AI in Drug Discovery and Development

Prof. Rajnish Kumar
Dept. of Pharmaceutical Engineering and Technology
IIT-(BHU), Varanasi

Week-01
Lecture-02

Welcome to the course AI in Drug Discovery and Development. In this lecture, we will be talking about the drug discovery workflows. So, by the end of this lecture, you will be able to understand the key stages of drug discovery, explain target identification and validation process, differentiate between hit, lead and clinical drug candidate, and also recognize the importance of clinical trials and their different phases as well as explore the regulatory submission and drug approval pathways. So, we have seen in the last lecture that drug discovery and development pipeline is a very complex and it starts with the target identification, followed by hit discovery, followed by lead optimization followed by preclinical testing of the lead compounds, followed by preclinical trials of the drug candidate, followed by a regulatory approval and then the drug becomes available in the clinic which can be used by the patients. So, we start the initial drug discovery phase with a large number of compounds. For example, in this case I have shown that we start with 10000 compounds which are reduced to around 250 compounds during preclinical phase and further reduced to 5 compounds at the at the clinical phase.

And then after FDA review, so we get only one compound that is approved as a drug for the treatment of the disease. And you can see that it takes large amount of time. So, the whole drug discovery and development can take around 10 to 15 years on an average. So, how do we identify the disease? So, usually the companies the pharmaceutical companies they identify a disease target through the market driven approach where the companies they often prioritize the diseases which are prevalent in wealthier regions where there is high return on the investment.

So, some of the common disease target includes osteoporosis, hypertension, hypercholesterolemia. Once, we have identified a disease target, the next is the identification of a drug target. So, to identify the biological target or the drug target, which is either a receptor or an enzyme or a nucleic acid is involved in a disease process. So, the target should be accessible to small molecules or biologics that can modify its activity. It means that they should be druggable. And for that we usually use our different methodologies.

For example, the genome sequencing or proteomics or high throughput screening or bioinformatics approaches are being used for identification of the targets. So, nowadays people are using multi-omics techniques. For example, they are combining all this

metabolomics, proteomics, epigenomics, transcriptomics, etc. And then utilizing that information, we try to identify novel drug target which can be used for developing a drug against a specific disease. So, once we have identified several drug targets, so how do we choose a target among several plausible targets? So, we will take an example of Alzheimer's disease here.

So, there are multiple pathologies which have been reported in Alzheimer's disease like you have the amyloid beta protein, you have the tau pathology where the hyperphosphorylated tau it is being aggregated in the form of neurofibrillary tangles inside the neurons. So, there is a pressing medical need, but there is no viable drug target for Alzheimer's disease. So, there are several theoretically relevant targets for Alzheimer's disease, but they have failed in clinic. One of the examples is gamma secretase, where researchers have identified a highly potent inhibitor for gamma secretase, which is semagacestat, but it has failed in the clinical trials. So, overall identifying a viable drug target is a very challenging task.

And sometimes the companies they identify that they select a target by through the competition as well. For example, they have to stay in the race. So, terfenidine and fexofenidine are the drugs which were developed by the company Sanofi in this competition actually. So, these are the drugs are antihistaminic in nature and they are being used to treat the allergic conditions. So, once we have identified a target, drug target, so next is the target validation.

So, we have to confirm that modifying that target can lead to therapeutic benefits. Otherwise, that target will be of no use. So, some of the techniques which we use for validating a target are RNA interference or CRISPR gene editing or small molecule inhibitors which are called as probes as well. So, one of the examples of a validation of a drug target is validation of HER2 in the case of breast cancer therapy. So, we use several techniques as I said like we use RNA interference, CRISPR gene editing, small molecule inhibitors, beside that we also use genetic studies, we do the pharmacological evaluation and we have some we have access to some biochemical studies as well which can lead us to validate a target.

So, next step is to identify molecules which can inhibit that target, which can engage that target and give us the desired therapeutic effect. So, basically hit refers to a compound that shows desirable activity against a biological target in an initial screening assay. So, they exhibit a measurable effect either the binding affinity or inhibition in a biochemical or cell-based assay and they must be selective to the target otherwise those hit molecules or hit compounds they are called as they are of no use. So, there are multiple ways to identify hit compounds, but to the most common methods which are being used for identification of

hits is first one is the physical high throughput screening methods where we use where we experimentally screen large chemical libraries against a biological target. In a assay actually, for example, we develop a fluorimetric or a calorimetric assay or a FRET based assay where we screen library of 50,000 or 100,000 compounds and then identify the compounds which are inhibiting that which are showing effect in that assay.

