Mind and machine in drug design

Originally published in:
Nature Machine Intelligence 1, https://doi.org/10.1038/s42256-019-0030-7

Mind and machine in drug design

After a difficult start, medicinal chemists are now ready to embrace AI-based methods and concepts in drug discovery, explains Gisbert Schneider.

Gisbert Schneider

Commercial drug discovery faces a constantly decreasing return on investment. Among several challenges, this is due to increasing drug resistance and safety issues, and the desire of the industry to adjust to a new era of personalized healthcare, with shrinking target patient populations[1,2]. Only 10–14% of drug candidates entering clinical trials actually reach the market as medicines, with an estimated US$2–3 billion price tag for each new treatment[3]. Without question, there is a need for fresh thinking, new and revised conceptions of the drug discovery process, and innovative approaches to deliver higher quality drugs at a lower cost to market. Will artificial intelligence (AI) come to the rescue[4,5]? Already, a variety of machine learning methods have been used to design prototypical compounds with desired drug-like properties and bioactivities, and to search for solutions to various tough problems in medicinal chemistry[6–9]. While the immediate impact of these advances on drug approvals may be limited, there is proof of concept for early recognition of potential side-effects, successful drug repurposing, improved accuracy for drug property predictions, and the autonomous generation of drug candidates by machine intelligence.

In fact, we are witnessing a renaissance of AI technology in medicinal chemistry. The previous wave of excitement crested roughly 30 years ago with the first applications of neural networks, inference systems and other machine learning models[10–12]. After much debate, and possibly too little in the way of convincing applications, enthusiasm faded. Instead, automated experimental approaches like combinatorial chemistry and high-throughput compound screening became the primary drivers of the industry[13]. Now AI is back — this time, apparently, for good. The combination of laboratory automation and innovative software solutions for process planning and drug design promises better drugs, discovered and delivered faster[14,15]. Medicinal chemists are now exploring, adopting and adapting this technology. It is a question worth pondering what took them so long.

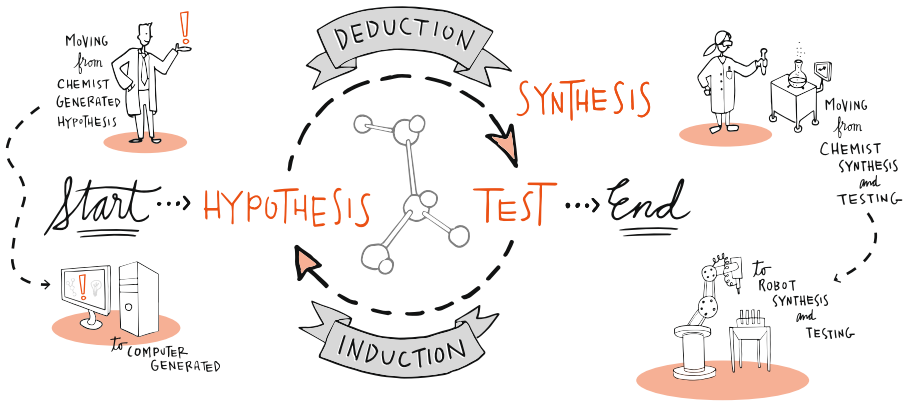

Fig. 1 | Cartoon representation of the drug design cycle as the interplay of inductive and deductive reasoning. The individual tasks along this process can be performed by humans or machines. Automating discovery processes with the aid of laboratory robots and artificial intelligence is a dynamic field of research and actively pursued in the biotech and pharmaceutical industries. Credit: Jack Burgess.

At the edge of chaos

To answer this question from a scientific vantage point, one must concede our incomplete understanding of the domain. The world of chemistry is largely based on explicit knowledge as encoded in written text, molecular models and the formulation of underlying physical principles. Although the rules of chemistry and physics are equally valid in medicinal chemistry, we face a particularly challenging situation in drug discovery that, to date, has precluded the possibility of exhaustive problem representation using chemical terminology. The systemic pharmacological effect of a drug is governed by poorly defined, and highly nonlinear, relationships between many contributing factors, leading to often unpredictable system behaviour. In other words, medicinal chemists work 'at the edge of chaos'[16].

Consequently, drug design is largely based on heuristic approaches, accompanied by decision making based on 'gut feeling'. The use of associative reasoning and pattern recognition for molecular structures, based on implicit knowledge, is commonplace. Prominent examples of such narratives are empirically derived guidelines for 'drug-likeness', and the notion of 'privileged' chemical (sub) structures[17]. It takes years of on-the-job training to become a knowledgeable expert in medicinal chemistry. Still, there remain huge differences of opinion as to which compound from any given series should become the next drug candidate.

