DISCOVERY REPORT

An ACS member exclusive

ACS Chemistry for Life®

ARTIFICIAL INTELLIGENCE IN DRUG DISCOVERY:

Hope or hype?

AI promises to revolutionize drug discovery and development. But what is it really capable of? And how can chemists help it succeed?

Published by c&en

The future of AI in drug discovery

At the world's pharma and biotech companies, artificial intelligence is increasingly critical to the drug-discovery engine. A 2019 survey by the Pistoia Alliance found that 70% of life sciences researchers use AI, including machine and deep learning, in their work, up from 44% just 2 years earlier. But where along the road to a new drug will AI have the most impact? And can it possibly live up to all the hype that surrounds it?

This Discovery Report explores how AI is going to affect drug discovery and how chemists can prepare. As a member of the American Chemical Society, you will receive four of these reports annually as part of your membership. The reports will analyze the new science and technology defining the chemical sciences, and should be of particular interest to members in the industrial sector.

C&EN associate editor Leigh Krietsch Boerner, who covers organic and medicinal chemistry, edited this report. It kicks off with an overview of what AI and machine learning are and then examines how discovery chemists are already using AI and machine learning to predict the activity of potential drug candidates, where pharma and biotech companies are investing, and what's coming next.

The report also includes an exclusive guide to the top 20 start-ups and small companies offering AI services to the drug-discovery sector, a profile of a new MIT–drug company consortium that's rethinking how chemists design drugs, and more.

Look for your next Discovery Report in early 2020.

Amanda Yarnell
Editorial director, C&EN
@amandayarnell

INSIDE

Published by c&en

Editor: Leigh Krietsch Boerner

Contributing writers: Lisa Jarvis, Sam Lemonick, Emily Mullin, Rick Mullin

Creative: Rob Bryson

Copyeditor: Sabrina Ashwell

CHEAT SHEET

5 questions and answers about AI in drug discovery and development

Q. What are the differences between machine learning, artificial intelligence, and deep learning?

» **Examples of artificial intelligence (AI) include** robots, computer vision, and natural language processing—agents that can perceive their environment, then respond to that environment to achieve a certain goal.

» **Machine learning is a type of AI** in which computers learn on their own. They identify patterns and then build models and make predictions based on those patterns.

» **Deep learning is a kind of machine learning.** The term deep learning refers to deep neural networks, which are algorithms that have multiple layers and are modeled on the human brain, and can learn from large amounts of data. These systems can be more accurate than machine learning because they're more complex, and they have had success in imaging and sound recognition.

Q. How is AI being used in drug discovery and development?

» **Researchers employ AI** to look for new drugs.

» **This involves using machine learning** as a tool to automate searching and some decision-making for well-defined questions with a specific set of data.

» **Some scientists try to find** new uses for existing drugs. Others design new drugs.

» **Some programs can predict** the properties of potential drug candidates or optimize the best candidates.

» **Others try to predict what will happen** when a potential medicine is used to treat a patient. How will the compound act?

Q. Why are drugmakers turning to AI?

» **AI programs may be able to find** patterns in sets of data that scientists might not see.

» **Using AI may help medicinal chemists find** drug candidates faster.

» **AI may also help scientists find** new, more efficient ways to synthesize drug candidates.

» **Machine-learning programs can help find** both disease targets and biomarkers.

» **AI can design and recruit** for clinical trials.

» **It can help chemists analyze** real-world evidence, such as information on how patients react to certain drugs, more quickly and thoroughly.

Q. What are the challenges of using AI in drug discovery?

» **There have not been many results yet,** and breakthroughs don't happen often.

» **Until scientists have more results,** the traditional R&D approach to discovering drugs is probably not going to change.

» **In many research areas,** there are not enough high-quality data for AI to be applied to. For example, AI cannot predict what drug candidate would fit into the active site of a protein if scientists don't know the exact structure of the protein.

» **AI relies on existing data sets,** so the quality of AI's output will match the quality of the data.

» **Any potential drug targets** have to be well characterized, as AI can't select for something that's not known.

» **AI costs money to implement,** and it's unclear when this investment is going to pay off.

Q. What's in the future of AI in drug discovery and development?

» **Awareness of the problems needs** to increase.

» **There are many opportunities** to speed up the drug-discovery process and reduce failure rates in clinical trials.

» **In all areas,** systematic and comprehensive high-dimensional data still need to be generated.

» **Ultimately, machine learning in drug discovery is limited** by what we know. AI can tease out patterns, but it can't invent things. A chemist's creativity is still needed and will be going forward.

8 scientists working on the cutting edge of AI weigh in on where drug hunters will find AI's true promise

Bin Chen

» **Assistant professor at Michigan State University**

Bin Chen thinks that drug-discovery scientists may have a cell-line mismatch problem. Generally, one of the first steps in drug discovery is to use cell lines to screen compounds, Chen says. But how well do cell lines capture the biological makeup of tumors from real patients?

Chen's lab uses artificial intelligence to compare tumor samples taken from people with metastatic breast cancer with the breast cancer cell lines generated for use in such screens. The results have indicated large genomic differences between the two models. In particular, the commonly used synthesized cell line MDA-MB-231 shows few similarities to patient samples. If the cell line is not an accurate representation of the patient's biology, "the drug might work in the cell line but fail in the real patient," Chen says.

His team also uses AI to examine organoids, 3-D models of human organs that are grown in vitro from stem cells. The researchers found that such organoids are more similar to patient tumors than cell lines. This is important for drug testing, Chen says, because if you can rule out which drugs aren't going to work earlier on, you could eliminate some of the costs associated with drug development. "We have to understand the differences and the similarities between the models," Chen says. "That's going to help us to better interpret our results."

> **"We have to understand the differences and the similarities between the models. That's going to help us to better interpret our results."**

Izumi Hinkson

» **Scientific project manager at the Accelerating Therapeutics for Opportunities in Medicine (ATOM) consortium**

It's Izumi Hinkson's job to bridge R&D and operations at ATOM to advance new treatments. The consortium is using AI not only to expand the number of druggable targets and aid in drug design but also to reduce the safety risks of new drugs, reduce clinical trial failures, and "most importantly, drastically cut the cost and time it takes to get new drugs to patients in need," she says.

The consortium, launched in October 2017, is creating a novel preclinical drug-discovery platform designed to get drugs to patients faster. The result of an agreement signed in June 2016 by GlaxoSmithKline, the National Cancer Institute, and the US Department of Energy, the group is combining publicly available data with data provided by GSK and other consortium members to generate new machine-learning models that can better predict how drug candidates will behave throughout the body. It plans to eventually make its platform available to researchers in both academia and industry.

