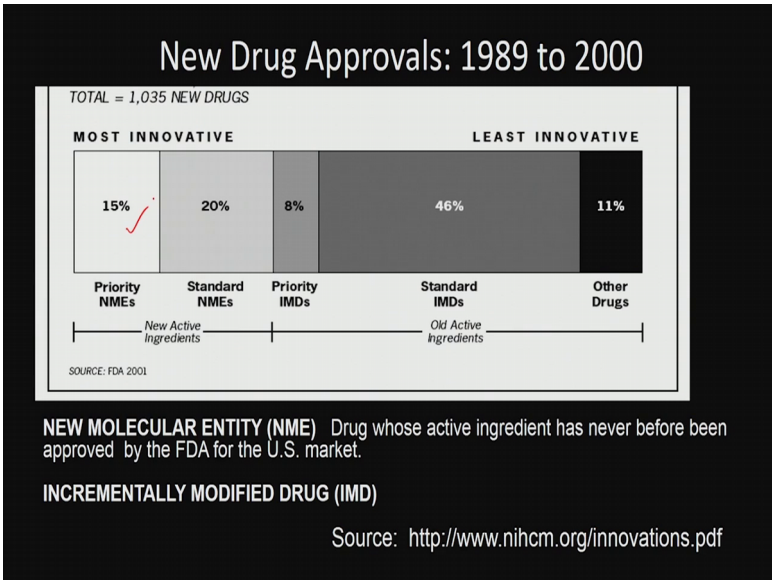
(Refer Slide Time: 00:33)



Yesterday, I showed you nice interesting slide okay this particular slide where we use a lot of in silico methods, where we have the target identification that is the target protein or enzyme which are looking at okay, the target protein enzyme and then trying to know the active site and so on. And then once you have got a target we go into the lead, lead is your molecule has to have certain activity that means it goes and bind to the enzyme and make it inactive.

It goes and binds to protein and denatures the protein so on actually, that is called lead identification. So we are going to talk about the various steps involved in the target identification and lead identification okay. And then later on we talk about the drug likeness property and so on.

(Refer Slide Time: 01:27)



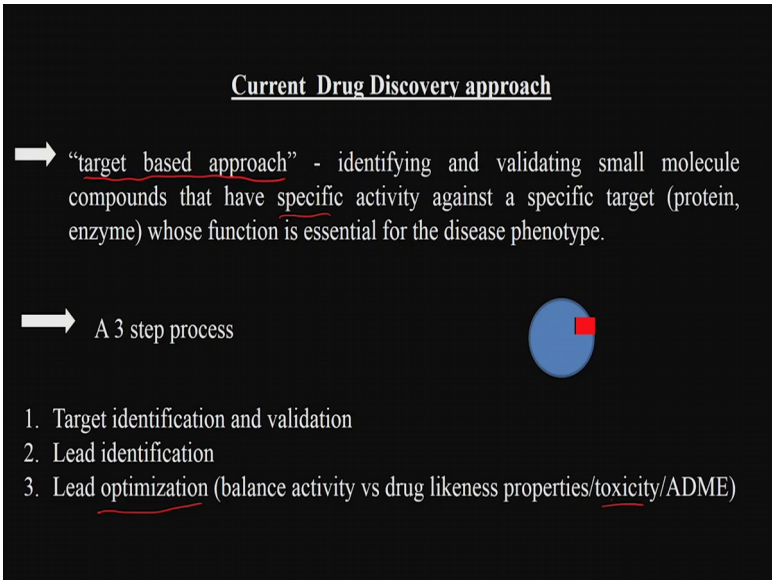
So if you look at some statistics 1989 to 2000 statistics, as you can see NME is new molecular entities that means totally new species, new compounds, which has gone for approval and only about 15% has gone for approval as a priority, and about 20% as standard new molecular entities for approval from the FDA okay. Then we have about 6% that is called IMD that is incremental modified drug, already a drug has been approved.

You make certain changes maybe you convert it from a liquid form to solid form, or change the crystallization procedure and so on okay, so that is called the incremental. So most of it as you can see here incremental, standard IMDs, priority IMDs, almost 60 70% is that, so you have an old active ingredient and you make some modification. So here there is not much invention, whereas these new molecular entities are inventive that means totally new compounds.

So about only 35% of the drugs that come into FDA approval or a new molecular species, whereas rest of them are incremental improvements only. So the traditionally how the drug discovery was followed, if you look at it what happened was researches in universities, Biotech companies or pharmaceutical companies, they started screening small molecules for disease. So there is a pharma company which is interested in colorectal cancer.

So they started screening lot of compounds, if there is academic institution they synthesized lot of molecules and then test it out for some inflammation, so that is how traditionally drug discovery was followed that is a very traditional approach.

(Refer Slide Time: 03:32)



Whereas the current approach is totally different they are all based on target based approach, that means we try to look at the target identify the target and then validate small molecules that has specific activity against the specific target protein. So that is very, very important specific the molecule which design should be very specific to the particular target protein or enzyme, it should be not binding to other targets or other protein enzymes.

Because then we will say it may cause side effects, so the specificity is low. So the most important thing is I need know the target, if I am looking at assay inflammation I am looking at an enzyme called cyclooxygenase 2, so I am designing molecules for cyclooxygenase 2, that is what it means actually okay. So it is very, very specific, so drug targets have to be known very well. So for example, you want the particular drug go and bind to the target and make the target inactive.