Only when confronted with a solvable task will a learning algorithm find causative input–output relationships. Therefore, applying machine learning to drug design requires further thought regarding the definition of the problem domain. While certain properties of molecules can be automatically learned from basic (sometimes referred to as 'fundamental') molecular representations, such as the atom connectivity and certain quantum chemical properties, the physicochemical patterns of a drug molecule alone rarely account for the observed pharmacological effect in a simple fashion. Most drugs have multiple biological targets and activities, and their relative importance is highly dependent on the individual genetic profile of patients, the impact of formulation and administration for a given drug's bioavailability, and a range of other factors. In other words, in certain areas of drug design we are confronted with inherently ill-posed problems owing to unknown contributing factors and many-to-many nonlinear relationships. Beyond a

certain level, the human mind struggles to retain and process such complex networks of variables.

Design, make, test, repeat

Consequently, drug design is non-deterministic. The molecular design–make–test cycle essentially represents an adaptive stochastic search. It alternates between the generation of new chemical matter by hypothesis-based deductive reasoning and inductive (or abductive) extraction of insights from the available data (Fig. 1)[14,15]. Machine learning can support chemists in situations of uncertainty by means of unbiased pattern recognition and feature extraction from chemical and biological data. However, machine intelligence — like human intelligence — often builds on implicit knowledge representations that elude our immediate understanding ('black box' models). The inherent difficulty in explaining the behaviour of a nonlinear, multiparameter model in terms of the established explicit chemical vocabulary has restricted more widespread use of advanced machine learning approaches in drug design to date. With good reason, chemists wish to understand why a certain suggestion is made and how it translates into the design of new molecules, in a manner consistent with their own understanding of the problem. Therefore, it is unsurprising that the majority of existing machine learning models deployed in chemical pursuits rely on molecular representations with an immediate, tangible basis in the substructural hierarchies, functional groups, molecular graphs and computed physicochemical properties familiar to chemistry. The field will unquestionably benefit from quantifying the uncertainty of deep models[18]. At the same time, we need to develop innovative molecular models that are sufficiently flexible to allow machine learning algorithms to find more intricate causative input–output relationships. Understanding the nature of the relationship between the structures and functions of bioactive compounds is key to successful drug design. The applicability domain of graph-based molecular models in AI-assisted drug design may thus be limited. Molecular pattern recognition could benefit from 'low-level' theoretical conceptions and approximations to describe, explain and predict a particular domain of phenomena[19], for example molecule representations in terms of electron density and shape distributions. To quote Richard Feynman in this context, "There's plenty of room at the bottom".

Toward the virtual chemist

Organic synthesis remains a rate-limiting factor in drug discovery projects[20]. Synthesis planning and chemical reactivity prediction, fast and accurate calculation of binding energies, and the de novo design of molecules with desired properties are examples of notoriously hard problems that might be addressable by a chemistry-savvy AI. The availability of representative training data, affordable high-performance computer hardware, and free access to software libraries has facilitated the straightforward application of machine learning algorithms in tackling these issues[21–23]. However, the decisive factors for the success of AI in drug design will be the ethos, attitude and willingness of chemists to apply these computational models and autonomous robots in their own research projects.

Importantly, the applicability of machine intelligence exceeds data analysis and model learning. Generative machine learning, which aims to find the distribution in the training dataset to generate new samples, can take the role of a chemist not only in the formulation of testable hypotheses, but also in the creative aspect, in the assembly of innovative molecules[24]. Prominent applications of generative models, which compose music or create new cooking recipes for today's dinner, have inspired chemoinformaticians to adapt these underlying methodological frameworks to chemistry, for example with recurrent neural networks, variational autoencoders and generative adversarial networks[25,26]. Admittedly, these current realizations of somewhat antiquated ideas are baby steps considering the breadth and depth of the machine learning arsenal that is available today, and have already been applied to good effect in other fields, specifically in image and speech recognition and modelling. Many of the underlying methodological concepts that are productively applied in drug design today had their first heyday in the 1990s, for example autoencoders and adaptive deep networks[27–29]. Fruitful crosstalk between expert medicinal chemists and computer scientists will therefore be warmly welcome, because only subsequent prospective testing will decide whether the contemporary design concepts are practical and sustainable for their intended purpose. After all, there is no learning without feedback, and shortening the timelines for the delivery of these critical data will be key for sustainable drug design, as powered by machine intelligence[14,30].

Integrating domain-specific machine intelligence in the pharmaceutical industry is the litmus [*sic*] test for AI in healthcare. The challenge is not only to generate novel drug candidates that are — practically — optimal in terms of their pharmacokinetic and -dynamic properties; the integration of AI will also require significant investment of time, money and the reorganization of laboratory structures and discovery processes. In consequence, the envisaged automated drug design engine may not only imitate but exceed human decision making as a core aspect of the drug discovery process. If successful in the long run, the approach will combine a continuously learning, chemistry-savvy AI with the synthesis and testing of pharmacologically relevant chemical matter. While many medicinal chemists are prepared for this transition, one should bear in mind that machine learning per se is neither a quick fix for the problems of the industry, nor does it provide immediate answers to the underlying scientific questions. We should be aware of the limitations of AI when it comes to modelling human cognition and avoid repeating the mistakes of the past[31]. Partial predictability in drug design is inevitable as a consequence of incomplete domain representation, and the fact that we are interfering with living organisms. Therefore, it would be wise not to place all one's eggs in the machine learning basket, but to expect successful, creative solutions from the collaborative efforts of human experts, process automation and advanced computer-assisted decision making[32,33]. Academic institutions and not-for-profit organizations can offer the necessary leverage, and leeway, to explore unconventional thinking and challenge machine intelligence models to generate novel drug candidates. The prospects for this mixture of machine and mind in drug discovery are huge. ❐

