AI in Drug Discovery and Development

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

Week-06
Lecture-26

Welcome to the course "AI in Drug Discovery and Development." In this session, we will talk about the basics of lead optimization. So, by the end of this lecture, you will be able to understand the role of lead optimization in drug discovery. Explore the medicinal chemistry and computational strategies used for lead optimization. Analyze molecular obesity and structural simplification for improved drug properties. And discover AI/ML-driven approaches for accelerating lead optimization, as well as identifying key challenges and solutions for optimizing lead compounds.

So, as we have seen earlier, like in drug discovery, lead optimization is a very important step in which we start by identifying a hit molecule. And once we have obtained a hit molecule that has some properties, activity, and bioactivity, this hit may not be optimal. For example, for ADMET properties: toxicity, solubility, and physical and chemical properties, we now have to optimize this hit; we have to convert this hit into a lead molecule and then optimize this lead molecule. So, the optimization of lead is a process of modifying and refining lead compounds to improve their potency, selectivity, pharmacokinetics, and safety while maintaining or enhancing their biological activity.

So, you can see how difficult and challenging this lead optimization step is because you have to optimize this molecule across multiple parameters, which is also known as multi-parametric optimization while retaining biological activity. Most of the time, it happens that whenever you make a very small change to this molecule to improve, for example, its solubility, you lose biological activity. Once we have this lead molecule converted into the preclinical candidate, we move to the clinical candidate and finally to the marketed drug after the clinical trials. So, it is a crucial step in drug discovery aimed at identifying the best candidates for preclinical and clinical studies. So, the objectives, as I said, are, you know, lead optimization, for example, improving potency and efficacy and enhancing selectivity in many targets.

So, there are, you know, multiple similar targets that can also be engaged by the molecules, leading to side effects or undesired consequences. So, in that case, we wanted to selectively target our target of interest, and we needed to enhance the selectivity profile of those lead molecules. And we also need to optimize the pharmacokinetic properties, where we try to improve the molecules' absorption and distribution, control metabolism, and also change

or optimize excretion. The toxicity is due to the fact that we have seen that one of the major reasons for the failure of drugs in phase three clinical trials is the lack of safety, which means toxicity. So, we need to reduce toxicity, and one objective of lead optimization is to improve synthetic feasibility.

Okay, and usually how it happens is we have, you know, our DMTA cycle where we design a molecule using either computational modeling or, you know, medicinal chemistry insights, such as structure-activity relationships or other techniques. We design, make changes to the structure, and then synthesize the molecule. After synthesizing, we evaluate the molecule; this test is done, for example, in vitro DMPK and biological testing—drug metabolism and pharmacokinetic studies, with DMPK standing for that. So, we test that molecule, and then the data are analyzed. So, until we are satisfied with the, you know, the properties or the profile of that molecule, we keep doing this.

So, this DMTA cycle is, you know, an iterative process. So, until we get an optimized lead, we have to go through several rounds of this DMTA cycle. Some strategies for optimizing lead compounds involve discussing medicinal chemistry approaches. So, we have the bioisosteric replacement; we use prodrug design, and we use structural simplification. We use structure-activity relationship studies to make changes to the molecular structures.

And then we have computational approaches like molecular dynamics simulations, AI/ML-based optimization, free energy calculations, and QSAR modeling. To gain insight into how that molecule is interacting with the target, either through a structure-based approach or by using QSAR models if it is a ligand-based approach. And then we make a decision based on the outcome from these models, design the molecule, synthesize that molecule, evaluate it, and then it goes back into that DMTA cycle. So, let us have a look at all these med chem approaches, such as the first one, which is Structure-Activity Relationship studies. So, what we do involves systematically modifying a compound structure to understand how changes influence biological activity.

And it is a cornerstone of rational drug design, guiding modifications that improve drug-like properties. You know we conduct systematic chemical modifications and then identify the key structural features that lead to improvements in the properties of that molecule. So, there are multiple approaches to SAR studies. One is the functional group modification, where we add, remove, or substitute the groups to assess their impact on activity. For example, we have a hydroxyl group in one of the molecules, then we convert this hydroxyl into a methoxy group, and we see how this change leads to a change in the bioactivity of that compound.

