VIAL

Reimagining Drug Discovery: Innovations in Technology and AI

White Paper

Reimagining Drug Discovery: Innovations in Technology and AI

1. Introduction

The process of drug discovery has long since been a sporadic and laborious endeavor. On average, a new drug takes 10-15 years and US$1-2 billion before it is readily available in the clinic [1]. Even with the substantial amount of time and money invested in drug discovery, approximately 90% of targets never make it out of the laboratory [2]. While there are many reasons that a prospective drug fails preclinically, there is a significant need for a more efficient, accurate, and comprehensive process to identify, characterize, and test prospective molecules for drug discovery. The emergence of new technology and artificial intelligence (AI) may be the key to identifying new drugs more productively and with a higher success rate.

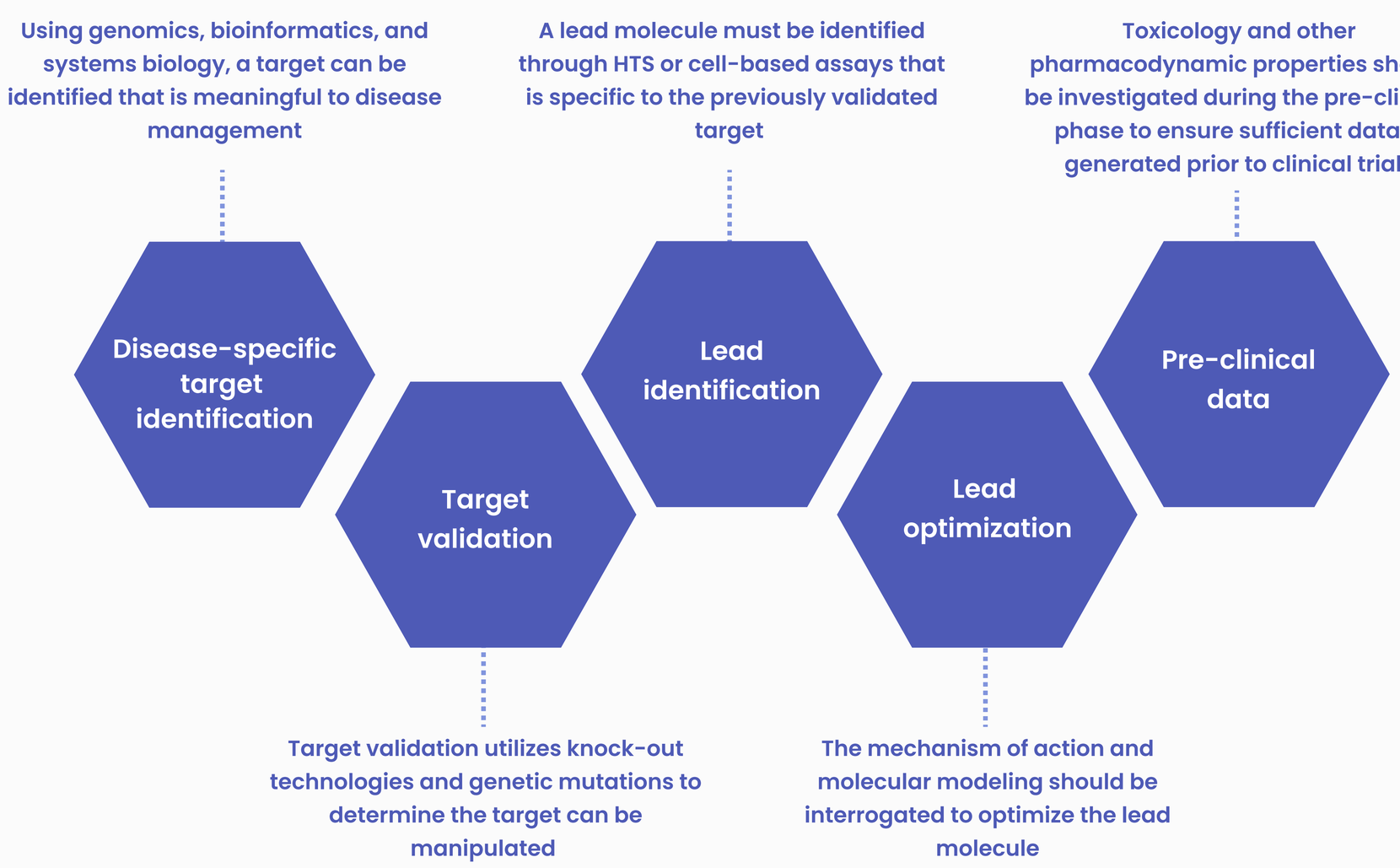
> **Figure 1: The large majority of prospective drug candidates fail before being investigated in a clinical trial. The small percentage of molecules that are deemed clinically relevant take over a decade and cost billions of dollars before they are approved for use.**