An offshoot of the Cancer Moonshot, an initiative launched by the Obama administration to make more therapies available to people with cancer, ATOM will initially use its platform to focus on oncology as a proof of concept. But overall, the group is taking a disease-agnostic approach, Hinkson says. "A lot of the machine-learning models we're developing can be applied broadly."

Jackie Hunter

» **Chief executive of clinical programs and strategic relationships at BenevolentAI**

Jackie Hunter thinks machine learning's biggest potential is in speeding up how quickly scientists can find new drug candidates for diseases that don't currently have a lot of treatment options.

BenevolentAI has active R&D drug programs that span

from target identification through Phase IIb clinical trials for Parkinson's disease, bone loss, inflammatory bowel disease, and amyotrophic lateral sclerosis. The company is also using its AI-driven platform—called the Benevolent Platform—in a partnership with AstraZeneca to identify new targets for chronic kidney disease and idiopathic pulmonary fibrosis, a progressive lung disease.

For the UK-based company, meeting this challenge means proposing drug targets that are most likely to be important in the disease, designing fewer but better molecules in a shorter period of time to treat the disease, and defining the patient population that will derive the clearest clinical benefit from those molecules. Researchers can apply the Benevolent Platform to a range of tasks in drug discovery and development, from target identification, chemical design, and synthesis to patient stratification and clinical trial design.

AI "gives us greater insight into what not to progress, which patients not to enroll into a clinical trial, and where the drug is most likely to be effective," Hunter says. "This can make a huge impact on the economics of our industry."

Daphne Koller

» CEO and founder of Insitro

For Daphne Koller, data are king. Her San Francisco start-up is taking a different approach to drug discovery than other AI companies: it's generating massive amounts of data on its own instead of using existing data sets that may be low quality or lacking in data.

Insitro is working with induced pluripotent stem cells, adult stem cells that have been reprogrammed into an embryonic-like state. Its researchers edit the cells with CRISPR to understand the phenotypic effects of tweaking certain genes. All the resulting information is stored in the company's data factory, which is specifically geared toward driving machine-learning models.

"We've elected to focus on early target discovery because fundamentally, we think the biggest bang for our buck today is that most drug candidates just don't work," Koller says. "And when they work, we don't know who they work for." She and her colleagues have created a scalable platform for identifying which drug candidates are likely to have an effect and in which set of the population.

"We've elected to focus on early target discovery because fundamentally, we think the biggest bang for our buck today is that most drug candidates just don't work. And when they work, we don't know who they work for."

"How do you develop an AI or predictive tool that can do something more than just tell you what you already know?"

In its first big pharma partnership, with Gilead Sciences, Insitro will work to discover and develop therapies for patients with nonalcoholic steatohepatitis, a progressive liver disease. The company will use its platform to provide insights into disease progression, suggest candidate targets, and predict patient responses to potential therapeutic interventions.

Michael Varney

» Executive vice president of research and early development at Genentech

Michael Varney sees a few specific places where AI is likely to make a near-term impact on how drugs are discovered at Genentech. But the technology is still maturing, he says, and the critical question remains: "How do you develop an AI or predictive tool that can do something more than just tell you what you already know?"

Varney points to using AI to find small-molecule inhibitors. Experienced medicinal chemists know that their chances of finding a molecule that safely and potently blocks a target are higher when they explore a diverse array of chemical backbones. Yet current AI sifts through data sets of only molecules known to fit into the target's active site and not those that "jump out of that groove," Varney says. In other words, there's still work to be done to coax computers into offering up an unexpected new class of structures.

A more likely place for AI to improve productivity is the vetting of molecules designed by chemists, he says. For example, industry already has reams of well-organized data on how existing drugs break down in the body, and computers should soon be able to draw on that information to predict how a human will metabolize a new small molecule. Varney also says AI will make a near-term difference in the design of synthetic routes for building molecules.

Longer-term goals, Varney says, include building tools that can sort molecules into classes and developing databases of antibodies that are robust enough to allow computers to relate antibody sequences to physical properties.

While the computers catch up, drug hunters at Genentech continue to apply their decades of knowledge to finding novel drugs, Varney says. "Molecule making is probably more valuable now than it's ever been."

CREDIT: ALL IMAGES COURTESY OF INSTITUTIONS

Rong Xu

» Associate professor at Case Western Reserve University

What if a drug to treat Alzheimer's disease already exists? If it does, Rong Xu might be the one to find it. Xu is using machine-learning algorithms to search existing drug databases for drugs that can cross the blood-brain barrier and bind to specific molecular targets implicated in Alzheimer's disease.

To accomplish this, her algorithm first scans the database to pinpoint which candidates will likely work in humans. Then the researchers test promising candidates in animal models to understand how they work and to confirm the drug efficacy. If these candidates are US Food and Drug Administration–approved drugs, the team can then can conduct AI-driven virtual clinical trials, which use patient health records and other digital tools instead of monitoring patients in person at trial sites.

Some of the failure seen in drug development related to Alzheimer's and related diseases is that models about how these drugs work in mice don't necessarily translate to humans, Xu says. But an AI drug-discovery engine can help minimize that gap by identifying drug candidates that are likely to work in humans, she says.

Xu's AI drug-discovery method has predicted that certain FDA-approved antiviral drugs can treat Alzheimer's and related dementia in humans. Her team performed an AI-driven virtual clinical trial with electronic health records of 63 million patients and found that patients on these antivirals appear to have a lower risk of Alzheimer's and other forms of dementia.

Alice Zhang

» CEO and cofounder of Verge Genomics

Alice Zhang left her graduate program to start Verge Genomics, a company focusing on using AI to find new drugs for neurological diseases. Her company has its own proprietary patient data sets and is using algorithms it developed to sift through that data to find new targets for central nervous system diseases.

"Currently, drug discovery is still largely a guessing game," Zhang says. "Drug companies typically use brute force to screen millions of compounds in hopes of finding just one that works, or they look at single genes in diseases caused by hundreds of genes."

Instead of screening a huge amount of compounds at once, the company tries to pinpoint the causes of disease, then predict which compounds might reverse those root causes. And instead of looking at one gene at a time, it's using in-house algorithms on human genomic data to map out the hundreds of genes that cause disease. Verge Genomics is already moving multiple drug candidates toward clinical testing.

Zhang's company is also trying to address what she sees as a major barrier to the use of AI in drug discovery: data. She says there are a lot of missing data in the field because scientists still don't fundamentally understand most of human biology. To address this gap, her company has built a huge collection of genomic data from patients with Parkinson's and Lou Gehrig's disease, which she says provides a glimpse into early disease progression. "I believe we're at the precipice of a technological and scientific renaissance in neuroscience."

