AI in Drug Discovery and Development

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

Week-01
Lecture-05

Welcome to the course "AI in Drug Discovery and Development." In this session, we will be talking about riddles or challenges in drug discovery. So, by the end of this lecture, the candidates will be able to understand the key challenges in the drug discovery realm. Analyze the difficulties in target selection, hit discovery, lead optimization, and ADMET considerations. Also, evaluate the trade-offs in drug formulation, bioavailability, and preclinical to clinical translation, as well as identify regulatory, economic, and ethical hurdles in drug development. So, as we already know, drug discovery is a highly complex multi-stage process with high failure rates, and despite the scientific advancements, many drugs fail at later stages due to unforeseen challenges.

So, these challenges can be related to anything from starting from the target identification to the phase 4 clinical trial. So, these riddles represent scientific, technical, and economic paradoxes that researchers must solve in order to achieve a high success rate in drug discovery and development. Sometimes you can see that a drug is working perfectly in animal models, but it fails in human trials due to unknown metabolic pathways in different species as well. So, we will be discussing all the challenges we face in drug discovery and development one by one.

So, let us start with the target identification riddle. So, choosing the right biological target, which can be an enzyme, receptor, or protein, is critical but uncertain. So, several challenges exist; one of them is the complex disease pathway, where multiple mechanisms make it hard to pinpoint a single target. One of the examples is Alzheimer's disease, where multiple pathologies are present in the patient. So, such as amyloid beta pathology, tau pathology, and inflammation.

Another challenge is the undruggable target. There are some proteins that lack suitable binding pockets for drugs, and that is why they are difficult to target. And then there are off-target effects, which are highly prevalent in drug discovery. So, the compounds can lead to potential toxicity due to unintended interactions of the drug molecules with other proteins. So, there are challenges with the target validation.

So, most of the time we lack a reliable animal model. So, next are the challenges with target validation. So, most of the time we lack reliable models. For example, we are using animal or cell models that may not fully replicate human diseases, and that is how sometimes a

drug candidate works very nicely in an animal model, but when we translate it to humans, it has no effect. Another challenge is the off-target effects.

So, sometimes when we are using CRISPR or RNAi, we are silencing the targets during the validation. So, they may impact unintended pathways, and then we can get the wrong results. So another challenge is biological redundancy. So, multiple proteins perform similar functions, and that is how making a single target drug ineffective. So, if we are blocking one pathway, the cell is depending on another pathway for the same function.

So, then that blockage becomes ineffective. And then the next challenge is translational failure. Validated targets in preclinical studies may not work in humans. And then there are many challenges with hit discovery. So, as we know, the chemical space is very large.

So, finding bioactive hits among millions or billions of compounds is challenging, especially with high-throughput screening. So, there are screening limitations because we want to screen a library of 1 million compounds or maybe 1 billion compounds. So, it is almost impossible to screen 1 billion compounds in high-throughput screening assays because we need a lot of investment for that. So, another challenge is the false positives and negatives when we identify hit compounds. So, some of those compounds show known specific binding or aggregation in the assay and other assay artifacts that mislead the hit identification.

So, usually we use this screen called pan-assay interference compounds screening. So, to avoid those compounds that can have a false positive outcome. And then there is a challenge with the hit potency versus drug likeness. So, sometimes those potent hits may have very poor ADMET properties, and optimizing them into an optimal ADMET properties-containing compound. Another challenge is the balance between hit potency and drug likeness.

So, many times the potent hit compounds may have very poor ADMET properties. While optimizing those ADMET properties, the hit compound may lose its efficacy. Another challenge is the computational versus experimental gaps, so most of the time the predicted hits may fail in biological assays; therefore, we need computational methods with high reliability. And then there is a reproducibility issue as well, so the hit compound must show consistent activity across different tests. So, that is how we need to use a primary screening assay followed by an orthogonal assay to validate the hit compound.

So then there are these Lipinski's Rules of Five paradoxes. For example, large molecules like antibiotics, peptides, and PROTACs break the rule of five and are still effective drugs. For example, cyclosporine, which is an immunosuppressant, violates a rule of five but

works due to active transport mechanisms. And then, PROTACs also have a large molecular weight, but they are highly potent due to targeted protein degradation. They have a large molecular weight, but they are highly potent due to targeted protein degradation.

And then there are problems with the hit-to-lead optimization stage, as well. So many of the compounds have poor ADMET properties. So the hit compounds may have low solubility, permeability, or metabolic stability. And then there is a selectivity challenge as well, such as optimizing for target specificity without affecting similar proteins. Many times, the compound is not highly selective; it is active on similar proteins.