Another way is the virtual high throughput screening. So, where we use the computational approaches that predict potential hits before experimental testing. So, this approach reduces the time as well as it reduces the cost and the manpower actually. So, some of the methods which we use in high throughput screening are for example, HPLC or LCMS based methods, HPLC or NMR based methods, absorbance or fluorescence-based assays, enzyme inhibition assays, surface plasmon resonance-based assays and isothermal titration calorimetry-based assays. So, some of these are biophysical methods which can help us to determine the binding affinity of those molecules, those hit compounds against our drug target.

So, in the case of virtual high throughput screening, so we can use molecular docking, we can use machine learning or AI and we also use pharmacophore modeling and all these technologies we will discuss later in the course. And for identification of those hits, one idea is to use the commercially available compound libraries. So, for example, Mabridge, Enamine, Sigma-Aldrich or Vitas-M, Acinex. So, these are commercially available compound libraries which anyone can screen computationally or they can be ordered and can be screened in physical assays as well. So, now once we have identified a hit molecule.

So, now the next step is to convert that hit molecule into the lead molecule and lead is a compound that has been optimized from an initial hit and shows improved potency, selectivity and druglike properties and is suitable for further development. So, it serves as a starting point for preclinical drug development. So, as I said like a hit compound is having some activity, but it is not optimal from that initially identified hit molecule if we want to develop it into a molecule which is optimal in all the aspects like the better pharmacokinetic properties, higher selectivity, favourable toxicity profile and high potency that molecule is called as a lead molecule. And once a lead compound is identified, it undergoes lead optimization where through medicinal chemistry approaches, we further refine the properties of those molecules. So, these properties may be the lipophilicity or solubility, metabolic stability and bioavailability etcetera.

And usually it is done through structure activity relationship studies where we synthesize large number of molecules around that initially identified lead and then try to optimize, try to convert it into a better compound with better properties such as metabolic stability and bioavailability, lipophilicity and solubility. So, these are some of the strategies for lead

optimization where we use as I said like we use structure activity relationship studies which is a systematic modification of functional groups to identify the key pharmacophore. And then there are other medicinal chemistry approaches where we use, for example, bioisosteric replacement or product design or scaffold hopping. And then we do the ADMET optimization by using Lipinski's rule of five through metabolic stability evaluation and hERG testing as well. And there are some computational approaches which can be used to optimize the lead compound which we will again discuss later in the course such as molecular dynamics simulation, AI/ML driven optimization and free energy calculations using methods such as MM/GBSA.

So, the lead optimization cycle looks like this where we start with the lead structure and then we perform structural change which synthesize that new molecule. Now, this new molecule is being tested biologically. So, after biological screening, if we are happy with the activity, if the activity is increased, so then we keep the structural changes. So, if the activity is increased, so then we keep the structural changes and that increase in activity is shown by this green square. So, then we keep the structural changes and if the activity is not increasing of that modified analog which is shown by the red square, then we discard the structural changes and then we perform the new structural changes by using this design make, test, analyse cycle.

So, we optimize the initial lead structure in order to obtain a better molecule, a molecule with better properties. So, followed by this, once we get an optimized lead compound, so the next step is to perform the preclinical testing where the objective is to assess safety, efficacy, toxicity and pharmacokinetics of the lead molecule before the clinical trials. So, there are some regulatory guidelines such as FDA, it says that you have to follow GLP good laboratory practices and we need to generate the safety data before clinical trials and we have the CPCSCA in India which regulates the ethical use of animals in experiments. So, there are some methods like the in vitro studies lab based experiments which can be used for preclinical testing. So, there are methods which can be used for you know in vitro studies like lab-based experiments like cell-based assays, cytotoxicity assays, metabolic stability studies and hERG testing and then there are multiple in vivo studies or animal models available for evaluating the acute and chronic toxicity.

those performing the dose response studies, performing the PKPD studies and modeling a disease by using even genetically modified animals. Following that the pharmacokinetic property shall be evaluated in case of preclinical testing where several parameters are evaluated like the membrane permeability, plasma stability, metabolic stability, solubility, CYP450 inhibition, and BBB penetration. So, the absorption studies are being performed using for example, CACO2 permeability assays and then solubility studies. The distribution studies are being evaluated using plasma protein binding assay, blood brain

barrier permeability, volume of distribution prediction. And the metabolic studies are performed using liver microsomes or hepatocytes or by using CYP450 inhibition assays.

