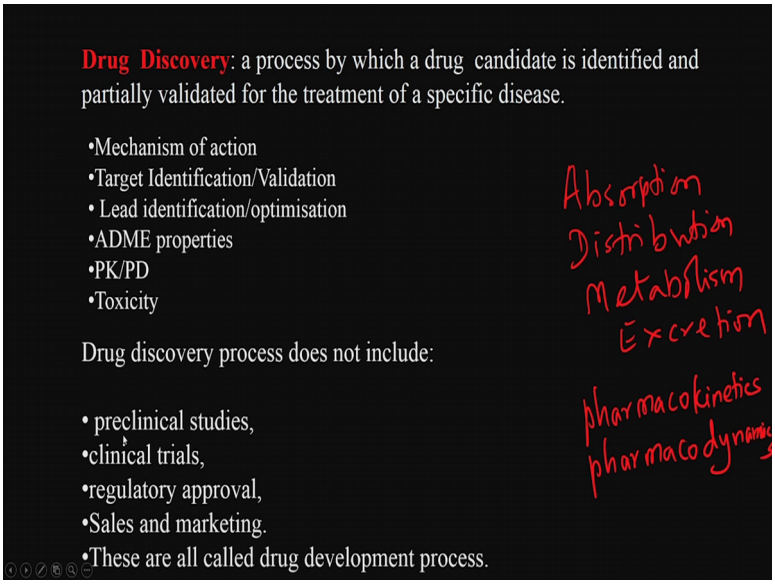
Computer Aided Drug Design
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology - Madras

Lecture - 01
Introduction

Hello everyone, this is a course on Computer aided drug design. I am going to take 40 lectures each lecture of half an hour duration. So we are going to talk about how to design drugs with the aid of computers, so computers will be an aiding tool, but you cannot replace a medicinal chemist or synthetic organic chemist or a physical chemist in helping in designing new drugs. So this particular talk is going to introduce this topic of Computer aided drug design.

(Refer Slide Time: 00:50)



So what is this drug discovery, so this is the process by which a drug candidate is identified, it is partially validated for the treatment of a specific disease. We are going to look at why partially validated because unless you actually do an animal trials, unless you actually do human volunteer trials validated compound does not become a drug which can be introduced in the market. So that is why we call it partially validated.

So you need to know the mechanism of action, you need to know how the drug acts, what are the enzymes on which it goes and binds to, or a protein there by inactivating the particular pathway of the disease, so we need to understand the mechanism of action. So we need to

identify that particular target, the enzyme or protein, and we need to validate the target that means we need to use tools to be sure that that is the target on which my drug is acting.

We need to identify a candidate generally before the compound is approved by FDA and introduced in the market, it is not called the drug it maybe it is called a lead compound that means a lead, a compound which has shown very good activity, a compound which looks very, promising, so that is called the lead identification okay. And then we need to optimize the lead because just because the compound acts very well in my lab does not mean it has all the properties of a drug.

As you know drug is something that is taken inside the human body, so it should not be toxic, it should not cause side effects, it should have a good absorption in the stomach and so on. So they are called drug likeness property, so we may have to modify those properties, so that it not only has good activity but also has good drug likeness property. So we are going to introduce this term drug likeness property for a many, many lecture okay.

So it should have these drug likeness property is called ADME, that means A refers to absorption, D refers to distribution, M refers to metabolism and E refers to excretion. Let me write down absorption, distribution, metabolism and excretion, that means the drug has to get absorbed into my stomach okay, mostly if it is an oral drug it has to be taken it goes to GI track that means gastrointestinal tract gets absorbed.

Then it gets distributed inside your bloodstream, it gets metabolized because there are many enzymes your liver is a gatekeeper which tries to keep on degrading any foreign compounds coming into it, and then finally it has to get excreted that makes thrown out of your body. So the drug which is taken in has to have good ADME properties that means absorption, distribution, metabolism and excretion properties.

So it may be a very good active compound but maybe gets metabolized and becomes inactive inside your stomach, so then it is of no use okay, so ADME is also very important when you are designing drugs. Then it should have good PK and PD, what is this PK and PD? PK is nothing but pharmacokinetics and PD is nothing but pharmacodynamics. That means how the drug is excreted as a function of time?

How the drug is absorbed as the function of time into the plasma? That is all given by pharmacokinetics. And how the drug which is inside the body acts on the disease that is called pharmacodynamics. We are going to talk about each one of them much more in detail, and it should not have toxicity, so we need to know the toxicity of the particular compound which we are testing that is the lead we are testing, it should be non-toxic short-term toxicity as well as long term toxicity.

