



Computer-calculated compounds

Researchers are deploying artificial intelligence to discover drugs.

BY NIC FLEMING

An enormous figure looms over scientists searching for new drugs: the estimated US\$2.6-billion price tag of developing a treatment. A lot of that effectively goes down the drain, because it includes money spent on the nine out of ten candidate therapies that fail somewhere between phase I trials and regulatory approval. Few people in the field doubt the need to do things differently.

Leading biopharmaceutical companies believe a solution is at hand. Pfizer is using IBM Watson, a system that uses machine learning, to

power its search for immuno-oncology drugs. Sanofi has signed a deal to use UK start-up Exscientia's artificial-intelligence (AI) platform to hunt for metabolic-disease therapies, and Roche subsidiary Genentech is using an AI system from GNS Healthcare in Cambridge, Massachusetts, to help drive the multinational company's search for cancer treatments. Most sizeable biopharma players have similar collaborations or internal programmes.

If the proponents of these techniques are right, AI and machine learning will usher in

an era of quicker, cheaper and more-effective drug discovery. Some are sceptical, but most experts do expect these tools to become increasingly important. This shift presents both challenges and opportunities for scientists, especially when the techniques are combined with automation (see 'Here come the robots'). Early-career researchers, in particular, need to get to grips with what AI can do and how best to acquire the skills they need to be employable in the job market of tomorrow.

The AI pioneers of the 1950s discussed ►

► building machines that could sense, reason and think like people — a concept known as ‘general AI’ that is likely to remain in the realms of science fiction for some time. However, the continued rapid growth in computer-processing power over the past two decades, the availability of large data sets and the development of advanced algorithms have driven major improvements in machine learning. This has helped to bring about ‘narrow AI’, which focuses on specific tasks. These include improved abilities to analyse, understand and generate text and speech through an AI technique called natural-language processing, and artificial neural networks designed to mimic the way our brains make sense of the world. Such techniques are already in widespread use in fields such as computer vision, voice analysis and route selection. This progress has also triggered a wave of start-ups that employ AI for drug discovery, with many of them using it to identify patterns hidden in large volumes of data.

For example, researchers at biotechnology company Berg, near Boston, Massachusetts, have developed a model to identify previously unknown cancer mechanisms using tests on more than 1,000 cancerous and healthy human cell samples. They modelled diseased human cells by varying the levels of sugar and oxygen the cells were exposed to, and then tracked their lipid, metabolite, enzyme and protein profiles. The group uses its AI platform to generate and analyse immense amounts of biological and outcomes data from patients to highlight key differences between diseased and healthy cells.

The aim of Berg’s approach is to identify potential treatments on the basis of the precise biological causes of disease. “We are turning the drug-discovery paradigm upside down by using patient-driven biology and data to derive more-predictive hypotheses, rather than the traditional trial-and-error approach,” says Niven Narain, Berg’s co-founder and chief executive.

Using this approach, Narain’s team identified the importance of certain naturally occurring molecules in cancer metabolism. This led the group to discover how a new cancer drug works, and indicated some possible therapeutic uses. The drug, BPM31510, is currently in a phase II clinical trial involving people with advanced pancreatic cancer. The company is also using this AI system to look for drug targets and therapies for other conditions, including diabetes and Parkinson’s disease.

London-based start-up firm BenevolentBio has its own AI platform, into which it feeds data from sources such as research papers, patents, clinical trials and patient records. This forms a representation, based in the cloud, of more than one billion known and inferred relationships between biological entities such as genes, symptoms, diseases, proteins, tissues, species and candidate drugs. This can be queried rather like a search engine, to produce ‘knowledge graphs’ of, for example, a medical condition and the genes that

are associated with it, or the compounds that have been shown to affect it. Most of the data that the platform crunches are not annotated, so it uses natural-language processing to recognize entities and understand their links to other things. “AI can put all this data in context and surface the most salient information for drug-discovery scientists,” says Jackie Hunter, chief executive of BenevolentBio.