And then we can use bioisosteric replacement, where we substitute groups with bioisosteres to improve the drug properties while retaining the activities of those molecules. And then we can modify the ring system by changing or replacing ring structures to enhance metabolic stability and receptor interaction. Or we can change the hydrophobic or hydrophilic nature of the molecule, where we can modify the polarity to optimize the solubility and membrane permeability of those lead molecules. So one of the case studies I wanted to discuss here is, you know, SAR for this molecule, where you can see the benzimidazole ring. So it states that the electron-withdrawing groups at the five and six positions are well tolerated, and the introduction of nitrogen into the ring is well tolerated.

So if you introduce nitrogen into this ring, you can put an electron-withdrawing group at the 5 or 6 positions. So, these are well tolerated, meaning they do not lead to a decrease in the activity of this molecule. And then at the N-alkyl chain, if you add a hydrophobic group, it is essential for the activity. And then the trifluoromethyl group showed the best balance between potency and metabolic stability improvement; thus, the trifluoromethyl group was metabolically stable. So, it was optimal for you to know about substitution here on an alkyl chain.

And then, for the acyl fragment, the hydrophobic group is important for activity, and pyridine and pyrazole derivatives improve potency and ADME properties. So, this is how the structure-activity relationship is performed for a scaffold or a series of molecules. So, you get insights about which kinds of substitutions you can make or which kinds of substitutions are required for the activities. Based on that, you design a molecule and then use the designed molecule to synthesize and further evaluate its activity. So, they started with this initial hit compound, which had an IC_{50} value of 1.89 micromolar.

And you know the Pampa permeability of 6.7 to 10 raised to the power of minus 6 centimeters per second, and the logD value of 4.6. So when they conducted the multi-parametric structure-activity relationship studies, they optimized and obtained new compounds. So the compounds A and B, where compound A had this substitution and compound B had this substitution, showed that the IC_{50} improved from 1.8 micromolar to 0.28 micromolar for compound A. The PAMPA permeability was 2.3×10 to the minus 6 centimeters per second, and the ELOG D value was 3.1, while for compound B, the IC_{50} value was 0.23.

So, it means that they converted it into a potent molecule while retaining its activity and improving its other properties. So, the second technique is called bioisosteric replacement, which involves substituting functional groups in a lead compound with bioisosteres that are chemically distinct but functionally similar while maintaining the biological activity. So, what we do in this approach can help us enhance potency by improving target binding affinity and increase selectivity by reducing off-target interactions. Optimize

pharmacokinetics by modifying metabolism, solubility, and stability; reduce toxicity by eliminating reactive or toxic functional groups. Enhance bioavailability by improving absorption and distribution properties.

So these are some of the bioisosteric groups. So the idea of bioisosteric replacement is that you replace, for example, hydroxyl with NH_2, SH, and OCH_3, which are in chemically similar environments. They will be producing, but they can alter the biological activity, or they can alter other properties, such as polarity, hydrogen bonding, and the stability of these molecules. And then a methyl group can be replaced by fluoro, chloro, or hydroxy, which can modify the lipophilicity, steric properties, and stability of these molecules. And the carboxylic group can be replaced by bioisosteres of carboxylic groups, such as tetrazole and sulfonamides, which improve stability and bioavailability.

An amide can be replaced by an ester, a urea, or a ketone, which enhances stability, solubility, and hydrogen bonding. A phenyl group can be replaced by pyridine, thiophene, or furan, which increases stability, solubility, and diversity. So, these are, you know, the key bioisosteric replacements that are being used extensively in drug discovery at the stage of lead optimization. So, this is one of the case studies where the impact of bioisosteric replacement of a phenyl ring with an ethynyl moiety on developmental ability parameters in the context of HRV 3C inhibitors is analyzed. So, you can see that here the authors reported one phenyl group and a difluoro-substituted phenyl group.

which had this molecule with an activity value of 0.011 micromolar EC_{50} and a clogP of 2.4 Human plasma t half was 1.4 hours, and the aqueous solubility was 25 micrograms per mL when they replaced it with an ethynyl moiety. So, the activity did not change much, but the human plasma half-life improved, and the aqueous solubility also improved significantly.

