AI-Driven Innovation in Protein Drug Development

By: Haile Teshome

December 3, 2023

This review aggregates the current developments in the role of Artificial Intelligence (AI) and Generative Biology in the domain of protein drug development. By leveraging AI and machine learning, scientists have been able to expedite the drug discovery, development, and understanding process by creating drugs that can target a specific domain with greater precision all while reducing the timeframe and cost for the development and discovery process. The focus of this review will be on the approaches used in a case study by Amgen in integrating AI into the drug development process but will also look at the broader scope of work that has been accomplished with an emphasis on the potential of these technologies to dramatically alter the standard pharmaceutical industry processes.

Artificial intelligence is the usage of statistical algorithms that are capable of learning complex tasks that are not preprogrammed and perform them just as a human would. Generative biology is the use of artificial intelligence and machine learning methods to better optimize the drug discovery and development process [5, 12]. With the advent of artificial intelligence's widespread application in numerous fields, the pharmaceutical industry has taken an interest in its capabilities by utilizing massive amounts of information for predictive and optimization purposes in drug development and discovery processes [11]. The drug development process is a multistage, time and resource-intensive, as well as rigorous process that typically spans over a decade on average [6,9]. The first stage of this process is the preclinical phase which encompasses in vitro and in vivo studies to evaluate a drug's safety profile, which typically involves screening anywhere from 5000 to 10000 compounds. If a compound successfully passes preclinical trials, the drug then enters a three-phase clinical testing

which progressively involves a larger number of participants, from 20-80 in Phase I, 100-300 in Phase II, and 1,000 to 3000 in Phase III, with phase IV being the public release and close monitoring [6,9]. Throughout these stages, only about 12% of drugs make it to market and on average 350 million dollars is spent per drug on development but there has also been up to 2.6 billion dollars spent [6,9,10]. The way artificial intelligence can integrate to improve this pipeline is using predictive modeling in protein shape determination which helps understand functionality, target discovery to best treat specific conditions, and the development of new lab techniques like simulating the human immune system to predict unwanted side effects earlier in the development process [11].

Generative biology has the potential to significantly accelerate the process of drug discovery by providing a platform that integrates artificial intelligence and machine learning with high-throughput lab-based science to maximize efforts on the therapeutic targets and drugs that have the highest likelihood of success [7,12]. Researchers can design proteins that are more suitable as drug candidates compared to those found naturally and through trial and error [9]. Machine learning has facilitated the creation of computer models capable of recognizing and learning from patterns in the data they are trained on that allow the prediction of protein structures to become more accurate than traditional laboratory methods [12]. Systems such as AlphaFold utilize deep learning that has a neural network architecture to train on large datasets of known protein structures from sources such as the PDB to predict the spatial arrangements of amino acid residues in a protein [8]. The system can then create a model that takes into consideration the most salient features in an amino acid sequence to correctly predict

the shape of a protein such as the distances between different amino acid pairs and the angles between bonds connecting [8]. This method has proven to be highly accurate as the CASP14 assessment showcases it had an accuracy of .96 Å rmsd whereas the next best method had a 2.8 Å rmsd (for reference, the average width of a carbon atom is approximately 1.4 Å) [8].

Additionally, target discovery is a very laborious task that could be significantly optimized with generative AI to find the best biochemical signature a drug can bind to. Traditional experimental methods used to do this include affinity-based biochemical profiling, genetic screening, and SILAC are used for target discovery, however AI-enabled multi-omic approaches that integrate diverse datasets from genomics, transcriptomics, proteomics, and metabolomics can provide superior results in significantly less time [1,2]. Additionally, other traditional methods such as pharmacophore screening and reverse docking can be augmented by machine learning to provide substantially more optimized approaches for predicting biological targets. AI and particularly deep learning have been able to process complex biomedical data, identify biomarkers, design drugs at the atomic and molecular level, and predict pharmacokinetics [2,3]. More recent methods have even created synthetic data to simulate biological scenarios [2,5].

To predict immune responses earlier in the drug development process, scientists have created 3D tissue models that emulate the human immune system. The research utilizes human tonsil and 3D bio-printed tissues to create immune tissue platforms, which can potentially test a drug's potential to invoke an immune response prior to the advancement to clinical trials [13]. This approach aims to generate data that can

enhance machine learning algorithms which can help potentially expedite the drug development process by focusing on specific drugs that showcase the best early signs of success by fostering the production of new antibodies while minimizing the cost of spending a lot of resources on drugs that don't showcase good data from simulated environments [11,13].

