



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## A new paradigm for drug development

In January, 2020, British pharmaceutical company Exscientia announced that its compound DSP-1181, which is intended to treat obsessive compulsive disorder, was starting a phase 1 clinical trial. DSP-1181 was created using an artificial intelligence (AI) platform to trawl through chemical libraries and identify the most relevant compounds. Reportedly, DSP-1181 is the first such drug to enter clinical trials.

Exscientia, which developed DSP-1181 in partnership with Japan's Sumitomo Dainippon Pharma, noted that it had taken less than 12 months from initial screening to the end of preclinical testing. The global industry-wide average is 4–6 years, with 1 in 1000 screened molecules progressing to clinical trials. Still, there is no guarantee that DSP-1181 will obtain regulatory approval. 90% of compounds that start phase 1 trials do not end up on the market. A 2018 analysis of the 12 leading biopharmaceutical companies by Deloitte found that each firm spends an average of US \$2.168 billion on research and development for every drug that comes to market, roughly one-third of which goes on drug development. It takes 10–12 years from discovering a molecule to launching it as a medicine. The hope is for AI to accelerate the development process and drive down costs. This change would also reduce the barriers to entry, which could lead to an expansion of the industry and an increase in the drug pipeline.

AI models can find patterns in vast quantities of data from different sources, far beyond the capacity of humans. Proponents of AI-driven drug discovery believe that such approaches can find targets, discover and optimise potential treatments, and even design drug candidates from scratch. The possibility also exists for improving the likelihood of successful preclinical

testing by using AI to identify the most appropriate animal model for a given disease. According to a 2019 analysis by Deloitte, 40% of drug discovery start-ups use AI to screen chemical libraries to find new candidates, 28% use AI to find new targets, and 17% use it for *de novo* drug design.

"The time is right [for AI] because of the amount of data we are collecting at the patient level", said Alix Lacoste, vice-president of data science at BenevolentAI, a London-based biotechnology start-up founded in 2013. She believes that AI is primed to take advantage of the increasing availability of a wide range of data, such as electronic health records, and omics data. As more and more datasets are released, the biology of different segments of the population might start to clarify. For example, the data that emerges from national biobanks could help define the prevalence of genetic predispositions to particular diseases—which could allow researchers to better understand why some drugs fail in a subset of patients.

In February, 2020, a paper in *Cell* outlined how a deep learning model had been used to discover a new potential antibiotic. The researchers used a dataset of 2335 unique compounds to train a neural network to find molecules that stop *Escherichia coli* growth. They then put the model to work on a library of 6111 molecules under investigation for human diseases, with the aim of identifying compounds that target *E. Coli*. The model correctly predicted antibacterial activity in 51 compounds, based on their structure. One compound was especially promising. c-Jun N-terminal kinase inhibitor SU3327 proved deadly to *E. Coli*. In mouse models, the compound was effective against *Clostridium difficile* and pan-resistant *Acinetobacter baumannii* infections. The investigators renamed the compound

"halicin" after HAL, the fictional AI system in *Space Odyssey 2001* (1968).

AI approaches to drug discovery centre on the ability of the system to predict how well a drug will bind to its target. Although DSP-1181 was discovered by AI, its target is well established for obsessive compulsive disorder. Halicin's structure is unlike conventional antibiotics, which hints at the possibility of AI models opening up entirely new classes of drugs. The research in *Cell* also described screening of additional databases containing more than 100 million molecules. The AI discovered two further compounds with broad-spectrum antibacterial activity, including against drug-resistant strains of *E. Coli*.

James Collins (Broad Institute of MIT and Harvard, Cambridge, MA, USA) co-author of the *Cell* paper points out that the advent of deep neural networks, fed by huge datasets and increasingly sophisticated algorithms, has encouraged experts in the field to turn their attention to drug discovery. The authors of the *Cell* paper stated that "novel approaches to antibiotic discovery are needed to increase the rate at which new antibiotics are identified and simultaneously decrease the associated cost of early lead discovery". Researchers have spent many years attempting to hone the ability of computer systems to predict the properties of molecules. Developments in molecular modelling have raised hopes that the field is on the verge of significant progress. "We are beginning to have the opportunity to influence the paradigm of drug discovery", noted the *Cell* paper.