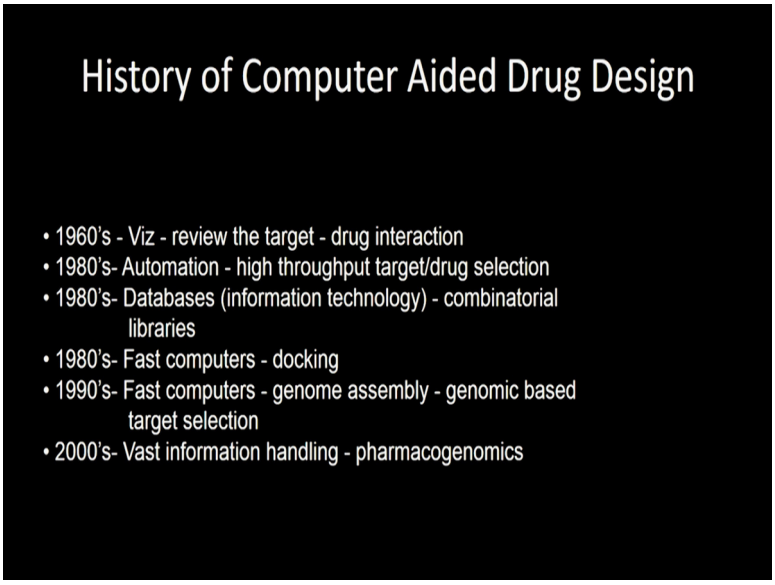
For example, if somebody has an arthritis problem they may have to take the anti-arthritis drugs for a very, very long time, so it should not have long term side effects or toxicity okay. So all these properties need to be looked at when we are designer drugs, it is not only the activity in my lab, but I need to understand what is the mechanism of action, which target it goes and binds to, and prevents the disease progression.

It should have good properties, drug likeness properties like absorption, distribution, metabolism and excretion, it should have good pharmacokinetics and pharmacodynamics, it should not have any toxicity and so on. So a drug discovery process does not include many steps: preclinical studies, so I have a nice lead in my lab, so I do not introduce that immediately into the market, they need to be something called a preclinical studies.

Normally, that is done in with the help of animals maybe rats, rabbits, dogs or monkeys and so on that is called preclinical studies. Then comes clinical trials where we use a human volunteers, then comes the regulatory approval you need to get approval from the regulating body like food and drug administration authority, and then it goes into sales and marketing. So these are entire drug development process actually.

But we are not going to talk about all these, our focus is more on these rather than these okay. So it is much more preclinical studies we are going to stop at.

(Refer Slide Time: 07:13)



So let us look at the history of computer aided drug design, it started in 1960s okay. So everybody started looking at the target okay, how the drug which target it went and bound to before those 60s yet drug was given to the patient, and there was a reduction in the pain or reduction in the fever or reduction in the infection, and that is it. Nobody, bothered about how the drug reacted, which targets it went and found to okay.

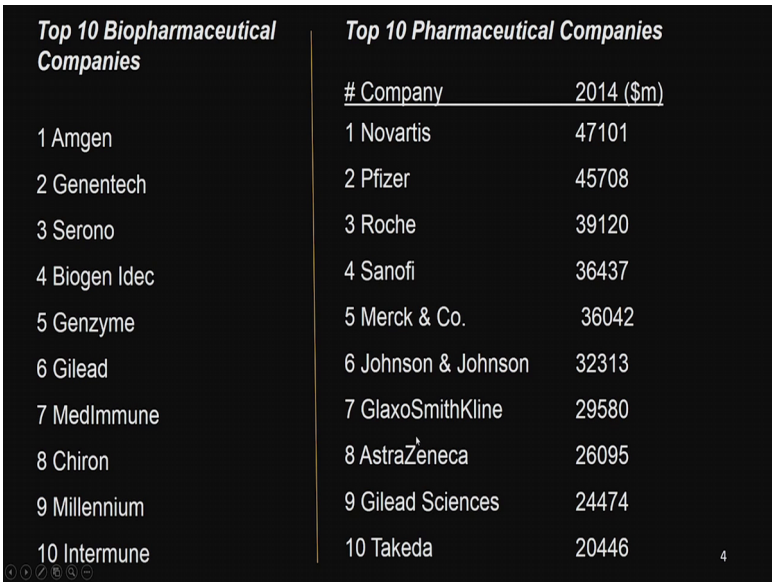
Then came the 80s where they started using high throughput screening by which many, many candidates where tested for certain targets okay in an automated way 10000-20000 compounds were tested in one shot by pharmaceutical companies okay, that is in the 80s because it is not possible to take one compound and keep testing one after another which may be a very time consuming process.

Then in the 80s okay again database is started coming, lot of databases are created combinatorial databases, database on activity of large number of compounds, activity of compounds on different targets, they are called combinatorial libraries. Then again fast computing started coming into computing power increase, parallel processes came, so scientists started using different parallel processing machines for performing ligand protein docking, drug protein docking, drug target docking okay.