"Currently, drug discovery is still largely a guessing game. Drug companies typically use brute force to screen millions of compounds in hopes of finding just one that works, or they look at single genes in diseases caused by hundreds of genes."

Alex Zhavoronkov

» CEO and founder of Insilico Medicine

Alex Zhavoronkov sees aging as akin to tiny cracks developing in our bodies over time. His Hong Kong–based company is building biological clocks using artificial intelligence to detect these cracks. Insilico's clocks are designed to predict a person's biological age using biomarkers in the blood, microbiome composition, genetics, and even a person's facial features.

Using data from South Korean, Canadian, and eastern European patient populations, the company has trained machine-learning models to recognize what healthy chronological aging looks like.

When presented with new patient data, the models can pick out which people appear to be aging more quickly or slowly than the typical population. The idea is to identify new biomarkers of aging and potential drugs that could be used to improve the health span—the time in people's lives when they're in good health.

The model's deep neural networks, which are trained to predict a person's biological age over time, can be helpful to drug developers because they can show researchers what healthy aging looks like, Zhavoronkov says.

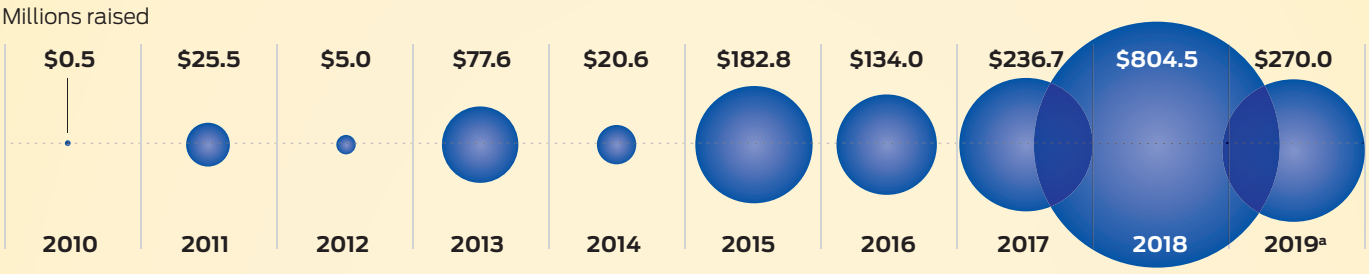
Such models would give researchers a baseline for testing interventions in clinical trials for longevity. Currently, it's almost impossible to conduct clinical trials for aging because the human life span is so long that it's difficult to predict if an intervention is slowing down the aging process. If you knew how a healthy person is supposed to age, Zhavoronkov says, "you can figure out if you're having an anti-aging effect."

CREDIT: ALL IMAGES COURTESY OF INSTITUTIONS

Follow the money that's headed into AI drug-discovery companies

Piles of dollars

Our top 20 (see page 12) AI drug-discovery companies raised $1.8 billion in the past 10 years.



Recent partnerships

This year has seen more and more deals between big companies and AI drug-discovery companies. Here are some of the most notable.

Month	Big company	AI company	Deal amount
September	Hansoh Pharmaceutical Group	Atomwise	Potentially up to $1.5 billion for up to 11 undisclosed target proteins in a range of therapeutic areas
August	Novo Nordisk	E-Therapeutics	Not disclosed
June	Google	Sanofi	Not disclosed
June	Eli Lilly and Company	Atomwise	$1 million per target and up to $550 million in development and commercialization
April	AstraZeneca	BenevolentAI	Not disclosed
April	Gilead Sciences	Insitro	$15 million up front, plus near-term payments of up to $35 million for hitting milestones, up to $200 million for milestones for each of the five Gilead targets, and up to low double-digit tiered royalties on net sales
April	Janssen	Iktos	Not disclosed
April	SK Biopharmaceuticals	twoXAR	Not disclosed
March	Merck KGaA	Iktos	Not disclosed
March	Ono Pharmaceutical	twoXAR	Not disclosed
March	Celgene	Exscientia	$25 million up front and eligibility to receive more for hitting milestone targets; Exscientia is also eligible for tiered royalties on net sales on any product resulting from the collaboration
January	Servier	Iktos	Not disclosed
January	Takeda Pharmaceutical[b]	Recursion Pharmaceuticals	Not disclosed
January	Lundbeck Pharmaceutical	Numerate	Not disclosed

Sources: Crunchbase, company websites.

a Year to date.

b Recent date is extended partnership; original is from December 2018.

ANALYSIS

As AI and machine learning ignite in pharmaceutical labs, they also exhibit their limits

MICHELANGELO'S DEPICTION OF THE creation of Adam on the ceiling of the Sistine Chapel pivots on the iconic image of two fingers nearly touching—the finger of the creator and that of the creature he made in his image. That touch marks a crucial turn in the Genesis story: the beginning of a uniquely intelligent species on earth, though one that overreaches badly in the field of intelligence not much further on.

A seemingly inexhaustible inspiration for art and literature, that imminent touch, with its implication of trouble ahead, has become a metaphor for artificial intelligence (AI).

A machine endowed with the cognitive capabilities of its creator has been a goal in computer development since World War II. Rudimentary versions of mechanized human intelligence are now able to process data and analyze images that would overwhelm whole laboratories of human brains, and there is a general perception that the intelligent machine has arrived. With it comes enthusiasm as well as foreboding.

Both feelings are present in laboratories dedicated to drug discovery and development, where the jobs include compound screening, experiment design, image processing, and patient data analysis.

As in other industries that have practical experience with intelligent machines, researchers in the drug industry have taken off the table the prospect of robots making humans obsolete. Many drug researchers consider the technology an indispensable aid and enabler.

CREDIT: SHUTTERSTOCK

AI, however, has also shown that adding decision-making and the ability to "learn" to a computer's traditional number-crunching role is changing the work done by the research scientist. Uncertainty regarding the extent and nature of that change is the source of some anxiety. That's certainly true for medicinal chemists.

Hugo Ceulemans, scientific director of discovery data science at Janssen Pharmaceuticals, is an advocate of AI.

"First of all, AI is never replacing the traditional researcher," he says. Instead, the technology will allow chemists to focus more effort on innovative science.

"It is making more data available," Ceulemans adds, noting that discovery labs are filling with quantities and varieties of important data that defy human processing for two reasons. "Humans have a limited capacity for dealing with data in decision-making. And it's also really boring. Machines excel at it because they are not afraid of boredom."

AI's grunt work also opens windows in the discovery lab, Ceulemans says, by "asking researchers to be more open than they traditionally have been to more exotic solutions they might not have anticipated or ever thought of."