“AI IS GOING TO LEAD TO THE FULL UNDERSTANDING OF HUMAN BIOLOGY AND GIVE US THE MEANS TO FULLY ADDRESS HUMAN DISEASE.”

When the company asked this system to suggest new ways to treat amyotrophic lateral sclerosis (ALS), also known as motor neuron disease (MND), it flagged around 100 existing compounds as having potential. From these, scientists at BenevolentBio selected five to undergo tests using patient-derived cells at the Sheffield Institute of Translational Neuroscience, UK. The research, presented at the International Symposium on ALS/MND in Boston, Massachusetts, in December 2017, found that four of these compounds had promise, and one was shown to delay neurological symptoms in mice.

PATTERN RECOGNITION

Despite these promising applications, many scientists are unaware of the capabilities of AI. A survey published in February by BenchSci, a start-up in Toronto, Canada, that provides a machine-learning tool for scientists searching for antibodies, found that 41% of the 330 drug-discovery researchers who took part were unfamiliar with the uses of AI (see go.nature.com/2xarpt3).

Leaders in the field think that researchers should brush up on this knowledge as soon as possible.

“AI is going to lead to the full understanding of human biology and give us the means to fully address human disease,” says Thomas Chittenden, who leads a team at Wuxi NextCODE in Cambridge, Massachusetts. Wuxi NextCODE was formed in 2015 after drug-discovery firm WuXi AppTec in Shanghai, China, acquired NextCODE Health, a spin-off from Icelandic company deCODE Genetics. “The way we develop drugs and assess them in clinical trials will all come down to very sophisticated pattern recognition,” he says.

In May 2017, a group including researchers

at Yale University in New Haven, Connecticut, demonstrated the role of a family of proteins called fibroblast growth factors (FGFs) in blood-vessel development (P. Yu *et al. Nature* **545**, 224–228; 2017). This process is key to both tumour growth and cardiovascular disease. Wuxi NextCODE uses AI as part of its approach of classifying genes according to their roles and other attributes, to look for connections between RNA-sequence variations, expression levels, molecular function and gene location. Using this approach, Chittenden’s team discovered that FGFs exert their influence through the control of glucose metabolism.

Some think the potential of AI to pinpoint previously unknown causes of disease will accelerate the trend towards treatments designed for patients with specific biological profiles. “Personalized medicine has been talked about for a long time,” says Hunter. “AI is going to enable it.”

Sceptics point out that some of these more enthusiastic claims echo the excitement over computer-aided drug design, which began in the early 1980s. Although such *in silico* modelling techniques are important in modern drug research and development (R&D), they have not halted a decline in pharmaceutical-industry R&D productivity dating back to the mid-1990s.

MOVING GOALPOSTS

Whatever happens, industry leaders agree that drug-discovery jobs and the skills needed to do them are unlikely to remain the same. Some think that broader training is needed. Narain says that “there needs to be a radical shift” in the way PhDs and other graduate courses are conducted, and that this should extend to medical-school and undergraduate teaching. He adds, “The years of students focusing solely on — and learning more than anyone else about — a particular gene mutation, say, are over.” Chittenden agrees: “The PhD is going to look very different ten years from now. Academic curricula will be broader. The next generation needs, first and foremost, the understanding of human biology, but coupled with computer science, computational statistics and statistical machine learning.”

Others think it is more a case of picking up the basics without diverting attention from core areas of expertise. “Undergraduates in biology need to move towards basic competency in statistics and computational ideas,” says Russ Altman, a biomedical AI researcher at Stanford University in California. “But at PhD level, people need to acquire deep, technical skills. They will be paid for depth, not breadth.”

In 2003, Altman co-launched an undergraduate degree in biomedical computation for students who want to delve deeply into both disciplines. It was relaunched within his institution’s bioengineering department in March. “I think that at Stanford we’re getting an early look at what is going to be happening at campuses worldwide,” he says.