Then again 90s genome assembly genomic based target selection, that means we try to understand what are the genomes involved in the disease processes okay. Then in the 2000 we went into pharmacogenomics, that means combining genomics with pharmacological, so

which genes affected, which pharmacological is used. So this I would say approximately the history of a Computer aided drug discovery.

(Refer Slide Time: 09:25)



Now let us look at some names of companies which are involved in drug discovery, and of course this is more as a overall knowledge or picture. The top 10 pharmaceutical companies you might have heard about it, this is 2014 data: Novartis, Pfizer, Roche, Sanofi, Merck, Johnson and Johnson, GlaxoSmithKline, AstraZeneca, Gilead, Takeda okay. And look at their annual sales in million dollars okay, this is almost 47 billion dollars 45 billion dollars.

So if you add up all these you are talking in terms of 200 to 300 or even 500 billion dollars okay that is a big business. So a drug discovery drug business is very, very large, these are some bio-pharmaceutical companies. Now many drugs are being manufactured using biological route rather than chemical route, and so many bio-pharmaceutical companies have come selectively using biological routes.

Whereas now these companies which were making dugs through chemical routes have also started making some products using biological routes, some of these companies are Amgen, Genentech, Serono, Biogen, Genzyme and so on actually. So a drug discovery if you look at it the global scenario, the market is very large US is the biggest market followed by Europe, Japan and so on actually.

(Refer Slide Time: 10:59)

Let us look at Indian companies they are much smaller when compared to for example Novartis or Pfizer maybe 100 size, but still they play a very important role especially in generic, Ranbaxy Dr Reddy's lab, Torrent Cadila, Cadila Healthcare, Cadila Pharma, Lupin, Sun Pharma, Nicholas Piramal, Dabur. So all these are Indian companies, and as I said India plays a very important role in generics.

That means drugs which have come out of the patent regime which anybody can make Indian industries are extremely competent in making compounds in a much cheaper way, so they become the products become very competitive and they are able to export to US, Europe, third world countries and so on actually. So Indian companies are not into new drug discovery, but they are into the generics and over the counter drugs, and they are really completing globally.

And their products are approved by FDA that means food and drug administration authority and so on actually okay.

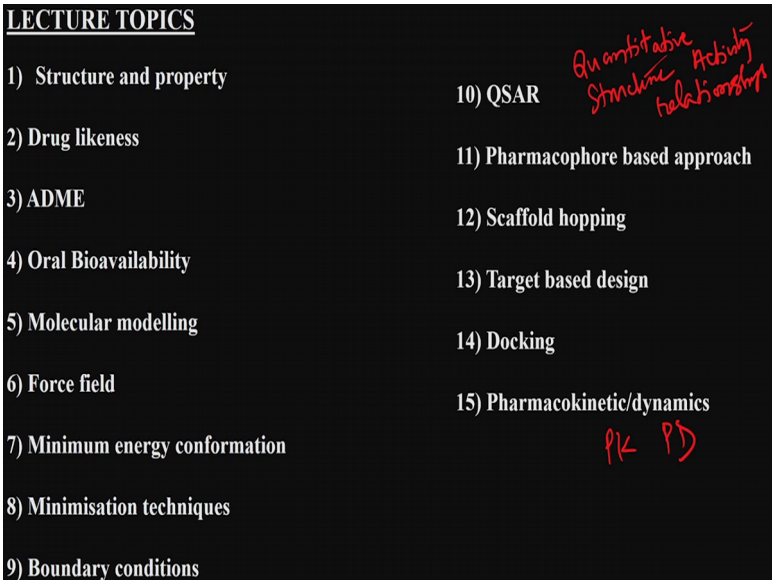
(Refer Slide Time: 12:20)

So let us look at the top 10 selling drugs year 2013 okay, just gives you an idea of what does drugs currently which are very, very important okay. So you look that these the top selling drugs in 2013 is an anti-inflammatory drug for rheumatoid arthritis, 14.8 billion sales okay that is a big money. Then this is also here something to do with arthritis autoimmune manufactured by these companies that is also very big.

Then comes another third compound which is also for rheumatoid arthritis, then we have a drug for asthma, diabetes, cancer, so several drugs for cancer, and then cholesterol lowering drugs and so on. So all these top selling drugs as of 2013, they are all in the range of 14 billion going right up to 5.5 billion sales. And the other drugs which are marked in red okay are manufactured using chemical routes, whereas other drugs are manufactured by biological routes.

So what it means is slowly the pharma companies are moving towards making drugs using by a bio-process approach rather than chemical approach, when you say bio-process we are using either an animal cell or a plant cell or a bacteria or virus or yeast cell and so on to make the product okay. Only 3 compounds are fully manufactured using a chemical route, I think as time progress all these molecules maybe manufactured using biological route okay.

