Early Development Medicinal Chemistry: Utilizing Data and Artificial Intelligence

April 3, 2022
Meg Rivers
Pharmaceutical Technology, Pharmaceutical Technology, April 2022, Volume 46, Issue 4
Pages: 21–22
Experts discuss what data to consider when selecting a high-potential drug candidate and how AI can be harnessed in medicinal chemistry drug discovery.

In early development of medicinal chemistry, there are a lot of considerations, such as determining promising agents and dosage form. *Pharmaceutical Technology* interviewed ***Chase Smith, PhD, senior application scientist at Optibrium*** (a software company for drug discovery), and ***Kevin Short, director of medicinal chemistry at Verseon International*** (a clinical-stage pharmaceutical company), who discuss key considerations for medicinal agents in early development, challenges and opportunities in medicinal chemistry, what data to consider when selecting a high-potential drug candidate, and how artificial intelligence (AI) can be harnessed in this process.

Key considerations in early development

PharmTech: What are key considerations when working with medicinal agents in the early development phase?

Short (Version): The most obvious general consideration is whether or not there are multiple paths forward. Since the medicinal chemist will inevitably synthesize multiple rounds of compounds in order to optimize physicochemical properties, pharmacologists will need to ensure there are easily accessible and relevant pharmacokinetics and

disease models, which will interrogate the compound candidates. There needs to be a functional link between potency in *in vitro* assays and *in vivo* efficacy in humans. During all the above, the medicinal chemist constantly needs to pay attention to physicochemical properties, and [consider] whether further optimization retains the original plan to deliver the drug by the desired method (i.e., oral, intravenous, intraperitoneal, etc., route of administration).

Smith (Optibrium): Ideally, one would be able to test any potential drug candidate in a well-proven disease model, allowing you always to identify the best possible compound. Realistically, chemical space is too vast, and good early-stage disease models are often not practical to run due to costs or being too low in their throughput.

Consequently, a phased approach to the different discovery stages is taken where the key criteria shift as the project evolves. Early on, the focus is often only on activity and perhaps selectivity. As things progress, the scope is expanded to include a broader range of ADMET [absorption, distribution, metabolism, excretion, and toxicity] properties before ultimately focusing on PK/PD [pharmacokinetics/pharmacodynamics], safety, and manufacturability at scale.

Early discovery has made great strides in strategies and methods for increasing throughput and maximizing the sampling of chemical space. However, the greatest challenge is moving some of the later-stage indicators (downstream ADMET, PK/PD, safety) earlier in the process. Computational tools are now increasingly contributing to this, with the ever-improving performance of predictive modeling synergizing with idea generation and guided library enumeration capabilities. These tools are facilitating improved virtual screenings, thereby generating higher-quality predictions. Overall, this helps to more rapidly identify high-quality chemical matter while simultaneously avoiding false leads before investing significant resources.

Challenges and opportunities

PharmTech: What do you see as the greatest challenge in medicinal chemistry in early drug development? The greatest area for opportunity?

Smith (Optibrium): One of the primary difficulties is that findings from early-stage assays and measurements often fail to translate to later-stage and vastly more complex systems, such as animal models or even humans. Another problem is that early drug discovery data is inherently noisy due to the uncertainty of the measurements and experimental error. This can lead projects in the wrong direction, wasting time and resources that could be better allocated or lead to missing opportunities.

While there is a lot of hype and promise around AI in drug discovery, much remains unsubstantiated. Still, several proven technologies are pushing the boundaries of what is possible. [Particular] methods are performing well that effectively account for the nature of data in early drug discovery, the relatively high variability and uncertainty in measurements, and the fact that most compounds have only been measured in subsets of assays due to time or cost constraints (sparse data). Furthermore, to build trust and acceptance by scientists, AI technologies need to align with how discovery scientists think and work to achieve greater adoption and significantly increase the rate of success and speed of drug discovery.

Short (Version): The most obvious challenges relate to diseases with additional obstacles to drug delivery, such as crossing the blood-brain barrier. Finding readily deliverable drugs for Parkinson's, Alzheimer's, multiple sclerosis, and other central nervous system (CNS)-related diseases is much more complex than typical non-CNS drug development. At the same time, the challenges in CNS drug development mean treating these conditions is also an arena of great opportunity. As for other types of systemic opportunities in medicinal chemistry, the use of all available data and the application of AI for lead optimization or repurposing existing drugs come to mind.