2. Traditional Drug Discovery Process

The effort in drug discovery has largely been focused on selecting the best lead drug candidate to achieve adequate clinical efficacy. Unfortunately, selecting the best molecule can oftentimes be ambiguous and inconclusive. When a disease is initially being characterized, oftentimes there can be a single protein or process that is attributed to its cause. The first step in drug discovery is identifying that disease-specific protein or process that would lead to meaningful management. Once identified, that malfunctioning protein or biological process becomes the target because, if corrected, the disease or symptoms could be minimized or even eradicated. In order to identify the target, genomics, bioinformatics, and systems biology are often utilized.

Next, the target must be validated to ensure that correction of the defect results in a normal or less severe biological phenotype [3]. Corrections can be simulated using knock-out technology or genetic mutations. By knocking out a protein in a 2D cell model or mouse model, researchers can understand how the biological model functions with the absence of a protein. For example, if a certain protein is hypothesized to drive cancer progression, removing that protein may decrease or even reverse tumor growth. Proteins can be knocked out using CRISPR-Cas9 technology or small interfering RNA (siRNA). It is ideal that the validation process generates data that suggests minimal change in normal cellular function while reversing the negative effects of the disease. For example, removing a central protein may inhibit a disease, but it also may inhibit vital processes that are necessary to sustain normal function. Therefore, there is a balance between identifying a protein that is easily targeted but also must be disease-specific.

Figure 2: Historically, the drug discovery process has consisted of trial-and-error with labor-intensive validation and optimization.



One of the most difficult aspects of drug discovery is lead identification. The target protein must have an appropriate set of characterizations that make it druggable. Unfortunately, up to 85% of all human proteins are considered undruggable, which means it is notoriously difficult to identify a molecule that can act upon a particular target [4]. In order to find a lead hit for a target protein, the molecular structure of the target likely needs to have a well-defined binding pocket or activation site, function independently of other proteins, or have a protein-specific motif that can be exploited. The difficulty arises when a target protein does not have one of these characteristics. Then, leads that can potentially bind to the target protein can be non-specific, unreliable, or only adhere under certain conditions.

Once an adequate lead molecule is identified, it must be modified to ensure it is optimized prior to preclinical testing.

Optimization techniques can be lead-, target-, or disease-specific but examples include minimally altering the molecular structure through chemical modifications, ensuring adequate localization by introducing a location-specific tag, adjusting the degradation mechanism, or expanding the half-life. However, there must be a balance during the optimization period, since it is important not to alter the molecule so much that the mechanism of action is changed.

The last phase of drug discovery is in the preclinical testing phase. Preclinical testing consists of mechanistic assays that identify how a potential drug interacts with its target and affects the organism as a whole. Similarly, efficacy and initial dosing experiments should be generated in a relevant disease model. Lastly, preclinical data should ensure the toxicology profile and pharmacodynamic properties of the molecule are sufficient to meet regulatory requirements. It is important that these data undergo the rigor and reproducibility that is necessary to gain approval of an Investigational New Drug (IND) application [5].