(Refer Slide Time: 14:15)



So the lecture which I am going to give in the next 40 lectures will cover topics like structure and property, so what are the structural features that a drug should have which will lead to certain property okay. For example, property could be solubility, property could be is lipophilicity and so on actually, even at properties like toxicity degradation, stability at different pH condition, they are all called properties.

And then drug likeness does it have the drug likeness property, that means that is because I design a molecule and it shows very good activity does not mean it can be taken in by human, because like I said it should be less toxic it, should be not have side effects, it should be coming out of the body after it is done its function, it should be stable at various pH conditions, so these are all called drug likeness property, so we are going to talk about that.

Then of course absorption how is it getting absorbed into the body, distributed, metabolized excreted. And then oral bioavailability, so we take the drug for orally, so that means it goes into the stomach where the stomach is highly acidic we are talking in terms of pH=2, and then it gets absorbed into the stomach okay, it goes to the blood and then from the blood it gets circulated into the body, and the liver which tries to degrade whatever a foreign objects there.

So it has to overcome that and finally has to reach the target site, so that is called oral bioavailability. So just because I take a 10 milligram tablet does not be at the target site the concentration of the drug will be 10 milligrams, because it might not be absorbed properly in

my stomach or it may get degraded in the acidity in my stomach, or it may get degraded by the liver enzymes.

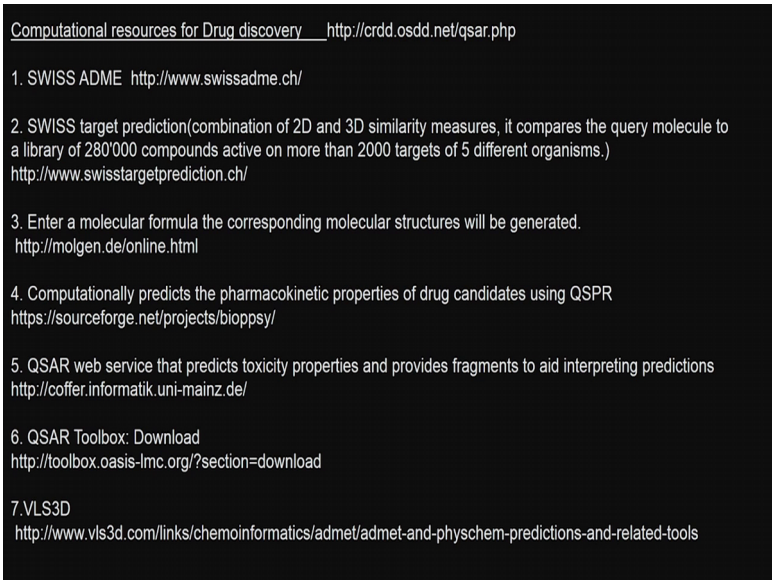
And so it may whatever concentration that is reaching my target site could be very, very small okay, so that is called oral bioavailability, so we are going to talk about that. Then we come to force fields, what are the different force fields we use? Well, molecular modelling okay, and then how do we calculate minimum energy confirmation? How do we minimize the structures of molecule? What are the boundary conditions we use? Okay.

QSAR that is quantitative structure activity relationship, then we are going to look at the structural features of the drug that is called pharmacophore, then we are going to how to modify the structure that is called scaffold hopping, and then we look at how targets are been identified, and how do we identify how the drugs goes and binds to a particular target. So we are going to look at docking, and then we are going to look at pharmacokinetics and pharmacodynamics that PK and PD okay whatever I said.

So in the next 40 lectures, we are going to look at all these topics, we are also going to look at a certain software's which are freely available, which we can use for performing all these operations. There are many, many software's which are freely available, so we do not have to buy any commercial software, I am not going to talk about any commercial software, but all the freely available software's we can use to perform several of these operations okay.

There are very good freely available software's in the web, we are going to look at all of them okay.

(Refer Slide Time: 18:16)