Amassing and utilizing data

PharmTech: What data needs to be considered in selecting a high-potential drug candidate? In what ways can that data best be organized and utilized?

Short (Verseon): A great deal of clinical data on marketed drugs and drug candidates is publicly available. There are also some other published data on compounds from both successful and failed medicinal chemistry campaigns. Unfortunately, data from failed campaigns are relatively sparse because corporations have many disincentives to publish that information. Whether or not the volume of data from failed programs exceeds that for successful ones, most agree that all data—including data for failed programs—are valuable to advance the field. The complete data need to be aggregated and anonymized in a usefully available manner to provide incentives for publication and allay concerns that releasing the data will compromise intellectual property.

Smith (Optibrium): The types of data currently collected (such as activity, solubility, metabolic stability, etc.) will most likely not significantly change within the near future. Most early drug discovery programs start from a basic set of early assays and compound measurements designed to provide proof of principle that the idea under consideration has merit. A complicating factor is that the targeted therapeutic area will also influence the perceived importance of one type of data over another. So, it is not the type of data being generated, but instead, it is the quality and quantity of the early data that, if improved, should benefit future programs. Computational tools support drug discovery scientists to make effective decisions based on complex data. This helps to target high-quality compounds for synthesis and testing and quickly focus on the best compound for progression to later-stage studies. In particular, computational approaches that enable translational insights from early data to better predict more complex, later-stage data, such as PK/PD, *in vivo* efficacy, and safety outcomes, increase the success rate and speed of discovery.

A requirement is that data are stored in an easily accessible format. Data that are not reliably and easily accessible to computational processes represent a barrier to progress. Essentially, the data become siloed or walled off from deeper analysis, thus limiting their value and impact on the drug discovery process. Inaccessible data also impede potential collaborative efforts between programs, whether internal or external to their organization.

Artificial intelligence in medicinal chemistry

PharmTech: How can AI be utilized in the medicinal chemistry space?

Smith (Optibrium): There are three key applications of AI in the medicinal chemistry space: 1) prediction of compound properties, based on learning from existing compounds and data; 2) generative chemistry, which generates new compound ideas (structures) to consider in the context of a drug discovery project; and 3) reaction prediction, or retrosynthetic analysis, which can efficiently evaluate synthetic feasibility. Outside of drug discovery, a key development in AI has been deep learning methods that benefit from large, precise, and complete data sets (Big Data) to generate predictive models of unprecedented quality. In contrast, the available data in early drug discovery projects are orders of magnitude smaller than in other fields. Furthermore, drug discovery data are highly variable and sparse due to the fact that most compounds have only been measured in subsets of the full spectrum of assays. However, recent developments in AI methods enable them to generate insights even from these challenging data to highlight opportunities that may have been missed due to uncertain, missing, or inaccurate data. However, the expert medicinal chemist, with experience and knowledge of the chemistry and biology underlying a project, is ideally placed to make educated decisions that can also guide the AI engine. This is the concept of augmented intelligence, whereby human experts and AI algorithms combine to achieve the best outcomes.

Short (Version): As we collect much larger and more complex data sets from increasingly advanced functionally predictive assays, machine-learning tools could potentially make medicinal chemistry operations more efficient. There are other areas where AI could be useful. For example, functionally interrelated proteins (e.g., kinases, proteases, and G-protein-coupled receptors) and their structurally interrelated inhibitors lend themselves to analysis by AI tools. AI that examines complex structure-activity relationships can also aid in lead optimization. Other AI tools can help integrate the medicinal chemistry development cycle with process chemistry and manufacturing at scale. Although there is a great deal of exuberance around AI today, one significant unanswered question is whether AI will enable medicinal chemists to solve problems that would have been intractable otherwise. That remains to be seen.

About the author
Meg Rivers is a senior editor for *Pharmaceutical Technology* Group and *BioPharm International.*
Article details
Pharmaceutical Technology
Vol. 46, No. 4
April 2022
Pages: 21–22
Citation
When referring to this article, please cite it as M. Rivers, “Early Development Medicinal Chemistry: Utilizing Data and Artificial Intelligence,” *Pharmaceutical Technology*, 46 (4) 2022.