So what they did led to a reduction in lipophilicity, which improved metabolic stability; it also improved aqueous solubility. So, the compound had a longer half-life, and it was soluble as well. And then in the next technique, we can perform functional group or ring modification. So, the functional group modifications involve adding, removing, or replacing specific functional groups to enhance drug activity, stability, and pharmacokinetics. So, by systematically modifying functional groups or ring structures, the chemists can develop more potent, selective, and bioavailable drug candidates.

For example, fluorine improves metabolic stability by blocking oxidative metabolism, and adding a pyridine ring in place of benzene also improves metabolic stability. So this is an example of lead optimization in the discovery of Gleevec, which is the anti-cancer drug imatinib. So they had this pyridine ring here, which enhanced cellular activity, and then

they had this side chain amide that was selective for tyrosine kinase. And then they had this, you know, methyl substitution on this benzene ring, which eliminates the PKC affinity. So, overall, making this molecule leads to increased solubility and oral stability by using all these medicinal chemistry lead optimization approaches.

And then there is another concept called molecular obesity and structural simplification. So, molecular obesity describes drug candidates with excessive molecular weights and structural complexity. So there are some molecules that are really bulky; they are very big. They have a very high molecular weight, and a higher molecular weight leads to many problems. For example, it reduces solubility and permeability; it can also affect metabolic stability.

So usually when we talk about molecular obesity, it means that if the molecules have high molecular weight, they typically exceed drug-like thresholds. And then they have increased lipophilicity or a higher log P/log D, which leads to poor solubility and metabolic instability. And then they have the complex molecular architecture that can contain multiple rings, excessive functional groups, and high steric bulk. Then reduce synthetic accessibility because it is complex and multifunctional. So, it is quite difficult to make them synthesize in the lab.

So, the impact on drug development, as I said, leads to reduced oral bioavailability due to poor absorption and permeability. It has a higher clearance rate as well, leading to a shorter half-life and increased dosing frequency, and it has a greater risk of toxicity due to off-target interactions. So, the solution to this molecular obesity problem is the structural simplification that involves reducing molecular obesity by removing non-essential groups and simplifying complex structures. So they can be easily synthesized. So, the pharmacokinetic properties might have better absorption, metabolism, excretion, and other characteristics.

And then drug development speeds up because we can perform lead optimization faster. So, one example of structural simplification is the discovery of the mu-opioid agonist methadone from morphine. So, morphine is a natural agonist of the mu-opioid receptor, which is obtained from the opium plant. So, it has mu-opioid receptor agonist activity and a Ki value of 1.8 nanomoles per liter. So, it contains five rings: A, B, C, D, and E.

In the first step, ring E was removed. And then the compound that was obtained was named butyphenol. So, which is a kappa opioid receptor agonist with a Ki value of 2.5 nanomoles per liter? And then further, rings C and E were also removed, leading to a compound called pentazocine. So, which had kappa-opioid receptor agonist and mu-opioid receptor antagonist activity. And then further removal of rings B, C, and E leads to the compound pethidine, which is, you know, a mu-opioid receptor agonist. And then, further removal of

rings B, C, D, and E led to compound 5, which is named methadone and has mu-opioid receptor agonist activity. So you can see that you know simplification because now these pethidine and methadone are structurally simplified molecules that can be easily synthesized in the lab during lead optimization. So, molecular simplicity or simplification is another approach to lead optimization.

Then we have the next approach, which is called prodrug design. So, it is a strategic approach in lead optimization where an inactive or less active precursor compound is chemically modified to enhance its pharmacokinetic and pharmacodynamic properties. The prodrug is converted into an active drug in vivo through either enzymatic or chemical processes, improving drug performance while overcoming key development challenges. So, we can have ester prodrugs that improve lipophilicity and absorption. We can have phosphate prodrugs that enhance water solubility.