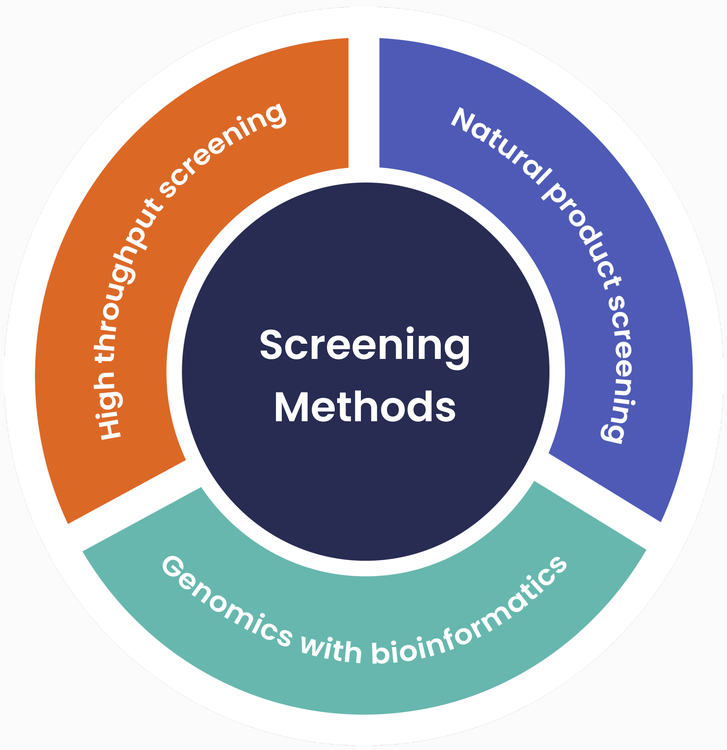
3. Screening to Identify a Lead Molecule

A large portion of the drug discovery process is spent identifying a lead molecule. Since there are, theoretically, an infinite number of potential molecules that need to be sifted through [6], understanding the most efficient screening processes is essential in developing the next blockbuster drug. One of the more recent screening techniques is high throughput screening (HTS) in which hits are identified from a compound library [7]. HTS consists of testing approximately 10,000 to 100,000 different compounds per day to evaluate the affinity and expected reaction of the compounds with the target molecule. These massive experiments are completed with robots that perform microscopic assays for each compound and collect data regarding the association between the screening compounds and the target molecule. Statistical analyses identify hits, and ultimately a lead molecule, that narrows down future tests.

The practice of screening natural products has also been used when beginning the search for alternative molecules that are not found in a biological system. Natural product libraries are comprised of plant extracts, marine invertebrates, and other microorganisms [9]. Introducing natural products creates a level of uncertainty given that they can be recognized as foreign to the body and lead to unexpected results. This level of uncertainty increases when using crude natural product samples since the purity of the sample is lower but is oftentimes done in early drug development stages to keep costs down. Once a natural product library is developed, it can be incorporated into HTS and tested against the target in question.

Another way to screen for potential lead hits is through utilizing bioinformatics with genomics, epigenetics, or proteomics [8]. By gleaning information from relevant databases, researchers can virtually screen for small molecules and identify pertinent protein-protein interactions. Since the characterization and sequencing of the entire human genome was completed, informatic experts have exploited that data to understand, theoretically, if a certain molecule would be ideal to bind to a specific target. However, the limitations of bioinformatics is that it can only utilize known, publicly available compounds. Similarly, the information gathered from these resources is only theoretical, therefore, significant testing must be done to validate the hypotheses developed using bioinformatic analyses.

Figure 3: Screening methods such as high throughput screening, natural product screening, and genomics with bioinformatics have traditionally been most often used to identify a target and lead molecules.