And despite its data-parsing power, dot-connecting skills, and ability to improve performance as it is exposed to more data, AI in its current state falls well short of emulating human intelligence, Ceulemans says. He ranks current technology at the level of "idiot savant": alarmingly good in a narrow field of endeavor, yet utterly naive outside it.

"True artificial intelligence would not only mean that a machine can learn but that it can also reason and actually decide," he says. "Currently, what the machine does is boring, mind-numbing explorations and correlations, offering the scientist suggestions that are not set in stone. It is the scientist that makes the ultimate decision."

"Right now, we don't have any machines to which we can say, 'Hey machine, go find me a compound that will affect the so-and-so receptor so I can cure pancreatic cancer.'"

But some researchers see a new world coming into focus, one where the line blurs between data scientist and traditional research scientist. It's a world that will include chemists with a skill set that is not yet clear.

"As machine intelligence kicks in, we may end up with fewer chemists doing this kind of work," says Derek Lowe, a drug researcher with experience at three large drug companies and a major biotech firm, referring to basic research chemistry and biology. "But they are going to be doing it at a much higher level."

Lowe, who has touched on AI in his In the Pipeline blog, notes that the technology has not yet kicked in. "The most useful and reliable use of machine learning in drug discovery is probably related to imaging," he says. "Right now, we don't have any machines to which we can say, 'Hey machine, go find me a compound that will affect the so-and-so receptor so I can cure pancreatic cancer.'"

But such machines may be on the horizon, Lowe says, given the pace of AI evolution and the range of possible applications in the pharmaceutical lab. "I don't see any reason why we are not going to be there."

Indeed, the technology appears to be at a kind of tipping point. Michael Shanler, a research vice president covering life sciences for consulting firm Gartner Inc., says the mainstream media is largely to blame for positioning AI as simultaneously delivering the ultimate promise in computing and "the end of the world" as robots take over.

Understanding AI begins, Shanler says, with realizing it is composed of many technologies, creating a challenge in implementation. "There is a zoo of machine-learning approaches out there, which have a variety of computing requirements," he says.

Lab managers must take traditional science, which varies from lab to lab, into consideration when implementing AI systems. And traditional scientists need to configure AI systems to meet their needs.

"You need process owners, people who understand the science or business process," Shanler says. "You need them to ask the right questions to guide the machine learning."

That guidance, however, must steer clear of researchers' bias. Just as a bad researcher is one who finds what he or she is looking for, there is a chance that an intelligent machine may operate as a mechanized bad researcher.

But implemented correctly, AI steers in the opposite direction of myopic research, according to Shanler. "That is where smart machines and AI kind of can excel," he says. "They can deliver an unanticipated result, one that somebody might have overlooked because of their own bias. This is one of the promises of AI, and I've seen it with some of my clients."

AI is making particular headway in the field of chemical synthesis planning. Chematica, chemical synthesis planning software developed by Grzybowski Scientific Inventions and acquired by MilliporeSigma in 2017, for example, passed the test of establishing a synthetic route to eight target molecules selected by MilliporeSigma and academic researchers in the U.S. and Poland.

CAS, a division of the American Chemical Society, is also getting into the game. In 2017, CAS licensed ChemPlanner, a retrosynthesis engine developed by the scientific publisher John Wiley & Sons. It used it to develop a new retrosynthetic engine for SciFinder^{n} by incorporating a trove of new chemical reaction data, including its collection of more than 100 million human-curated chemical reactions.

The partnership is an early move into the AI market for CAS, which is also developing programs for analyzing data clusters, neural networks, and other information with the goal of aiding chemical synthesis.

Meanwhile, AI is making inroads elsewhere. In early 2018, Google reported that an AI algorithm it developed can scan patients' eyes to predict heart disease, and researchers at Stanford University successfully put AI's image-reading capabilities to work screening moles for melanoma.

On the other hand, one major AI project—the installation of IBM's Watson for Oncology at the University of Texas MD Anderson Cancer Center—collapsed at the end of 2016 when the university system's audit office pulled the plug, citing improprieties in the procurement process. News of the failed project, which had run up a price tag of $62 million by the time it was stopped, was a black eye for AI in drug discovery.

The university's report on the project, however, stated in boldface type that the termination should not be taken as a positive or negative assessment of AI technology or IBM's Watson computer, which is being used by Pfizer in immuno-oncology and elsewhere in pharmaceutical research.

"That is where smart machines and AI kind of can excel. They can deliver an unanticipated result, one that somebody might have overlooked because of their own bias. This is one of the promises of AI, and I've seen it with some of my clients."

As AI works its way fitfully into the drug laboratory, traditional tools—laboratory information management systems (LIMS), databases, and scientific instruments—are evolving to accommodate it. Thermo Fisher Scientific, a major supplier of laboratory IT, is pivoting from big data to AI, positioning its Platform for Science LIMS as a foundation for machine learning and other intelligent machine functionality.

Clarivate Analytics, a life sciences data and analysis company, added AI engines as adjuncts to its MetaCore and Integrity data and analytics products in recent years.

Executives at Thermo, Clarivate, and IBM agree that the drug discovery sector is at an early stage of experimentation with the technology, even as research labs feel out new roles for scientists interacting with technology.

"I don't think organizations know yet how they are going to implement AI," says Trish Meek, director of commercial operations for digital science at Thermo Fisher.

One of the first things companies are learning, she says, is that they need a well-managed database on which to build it. "Everyone knows AI offers opportunities, but they realize laboratory IT infrastructure doesn't support it without a platform approach to informatics."

That said, Jeff Noonan, business development director for Thermo Fisher's digital science division, notes that IT platforms are becoming less monolithic—and researchers more empowered in their access to and use of data—partly because of cloud service software applications.

Noonan sees a shift from "a place where organizations require their own IT departments and their own coding departments to develop software applications to where the creation of those applications happens in the laboratory, giving scientists the ability to create solutions to address the needs of their specific laboratory."

Clarivate is also working with drug companies that are looking for an entrée into intelligent systems, says Roger Willmott, the firm's vice president of technology. "They are using AI in modeling, but it is at an experimental state." Data quantity and quality are interrelated hurdles. "Researchers clearly have a lot of data," he says. "They are all trying to work out how they can use it."

Louisa Roberts, associate partner at IBM Watson Health-Life Sciences, adds that data, once processed by an intelligent machine, need to be presented in a clear and navigable visual format. Data visualization, she says, needs to be customized by researchers to meet the specific needs of their labs.

Drug companies are also pursuing AI on other avenues. Several of them have joined a consortium called Machine Learning for Pharmaceutical Discovery and Synthesis. Hosted by Massachusetts Institute of Technology, the group seeks to replace labor-intensive trial-and-error work in molecule synthesis with a computational reaction-design process.