So there are 3 steps target identification and validation: so you need to identify there could be many enzymes and proteins in the pathway of inflammation but you are interested in one particular enzyme which you want inhibit okay. So you identify that there good be pharma companies working on different targets in the same inflammatory pathway that is also possible, and then you need to validate, you have to be very sure that is the target by doing different types of experimental studies as well as modeling studies we will talk about some of them.

And the new design lead compounds like I said compound does not become a drug unless it goes through all the clinical trials, preclinical as well as human volunteer trials. So the lead

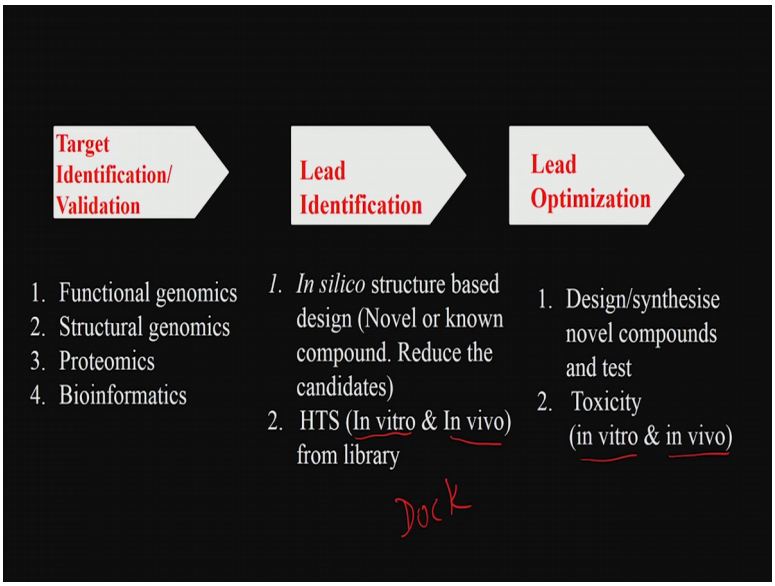
what you do is your test out thousands and thousands and thousands of compounds on that particular target and try to find one or two which seems to be very promising. And then you need to optimize the lead, that means what does that mean optimize the lead.

It is not that the only activity is important like I said that drug has to be soluble in the stomach, it should be stable at that pH, it should have good absorption through the GI into the bloodstream, it should have a good distribution throughout, it should not metabolize degrade inside because of the enzymes in the liver. So the active ingredient should be sufficiently of higher concentration to reach the target and do its job.

And finally it should get excreted from the body we should not stay in the body for a long time at the same time it should not cause toxicity, short term toxicity and long term. So when you do lead optimization, you need to balance the activity as well as the various drug likeness properties, so the very best compound from your lab might not have all these drug likeness properties maybe it is not very stable at pH of 2 that is your stomach pH.

Whereas the second best compound may satisfy all these properties, so you may try to select the second best compound and take it further into the preclinical and clinical trial. So this entire process is called lead optimization. So all these have to be done, so it is like a balance between activity versus other properties like drug likeness, ADME, toxicity, side effects, stability all these things. We are going to look at all of them in detail later.

(Refer Slide Time: 07:12)



So target identification and validation it could be broken down into several steps okay genomics, so I need to know what are the genes which get up regulated if I have a disease or the structural features of that, and then try to identify what is the protein that is involved. So I need to purify the protein, I need to get the structural details 3D structural details of the protein, I need to get the active site detail.

So I need to do all these that is what is target identification and validation means actually, so we first look at the gene that is involved in that particular place, we look at the proteins and do I have a 3 dimensional structure of the protein or can I get the3 dimensional structure using x-ray crystallography, do I know the active site I need to confirm or validate whether that is the active site, then all these are called target identification.

Once I have a target, and once I have validated that particular target then you need to identify a lead, so you may test millions and millions of compounds that is the in silico structural based design, so I take millions and millions of compounds and then try to docking okay, docking means we try to see whether the compound binds to the active site on the target, how it binds? It is binding very good, and you do all those things that is called in silico approach okay.

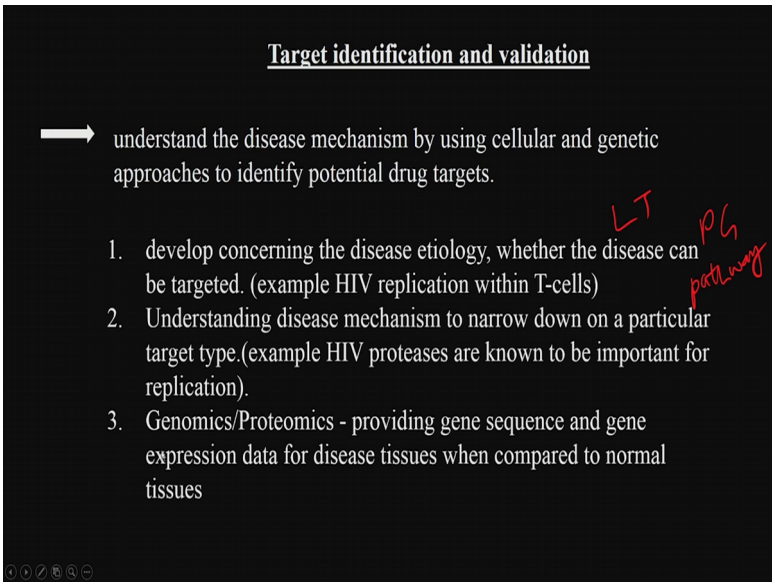
We can also do in vitro approach, I take a thousands of compounds and test out on that particular protein, and see which compound or compounds have very, very high inhibitory activity against the protein, that is called in vitro approach okay. Once I am satisfied with the one or two which has the drug likeness property, then I may go into the in vivo that is the animal studies. Of course I need to simultaneously do the lead optimization do not forget that.