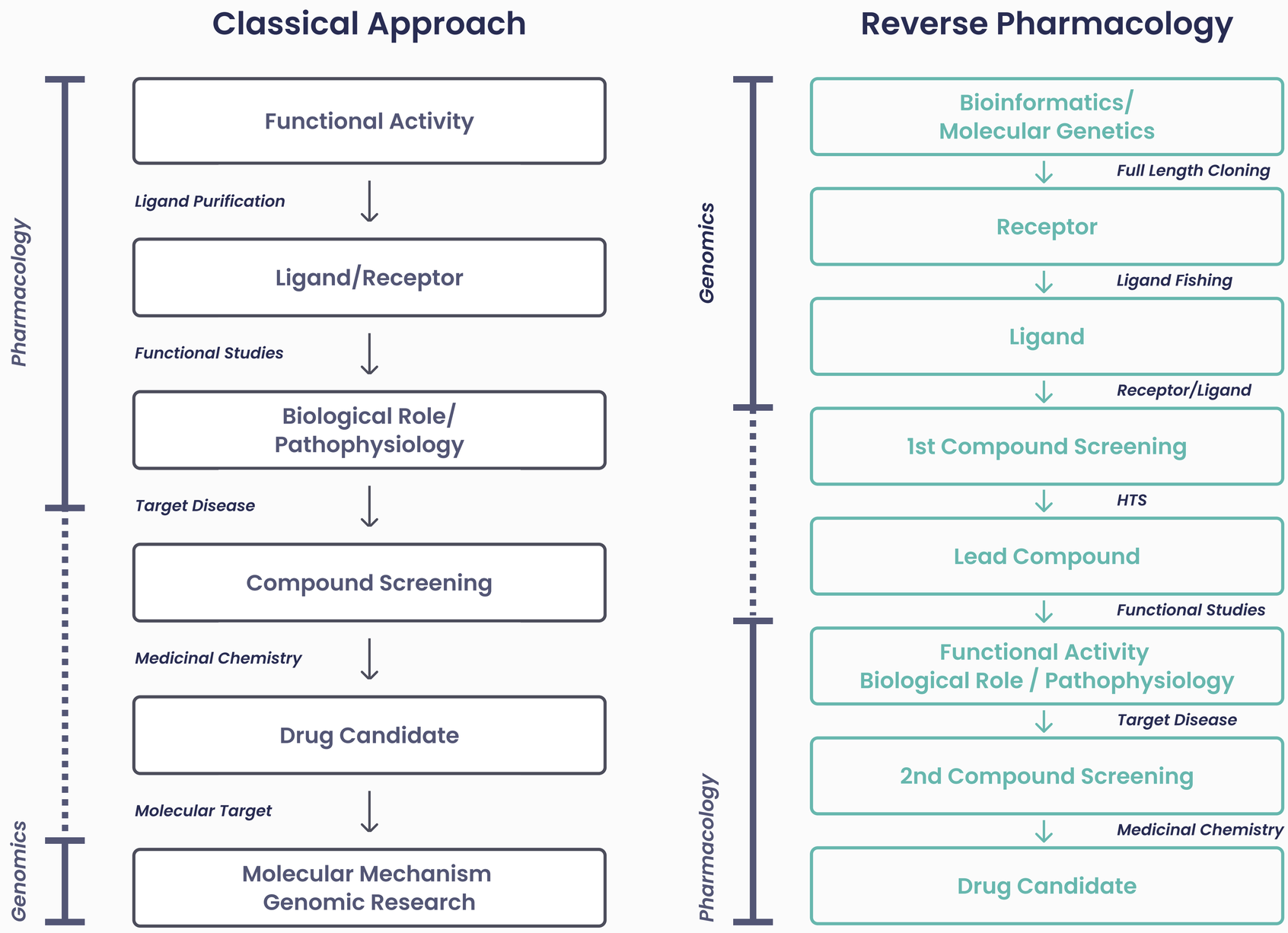
4. Lead Molecule Validation

Once a lead molecule has been selected, it is vital that it undergoes rigorous and reproducible pharmacological validation. First, the molecule must be able to be purified and have the ability to be manufactured reliably on a large scale. Then, with the purified molecule, the functional activity of the molecule must be determined. The biological role should be established in a relevant model such as an appropriate murine model. Toxicity studies should then be performed to ensure that there are not any significant off-target effects that cause a change in efficacy, mechanism of action, or safety.

In drug discovery, there are two types of pharmacology processes: the classical approach and the reverse approach [10]. The classical approach identifies the pharmacological activity and phenotypic profile of a molecule first and then utilizes medicinal chemistry and genomics to determine the molecular mechanism and a potential target that it could affect. This process can be beneficial because once a molecule is optimized, it can be introduced to the disease state more efficiently. However, this process can also take a much longer period of time because determining the biological target can be an ambiguous and haphazard process.

In contrast, reverse pharmacology is a more recent approach to drug development. In this process, the targets are identified first and then lead molecules that can interact with the target are investigated. While this can be a more direct approach, and take significantly less time, there are instances where the target is not easily druggable, making screening for lead molecules unproductive. For example, it was long since understood that KRAS, a protein that, when mutated, can lead to significant cancer formation and progression, was undruggable [11]. Due to its unusual shape and relatively smooth protein structure, designing inhibitors to bind to it was seemingly impossible. Although it was understood that KRAS was the root of the problem, researchers were forced to inhibit its various downstream and upstream effectors due to limitations in drug design.

Figure 4: Despite the advancements of pharmacology from the classical approach to the reverse approach, the process to identify a drug candidate remained time-consuming and costly.



