Conclusions

Drug design and development will continue to be an early adopter of new and evolving technologies, both experimental and computational. Among the challenges is whether to apply these technologies to enhance the existing pipeline and processes or is the real opportunity to re-engineer the processes in light of these technologies? Big data, digital healthcare, remote monitoring, and genomics will drive the need to explore how computational and reasoning approaches can be applied to enhance the process both in terms of clinical significance and cost reduction. Artificial intelligence methods hold great promise towards these goals but their success will depend on aligning the right question with the right technology. 🍷

References:

1. A Complete History of Artificial Intelligence, May 25, 2021 Rebecca Reynoso, <https://www.g2.com/articles/history-of-artificial-intelligence>
2. Artificial intelligence in chemistry and drug design, Nathan Brown *et al*, *Journal of Computer-Aided Molecular Design* (2020) 34:709-715; <https://doi.org/10.1007/s10822-020-00317-x>
3. Artificial intelligence in drug discovery and development, Debleena Paul *et al*, *Drug Discov Today*. Jan 2021 Jan; 26(1): 80-93; <https://doi.org/10.1016/j.drudis.2020.10.010>
4. History of artificial intelligence in medicine, Vivek Kau, Sarah Enslin, and Seth A.Gross, *Gastrointestinal Endoscopy*, October 2020, 92(4): 807-815; <https://doi.org/10.1016/j.gie.2020.06.040>

Key points to succeed in Artificial Intelligence drug discovery projects

by Quentin Perron, Vinicius Barros Ribeiro da Silva, Brian Atwood, and Yann Gaston-Mathé

Drug discovery and development is an expensive, complex, and time-consuming task [5]. Recently, the development of artificial intelligence (AI) approaches to drug discovery, specifically *de novo* drug design through the use of deep generative models, has triggered a lot of interest in the drug hunter community, especially as an important tool to speed up the process [6].

Since 2017, we at Iktos have worked in collaboration with industry and academia in many different projects,

1980: WABOT-2 was built at Waseda University. This inception of the WABOT allowed the humanoid to communicate with people as well as read musical scores and play music on an electronic organ. (Reproduced from [1])

from hit discovery to lead optimization, using AI with ligand and structure-based techniques. In a recently published preprint work, we described the results of a successful collaboration between Iktos and Servier in a late-stage lead optimization project [7]. At this occasion we described, for the first time, the successful application of deep learning to *de novo* design for solving a Multi-Parameter Optimization (MPO) issue in an actual drug discovery project. Using the initial dataset of the project, with 881 molecules measured on 11 biological assays, we built 11 QSAR models and used them in combination with our deep learning-based AI *de novo* design algorithm. We were able to automatically generate 150 virtual compounds predicted as active on all 11 objectives. 20 molecules were selected as the most promising, and 11 were synthesized and tested. Interestingly, the 11 AI-designed compounds that were synthesized and tested displayed functional groups that were either rare in the initial dataset or never tried earlier in the project. Ultimately, one of the 11 AI-designed molecules met all the objectives of the project at the same time, suggesting that this method can propose innovative new molecules to solve MPO

problems, via its ability to identify favorable modifications, even with few data points to learn from [7].

With our experience in this and around 40 other projects completed or in progress here at Iktos, we have been able to identify some key points for a successful application of AI in drug discovery projects. Here they are:

1) One key thing to have in mind, AI needs data to feed upon, and the higher the quality of the input data, the higher the probability of obtaining good results. Since the beginning of the QSAR era with Hansch and Fujita [8], we know that the quality of the input data is one of the major requirements for obtaining good results, and with AI it is no different. Be sure you are collecting and storing the highest quality data you can. Good data produces trustable models. We are often asked: how many data points do you need to apply your AI technology to a given project. There is no definitive answer to that question (it varies from case to case, depending on the diversity or consistency and the level of “contrast” in the initial data set), but one thing is certain, the more, and the higher the quality, the better. Sometimes we are asked whether AI can help for undruggable targets with no known 3D structure, binding site or ligands. Clearly, this is a hard case for AI, as for anybody else! Conversely, AI can definitely help getting the best out of existing data and accelerate the optimization process.

2) Know when to trust your models, and when not to trust your models. Sometimes even with good data, models are not able to correctly predict the activity of molecules. Models built to predict the pIC₅₀ of pyrrole molecules in a given target may not be the best to correctly predict the pIC₅₀ of pyrimidine molecules in the same target. In other words, understand your applicability domain (AD) [9], an important tool for careful use of QSAR models. The AD for a chemical space is a theoretical region comprising both the model descriptors and modeled response which allows estimating the uncertainty in the prediction of a particular compound based on how similar it is to the training compounds used to build the model [9].

3) Synthetic accessibility is key to testing hypotheses. Even good models, built on quality data, respecting the AD, can make mistakes. Ultimately, they are just models. Thus, while selecting the molecules after an AI-generation, it is important to focus on the molecules with the easiest synthetic routes. In the end, the more molecules you are able to synthesize and test, the higher the probability you will identify active compounds.