Because I may find the best lead molecule does not have the drug likeness property, so I will make a modify the structure, and so that the drug likeness properties are improved, or maybe the toxicity is reduced, or maybe the side effects are reduced, maybe become more specific or solubility is increased, so that is all the lead optimization. Simultaneously, I will also have to know that toxicity I can do it on animals, or I can do it cell lines.

Different types of cell lines are available on which I can study this toxicity of the compound, so that is lead optimization, in fact the lead optimization and lead identification sort of get

merged as soon as you find one or two compounds identified you have to keep on optimizing around that, so that all the properties are also satisfied in addition to the activation okay.

(Refer Slide Time: 10:23)



Let us go more in detail, so if I want to know the target first I need to understand the disease mechanism okay. So I am saying I am interested in inflammation, how does the information progress? There is something called prostaglandin, there are something called leukotrienes, so there are many enzymes in the prostaglandin pathway, there are many enzymes in the leukotrienes pathway, so I am trying to inhibit one of those enzyme.

So I need to understand the disease mechanism by based on cellular genetic approach okay, so that I can decide on a particular target. There could be many targets, because if you look at the prostaglandin pathway in inflammation there are many enzymes, cyclooxygenase 2, prostaglandin each in ts1 and so on actually okay. Similarly, the leukotrienes you may have lipoxygenase, so there are many enzymes which we can target okay.

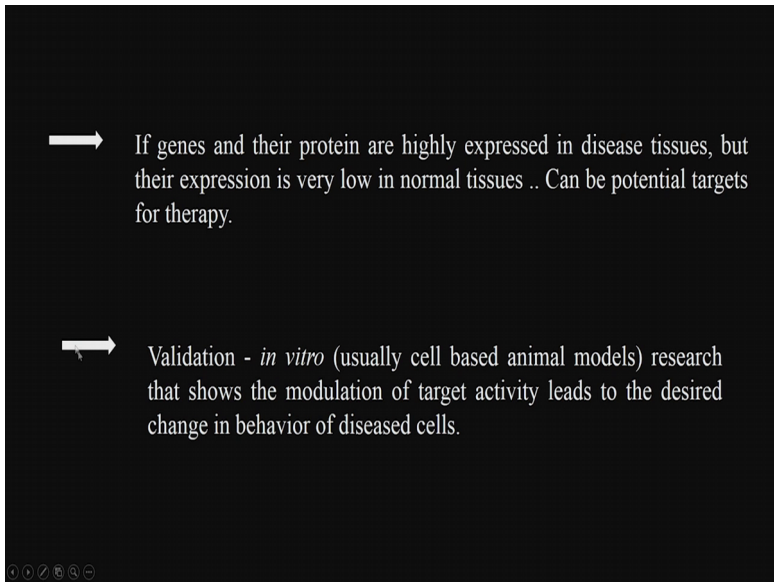
Then I need to know the develop concerning the disease etiology, where the disease can be targeted that is also very important, so there could be some places where the disease can be targeted. For example, HIV replication within T-cells cells so I could get there okay. So there could be some diseases you might not find the target okay, then you are in trouble. Understand the disease mechanism to narrow down in a particular target okay.

So once you understand the disease mechanism there could be many targets, you may have to focus on target. For example, HIV proteases are known to be important for replication, so

maybe this could be my target that is an example. And like I am talking about inflammation there could be enzymes in the prostaglandin pathway which lead to inflammation, so those could be the target. Then I need to now do genomics and proteomics.

I need to get the gene sequence and gene expression data for the disease tissue when compared to normal tissue, I need to know which genes get upregulated, which genes that downregulated when there is a disease condition vis-a-vis the normal condition okay.

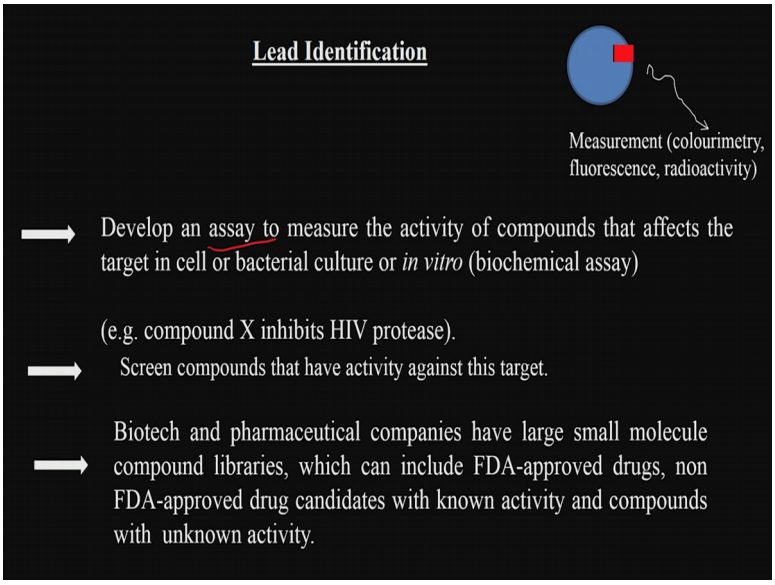
(Refer Slide Time: 12:44)



So if the genes and proteins are highly expressed in disease tissues, but their expression is very low in normal tissues, then that could be the target for therapy okay. So I could focus on those genes which get okay. Then I need to validate, so I need to validate means I need to do some in vitro studies experimental studies, so I can do some cell based study, I can use different types of cells cancer cells, if I am looking at cancer inflammation.