Takenaka, T. (2008). Classical vs reverse pharmacology in drug discovery. BJU International, 88, 7–10. doi:10.1111/j.1464-410x.2001.00112.x

Validating a lead molecule can certainly be a drawn out process and there are multiple avenues that can ensure a lead is ready for preclinical testing. However, each method is quite antiquated and requires laborious and tedious work. Between screening, validating, and ultimately generating preclinical data, methods of drug discovery can be taxing and costly. Because of that, researchers have spent decades leveraging new technology to ease the drug discovery process and bring new and better therapies to the market faster.

5. AI-Driven Drug Discovery

With the rise of AI in the healthcare industry, many pharmaceutical and biotechnology companies are using the new-generation tool to enhance drug design and development. Initially, AI-driven drug discovery was proposed as a method to cut back the expenditures of traditional drug discovery and was met with significant skepticism. However, as the technology has developed and low-risk prototypes have become successful, that skepticism has waned and more corporations are not only utilizing AI in drug discovery, but contributing to its advancement.

AI has changed the game of drug discovery. There are several ways in which AI-based drug discovery differs from the traditional approach [12]. While traditional drug discovery is largely target-driven that is ideal for well-understood and druggable targets, AI drug discovery is data-driven with complex machine learning processes that can identify compounds that may bind to undruggable targets. Other advantages of using AI in the drug discovery process include reduced bias since AI does not rely on predetermined target or prior knowledge, a level playing field in drug development due to the rapid increase in publicly available processing power, and the higher predictive capabilities to streamline the screening process of prospective molecules. In contrast, it seems the only disadvantage of introducing AI to the drug discovery process is cost and access. However, as competition in the AI field rises, it is likely that those factors will be less of a barrier.

6. Biggest Players in AI-Driven Drug Discovery

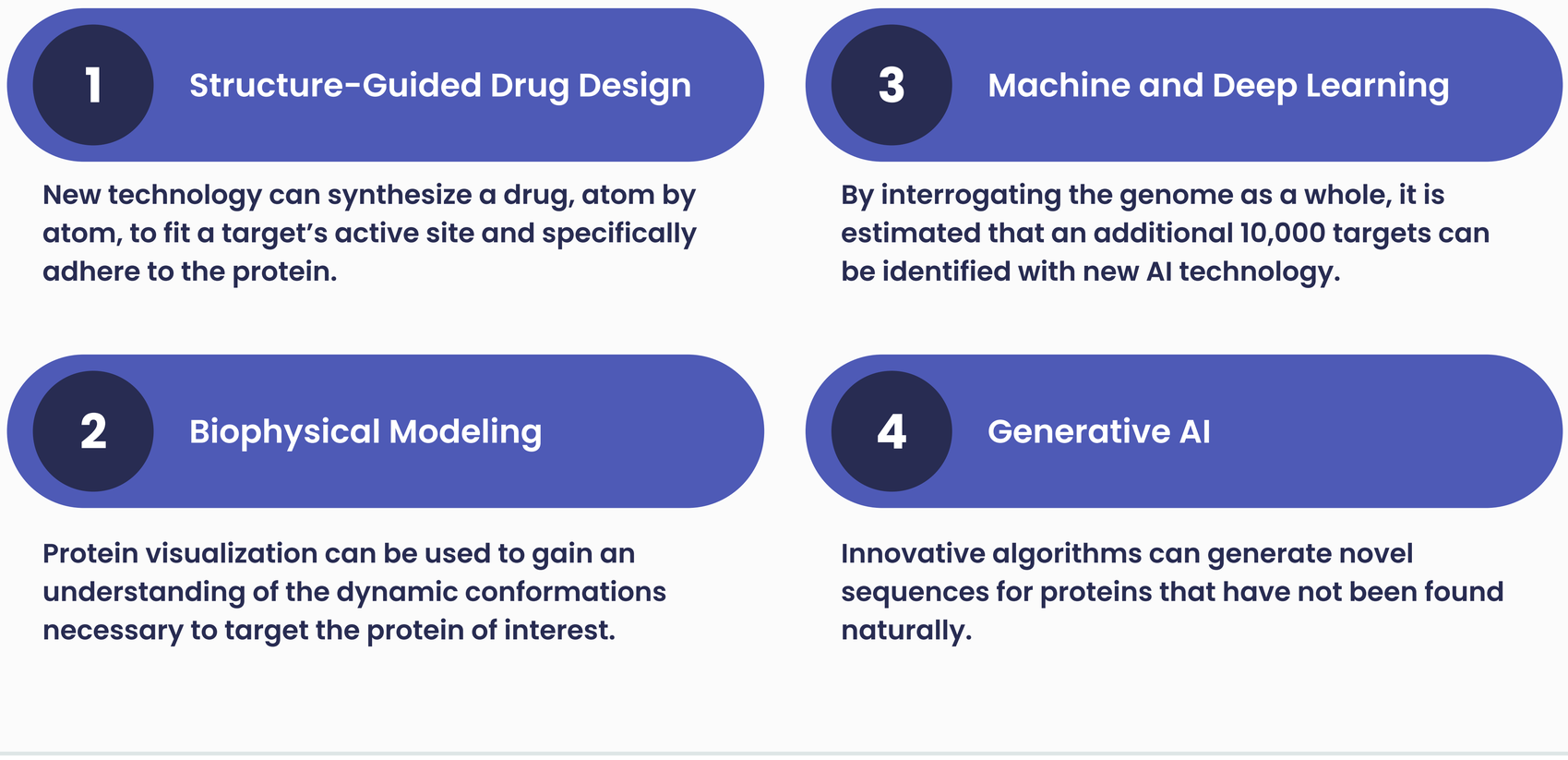
Several companies have hit the ground running in terms of AI-driven drug discovery and there are currently a few big players in the field that are on the cutting edge of this new technology. BioCryst® is a North Carolina-based commercial-stage biotech company that is committed to finding novel therapeutics for patients with rare and serious diseases. BioCryst® utilizes structure-guided drug design to increase selectivity and enable efficient, rapid drug development. Their technology involves studying the active site of an enzyme to design a molecule that will bind to it. Through this program, BioCryst® developed BCX10013, which targets Factor D in complement-mediated diseases. Not only did this compound show promising preclinical results, but recent data from an ongoing Phase I clinical trial showed that a single dose of BCX10013 resulted in >97% suppression of the alternative pathway of the complement system within 24 hours. With these encouraging data, BioCryst® plans to advance the BCX10013 compound into patient studies and ultimately a pivotal trial in the near future.