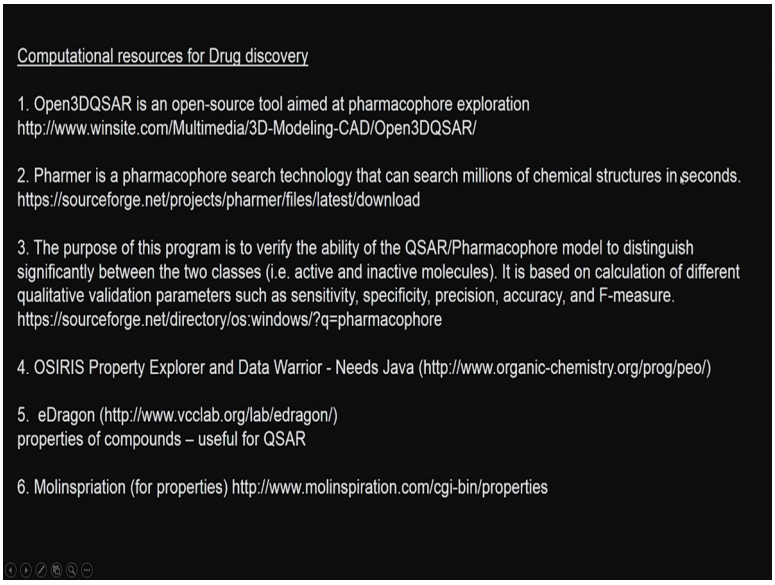


So what are the computational resources for this okay, there is q something called SWISS ADME I given this site address also, so you people can have a look at it. One is called SWISS ADME which can give when we draw a structure it can give properties of molecules; it can give what is the oral bioavailability. Then we have this SWISS target prediction, it can predict what are the possible targets if I give a structure okay, that is also very useful actually.

If I give this one, if I give a molecular formula it gives you a molecular structure, this software predicts the pharmacokinetics properties of a drug candidate using QSPR okay. Then this software predicts toxicity of a molecules, and also toxicity of the fragments of the molecules. Because when a drug goes inside the body as I said the enzymes in the liver breaks the drug into small fragments, so the fragments also could be toxic.

This software tries to predict whether the drug itself has toxicities, properties or the fragments that are created generated by the degradation of the drug also has. Then we have QSAR tool box, then we have a chemoinformatics, admet, physical chemical predictions, this this software called VLS3D okay.

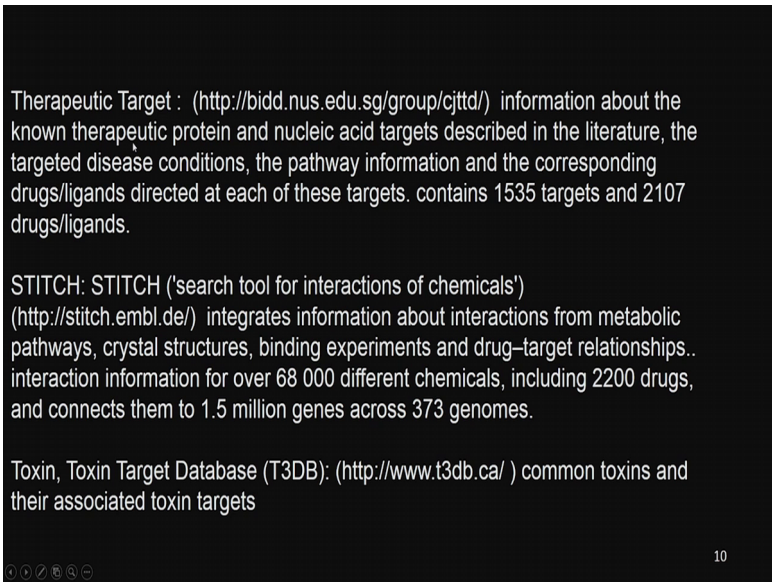
(Refer Slide Time: 19:48)



And then there is something called 3DQSAR 3 dimensional QSAR, there is a software for this software gives you the pharmacophore, details that means it can identify the features of the molecules and try to search millions of chemical structures okay, which have the same pharmacophore. This another software gives you the QSAR pharmacophore model. Then we have a property prediction OSIRIS.

Then we have a another property prediction that is called a eDragon, and there is another software for property protection that is called Molinspirational. So you see a lot of softwares available okay, and during the course I am also going to use many of these softwares, and I am going to demo to you people how to use some of them based on my experience and so on actually okay.

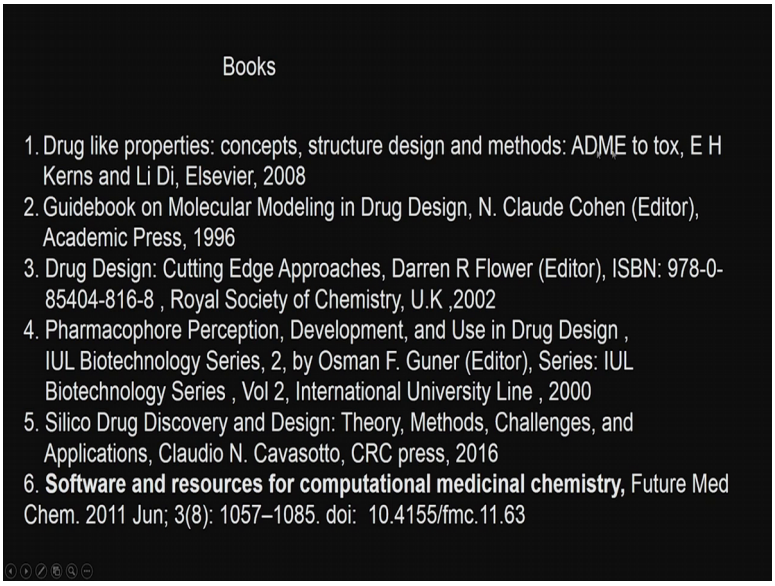
(Refer Slide Time: 20:50)



Then there is another software is called the therapeutic target, so it has data information about the therapeutic protein nucleic acid targets described in the literature including the targeted disease conditions, the pathway information and so on. Then there is another software called STITCH, so it connects the chemicals or the drug with the target, and genes so it creates a very complicated map okay it stitches basically between the genes the target the drugs, and creates a very complicated map connection.