I could take the cell and create inflammation or induced inflammation okay, and then check them out okay, whether those things happened, whether the genes that upregulated or downregulated based on disease vs normal cells. So that is called the validation, so I have to be very sure based on in vitro studies.

(Refer Slide Time: 13:31)



Now look at the lead once assume, I have done the target, so what do I do? I need to design this lead okay, how do I do? There are many, many steps involved in that, so for example if your molecules go and binds to a protein or enzyme, there could be changes maybe in the colour fluorescence, radioactivity, so I could measure that and that could be a measure of the activity of this lead with respect to the control.

So I could screen large number of lead and make measurements accordingly okay, that is called lead. So I need to develop an assay that means an experimental procedure, assay is nothing but an experimental procedure to measure the activity of the compounds that affects the target, so when the compounds goes and binds to an active site of protein, or it goes and binds to an enzyme there is a change in calorimetry fluorescence radioactivity.

Or some change biochemical change which I could measure either using a bacteria or I can use animal cell in my lab okay that is the biochemical assay. I need to develop that assay if I want to do experiments on testing out different molecules to the inhibition of a particular target. For example, so I can say this particular compound inhibits this protein okay, I can say aspirin goes and inhibits the cyclooxygenase 2 enzyme and affects inflammatory pathway like that.

So but then the most important thing is I need to have an experimental assay which can be done in my lab, so that means I need to have that particular protein purified then I need to test a large number of compounds incubate each of the compound with that particular protein and

measure changes, some physicochemical changes, biochemical changes that is happening and then identify the activity of the compound with respect to the control.

Control means without any of the compound okay, so that is how I test out large number of lead molecules, these are called screening, screen compounds that have activity against this particular target. So if you take Biotech and pharmaceutical companies they will have large small molecule libraries, they will have FDA approved drugs, non FDA approved drugs candidates. So, so many candidates may be there in their libraries.

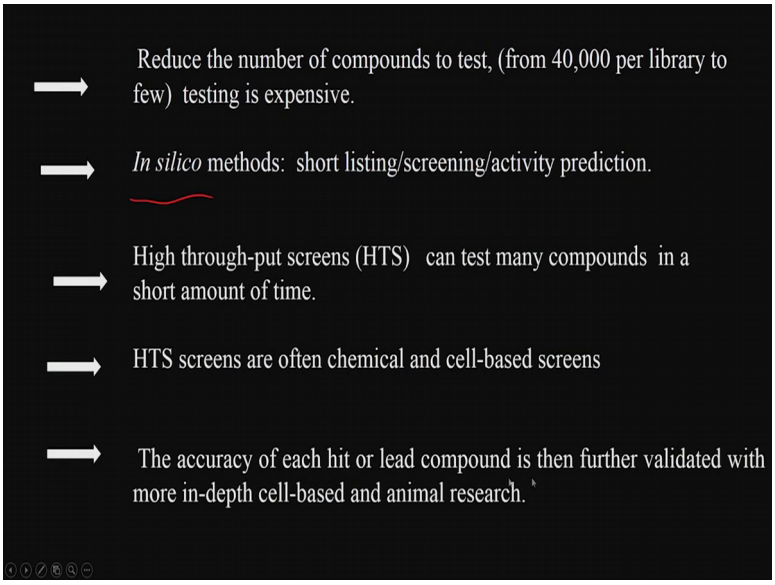
So the quickly start screening all of them in their lab, so that is an in-vitro type of approach, and see which ones look very, very promising that is the lead identification okay, which one gives the best sheet okay. So that is how you can do it in in vitro. You can also do in silico dock all this molecule to this particular target, and identify those which shows best binding to the target in the active site of the enzyme okay, that also can be done that is called in silico.

So I could combine in-silico and in-vitro, I take thousands and thousands of molecule structures and then I bind them or dock them to the active site of the target and see which ones bind very well, and take only those and then perform experiments in-vitro experiments in my lab to see whether the so good activity. So thereby I can reduce testing out millions of compound in the in-vitro studies which could be very expensive time consuming and so on.

I use computational in-silica method screen them dock them to the target, and select maybe hundreds of them which looks very promising, and then take those hundred take it to my wet lab and perform biochemical assays and identify a lead okay. So instead of a testing out millions of compounds in my wet lab which is time consuming expensive I could use this in-silico screening method, and then select only hundreds which looks very promising.

And then test it out in vitro using high throughput screening, that is call HTS high throughput screening, and then identify some lead molecules which seems to be very promising okay, that is called the lead identification.

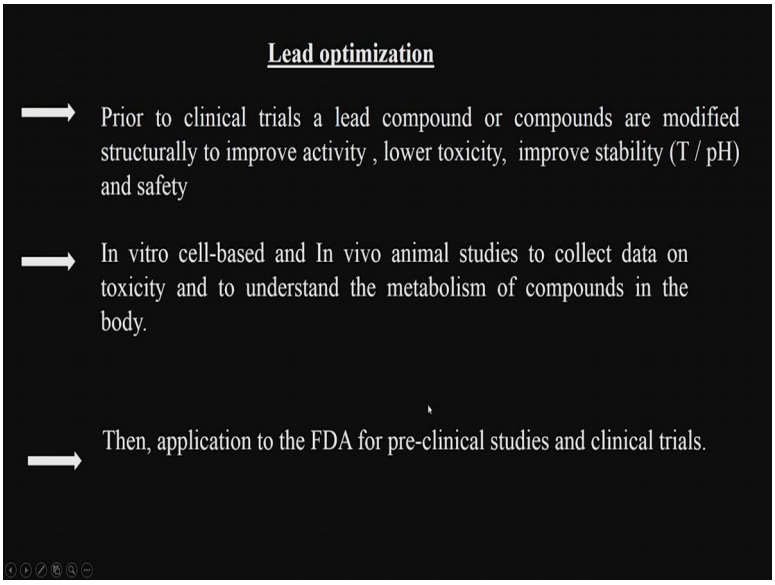
(Refer Slide Time: 18:26)



