De Novo Molecular Design with Machine Intelligence

Gisbert Schneider

Molecular Design Laboratory, ETH Zurich, Vladimir-Prelog-Weg 4, 8093 Zurich, Switzerland.

Molecular design can be viewed as a constructive process rooted in pattern recognition (Figure 1). Medicinal chemists excel in visually recognizing chemical structures and their association with (retro)synthesis routes and specific molecular properties. In this context, various "artificial intelligence" (AI) methods have emerged as enabling technologies for drug discovery and automation. These systems aim to mimic chemists' pattern recognition capabilities and elevate them by considering domain-specific data during the molecule construction process [1]. Similarly, predicting the pharmacological activity and other properties of small molecules also benefits from AI methods. Machine learning models, particularly deep networks, ensemble methods, and hybrid approaches, have propelled the field forward by offering increasingly accurate qualitative and quantitative predictions [2, 3]. One of the key attractions of employing generative AI methods in drug design is their potential to develop data-driven, implicit model building processes capable of navigating vast datasets without explicitly enumerating the chemical search space. The ultimate challenge in drug design with AI is to implement methods that autonomously generate new chemical entities with desired properties from scratch ("de novo") [4]. This approach has the potential to significantly reduce the need for costly experimental compound screening to identify suitable candidate molecules.

We will present selected AI methods for de novo drug design, with a focus on approaches proven effective and reliable in scenarios with limited data availability. Prospective case studies from the field of drug discovery will be discussed, emphasizing chemical language models, graph neural networks, and their combination [5–8]. Furthermore, we will provide a critical assessment of the capabilities and limitations of each approach, while also daring to forecast the future of drug design with machine intelligence.

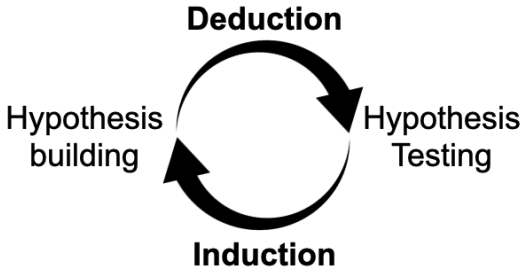

Figure 1. The molecular design process involves a dynamic interplay between inductive and deductive reasoning. Induction begins by gathering pertinent data relevant to the problem at hand. From empirical observations such as known molecules with desired properties, hypotheses (in this context, new molecules) are formulated, seeking to develop a theory that can account for these observed patterns. Deduction entails applying the theory to specific instances (hypothesis testing). The outcomes obtained from this deductive process inform the refinement or updating of the original hypothesis using the newly acquired data.

Bibliography:

[1] G. Schneider. Nat. Rev. Drug Discov. 17 (2018) 97–113.
[2] S. Allenspach, J. A. Hiss, G. Schneider. Nat. Mach. Intell. 6 (2024) 124–137.
[3] A. Tropsha, O. Isayev, A. Varnek, G. Schneider, A. Cherkasov, Nat. Rev. Drug Discov. 23 (2024) 141–155.
[4] G. Schneider, D. Clark. Angew. Chem. Int. Ed. 58 (2019) 10792–10803.
[5] K. Atz, L. Cotos Muñoz, C. Isert, M. Håkansson, D. Focht, M. Hilleke, D. F. Nippa, M. Iff, J. Ledergerber, C. C. G. Schiebroek, V. Romeo, J. A. Hiss, D. Merk, P. Schneider, B. Kuhn, U. Grether, U. G. Schneider. Nat. Commun. 15 (2024) doi: 10.1038/s41467-024-47613-w.
[6] M. Moret, I. Pachon Angona, L. Cotos Muñoz, S. Yan, K. Atz, C. Brunner, M. Baumgartner, F. Grisoni, G. Schneider. Nat. Commun. 14 (2023) 114.
[7] K. Atz, F. Grisoni, G. Schneider. Nat. Mach. Intell. 3 (2021) 1023–1032.
[8] D. F. Nippa, K. Atz, R. Hohler, A. T. Müller, A. Marx, C. Bartelmus, G. Wuitschik, I. Marzuoli, V. Jost, J. Wolfard, M. Binder, A. F. Stepan, D. B. Konrad, U. Grether, R. E. Martin, G. Schneider. Nat. Chem. 16 (2024) 239–248.