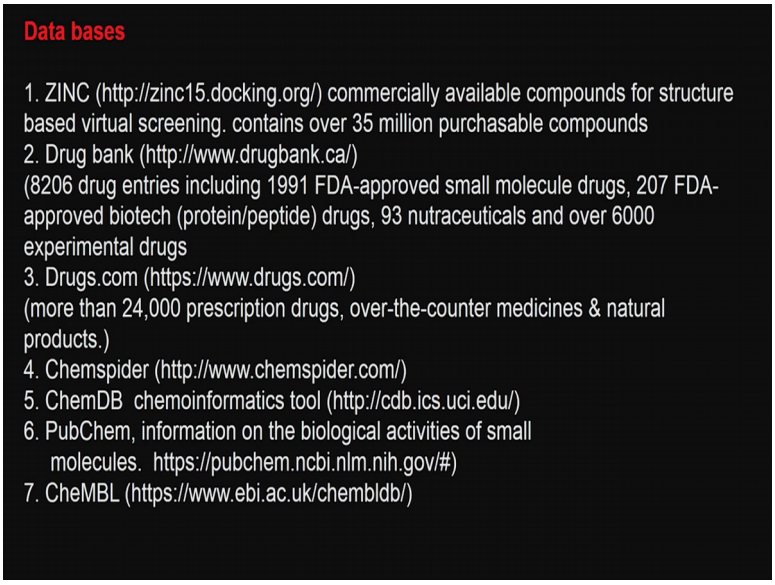
So it can connect almost 1.5 million genes. Then there is a T3DB toxin, toxin target database, so the common toxins and their associated toxin targets we can identify using this database okay.

(Refer Slide Time: 21:45)



So a lot of software's, and I am sure you guys maybe finding your own software as you keep searching the web. Books, there are some good books are there; Drug-like properties: concept structure design methods ADME to tox okay this is a very good book okay. And then you have guidebook on molecular modelling in drug design, Drug design, Pharmacophore perception, Silico drug design, Software and resources computation. So very good books are there you guys can have a look at them.

(Refer Slide Time: 22:19)

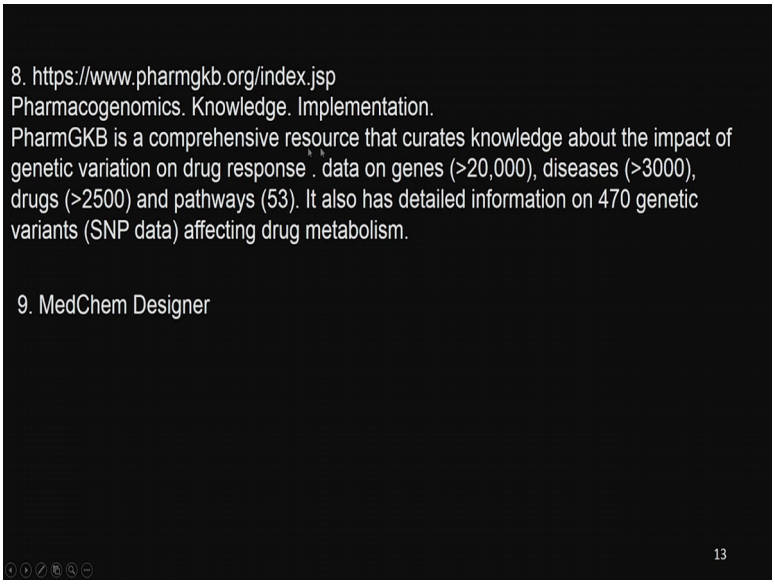


Databases, we have this ZINC database, ZINC has 35 million purchasable compound structures okay, so if I going to look at those structures those properties and I find one molecules looks very promising I can even buy from them okay for a cost. So this is a very useful database. Then we have the drug bank database okay, there are 8200 entries in that, that means drugs which are approved by the food and drug administration authority okay, that is FDA approved drugs okay.

Then Biotech drugs, nutraceuticals experimental drugs. So if I want to know is there are any drugs for a particular disease I would look at this particular database or another database called drugs.com, these got 24000 prescription drugs over the counter medicines and natural products. So first step if I want to know whether there is any drug available for a disease is that I will go through these 2 databases okay.