So the reduced the number of compounds to test from very large using computational in-silico methods okay, so that we can use this as a screening tool the docking of a large number of lead molecules to the active site and just picking up a few of them will look promising that is a in-silico screening that is called okay. Then we can use high throughput screens, that is HTS, so we can do large number of activity studies using that, it saves lot of time.

It could be get a biochemical based or a cell based screens okay, and then after that if there are some promising candidates we can do more experiments, take this and then we can do animal studies or we can do more assays. So initially you test out large number of compounds using high throughput screening, and then candidates which looks very promising you spend more time do in depth studies, biochemical in depth more biochemical assays or even validate with animals okay.

(Refer Slide Time: 19:37)

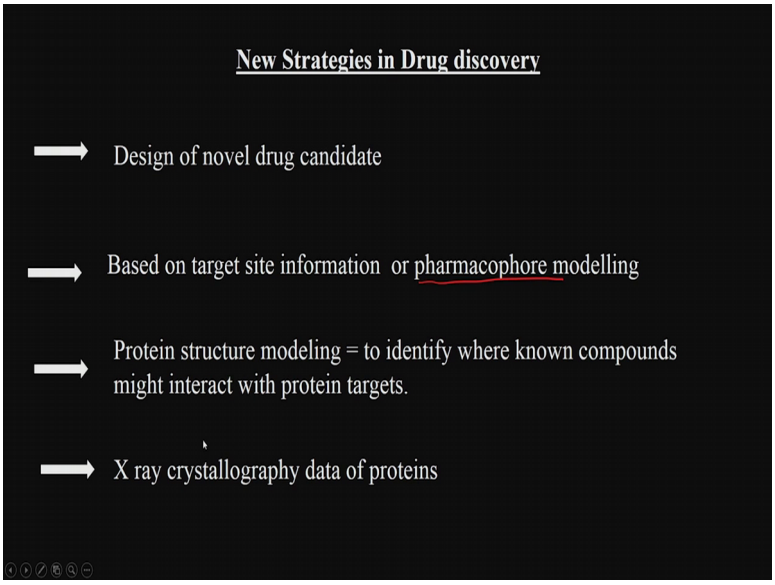


And then comes the lead optimization okay, so now we have looked at one or two candidates or maybe few candidates which look very promising, which has now good activity, but then as I have been telling it should also have lower toxicity, it should have good stability temperature pH safety. So that is where the lead optimization comes into the picture, so a compound which is very, very active may be very toxic that happens in cancer.

Because drugs for chemotherapy are not only very active in killing cancerous cells, but it can even kill normal cells okay, so it is very highly toxic. Whereas the second best compound that means compound which is not as active the first one, they have better stability or less toxicity or better pH stability and so on actually. So that is the lead optimization part which we do. So we can collect data from maybe cell based studies or maybe even animal studies.

Understand the metabolism of these compounds, whether the compound is stable at pH, whether the compound is stable in the liver region and so on, may be the best compound is very unstable, because of the action of the liver. Maybe the second best compound maybe more stable and so on actually, so we can either shortlist one or two which not only satisfies the activity, reasonably good activity but also quite stable okay, so that is called the lead optimization okay.

(Refer Slide Time: 21:29)

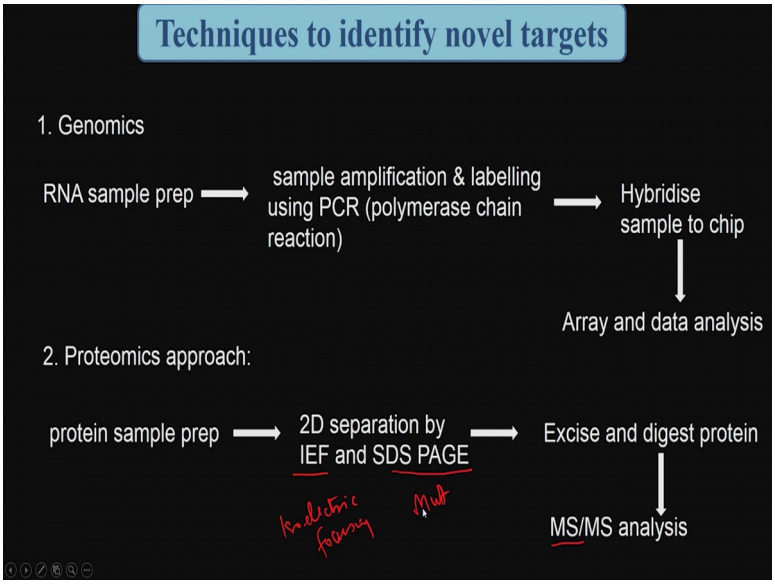


These information of course is very, very important before we go into preclinical, that means animal studies or before we go into human volunteers. So new strategies when compared to the older drug design strategies design of novel drug candidate based on the target site information, based on the structural features of the molecule, we are going to spend a lot of time on pharmacophore later.