4) User experience is key to allowing chemists the opportunity to take advantage of promising new technologies. Techniques such as docking, external (commercial) models, retrosynthesis software (i.e. our retrosynthesis technology <https://spaya.ai>), are frequently used to (re)score the molecules. An intuitive user experience is critical to making sure these techniques are used. Combined with a complex IT infrastructure to run everything efficiently and effectively, the barrier to using new technologies should be lowered as much as possible to ensure adoption.

5) Collaborate! Designing new drugs is a complex task and requires different capabilities like medicinal chemistry, machine learning, computational chemistry and more recently AI. Being able to make all those expertises collaborate efficiently and smoothly is challenging. Having access to a collaborative platform is key for success. Indeed, chemists are the best person to set up a generative AI because of their knowledge of the project and SAR (Structure Activity Relationship) and computational chemists are the best to build models. At Iktos we have built Makya (makya.ai) a platform which allows chemists and computational chemists to make the most of AI by generating easy to make molecules focusing on the team's expectations in a very simple and efficient manner.

6) Novelty comes with risk! It is important to balance the desire for new and novel ideas while understanding the limitations of the AD. In the previous example, where models constructed using pyrrole molecules may not correctly predict the activity of pyrimidine molecules, this balance is key. You might be interested in ring systems other than the pyrrole, but you'll need to acknowledge that the models have less predictive power outside of the AD.

7) Iterate! Success in getting the perfect molecule may not come at once, despite AI. It is often needed to run several AI-enabled Design Make Test (DMT) cycles before converging to optimized molecules meeting the project's success criteria. This comes as a consequence of the points raised above: the models are usually not perfect, especially if you are interested in generating novelty, but hopefully they will get better as you enrich them with new data points generated through an AI-guided process. What we have seen in our most recent successful experiences is that AI often enables substantial acceleration in improving the overall Multi-Parametric Optimization (MPO) profile of the molecules, over several DMT cycles, compared to standard human-driven approaches. But this requires trusting the technology over several iterations, rather than expecting a perfect solution at the first shot.

8) **There is no magic in AI.** Sometimes there is no solution in the project's chemical space. Instead of identifying the solution, in some cases with our AI technology, we were able to help our clients conclude that, according to our models, no molecule will meet all the expected criteria at the same time. This insight has enabled them to make the decision to stop the project (or revert to a more early stage back-up series) and focus resources elsewhere.

We are witnessing the beginning of AI being applied to drug discovery. The technologies being developed have a high potential, but they require significant effort to generate value in real-life settings. Have these key points in mind while collaborating on a drug discovery AI project will help you to avoid pitfalls and increase your rate of success.

References:

5. Steve Morgan, *et al.* The cost of drug development: a systematic review. *Health Policy*. 2011, 100(1): 4-17; <https://doi.org/10.1016/j.healthpol.2010.12.002>
6. Petra Schneider, *et al.* Rethinking drug design in the artificial intelligence era. *Nat. Rev. Drug. Discov.* 2020, 19, 353-364; <https://doi.org/10.1038/s41573-019-0050-3>
7. Quentin Perron, *et al.* Deep Generative Models for Ligand-based *de Novo* Design Applied to Multi-parametric Optimization. *ChemRxiv*. 2021, <https://doi.org/10.26434/chemrxiv.13622417.v2>
8. C. Hansch, P. Maloney, T. Fujita, and R. Muir. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature*. 1962, 194, 178-180; <https://doi.org/10.1038/194178b0>
9. Kunal Roy, Supratik Kar, and Pravin Ambure. On a simple approach for determining applicability domain of QSAR models. *Chemom. Intell. Lab. Syst.* 2015, 145, 22-29; <https://doi.org/10.1016/j.chemolab.2015.04.013>
10. Kit-Kay Mak and Mallikarjuna Rao Pichika. Artificial intelligence in drug development: present status and future prospects. *Drug Discov Today*. 2019, 24(3), 773-780; <https://doi.org/10.1016/j.drudis.2018.11.014>
11. A. Bender and I. Cortés-Ciriano, Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. *Drug Discov Today*. 2021, 26(2), 511-524; <https://doi.org/10.1016/j.drudis.2020.12.009>

<yann.gaston.mathe@iktos.com>; Iktos, <https://iktos.ai/>

*Drug Discovery and Development Subcommittee Newsletter

The Drug Discovery and Development Subcommittee (D3SC) is part of IUPAC Division VII—Chemistry and Human Health and the SC mission is to facilitate the understanding and public awareness of the topic of Drug Discovery and Development and its international impact with an emphasis on medicinal chemistry.

The D3SC NEWSLETTER is a web-based offering complementary to IUPAC's *Chemistry International* and focusing on advances in Drug Discovery & Development Chemistry. In addition to

posting current information, we will invite expert's opinions to introduce and/or critically comment on new technologies, therapeutic modalities or new approaches in Drug Discovery and Development Chemistry. These brief articles will focus on the key advantages, current status of maturity and its impact on Drug Discovery & Development. In addition, the D3SC Newsletter will highlight news from IUPAC Division VII and its Subcommittees and alerts on conferences and key publications in the field.

The first edition was released in Sep 2021 and covered the topic of Artificial Intelligence in Drug Discovery & Development and solicited the present feature. <https://iupac.org/d3sc-newsletter-sep2021/>