Crucially, this reported AI model is generalisable. The neural network developed by Collins' team could be applied to searching for drugs for other diseases. "We have so much data these days, we can go almost anywhere", said Collins. "We are currently using our deep learning platform to discover



For the 2018 analysis of 12 leading biopharmaceutical companies see <https://www.deloitte.com/content/dam/Deloitte/uk/Documents/life-sciences-health-care/deloitte-uk-measuring-return-on-pharma-innovation-report-2018.pdf>

For the 2019 analysis of drug discovery see [https://www2.deloitte.com/content/dam/insights/us/articles/32961\\_intelligent-drug-discovery/DL\\_Intelligent-Drug-Discovery.pdf](https://www2.deloitte.com/content/dam/insights/us/articles/32961_intelligent-drug-discovery/DL_Intelligent-Drug-Discovery.pdf)

For the *Cell* paper see [https://www.cell.com/cell/pdf/S0092-8674\(20\)30102-1.pdf](https://www.cell.com/cell/pdf/S0092-8674(20)30102-1.pdf)

novel narrow-spectrum antibiotics as well as design *de novo* novel antibiotics. We are also harnessing our platform in attempts to discover antiviral therapeutics for treating COVID-19." Models can be deployed in very specific circumstances—for example, in a quest to develop the antibiotics classified by The World Health Organization as most urgently needed.

For machine learning to be successful, much depends on the quality and breadth of the training sets. "It is very important for the systems to learn from the hypotheses", said Lacoste. "One of the greatest limits of AI is the data that is fed into the algorithms". Lacoste notes that the field is moving towards standardised training sets from which everyone can learn. "Where machine learning models have been used in other fields, they have used well-established training sets, and that is when great advances have been made", Lacoste told *The Lancet Digital Health*. "We are still at the beginning of that process, but the pharmaceutical industry is coming together, understanding the need for those sets and forming consortia to make progress".

Collins agrees. "The larger and more diverse training sets that are put to work, the more our capabilities

will expand", he said. "We want to eliminate the silos as much as possible". Nonetheless, drug discovery poses particular challenges. "The successes are few and they are very far in the future", explains Lacoste. "So you cannot do testing in the same way as you would if you were labelling images or pictures of animals".

Pharmaceutical companies are accustomed to forming partnerships. But to unlock the full potential of AI, academic specialties will also have to get used to working together. "These approaches rely critically on experimental systems to generate the data and to test the predictions of the modelling", explains Collins. "We need to create teams that span machine learning, chemistry, and biology; I do not think we have seen those kind of integrated efforts in academia, but given the promise of such approaches, we have to get there".

The issue of interpretability is likely to become increasingly intractable as AI models grow in sophistication. It is not always straightforward to explain how or why a system reaches a particular conclusion. If the right conclusion is made for the wrong reasons, then applying the model in a different situation will be problematic.

For example, a model trained to spot dangerous lesions using scans taken from a hospital with an unusually high mortality rate might start to codify images based on their source, rather than clinical features.

In *Space Odyssey 2001*, HAL eventually decided that he could do without his human co-workers. But we are long way from a fully automated drug discovery process. "You have to use machine learning in conjunction with the expert's knowledge to really refine the model", stresses Lacoste. Living, breathing supervisors will have to design the systems and make sure they are asking the right questions. They will have to tweak algorithms that, for example, recommend targets that are associated with significant adverse events. "Machine learning is not going to diminish the need for experts. You still need a team stocked with microbiologists and chemists and other specialists with relevant training and experience", said Collins. "That is not going to change in the foreseeable future."

*Talha Burki*

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.