This is called structural features maybe molecule has some special structural features like hydroxyl groups, or molecule has certain acidic groups or maybe it has got a heterocycle group okay which gives that activity. So we try to identify that. Of course the target protein we try to model that we also look at how compounds interact or bind to the protein, and how do we get the structure of protein? By x-ray crystallography.

So these are new strategies that have come into the drug design as against the traditional drug design in the past 15 to 20 years, a lot of this computational tools and new analytical tools have come into the drug design okay.

(Refer Slide Time: 22:49)

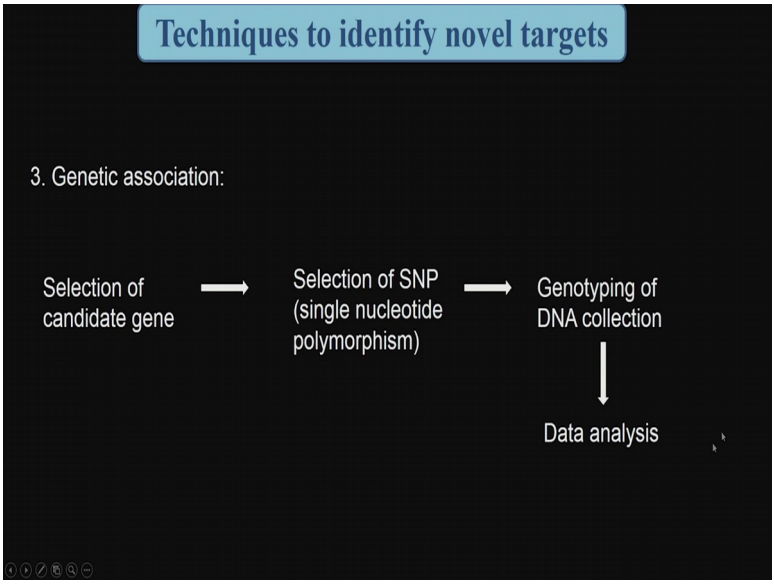


So let us go in detail on this identifying novel targets using genomic approach, using proteomic approach. So what do we do in genomics? We look at the RNA sample, we take an sample of the RNA, then we run a PCR using labelling, PCR you must have all studied polymerase chain reaction, and then look at the sample to chip, and then array and data analysis. Whereas in a proteomics approach, we are looking at the proteins.

So we take the protein sample, we separate out the protein mixture using a 2D 2 dimensional gel based on isoelectric focusing okay, that is the pH at which the protein charge is 0 that is the isoelectric point, and then the PAGE which is based on molecular weights, so the protein is separated the protein mixture is separated based on the isoelectric pH as well as based on the molecular weight okay.

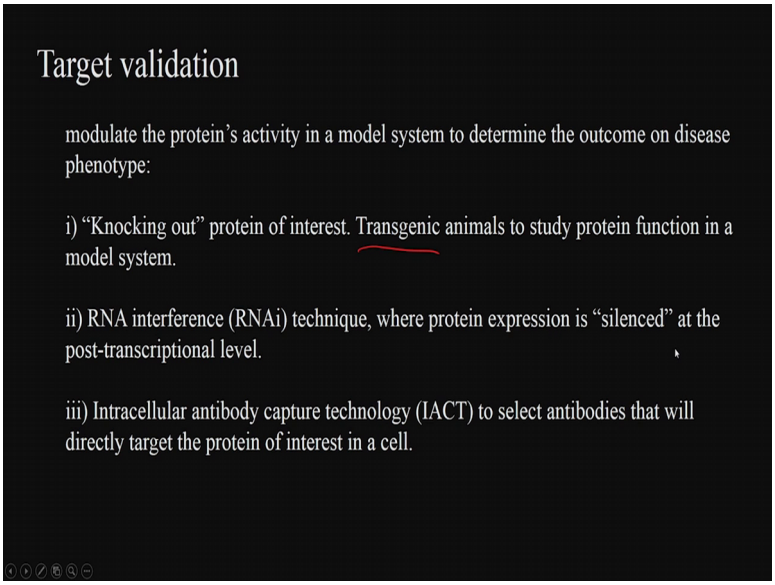
Then take the particular protein and then perform a mass spectrometer analysis, and then try to identify the protein molecular weight, protein sequence, that is called the proteomics approach okay. In the genomics approach you look at the gene, then go through a PCR make an array and then data analysis here, you isolate 1 protein from a protein mixture, and then try to get the protein sequence the amino acid sequence and the molecular weight using mass spectrometer analysis okay.

(Refer Slide Time: 24:29)



Then the genetic approach looks at the candidate gene, then look at a single nucleotide polymorphism, and then look at the genotyping of the DNA, and then you analysis the data. So there are different ways of identifying your target.