CREDIT: SHUTTERSTOCK

Regina Barzilay, a computational scientist at MIT, has worked on AI projects with consortium members in the Cambridge, Mass., area. "I can tell you that they totally realize it is a huge place of opportunity," she says. "They are trying to learn, and they are fast learners." Barzilay sees AI as a true turning point in technology, one that will free scientists from data drudge work and repetitive experimentation.

It is early days, however, and most discovery science researchers say they have seen only the first flashes of intelligent life in laboratory data analysis systems.

"Here everything is pretty manual," says a medicinal chemist at a major biopharmaceutical firm, who asked for anonymity because he's not authorized to speak publicly on the subject. He says researchers at his company are not concerned that AI will take over their jobs, but he does anticipate some pushback as it inevitably encroaches.

"There is concern about the human element. I think medicinal chemists value that most highly. Personally I think some of them value it too highly," he says. "Medicinal chemists are a little reluctant to adopt these things that take them out of the decision-making or idea-generating process."

"Even in a fast-paced drug discovery environment, it takes one week to get data back when you give the scientist a compound to test," says Ashutosh Jogalekar, a computational chemist at Strateos. "Maybe AI can have an impact on analyzing the results."

"There is concern about the human element. I think medicinal chemists value that most highly. Personally I think some of them value it too highly."

Other areas where Jogalekar sees possible change include reaction planning, analyzing results of phenotypic screens, and similar lab operations. "I place a lot of value in this stuff," he says. "I would like to see a lot more of it, and I would like to see it improve."

Lowe, who has decades of experience as a medicinal chemist, says there is little concern in discovery labs about smart machines turning traditional human scientific endeavor into an automated commodity. Automation is nothing new, he says, and science abides.

"There are big chunks of stuff now that are done as a kind of science as a service that used to be bespoke areas of research," he says. "Sequencing DNA, collecting NMR data—if you want to go back further, LC mass spec data; now, these are all walk-up machines. There has always been a tendency of things going from cutting edge to difficult to easy to automated."

The world in which one asks a machine to make a compound is upon us, Lowe says. There is little question that smart machines will gain traction. In effect, resistance is futile, according to Lowe.

"These machines are pretty good, and they are getting better," he says. "We're not."

This article is reprinted with permission from C&EN. A version of this article was published in C&EN on April 2, 2018, on page 20.

The 20 most promising companies using AI to improve drug discovery—and how they're doing it

AI Therapeutics

» **www.ai-therapeutics.com**
» **Based in:** Guilford, Connecticut
» **Founded in**: 2013
» **Money raised to date:** $98 million
» **Key partnership:** Genentech
» **Strategy:** AI Therapeutics searches for drugs for cancer and rare diseases and tries to predict which therapies will work best on patients according to their genetic makeup. With both public and proprietary data, the company uses information from treatments and patients' clinical responses to determine which patients to treat and how.
» **Why watch:** Formerly known as LAM Therapeutics, the company has raised $98 million from just one investor, Suzhou Industrial Park Bioventure Investment Management. AI Therapeutics has three drugs in clinical trials: LAM-001 for lymphangioleiomyomatosis (Phase II), LAM-002 for B-cell non-Hodgkin lymphoma (Phase II), and LAM-003 for acute myeloid leukemia (Phase I).

Atomwise

» **www.atomwise.com**
» **Based in:** San Francisco
» **Founded in:** 2012
» **Money raised to date:** $51.3 million
» **Key partnerships:** AbbVie, Bayer, Charles River Laboratories, Eli Lilly and Company, Merck & Co., Pfizer
» **Strategy:** Atomwise identifies novel drug candidates from a large organic chemistry data set. Atomwise's methods allow the company to quickly narrow down a few chemical scaffolds from millions that may bind to a target protein. The company's algorithms also screen drug candidates for toxicity and oral bioavailability and can help identify those candidates' mechanisms of action.
» **Why watch:** The company claims to have over 60 partners and to have multiple drug candidates already in development with these partners. As part of its partnership with Lilly, Atomwise could receive up to $1 million per target and could get as much as $550 million, depending on what the partners achieve. The company recently partnered with Chinese giant Hansoh Pharma to develop small molecules for up to eleven undisclosed target proteins in a range of therapeutic areas, potentially earning Atomwise up to $1.5 billion. Monsanto Growth Ventures and Y Combinator, a Silicon Valley start-up accelerator, have also put money behind the firm.

BenevolentAI

» **benevolent.ai**
» **Based in:** London
» **Founded in:** 2013
» **Money raised to date:** $202 million
» **Key partnerships:** AstraZeneca, Janssen
» **Strategy:** BenevolentAI analyzes drug targets for diseases such as amyotrophic lateral sclerosis (ALS), Parkinson's, glioblastoma, and sarcopenia. The company has used its platform to find novel drug targets and lead molecules for ALS; one lead molecule came from a breast cancer drug. Its platform can also predict preclinical and clinical success using the drug's target and mechanism and the genetic basis of the patient's disease.
» **Why watch:** The company has raised a huge amount of capital in only 6 years. It has partnerships with two large drug companies and counts Goldman Sachs and Lundbeck among its investors.

BERG

» **www.berghealth.com**
» **Based in:** Framingham, Massachusetts
» **Founded in:** 2006
» **Money raised to date:** Not available
» **Key partnerships:** AstraZeneca, Becton, Dickinson and Company, Sanofi Pasteur

» **Strategy:** Berg sifts through patient data to identify biomarkers in oncology, neurology, and rare diseases. The company uses its artificial intelligence–based machine-learning platform to investigate cause-and-effect relationships in systems biology and understand how specific diseases work. Berg then analyzes the data to identify specific genes, proteins, and genetic variants of disease to guide drug candidate selection and strategies.
» **Why watch:** Berg is one of the largest companies on our list. It has deals with Sanofi to develop flu vaccines that are more effective and with the drug giant AstraZeneca to winnow down targets for Parkinson's disease.