So, that is a big challenge: to optimize the compound into a highly selective and potent molecule. And then there are structural optimization issues as well. Where a small modification in the molecule can drastically alter the bioactivity, that term is known as an activity cliff. A very small change in the structure leads to a drastic change in biological activity. And then another challenge is synthetic feasibility, where complex molecules may be difficult to synthesize on a large scale.

And this is one of the limitations of generative modeling, which we will discuss later. And then there are challenges with lead optimization, as well. So, we have to balance potency versus drug likeness, where we have to enhance the activity without compromising drug likeness, such as solubility or permeability. And then there are ADMET issues; for example, addressing poor absorption, rapid metabolism, toxicity, and low bioavailability. Again, target selectivity is an issue.

We are improving specificity to reduce off-target effects and side effects. To ensure chemical stability and scalability, we must ensure that the lead compound is stable and synthetically feasible. And then, regarding metabolic liability, we need to prevent rapid breakdown while ensuring the safe elimination of the lead compound. So this is summarized in this picture. So you can see the problems in lead optimization.

So, they are related to potency versus solubility, where the increase in binding affinity often reduces aqueous solubility. And then there is an issue related to lipophilicity versus safety, where increased lipophilicity improves permeability but may also cause toxicity. So, there are issues with metabolic stability where reduced metabolism can lead to accumulation, and then that can lead to toxicity. And then there are selectivity issues where highly selective compounds may have poor permeability and absorption. And then there is an issue with the efficacy, where enhancing the target engagement may also increase the off-target effects.

So, there is an issue with the bioavailability that is known as the bioavailability bottleneck.

So, the bioavailability refers to the extent and rate at which the active drug ingredient is absorbed and becomes available at the site of action. So many promising compounds fail due to poor solubility and absorption. Up to 40% of the new chemical entities have poor bioavailability. Paclitaxel is one example of an anti-cancer drug that is highly effective but requires a solvent to increase its solubility, which can lead to toxicity.

So, some of the problems with bioavailability are related to solubility. So, there is a trade-off between solubility and permeability. So, if we enhance the solubility, it often reduces membrane permeability, and first-pass metabolism is another challenge involving the liver enzymes known as CYP450 enzymes. So, they degrade drugs before they reach systemic circulation. So, approximately 40% of new chemical entities face solubility issues; the amorphous forms are generally more soluble but less stable in nature.

So, the ADMET predictions are also challenging. For example, in the case of absorption, poor solubility or permeability can prevent drug uptake, and the gastrointestinal pH and enzyme activity vary among individuals, affecting absorption. And if we talk about distribution, predicting drug penetration into tissues, including the brain and blood-brain barrier permeability, remains very difficult. And the plasma protein binding affects the concentration of the free or active drug. And then, related to metabolism, liver enzymes like CYP450 metabolize these drugs, but predicting the exact metabolic pathways is highly challenging.

Some of the drugs they produce toxins or activate undesired pathways. Related to excretion, the drug clearance rates through the kidneys, known as renal excretion, and liver hepatic metabolism may vary among individuals, and some drugs may accumulate in tissues, leading to long-term toxicity. The toxicity prediction also faces challenges. For example, the off-target toxicity, the unintended interaction with other proteins that causes adverse effects, and the metabolic toxicity. So, the reactive metabolites can lead to hepatotoxicity or genotoxicity, and there are species differences as well, where the animal toxicity data may not accurately predict human toxicity.

And then there are idiosyncratic reactions, which are rare, unpredictable toxic responses in a subset of the population. Some of the challenges exist in the formulation development, as we discussed regarding the solubility. So, the solubility issue is very prevalent in formulation studies. So, poorly soluble drugs limit bioavailability and therapeutic efficacy. And the formulations they face stability challenges as well.

So, the formulations may degrade due to oxidation and hydrolysis. Drug-excipient compatibility is another challenge where there might be unwanted interactions affecting the potency and stability of the drug molecules in a formulation. Moving from the

preclinical to the clinical transition also faces some challenges. For example, the species difference is that the animal models do not always replicate human physiology and drug responses, which is why some of the drugs work well in animal models but do not show any effect in humans. And then there is a PK/PD mismatch, a pharmacokinetic/pharmacodynamic mismatch.

So, where difference in pharmacokinetic and pharmacodynamic parameters between preclinical and clinical studies. And then, toxicity discrepancies indicate that the safe doses in animals may cause adverse effects in humans because of some metabolic issues. And then there are toxicity discrepancies as well, where a dose that is safe in animals may cause adverse effects in humans. And then there are dosing challenges, such as scaling the dose from animals to humans, which is complex and often inaccurate, leading to difficulties in calculating a dose for humans based on the animal experiment. And then there are disease model limitations, such as the animal models, which may not fully capture the complexity of human disease.