(Refer Slide Time: 24:50)



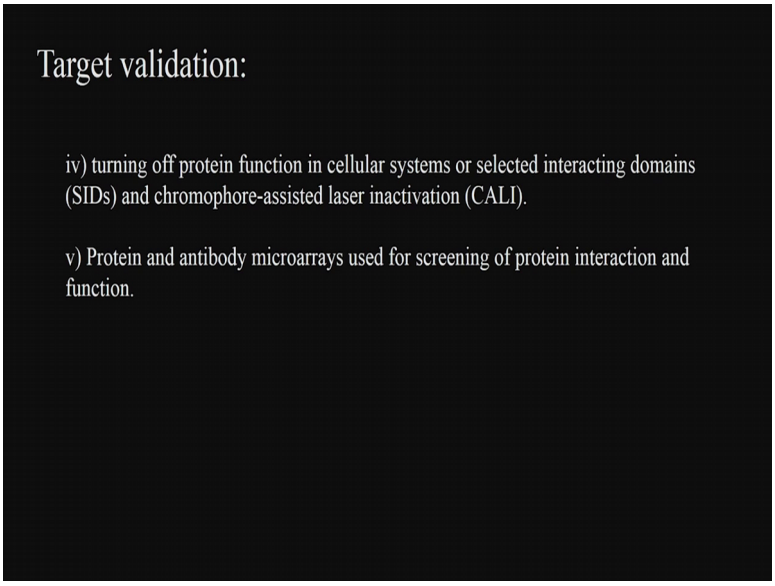
Once you have identified the target, we need to the target validation, so how do I do that? I need to modify that particular activity of the protein to be sure whether that is the protein which is involved in the disease phenotype okay. So what do we do? We knockout that particular protein, we have transgenic animals they are called transgenic animals, animals which have certain proteins of interest knocked out okay.

So for example, if I am interested in diabetes, I may have animals which have the insulin production knocked not okay, then I can be very sure that is the target. Or I can have RNA

interference technique, where the protein expression protein is after all expressed through the gene, so the protein expression is silenced at the post transcriptional level, that means the protein is not produced and then you study.

Because all these techniques are needed to be very sure that is the protein which is involved in particular disease. Another approach is called intracellular antibody capture technology, so you select antibodies that will directly target the protein, as you know antibody, antigen, antibodies are very, very selective to that particular target, and so it will go and bind to that protein of interest, so that protein is not available, and then see how the disease progresses so that is third approach.

(Refer Slide Time: 26:17)

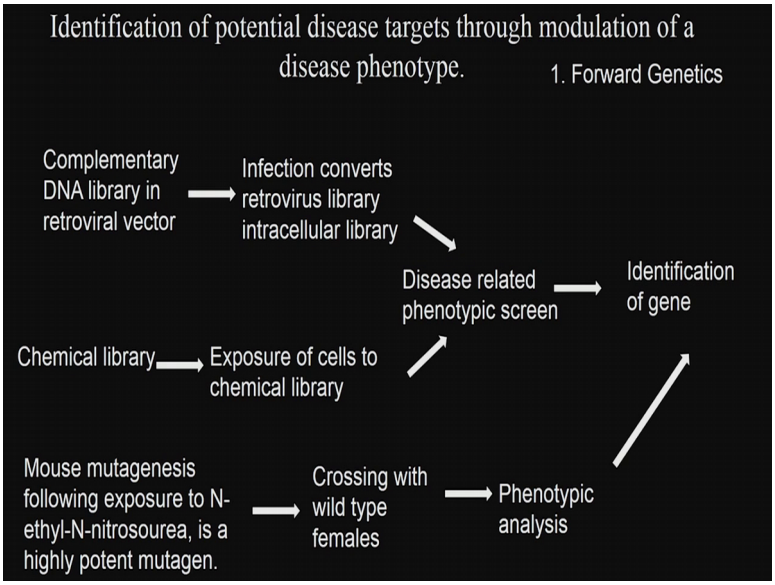


Then you have the 4th approach, so we can turn off protein function in the cellular system or in the domains by chromophore assisted laser inactivation, so we turn off the protein is produced but its function is turned off okay by this particular of approach. The 5th approach is protein and antibody microarray used for screening of proteins interaction and its function okay, this is 5th approach.

So large number of approaches are available for target validation, so first is target you have to decide on that target identification, but then you have to be very, very sure but that is your target of interest okay. So there are many steps knocking out the gene which stops the production of the protein, then the protein expression is silenced, and then protein it produces antibodies which will go and bind to that particular protein very specific, so that protein is not available.

You turn off that protein function by using external techniques, and then you look at a large number of microarrays so that you will get a protein antibody interaction, and that way you will validate your target. A large number of approaches are available.

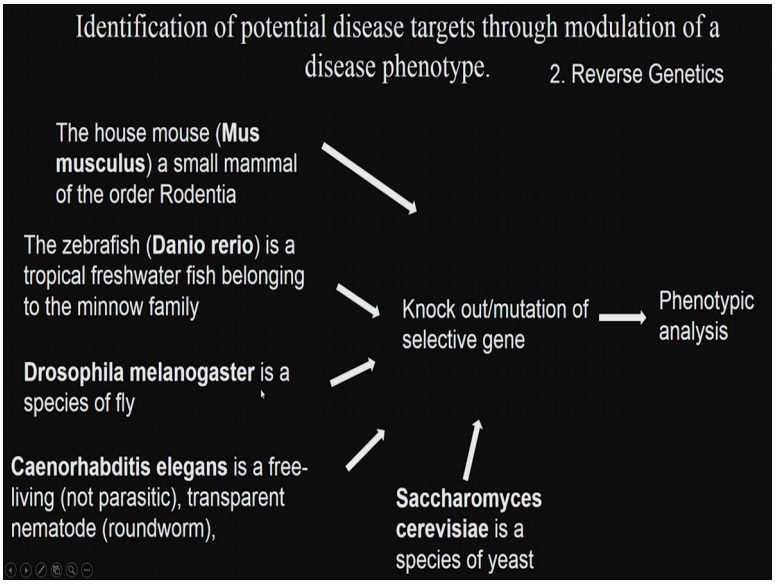
(Refer Slide Time: 27:43)



Another approach this is called the forward genetics and there is a reverse genetics by which we can also identify your disease phenotype okay. So what we do is we look at the complementary DNA library in a retroviral vector, infection converts this retroviral into intracellular library, so you get the disease related phenotype. Then we have a chemical library exposure of these cells to the chemical library, and then from there you identify the gene okay.