Another example of a larger scale AI-driven drug discovery process lies in the COVID Moonshot project [13]. This project is a consortium of the data generated to develop a SARS-CoV-2 protease inhibitor. After gaining contributions from several dozen researchers from around the globe, the COVID Moonshot project identified over 200 crystal structures of fragment-like and lead-like molecules and synthesized over 1000 compounds and characterized their corresponding activity. From these data, a covalent and non-covalent viral inhibitor was developed with an IC_{50} value under 150nM and viral inhibition detected under 5μM. While there has been significant characterization of this molecule, there are still considerable clinical gaps that remain. Studies to generate in-depth safety data and pharmacokinetic data are currently being conducted with the hopes of advancing the SARS-CoV-2 inhibitor to a clinical trial soon.

As mentioned previously, some targets have been dubbed simply ‘undruggable’ due to their structure or shape. The tech-enabled, AI-powered pharmaceutical company, Atomwise, is looking to change that. Atomwise has the goal of building a pipeline of small-molecule drug candidates, specifically for ‘hard-to-drug’ targets. They do this through tremendous computational efficiency that can screen trillions of compounds in silico using a global model with exquisite accuracy. Their approach does not require crystal structures or local ligand training data and instead, is constantly evolving their algorithms to generate more robust data. To date, Atomwise has developed several wholly-owned immunology drugs, co-developed several oncology, hematology, and immunology drugs, and is currently collaborating with over a dozen academic institutions throughout the world. Throughout these different programs, Atomwise hopes to move their most promising candidate drugs from the optimization phase into the preclinical and clinical phase quickly.