We can have amide prodrugs that can also modify metabolic stability. So, one of the case studies for this prodrug is triptolide, which is a diterpenoid epoxide from Tripterygium wilfordii. It exhibits unique bioactivity and has been clinically evaluated in China for rheumatoid arthritis and leukemia. However, the molecule suffers from poor water solubility and severe toxicity. So, to address this, omtriptolide, which is a highly water-soluble prodrug, was developed as a 14-succinyltriptolide sodium salt that converts to triptolide in vivo.

So, you can see here that this succinyl triptolide salt has formed. And this salt, when it enters the body during metabolic processing in the liver, is converted into triptolide, which is the active component of this drug. So, this is a prodrug of triptolide. Ok, coming to the computational approaches, we have seen the classical medicinal chemistry-based methods that are being used for lead optimization. So, now let us discuss some computational approaches, such as molecular dynamics simulations, which are powerful methods to study structural stability.

conformational flexibility and binding interactions of drug candidate with their target protein at an atomic level and provides insights beyond static docking studies. So, we can use it in you know defining the lead target binding interaction, understanding ligand induced conformational changes and validating the bioisosteric replacements. And then we have the free energy calculations; we can either use MMGBSA, MMPBSA, or we can use a relative free binding energy. Or we can also do the absolute binding free energy (ABFE), or there are several tools nowadays, such as free energy perturbation and thermodynamic integration; all of these methods can be used for determining the free energy. So, these free energy calculations they help us to determine how strongly a ligand is binding to the to the receptor or the target.

So that we can make alterations to the ligand structure, we can see how the binding affinity changes by calculating all these free energies using all these methods. So, unlike the docking scores, it provides more accurate and physics-based free energy estimates, guiding modifications to enhance potency, selectivity, and drug-like properties. And we can integrate MMGBSA with MD simulations, which accelerates the identification of highly potent and selective drug candidates. And, as I said, in addition to MMGBSA free energy binding calculations, we can use all those absolute binding free energies. Relative binding free energy and methods to determine the binding free energies of molecules with the targets of interest.

And how do we use them? We can use them to rank lead compounds based on binding affinity and to understand key binding contributions. And we can refine the molecular docking results. And we can also predict the effects of lead modifications on binding stability. And then we have another approach that is heavily used in lead optimization, which is QSAR modeling. So, which is a computational approach that establishes a mathematical relationship between the chemical structures of compounds and their biological activities? And then what we use here are the machine learning concepts, or deep learning concepts, to correlate the structure of a molecule with its biological activity.

We know how the structure correlates with biological activity. So we can alter the structure in order to improve biological activity. So, there are multiple types of QSAR methods, such as 2D QSAR, that use topological, electronic, and physicochemical descriptors. Like HANSH analysis and Free-Wilson models, which we will discuss later in this course. And then we have the 3D QSAR, which incorporates spatial features and molecular fields such as CoMFA and CoMSIA.

And then we have AI/ML-based QSAR, which uses deep learning, SVMs, and random forests for more accurate predictions. So, QSAR modeling helps in rational drug design by guiding modifications to improve efficacy, reduce toxicity, and enhance drug-like properties. And then we can use AI/ML-driven optimization, where we employ AI/ML algorithms to refine drug candidates by predicting their biological activity, pharmacokinetics, toxicity, and binding efficiency. So, it accelerates drug discovery by leveraging predictive modeling, deep learning, and generative algorithms to design optimized leads with enhanced potency, selectivity, and drug-like properties while reducing experimental costs and timelines. Some of the areas where we use AI/ML for lead optimization are AI-driven QSAR modeling and deep learning for property prediction.

Generative models for de novo drug design and AI-powered molecular docking virtual screening are being developed. So, we will actually discuss all of these in this course. So,

some of the examples have been deleted. It is called Deep Lead Optimization Enveloped in a Protein Pocket.

So, this is a recently released tool. So, which uses a unified deleting strategy for lead optimization. So, what it can do is handle fragment growth, linking, and replacement. So, for example, if you have one fragment that you have identified as binding to your target structure, So, you can grow this fragment by using this technology, and then if you have maybe two fragments, you can link those two fragments, and you can also replace those fragments with other scaffolds. So, this is, for example, scaffold hopping, or these are other classic techniques like scaffold hopping, or, you know, R-group analysis; but in this case, you can use deep learning to do all this. So, you can use both 2D and 3D structures, and it accounts for protein-ligand interactions based on the binding poses as well.