Likewise, the clinical trial faces some challenges. So, there are some challenges related to clinical trial design that can occur at the stages of patient recruitment and retention, as well as in placebo and control selection. Sample size determination, blinding and bias minimization, cost and time constraints, and adverse event monitoring are essential components of the study design. So, some of the challenges in clinical trials are summarized here. For example, in Phase 1, the first-in-human dose is selected.

So, there is uncertainty in selecting a safe starting dose based on preclinical data. Unexpected toxicity is another challenge in which the adverse effects on humans are not predicted by animal studies. Healthy and patient volunteers have differences in drug metabolism. In phase two, we may see challenges related to the limited patient population, where the results may not reflect a broader demographic. And we have challenges with the placebo effects as well, which are impacting perceived drug efficacy.

And then we have challenges with the efficacy endpoints, where we face difficulties in selecting meaningful and measurable outcomes. And we also have a very high failure rate in phase two. So, in phase 3, we face challenges with the cost and duration because these trials can take years and require massive funding. Patient recruitment and retention also face challenges because the long trial causes many dropouts during the trials. And there are some rare adverse events as well, which are difficult to detect in smaller studies.

The large-scale trial management and data analysis are also highly challenging. So we talk about the challenges in phase four studies. So monitoring long-term safety is highly challenging because it may reveal late-emerging adverse effects. Identifying new

indications for approved drugs is highly challenging as well. And if we talk about the regulatory and legal risks, there may be drug recalls or label modifications due to new safety concerns.

As we discussed earlier about the drug rofecoxib, which was withdrawn from the market after incidents of severe side effects related to cardiac arrests. And then there is a challenge: there are challenges with the pharmacovigilance compliance that we have to maintain in order to meet the ongoing monitoring requirements as well. And then, there are some challenges with biomarker development as well. We need to ensure that those biomarkers are reproducible and correlate with the clinical outcomes.

So, we have to validate those biomarkers. Another challenge is biomarker heterogeneity, where the variability across patient populations complicates the universal use of biomarkers. Biological complexity is another factor that poses challenges, such as complex diseases like cancer and neurodegeneration. They involve multiple interacting pathways, which makes the biomarker hard to pinpoint. And there are challenges in maintaining the cost and accessibility of biomarker development as well because we need to use expensive assays and have limited access to advanced testing facilities. So now we will discuss a case study on the development of a successful proton pump inhibitor called omeprazole.

And we will see how challenging it is to discover a drug. And so the company called AstraZeneca started by identifying a compound called CMN131, which was a highly potent lead compound that showed very good results in decreasing acid secretion, but it also showed liver toxicity. So they could not continue with this molecule. So, they had to identify another molecule. So, they identified another molecule, a benzimidazole derivative known as H124/26, but this scaffold had already been patented by another company.

So, they could not use that a molecule. So, then they identified another molecule which is an metabolite of H124. So, this compound was again potent called as timoprazole H83/69, but further it this compound was showing side severe side effects which was due to inhibition of iodine uptake by this compound. So, then the company discontinued that compound as well and then they identified another compound known as picoprazole H149/94. So, this compound picoprazole was showing excellent activity, but again this compound was showing side effects due to necrotizing vasculitis.

So, they further could not continue with that compound. So finally, they came up with this molecule known as Omeprazole H168/68 in 1979, which was a potent and non-toxic inhibitor of acid secretion, which is a proton pump inhibitor and was able to reduce acid secretion and became a blockbuster drug. So, later on when they identified that the S enantiomer of the omeprazole is more active. So, they identified and marketed S

omeprazole, got approval for S omeprazole which is the S enantiomer of omeprazole which is more potent than the R enantiomer. So, you can see from this case study that drug discovery overall is highly challenging and then it is a long process and it require continuous efforts to develop a drug and bring it into the market starting from the identification of drug targets. So, in summary we can say that drug discovery involves solving numerous scientific and technical riddles across multiple stages and the target identification and validations are fraught with uncertainties including off target effects and biological redundancies Hit to lead and lead optimization phases require balancing potency, selectivity and drug likeness while managing ADMET challenges and the transition from preclinical to clinical trials is complex due to species differences and unpredictable human responses.

So, you may go through these publications to know more about the challenges in drug discovery and development. And in the end, I have an open question for you. So, in hit discovery, how can we balance the vast possibilities offered by AI-generated molecules with the need for synthesizable and drug-like compounds? With that, thank you all.