CYCLICA

» **cyclicarx.com**
» **Based in:** Toronto
» **Founded in:** 2013
» **Money raised to date:** $7 million
» **Key partnerships:** Bayer, Merck KGaA
» **Strategy:** Cyclica's AI platform analyzes all the potential proteins that a drug candidate could interact with in the body. This allows its scientists to design ligands meant to minimize off-target interactions that could cause side effects and to develop medicines that accurately hit their targets.
» **Why watch:** While the Canadian company hasn't raised much capital compared with other start-ups, it has partnerships with pharma powerhouses Merck KGaA and Bayer.

e-therapeutics

» **www.etherapeutics.co.uk**
» **Based in:** Long Hanborough, England
» **Founded in:** 2003
» **Money raised to date:** $66.8 million
» **Key partnerships:** Biorelate, Intellegens, Novo Nordisk
» **Strategy:** Complex molecular networks control the activity inside both healthy and diseased cells. E-therapeutics analyzes how these molecular networks interact with one another to produce either disease or normal cellular biological function. It then uses the data to predict how diseased cells will interact with a drug candidate instead of just looking at how that candidate interacts with a single target protein. The company claims its approach will lead it to small-molecule drug candidates with high efficacy and low side effects.
» **Why watch:** The only company on our list that's jumped into the initial public offering realm, E-therapeutics has nabbed a partnership with Novo Nordisk and a few smaller AI companies.

Note: Companies were included based on the novelty and promise of their research methods, amount of capital raised, number of pharma and biotech partnerships, and number and identity of investors.

Exscientia
DRIVEN BY KNOWLEDGE

» **www.exscientia.co.uk**
» **Based in:** Oxford, England
» **Founded in:** 2012
» **Money raised to date:** $43.7 million
» **Key partnerships:** Celgene, Evotec, GlaxoSmithKline, Sanofi, Sumitomo Dainippon Pharma, Sunovion Pharmaceuticals,
» **Strategy:** Specializing in small molecules, Exscientia designs potential drug candidates and predicts their potency, selectivity, and pharmacokinetics. The company claims it can find candidates in about one-quarter the time of traditional techniques.
» **Why watch:** In addition to successfully joining with multiple large pharma companies such as Evotec and GSK, Exscientia acquired the smaller AI company Kinetic Discovery in 2018. It is currently expanding across Asia.

GNS HEALTHCARE

» **www.gnshealthcare.com**
» **Based in:** Cambridge, Massachusetts
» **Founded in:** 2000
» **Money raised to date:** $77.3 million
» **Key partnerships:** Amgen, Genentech, Zambon
» **Strategy:** GNS Healthcare uses its AI platform to compile large amounts of patient data, including clinical, genetic, and lab data, into models that attempt to explain cause and effect in disease. The company then uses the models to run "what if" simulations to find a disease's root cause. GNS says this allows it to more accurately predict what therapies will work and for whom.
» **Why watch:** GNS Healthcare has partnerships with several large pharmaceutical companies.

IKTOS

» **iktos.ai**
» **Based in:** Paris
» **Founded in:** 2016
» **Money raised to date:** Not available
» **Key partnerships:** Janssen, Merck KGaA, Servier
» **Strategy:** Iktos's platform can generate novel data points using data that researchers feed it. For example, when supplied with publicly available databases of molecular structures, activity assays, selectivity data, and more, the Iktos algorithm offers up unexplored lead compounds. The company

claims it can thus help researchers explore chemical space for new drug candidates more quickly and efficiently.

» **Why watch:** The company has partnerships with several European biopharmaceutical companies, such as Galapagos, Merck KGaA, and Pierre Fabre. Its latest partnership with giant Janssen shows that it's moving into the US.

Insilico Medicine

» **insilico.com**
» **Based in:** Hong Kong
» **Founded in:** 2014
» **Money raised to date:** $14.3 million
» **Key partnerships:** GlaxoSmithKline, Johnson & Johnson, L'Oréal, Novartis
» **Strategy:** Insilico Medicine focuses on cancer, aging, and age-related diseases. It uses its platform to predict the pharmacological properties of potential drugs and supplements as well as to find relevant biomarkers for disease.
» **Why watch:** The company claims to have over 150 partners.

insitro

» **www.insitro.com**
» **Based in:** San Francisco
» **Founded in:** 2018
» **Money raised to date:** $100 million
» **Key partnership:** Gilead Sciences
» **Strategy:** Insitro uses a range of large and high-quality data sets, including human cohort data sets that contain molecular and clinical data for hundreds of thousands of patients, to train machine-learning models to discover new drug candidates. In addition, the company claims to be able to develop and observe biological models in its labs, which it says will bring down costs.
» **Why watch:** Although it has existed for only a year, Insitro has already raised $100 million from nine investors, including Bezos Expeditions. Led and founded by Daphne Koller, a former computer science professor at Stanford University and cofounder of online learning platform Coursera, the company also recently signed a deal with Gilead to find treatments for the liver disease nonalcoholic steatohepatitis. The deal included an up-front payment of $15 million and potential milestone payments of up to $1 billion.

Sources: Crunchbase (accessed June and July 2019), company websites, news reports.

NIMBUS THERAPEUTICS

» **www.nimbustx.com**
» **Based in:** Cambridge, Massachusetts
» **Founded in:** 2009
» **Money raised to date:** $137 million
» **Key partnerships:** Celgene, Charles River Laboratories, Genentech, Gilead Sciences
» **Strategy:** Nimbus Therapeutics uses AI and other computational methods to find drug candidates that attack known targets of metabolic diseases, cancer, and immune-inflammatory disorders. It also designs allosteric inhibitors that fit in grooves beyond a protein's active site, targeting, for example, acetyl coenzyme A carboxylase in metabolic diseases and cancer.
» **Why watch:** Formerly known as Nimbus Discovery, the company is a computational partner of Schrödinger, a computational chemistry company whose platform is used by many biotech firms. In 2016, Gilead paid $400 million to license Nimbus's drug candidate firsocostat only 18 months after its discovery. The compound, used to treat the liver disease nonalcoholic steatohepatitis, is in Phase II trials.

Numerate

» **www.numerate.com**
» **Based in:** San Francisco
» **Founded in:** 2007
» **Money raised to date:** $17.4 million
» **Key partnerships:** Boehringer Ingelheim, Lundbeck, Merck & Co., Takeda Pharmaceutical, Servier
» **Strategy:** Numerate's platform can predict how a potential drug will be absorbed, distributed, metabolized, and excreted in the human body for diseases such as obesity, heart failure, Alzheimer's disease, and Huntington's disease.
» **Why watch:** The company has inked multiple partnerships with large pharmaceutical companies and is using its technology in other areas, including projects for the US Department of Defense.

NURITAS

» **www.nuritas.com**
» **Based in:** Dublin
» **Founded in:** 2014
» **Money raised to date:** $60.7 million
» **Key partnerships:** BASF, Nestlé
» **Strategy:** Nuritas uses AI to search for potential drug candidates among the huge number of food-derived bioactive peptides that have been identified to date.
» **Why watch:** In partnerships with BASF and Nestlé, the company is looking in an underexplored area for new drugs.