It uses a structure-aware network for improved prediction, and the GitHub ID, along with a GitHub link, is provided here. So you can go to GitHub, and then you can get more information about it. And likewise, we have Optibrium StarDrop, which is a comprehensive platform for small molecule design optimization that handles AI-guided lead optimization capabilities. So, they have these patented rule induction methods where the sensitivity analysis for developing optimization strategies is being done. And then it integrates multiple parameters for balanced compound design; the link to this is also provided.

So you can refer to it if you want more details. And then we have the Cresets flare, which is version 8, so it is again an advanced protein-ligand modeling tool for lead optimization. So, where you can do the FEP, as well as free energy perturbation, and then incorporate molecular mechanics. MM/GBSA methods for free energy calculations can also calculate the binding free energies of ligand-protein complexes. It can also provide an intuitive visualization of protein-ligand interactions, and the link is provided here.

So, then we have open eye lead optimization solutions. So, it empowers people in designing potent selective molecules offering flexibility through cloud and local machine access as well. So, we can either do it in the cloud or on a local machine. It includes tools for structure-based design and ligand-based design as well. And it also provides molecular property prediction and ADMET toxicity modeling, and it supports large-scale virtual screening and compound library design.

And then we have Atomwise's AtomNet. So, it is a deep learning-based platform for small molecule drug discovery that specializes in structure-based drug design and lead optimization. So, it utilizes convolutional neural networks for binding affinity prediction and is capable of screening billions of compounds rapidly. It also offers partnership

programs for pharmaceutical and biotechnology companies. Of course, it is not possible to cover all the tools that are available. So, I have just covered a couple of them, but there are plenty of tools that can be used for, you know, AI/ML-enabled lead optimization.

And then we have Schrödinger's Live Design, which is a collaborative platform for structure-based design. integrates physics based modeling with machine learning and it offers real time collaboration feature for team based optimization. That includes automated workflow capabilities for high throughput analysis as well and provides seamless integration with other Schrodinger's tools such as Glide etc. Okay, coming to the challenges.

As I told you, lead optimization is one of the most difficult challenges. So, the first challenge is the potency versus selectivity. Here, we enhance target binding while minimizing off-target effects. So, what we have to do is increase the potency while keeping the selectivity as well. Because if we are increasing the potency and that molecule is not selective to the target of our interest, and it is, you know, binding to other targets as well, then it will be of no use. And then we have the ADME issues, such as the compounds that suffer from poor solubility, permeability, metabolic stability, and clearance.

And then we have the toxicity risk, where eliminating harmful functional groups without losing activity is key. And then we have synthetic challenges as well; we are designing lead compounds that are easier and more cost-effective to synthesize. And then we discussed molecular obesity, where we need to reduce the excessive molecular weight and complexity of the molecules to make them more favorable and optimal. And then we have a very high attrition rate as well, even if we have an optimized lead compound. So, they fail in preclinical or clinical trials because they can face multiple challenges, such as toxicity issues or bioavailability issues.

Ok, summarizing all this, as we discussed, lead optimization plays a crucial role in drug discovery by refining lead compounds to enhance potency, selectivity, pharmacokinetics, and safety. And we have traditional med chem approaches that include SAR studies, bioisosteric replacement, and structural modifications, which are essential for improving drug properties. Then various computational tools such as molecular dynamics simulation, MMG BSA free energy calculations, and QSAR modeling can help accelerate lead refinement and optimization. And we have several emerging tools, such as AI/ML-driven optimization and generative AI, which are shaping the future of lead optimization. And we have plenty of those tools that can be used for this important drug discovery step, which is lead optimization.

And with that, I have an open question for you. So, if a lead compound shows excellent

potency in vitro but fails due to poor pharmacokinetics and toxicity in vivo, how would you approach its optimization? So, just think about it, and then I have given some literature here. So, you can just go through all these papers to get more information about, you know. So, especially in this deep lead optimization Zhang's paper, where you can see how GenAI can be used for structural modification in lead optimization. And with that, thank you.