Gisbert Schneider

Swiss Federal Institute of Technology, Department of Chemistry and Applied Biosciences, RETHINK, Zurich, Switzerland.
e-mail: gisbert@ethz.ch

Published online: 25 February 2019
https://doi.org/10.1038/s42256-019-0030-7

References

1. Lee, V. H. *Ther. Deliv.* **1**, 615–619 (2010).
2. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. *Nat. Rev. Drug Discov.* **11**, 191–200 (2012).
3. Smietana, K., Siatkowski, M. & Møller, M. *Nat. Rev. Drug Discov.* **15**, 379–380 (2016).
4. Mignani, S., Huber, S., Tomás, H., Rodrigues, J. & Majoral, J. P. *Drug Discov. Today* **21**, 239–249 (2016).
5. Jordan, A. M. *ACS Med. Chem. Lett.* **9**, 1150–1152 (2018).
6. Segler, M. H. S., Preuss, M. & Waller, M. P. *Nature* **555**, 604–610 (2018).
7. Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. *Nat. Commun.* **8**, 13890 (2017).
8. Gupta, A. et al. *Mol. Inf.* **37**, 1700111 (2018).
9. Merk, D., Grisoni, F., Friedrich, L. & Schneider, G. *Commun. Chem.* **1**, 68 (2018).
10. Pierce, T. H. & Hohne, B. A. (eds) *Artificial Intelligence Applications in Chemistry* (American Chemical Society, Washington, 1986).

11. Gasteiger, J. & Zupan, J. *Angew. Chem. Int. Ed.* **32**, 503–527 (1993).
12. Walters, W. P., Green, J., Weiss, J. R. & Murcko, M. *J. Med. Chem.* **54**, 6405–6416 (2011).
13. Sanderson, K. *Nat. Rev. Drug Discov.* **14**, 299–300 (2015).
14. Schneider, G. *Nat. Rev. Drug Discov.* **17**, 97–113 (2018).
15. King, R. D. et al. *Science* **324**, 85–89 (2009).
16. Schneider, P. & Schneider, G. *J. Med. Chem.* **59**, 4077–4086 (2016).
17. Leeson, P. D. & Springthorpe, B. *Nat. Rev. Drug Discov.* **6**, 881–890 (2007).
18. Begoli, E., Bhattacharya, T. & Kusnezov, D. *Nat. Mach. Intell.* **1**, 20–23 (2019).
19. Seifert, V. A. *Found. Chem.* **19**, 209–222 (2017).
20. Blakemore, D. C. et al. *Nat. Chem.* **10**, 383–394 (2018).
21. Merk, D., Friedrich, L., Grisoni, F. & Schneider, G. *Mol. Inf.* **37**, 1700153 (2018).
22. Coley, C. W., Green, W. H. & Jensen, K. F. *Acc. Chem. Res.* **51**, 1281–1289 (2018).
23. Smith, J. S., Isayev, O. & Roitberg, A. E. *Chem. Sci.* **8**, 3192–3203 (2017).
24. Schneider, G. *Mol. Inf.* **37**, 1880131 (2018).
25. Hessler, G. & Baringhaus, K. H. *Molecules* **23**, 2520 (2018).
26. Mak, K.-K. & Pichika, M. R. Artificial intelligence in drug development: present status and future prospects. *Drug Discov. Today*. https://doi.org/10.1016/j.drudis.2018.11.014 (2019).
27. Devillers, J. (ed.) *Neural Networks in QSAR and Drug Design* (Academic, Cambridge, 1996).
28. Lohmann, R., Schneider, G. & Wrede, P. *Biopolymers* **38**, 13–29 (1996).
29. Schneider, G. & Wrede, P. *Prog. Biophys. Mol. Biol.* **70**, 175–222 (1998).
30. Plowright, A. T. et al. *Drug Discov. Today* **17**, 56–62 (2012).
31. Graubard, S. R. (ed.). *The Artificial Intelligence Debate: False Starts, Real Foundations* (MIT Press, Cambridge, 1988).
32. Topol, E. *J. Nat. Med.* **25**, 44–56 (2019).
33. Goldberg, K. *Nat. Mach. Intell.* **1**, 2–4 (2019).

Acknowledgements

I am grateful to the many colleagues, co-workers and students with whom I had the privilege to work on drug design with machine intelligence. This work was financially supported by the Novartis Forschungsstiftung (FreeNovation: AI in Drug Discovery) and the Swiss National Science Foundation (grant no. 205321_182176 to G.S.).

Competing interests

The author declares a potential financial conflict of interest as consultant to the pharmaceutical industry and co-founder of inSili.com GmbH, Zurich.