Then you have Chemspider, ChemDB, chemoinformatics tool, PubChem, ChemMBL, all these databases give structures of molecules, structures of drugs, structures of prescription drugs or structures of nutraceuticals okay, and some properties like solubility and lipophilicity, molecular weight, number of rotatable bonds, number of hydrogen bonds and so on actually okay.

(Refer Slide Time: 24:01)



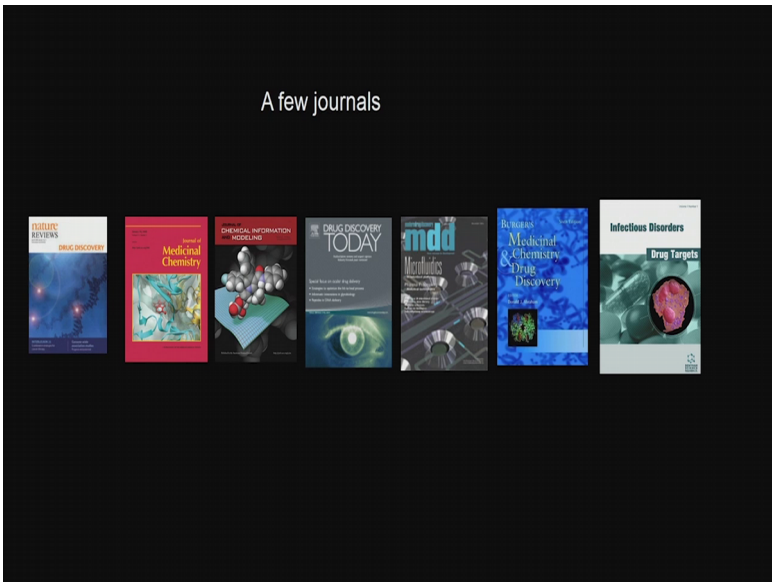
So these are I would call database is very useful, for if you want to look at structures more of it pharmacogenomics knowledge implementation, this another database is called PharmGKB, this resource that curates knowledge about the impact of genetic variation on drug response okay. So if there is a genetic variation because of different races or different communities, then how the drug will respond, so there is a lot of information okay, data on genes, diseases, drugs, pathway.

So we can look at this database also if you are interested. Then of course MedChem designer, so large number of databases structures of billions of molecules are available as a resource okay, more of this databases of interest for drug discovery okay.

(Refer Slide Time: 24:41)

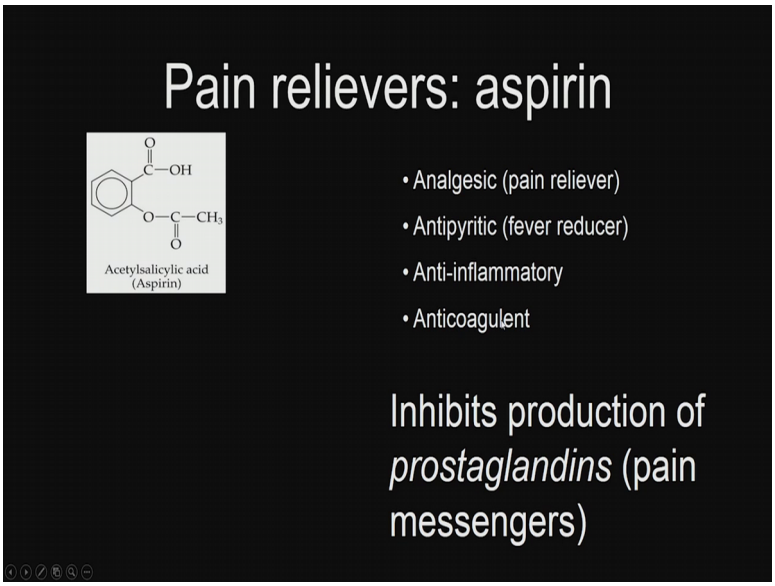
So we have of course as you can see some of them are commercial products okay, some are publicly available, freely accessible for academia, commercial versions also available, so we can have large number of databases on molecules of different structures okay, of course some of them are commercial also. That means if you make a payment you will be able to get databases for you which can be used for in silico analysis.

(Refer Slide Time: 25:18)



Journals, these are some few journals that are talking about Computer aided drug discovery, Medicinal chemistry, Drug discovery today, some of them deal with okay reviews and some of them deal with experimental studies and so on actually okay. So large number of resources I have introduced it to you, you can play around look at some of these structural databases as well when we have time okay, and see how useful they can be okay.