There is little consensus about how, even

HERE COME THE ROBOTS

A new type of scientist

When the time comes for the history of artificial intelligence (AI) to be written, the algorithm that gets the job is likely to flag 12 June 2007 as worthy of note. That was the day that a robot called Adam ended humanity's monopoly on the discovery of scientific knowledge — by identifying the function of a yeast gene.

By searching public databases, Adam generated hypotheses about which genes code for key enzymes that catalyse reactions in the yeast *Saccharomyces cerevisiae*, and used robotics to physically test its predictions in a lab. Researchers at the UK universities of Aberystwyth and Cambridge then independently tested Adam's hypotheses about the functions of 19 genes; 9 were new and accurate, and only 1 was wrong.

"Robot scientists using AI can test more compounds, and do so with improved accuracy and reproducibility, and exhaustive, searchable record-keeping," says systems biologist Steve Oliver of the University of Cambridge, a member of the group that developed Adam.

In January, the same team announced that Adam's more advanced robot colleague, Eve, had discovered that triclosan, a common ingredient in toothpaste, could potentially treat drug-resistant malaria parasites. The researchers developed strains of yeast in which genes essential for growth had been replaced with their equivalents either from malaria parasites or from humans. Eve then screened thousands of compounds to find those that halted or severely slowed the growth of the strains dependent on the malaria genes but not those containing the human genes — to target the parasites while reducing the risk of toxicity. Early results were used to inform the selection of later candidates to screen.

This identified triclosan as affecting malaria-parasite growth by inhibiting the DHFR enzyme — also the target of the antimalarial drug pyrimethamine. However, resistance to pyrimethamine is common. The researchers showed that triclosan could act on DHFR even in pyrimethamine-resistant parasites. **N.F.**

just a decade from now, AI will affect the skills needed to discover the therapies of the future. "Being able to code will be useful for at least the next 5–10 years, but my suspicion is that, beyond that, computers will largely do it for us," says computational medicinal chemist Anthony Bradley at the University of Oxford, UK. "In the lab, we might need a more highly trained, specialized workforce working with the automation and AI experts to fine-tune processes in particular reaction areas," he says. Or, he adds, it might be that wet-lab skills (those needed to perform practical chemical or biological experiments) "might be no use ten years from now".

Bradley uses the Diamond Light Source synchrotron near Oxford to screen compounds for small chemical fragments that bind to molecular targets, even if only weakly, with the aim of improving their binding strength to produce new therapies. He is a member of a group that is using artificial neural networks — an approach to training algorithms inspired by the way our brains process information — as part of a structure-based drug-design project with the Oxford Protein Informatics Group. The aim is to use publicly available data on the structural and chemical activity of small molecules to teach their system to identify those that will act on protein drug targets.

What can those hoping to work in drug discovery do to prepare themselves for this rapidly evolving environment? Taking steps to become informed and flexible are important, say those at the cutting edge of the field.

"My training gave me the groundwork so that I knew roughly where the field was, but to some extent it's down to students themselves to see the way technological trends are going," says Bradley. "Only by remaining versatile can you make the best use of the power of the available tools." He advises those seeking to enter the drug-discovery field to keep track of developments in AI by monitoring the latest articles in leading journals and technology-focused news sources and blogs.

Self-driven learning is especially important, Bradley says, because there are limits to how well universities can provide the skills that students need to be ready for the future role of AI in research. "Almost by definition," he says, "no one can really know what those skills will be."

Some of the more extravagant predictions being made about the ability of AI to revolutionize drug discovery might well turn out to be overblown. Critics point out that there are commercial interests at play, and that, as yet, there are no approved AI-developed drugs. Narain, who thinks the technology will drive major advances, agrees that overblown claims are being made, but says it won't be long before these are exposed for what they are. "The hype can't last very long because over the next five years or so, the truth will come out in the data," he says. "If by then we are creating better drugs, and doing it faster and cheaper, then AI will really take off" ■

Nic Fleming is a freelance science writer based in Bristol, UK.