AI and generative biology hold the promise to revolutionize the field of protein drug discovery by improving efficiencies in the selection and testing of drug targets and candidates as well as increasing the velocity of the drug development and discovery process. Many challenges impede the progress of this technology from the availability of large quantities of high-quality data, available computational power using large graphics processing units (GPU) computing clusters and the extensive costs associated with them, as well as algorithmic bias and ethical considerations of handling sensitive electronic health records [5]. Despite these challenges, the integration of AI and generative biology in the field of drug discovery offers a fresh perspective on the process of developing protein drugs more efficiently. The capabilities of AI and especially generative AI, have showcased a creative approach to utilizing both traditional methods augmented with machine learning as well as completely new autonomous approaches such as those used by Amgen to synergize the productivity of AI and large-scale lab-based to not only significantly reduce the cost to develop drugs, and ultimately their cost to consumers, but also the number of drug candidates that are being approved as well as increasing the accessibility to medications for those with rare diseases where the financial incentive is not in line with doing research in their condition [10]. The next steps in this field are to continue growing its applications into more steps

throughout the drug development process including manufacturing and distribution, as well as continuously improving the models to become more accurate, scalable, cost-effective, and inclusive to keep pushing innovating medicines out and ultimately provide a high quality and individualized healthcare system [4].

Citations:

1. You, Y., Lai, X., Pan, Y. *et al.* Artificial intelligence in cancer target identification and drug discovery. *Sig Transduct Target Ther* 7, 156 (2022). https://doi.org/10.1038/s41392-022-00994-0

2. Pun FW, Ozerov IV, Zhavoronkov A. AI-powered therapeutic target discovery. Trends Pharmacol Sci. 2023 Sep;44(9):561-572. doi: 10.1016/j.tips.2023.06.010. Epub 2023 Jul 19. PMID: 37479540.

3. Jussi Paananen, Vittorio Fortino, An omics perspective on drug target discovery platforms, *Briefings in Bioinformatics*, Volume 21, Issue 6, November 2020, Pages 1937–1953, https://doi.org/10.1093/bib/bbz122

4. Dugger, S., Platt, A. & Goldstein, D. Drug development in the era of precision medicine.*Nat Rev Drug Discov* 17, 183–196 (2018). https://doi.org/10.1038/nrd.2017.226

5. D. Martinelli, et al. “Generative Machine Learning for de Novo Drug Discovery: A Systematic Review.” *Computers in Biology and Medicine*, Pergamon, 13 Mar. 2022, www.sciencedirect.com/science/article/pii/S0010482522001950?via%3Dihub.

6. Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discovery Today. 2021 Jan;26(1):80-93. doi: 10.1016/j.drudis.2020.10.010. Epub 2020 Oct 21. PMID: 33099022; PMCID: PMC7577280.

7. Qureshi R, Irfan M, Gondal TM, Khan S, Wu J, Hadi MU, Heymach J, Le X, Yan H, Alam T. AI in drug discovery and its clinical relevance. Heliyon. 2023 Jul;9(7):e17575. doi: 10.1016/j.heliyon.2023.e17575. Epub 2023 Jun 26. PMID: 37396052; PMCID: PMC10302550.

8. Jumper, J., Evans, R., Pritzel, A. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

9. Kraljevic S, Stambrook PJ, Pavelic K. Accelerating drug discovery. EMBO Rep. 2004 Sep;5(9):837-42. doi: 10.1038/sj.embor.7400236. PMID: 15470377; PMCID: PMC1299137.

10. Neiser, Julius, et al. “Biopharma’s Path to Value with Generative Ai.” *BCG Global*, BCG Global, 18 Oct. 2023, www.bcg.com/publications/2023/biopharma-path-to-value-with-generative-ai.

11. “How AI and Life Sciences Research Are Revolutionizing Protein Drug Development.” *Amgen*, wwwext.amgen.com/stories/2023/10/how-ai-and-life-sciences-research-are-revolutionizing-protein-drug-development. Accessed 4 Dec. 2023.

12. “Generative Biology: Designing Biologic Medicines with Greater Speed and Success.” *Amgen*, wwwext.amgen.com/stories/2022/06/generative-biology--designing-biologics-with-greater-speed-and-success. Accessed 4 Dec. 2023.

13. “3D Tissue Models Mimic the Human Immune System to Inform Drug Development.” *Amgen*, wwwext.amgen.com/stories/2023/02/3d-tissue-models-mimic-the-human-immune-system-to-inform-drug-development. Accessed 4 Dec. 2023.

Articles:

https://wwwext.amgen.com/stories/2023/10/how-ai-and-life-sciences-research-are-revolutionizing-protein-drug-development

https://wwwext.amgen.com/stories/2022/06/generative-biology--designing-biologics-with-greater-speed-and-success

https://wwwext.amgen.com/stories/2023/02/3d-tissue-models-mimic-the-human-immune-system-to-inform-drug-development