So we can simultaneously have the mouse mutagenesis, and then from the phenotypic analysis we can identify the genes, this is called the forward approach that means you create a particular infection and then study.

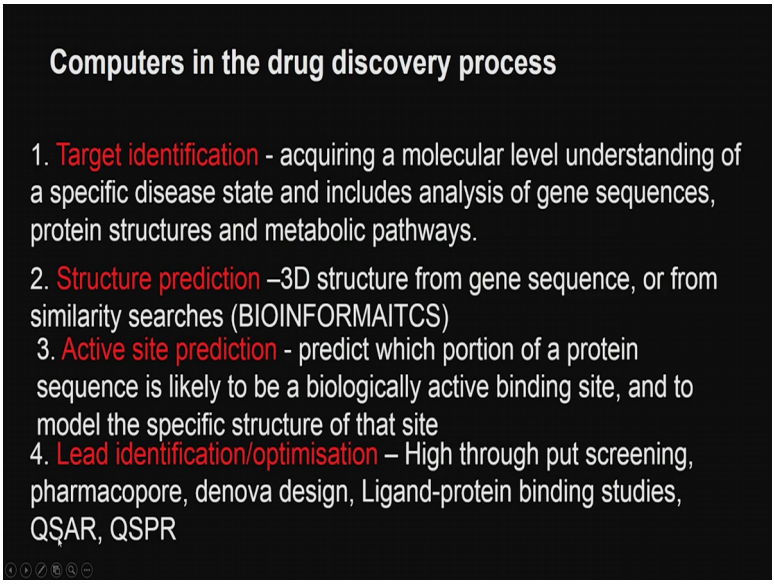
(Refer Slide Time: 28:33)



In the reverse approach okay, so you do the other way actually. So we have the mouse small mammal of the order of or a zebrafish or a drosophila or a sea elegans, saccharomyces or yeast and then we perform a knockout or mutation of a selective gene, and then look at the phenotypic analysis. So here we do not know which gene is needs to be knocked out to arrive at that particular phenotype, so you may have trial and error type approach.

This is the reverse approach, so there are different approaches one is called the forward approach genetic approach, the other is called reverse genetic approach for identifying potential disease target through the modulation of a disease phenotype. So this is also a long winding approach, as you can see you can test it out on small mammals, zebrafish, drosophila, sea elegans, yeast saccharomyces and get the phenotypic analysis done actually okay.

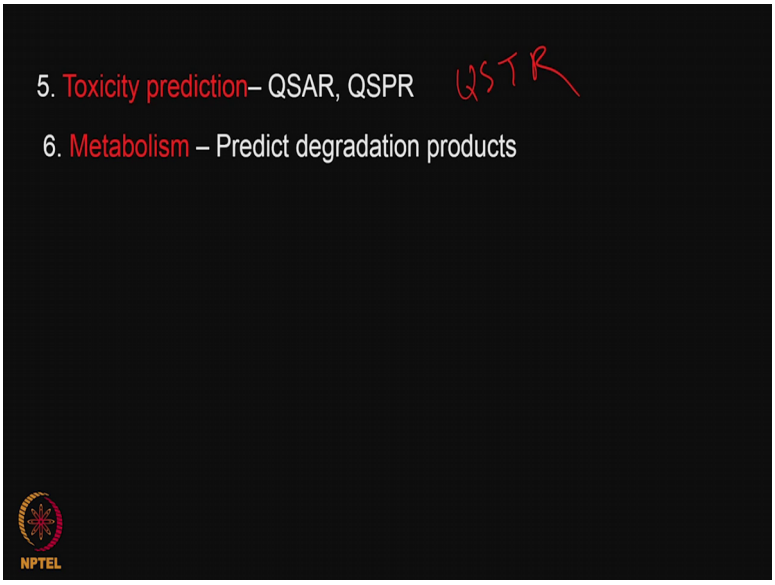
(Refer Slide Time: 29:36)



So computers can be widely used in the drug discovery as we are going to learn in the next set of lectures target identification: at the molecular level understanding of a specific disease state, then look at the gene sequence, protein structure, metabolic pathway. Then structure prediction we can predict the 3 dimensional structure of the target using different types of bioinformatics tool. Then we can look at the active site.

Because it is not only the 3 dimensional structure of the protein, I need to know the active site of the protein on into which my particular lead will go and bind and make the protein inactive, so I can use computational tools for active site prediction. Then lead identification and optimization using a pharmacophore model, denova design, ligand-protein binding studies, QSAR that is quantitative structure activity relation, quantitative structure property relation, all the lead identification optimization techniques.

(Refer Slide Time: 31:01)



Then toxicity prediction there are software's now which can predict toxicity based on QSPR, so they also have something called QSTR quantitative structure toxicity prediction. And then metabolism, can it predict the degradation of the drug into various products. So all these are possible using computational tools, so we can identify targets, we can predict the 3 dimensional structure of the target protein. We can identify or predict the active site into which my leads will go and bind to.

Then I can use computational tools for lead identification, lead optimization, then I can use it for toxicity prediction and then I can use it for metabolism that means I can predict the degradation products. Then I can predict its excretion from the body and so on actually. So many, many tools computational tools are available in the drug design process, and we are going to cover some of them as I said and in forthcoming classes okay, thank you very much for your time.