Lastly, there are companies leveraging the concept of generative AI. One of the pioneers of generative AI is Generate Biomedicines. Generate recognizes that the traditional drug discovery industry is grounded in trial-and-error and aims to provide novel therapies in a reliable, predictable, and instantaneous manner. They do this by training a platform on the entire compendium of protein structures and sequences found in nature. Then, machine learning algorithms analyze hundreds of millions of proteins and link statistical patterns to amino acid sequences, structures, and functions. Utilizing this technology, Generate developed an engineered resurfaced *E.Coli L-Asparaginase* variant that alleviates hypersensitivity responses in animal models [14]. In theory, by resurfacing proteins and harnessing Generate's novel technology, researchers may be able to reduce the effects of pre-existing immunity to protein-based therapies. Based on the success of this proof-of-concept study, Generate hopes to use computational design approaches to engineer life-saving drugs for diseases with limited therapeutic alternatives.

Figure 5: New technology that leverages artificial intelligence can lead to faster, cheaper, and more efficacious drug discovery.



7. Conclusion

Researchers are certainly at the forefront of a new era in drug discovery. In the past, is the traditional process for drug discovery that consisted of trial-and-error, laborious screening, and strenuous validation. And in the future, is the utilization of computational methods in AI-driven drug development, in which many biotechnology companies have already carved their niche. With more information and opportunities present, it is clear that drug development will soon become a more cost efficient and timely process and lead to more effective novel therapies than ever before.

References

1 Hinkson IV, Madej B, Stahlberg EA. Accelerating Therapeutics for Opportunities in Medicine: A Paradigm Shift in Drug Discovery. Frontiers in Pharmacology. 2020;11. doi:10.3389/fphar.2020.00770

2 2.Sun D, Gao W, Hu H, Zhou S. Why 90% of clinical drug development fails and how to improve it? Acta Pharmaceutica Sinica B. 2022;12(7):3049-3062. doi:10.1016/j.apsb.2022.02.002

3 Drews J. Biotechnology's metamorphosis into a drug discovery industry. Nature Biotechnology. 1998;16(S1):22-24. doi:10.1038/5405

4 *Yesylevskyy S. Undruggable targets: the major challenge for modern drug discovery. Receptor.AI. https://medium.com/receptor-ai/undruggable-targets-the-major-challenge-for-modern-drug-discovery-52a87ccb94f8. Published April 20, 2022. Accessed March 14, 2023.*

5 Reiner AT, Witwer KW, van Balkom BWM, et al. Concise Review: Developing Best-Practice Models for the Therapeutic Use of Extracellular Vesicles. Stem Cells Translational Medicine. 2017;6(8):1730-1739. doi:10.1002/sctm.17-0055

6 Artificial intelligence model finds potential drug molecules a thousand times faster. MIT News | Massachusetts Institute of Technology. Accessed March 16, 2023. https://news.mit.edu/2022/ai-model-finds-potentially-life-saving-drug-molecules-thousand-times-faster-0712

7 Mayr LM, Bojanic D. Novel trends in high-throughput screening. Current Opinion in Pharmacology. 2009;9(5):580-588. doi:10.1016/j.coph.2009.08.004

8 Xia X. Bioinformatics and Drug Discovery. Current Topics in Medicinal Chemistry. 2017;17(15):1709-1726. doi:10.2174/1568026617666161116143440

9 Wilson BAP, Thornburg CC, Henrich CJ, Grkovic T, O'Keefe BR. Creating and screening natural product libraries. Natural Product Reports. 2020;37(7):893-918. doi:10.1039/c9np00068b

10 Takenaka T. Classical vs reverse pharmacology in drug discovery. BJU International. 2001;88:7-10. doi:10.1111/j.1464-410x.2001.00112.x

11 The Lancet Oncology. Undruggable KRAS—time to rebrand? The Lancet Oncology. 2021;22(3):289. doi:10.1016/s1470-2045(21)00091-7

12 Gupta J. The power of Artificial Intelligence in drug discovery. Published June 2, 2021. Accessed March 21, 2023. https://indiaai.gov.in/article/the-power-of-artificial-intelligence-in-drug-discovery

13 Consortium TCM, Chodera J, Lee A, London N, Delft F von. COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning. American Chemical Society (ACS); 2020. Accessed March 21, 2023. http://dx.doi.org/10.26434/chemrxiv.13158218

14 Bootwala A, An HH, Franklin MW, et al. Protein re-surfacing of E. coli L-Asparaginase to evade pre-existing anti-drug antibodies and hypersensitivity responses. Frontiers in Immunology. 2022;13. doi:10.3389/fimmu.2022.1016179