OWKIN

» **owkin.com**
» **Based in:** New York City
» **Founded in:** 2016
» **Money raised to date:** $18.1 million
» **Key partnerships:** Amgen, Ipsen, Roche
» **Strategy:** Using mathematical models and machine-learning algorithms, Owkin aggregates and then interprets vast amounts of patient data, biostatistics, and biomedical images. The company helps health-care organizations navigate data-sharing challenges, predict how patients will respond to potential treatments, and organize efficient clinical trials.
» **Why watch:** Owkin has partnered with large companies, such as Roche, to try to solve the ongoing problem of how to share health-care data while maintaining data privacy and security.

RECURSION

» **www.recursionpharma.com**
» **Based in:** Salt Lake City
» **Founded in:** 2013
» **Money raised to date:** $226.4 million
» **Key partnerships:** Sanofi, Takeda Pharmaceutical
» **Strategy:** Recursion Pharmaceuticals combines experimental biology, automation, and AI to quickly and efficiently identify treatments for rare diseases, aging, inflammation, infectious diseases, and immunology. The company's AI platform scans diseased-tissue images to generate computational fingerprints for each target disease. This can help Recursion predict the safety of drug candidates and understand how the disease model and the potential drug candidate interact.
» **Why watch:** Recursion is second only to Relay Therapeutics in the amount of money raised for an AI-based start-up on our list, including a recent $121 million in further funding. The company is partnering with both Sanofi and Takeda and counts the Bill & Melinda Gates Foundation among its investors.

RELAY THERAPEUTICS

» **relaytx.com**
» **Based in:** Cambridge, Massachusetts
» **Founded in:** 2016
» **Money raised to date:** $520 million
» **Key partnerships:** Not disclosed
» **Strategy:** Relay Therapeutics uses AI to analyze how proteins move inside human cells and how those proteins' shapes and movements influence health and disease.
» **Why watch:** While it has disclosed no partnerships, the company is one of the most ambitious in the biotech AI space and has raised a whopping $520 million since its inception in 2016. The experiences of a number of the company's team members, including CEO Sanjiv Patel, who left the multibillion-dollar company Allergan to head Relay, suggest that they have the know-how to become a successful company.

twoXAR

» **www.twoxar.com**
» **Based in:** Mountain View, California
» **Founded in:** 2014
» **Money raised to date:** $14.3 million
» **Key partnerships:** Ono Pharmaceutical, Santen, SK Biopharmaceuticals
» **Strategy:** twoXAR screens publicly available libraries to find compounds with the highest possible efficacy for a disease. The company has focused mainly on rheumatoid arthritis and breast cancer, examining electronic health records and patient gene expression to determine new biological targets.
» **Why watch:** With partnerships mainly with Asian companies, twoXAR seems to be focusing on that overseas market, where its US-based competitors aren't necessarily looking.

VERGE genomics

» **www.vergegenomics.com**
» **Based in:** San Francisco
» **Founded in:** 2015
» **Money raised to date:** $36.1 million
» **Key partnerships:** Datavant, Genomics England, Johns Hopkins University, Scripps Research
» **Strategy:** Verge Genomics maps the genes that cause a disease to find potential drug candidates that can target all those genes at once. To start, the company has built its own proprietary genomic data sets from brain tissue samples of deceased patients with amyotrophic lateral sclerosis (ALS) and Parkinson's disease.
» **Why watch:** The company, cofounded by Alice Zhang, has partnered with multiple hospitals and academic centers to build up its neurodegenerative disease database, and it has recently started applying its algorithms to pharmaceutical databases to design clinical trials with partner Datavant.

XtalPi

» **www.xtalpi.com**
» **Based in:** Cambridge, Massachusetts
» **Founded in:** 2014
» **Money raised to date:** $67.6 million
» **Key partnerships:** Alibaba Cloud, Amazon Web Services, Pfizer
» **Strategy:** XtalPi, a joint US-China company, uses algorithms to predict what crystallized form a drug will have. The company's AI platform gives the researchers a better idea of the molecular packing in cocrystals, which allows XtalPi to predict the safety, stability, and efficacy of potential drug candidates on the basis of their physiological and chemical characteristics.
» **Why watch:** With unusual partnerships with Amazon Web Services and Alibaba Cloud, XtalPi is finding ways to get to the data it wants faster. The company also counts Google among its investors.

A new MIT–drug company consortium hopes to reprogram the way chemists design drugs

Regina Barzilay, a computer scientist at the Massachusetts Institute of Technology, was trying to figure out how to use machine learning to approach retrosynthesis when she and her colleagues had an idea. Why not invite their industry neighbors to contribute?

"We realized that there was a lot of potential to develop this technology for the pharmaceutical industry, not only in regard to retrosynthesis but also to a variety of other properties," she says.

Barzilay and coworkers started meeting with company representatives and presented software and technology that they developed under a Defense Advanced Research Projects Agency (DARPA) grant. In May 2018, MIT announced the formation of the Machine Learning for Pharmaceutical Discovery and Synthesis (MLPDS) Consortium. Since then, 13 industry partners have joined: Amgen, AstraZeneca, BASF, Bayer, GlaxoSmithKline, Janssen, Leo Pharma, Eli Lilly and Company, Merck & Co., Novartis, Pfizer, Sunovion Pharmaceuticals, and WuXi AppTec. On the MIT side, Barzilay leads the consortium with faculty members William Green, Tommi Jaakkola, Tim Jamison, and Klavs Jensen.

The scientists are generating models using publicly available data and handing those models over to member companies. Each company runs those models using its own private data, which MIT doesn't have access to. Barzilay says this is by design. "Since we're a neutral player, this is one of the advantages for the companies to pair with us because we can just put our models out there and we don't have a material interest to prove the numbers one way or another."

The ultimate goal is to help speed up drug discovery and development. The consortium is doing that in a few ways. One is in the area of molecular-property prediction, in which machine learning can predict the biological properties of unknown molecules. The other is using artificial intelligence models to generate molecules with certain desired profiles. This involves starting with a molecule and improving it in some way, such as making it more potent. Scientists can also use this technique to create new molecules altogether.

"If you're thinking about the process of drug design, there

MLPDS at a glance

- **Full name:** Machine Learning for Pharmaceutical Discovery and Synthesis Consortium
- **Founded:** 2018
- **What it is:** A collaboration among the pharmaceutical and biotechnology industries and MIT's Departments of Chemistry, Chemical Engineering, and Electrical Engineering and Computer Science
- **Its goal:** Design software that can automate small-molecule discovery and synthesis
- **Faculty leaders:** MIT's Regina Barzilay, William Green, Tommi Jaakkola, Tim Jamison and Klavs Jensen
- **Partners:** Amgen, AstraZeneca, BASF, Bayer, GlaxoSmithKline, Janssen, Leo Pharma, Eli Lilly and Company, Merck & Co., Novartis, Pfizer, Sunovion Pharmaceuticals, and WuXi AppTec

Regina Barzilay gives a tutorial on artificial intelligence to members of the Machine Learning for Pharmaceutical Discovery and Synthesis Consortium.