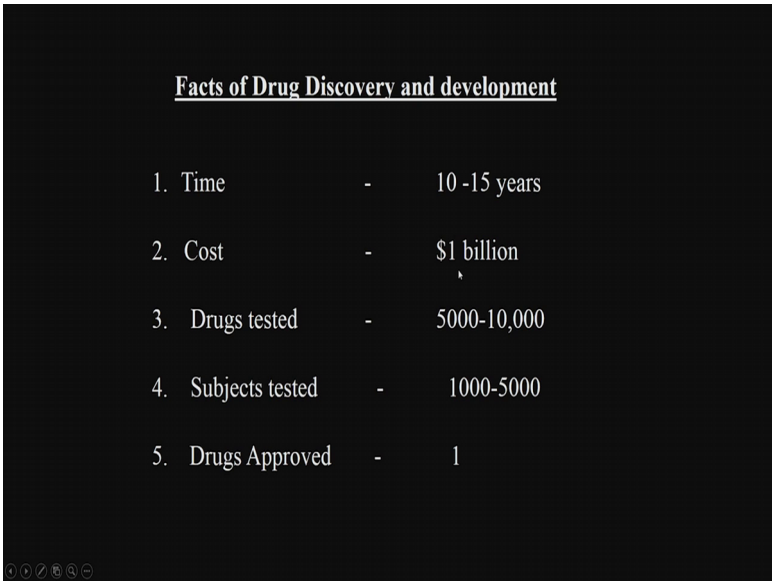
(Refer Slide Time: 25:54)



Let us look at some structures, aspirin he is a molecule we all must be knowing for a very long time, and we all would have taken this particular drug, it is an analgesic, it is a pain reliever, it is an antipyretic that means it reduces fever, anti-inflammatory that means it reduces inflammation and anticoagulant. Aspirin has become a wonder drug, nowadays it is being used somebody has a problem with the cardiovascular problem.

And if the blood viscosity has to be reduced they are asked to take aspirin, so that is also used as an anticoagulant, so it inhibits the production of prostaglandins okay, these prostaglandins are mediators which carry pain okay, this is the structure of aspirin is the acetylsalicylic acid okay, acetyl this is the acetyl group okay, this is the salicylic acid. There are many variations of aspirin because you can think about different situations and you can have different structural features.

(Refer Slide Time: 27:03)

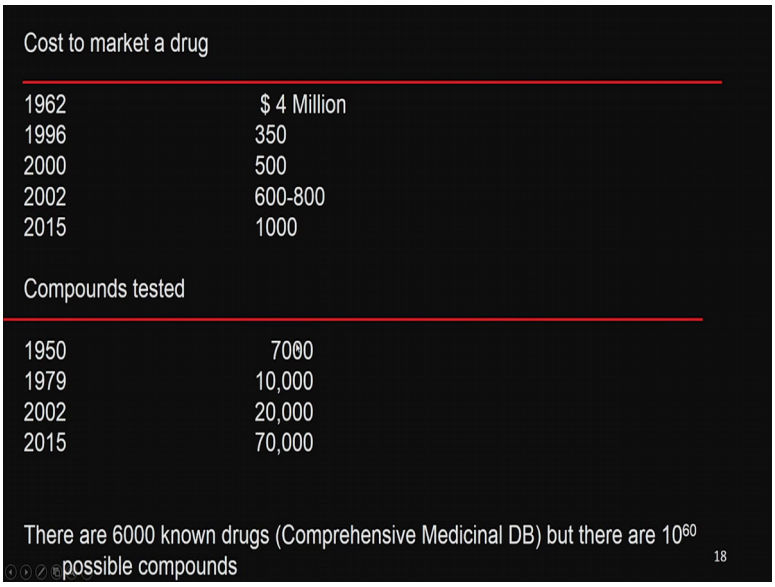


So if you want to go into a drug discovery, we are talking in terms of 10 to 15 years it cost about 1 billion dollars, you may have to test maybe 10,000 compounds, animal trials or clinical trials or in the lab, and then you may have to test it out on about 1000 to 5000 volunteers, and hopefully you may have one successful candidate okay. So drug discovery is like the lottery in computers we may have to test maybe 100 thousand compounds, in the lab we may have to test about 5000 to 10000.

And finally you may succeed with one drug, or some of the most of them will fail. So a lot of money is lost because of this particular exercise, that is why you when drugs are introduced, a

newly introduced drug is always very, very expensive, because there is lot of money involved, and there are many failures for every drug introduced into the market okay.

(Refer Slide Time: 27:58)



So in the 60s drug discovery used to be cheaper, talking in terms of 4 million US dollars, but as you can see you can go down it is become almost a billion dollar okay. Simultaneously, in 50s maybe only few compounds were tested, but as you go down we can see we are talking in terms of 10 times more compounds are being tested. So we started testing more compounds the cost of marketing and drug also keeps going up and up and up okay.

And there are only 6000 known drugs, but there are 10 power 60 possible compound. So obviously there are many drugs in this compound list which we do not know what type of activity it will have okay. So there is lot of scope for developing a new molecules and new compounds and so on. So we will continue in the next class of the concept of drug discovery, and what are the steps involved in target identification and lead identification okay, thank you very much for your time.