And the excretion studies are performed using the renal clearance assay or biliary excretion studies. So, once we have performed all these pharmacokinetic evaluations, so next step is to identify a clinical candidate. So, which is a most promising molecule optimized lead chosen from preclinical studies for advancement into human clinical trials. So, it shall be having low toxicity and having a high safety margin and then there shall be a possibility to manufacture it in large quantity. So, it should we have to assess the manufacturability and scalability of the lead compound, of that optimized lead compound. So, once we are happy with the results, so then we go for the investigational new drug application filing.

So, this IND filing is a formal request that shall be submitted to a regulatory agency. For example, FDA in the case of US to seek approval for testing the clinical candidate in humans. So, there are multiple kinds of IND applications like there is commercial IND, non-commercial investigator IND, emergency use IND and treatment IND. So, once we file the IND and once the regulatory body approves it, so then one can start the clinical trials. So the clinical trials, they assess the safety, efficacy and optimal dosing of a drug in humans before final regulatory approval.

So, the drugs they must pass clinical trials to prove their safety and efficacy in humans and there are multiple challenges such as the clinical trials cost is very high and then it takes a lot of time and then the patient recruitment and retention is also one of the biggest challenge because usually those clinical trials they run for a longer period of time. So, retaining the patients in that clinical trial is a big challenge. And then there are regulatory hurdles in approval of those clinical trials as well. So, the clinical trial phases they are there are three phases in the clinical trial. So, phase 1, phase 2, phase 3, the phase 1 is mean for safety and doses, phase 1 is mean to evaluate the safety and doses.

So, it is conducted on usually 20 to 100 healthy volunteers or patients and the objective is to evaluate safety, tolerability and pharmacokinetic and to determine the maximum tolerated dose. Likewise, the phase 2 is to evaluate the efficacy and side effects which is conducted on 100 to 500 patients with the target disease and the objective is to assess efficacy and short-term side effects and then it also identifies the optimal dose and the dosing schedule. So, it may be divided into two phases like phase 2a which is dose finding and phase 2b which is confirmatory efficacy studies. And then phase 3 is a large-scale efficacy and safety study where it is conducted on large number of patients around 1000 to 5000 across multiple sites and the objective is to confirm therapeutic effectiveness and long-term safety. So, the data from this phase is used for regulatory approval submission

through the new drug application.

So, once the new drug application is submitted after successful phase 3 trial to seek regulatory approval for marketing a drug. So, usually the regulatory agencies they review it takes around 6 to 12 months. So, usually the regulatory agencies they review the evidences for safety, efficacy and manufacturing standards and it takes usually 6 to 12 months and if approved the drug can be marketed and then it will be available in the market for the use by the patients for treating a disease. There is another phase that is called as phase four, which is known as post marketing surveillance as well. So usually, drugs are monitored for long term safety, rare side effects and new indications.

So it involves collecting data on adverse effects through database and patient reporting. So, some drugs may be withdrawn or relabeled if unexpected safety issues arise. So, one of the examples is rofecoxib which was marketed in the name of Vioxx. So, which is an anti-inflammatory drug was withdrawn from the market in 2004 because of increased risk of heart attack and stroke after long term uses. Another approach is drug repurposing where it is the process of identifying new therapeutic uses for existing drugs that are already approved or in clinical development.

So, the advantage of drug repurposing is that it reduces development time and cost because the compounds they are already having known safety profiles and having a high success rate. So, one of the examples of drug repurposing is the use of anti-parasitic drug ivermectin for the COVID-19 treatment. So, the collaboration and partnership in the drug discovery is the key to success. Recently, there has been an increase in public private partnerships where the pharma companies, academic institutions and government they are coming together and then they are working for a common cause to discover drugs for different diseases. So, in summary, we can say that drug discovery is a systematic multi-phase process that ensure the development of safe and effective treatments and it begins with target selection and validation followed by hit discovery and lead optimization and further preclinical testing which assesses the drug safety and effectiveness before human trial and then the clinical trials phase 1 to 4 they evaluate drug performance safety and optimal dosing And then the regulatory approval involves submission of agencies like the FDA before commercialization.

And another important aspect is this post-marketing surveillance or phase 4 trial which ensures the long-term safety of any drug. And I have an open question for you to think about how might the drug discovery workflow change if we could accurately predict the compounds efficacy and safety using only computational method. And these are some of the nice articles which you can refer to if you would like to know more about the process of drug discovery and development. With that, I would like to say thank you to all.