CREDIT: GRETCHEN ERTL

are prediction problems at every step of the way," Barzilay says. Even if you have high-throughput screening, you still need to decide what subset of molecules to start looking at, she says. Being able to narrow down which molecules to screen for can help speed up drug discovery by steering chemists down the right road. In manufacturing, artificial intelligence can be used to help chemical engineers determine the best sequence of reactions needed to synthesize molecules effectively, Barzilay says.

Along with colleagues from MIT's Computer Science and Artificial Intelligence Laboratory and Department of Electrical Engineering and Computer Science, Barzilay has developed a tool to select molecular candidates according to desired properties. The model can also suggest alterations to the molecular structure of these candidates to achieve higher potency.

Testing the tool on both public and private data sets, researchers found that it performed better than models currently used in the field (*J. Chem. Inf. Model.* 2019, DOI: 10.1021/acs.jcim.9b00237). It's a step toward automating the manual, time-consuming process of new-molecule design and lead optimization.

The consortium also aims to establish criteria for assessing how accurate machine-learning methods are. Barzilay sees machine learning as the Wild West of artificial intelligence. "Today we really need to have generalized, cross-industry standards on how tools work, which tasks they can solve, and which tasks they cannot solve," she says. Standardized data sets would also be a huge asset because data sets vary widely in terms of the types of information they include, Barzilay says.

> **"Today we really need to have generalized, cross-industry standards on how tools work, which tasks they can solve, and which tasks they cannot solve."**

Data sets in general are a major problem in the machine-learning field. Many companies use proprietary data sets to evaluate models, so some researchers don't have access to them. In addition, databases are problematic in that they don't all collect the same kinds of information, they can contain flaws or biases, and they may not have been assembled with machine learning in mind. These factors can make it difficult for researchers to compare one machine-learning model with another, from company to company or from industry to academia. That is where MIT's expertise comes in. Barzilay's team is running comparisons of many available models to see how much AI is actually helping the drug-discovery and design process.

Having more stringent standards will help researchers better understand the promise of machine-learning technology, Barzilay says. And it will also speed along new and improved drugs to patients—precisely the goal of AI-assisted drug design, she says.

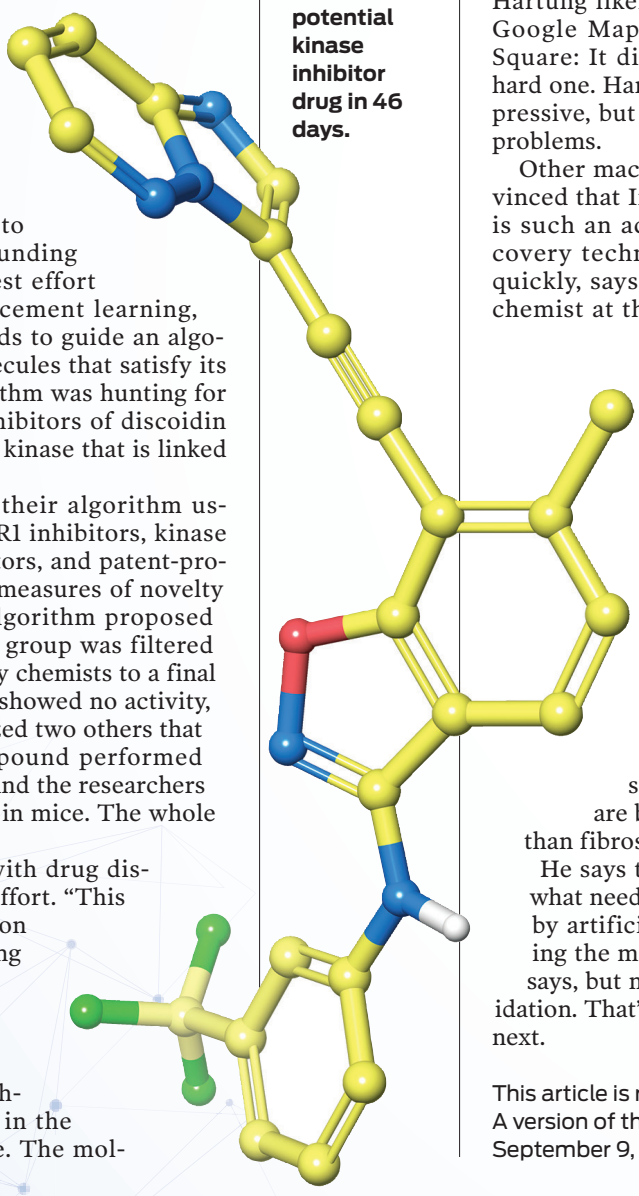
AI identifies kinase drug candidate in weeks. But can it do the same for harder targets?

The artificial intelligence start-up Insilico Medicine has used machine learning to find credible drug candidates in a matter of weeks (*Nat. Biotechnol.* 2019, DOI: 10.1038/s41587-019-0224-x). Experts say it's an important demonstration of what machine learning can do in drug discovery, but it isn't a revolution.

Insilico Medicine has been showing off its progress in teaching computers to find new drugs since its founding in 2014. The company's latest effort involves generative reinforcement learning, a technique that uses rewards to guide an algorithm as it searches for molecules that satisfy its goals. In this case, the algorithm was hunting for small molecules that are inhibitors of discoidin domain receptor 1 (DDR1), a kinase that is linked to fibrosis.

The researchers trained their algorithm using databases of known DDR1 inhibitors, kinase inhibitors, nonkinase inhibitors, and patent-protected molecules. Based on measures of novelty and DDR1 inhibition, the algorithm proposed 30,000 potential drugs. This group was filtered by computer programs and by chemists to a final six candidates. Two of them showed no activity, but the researchers synthesized two others that showed promise. One compound performed well against kinase screens, and the researchers tested its metabolic stability in mice. The whole process took 46 days.

Some chemists familiar with drug discovery and AI applaud the effort. "This is an impressive demonstration of rapid hit expansion starting from a pool of known kinase inhibitors," says Connor Coley, a computational chemist at the Broad Institute of MIT and Harvard. But he and others agree that the caveat is in the second half of that sentence. The molecules that the algorithm identified look similar to other kinase inhibitors, and Insilico Medicine had a lot of public information about similar compounds to train it on, notes Ingo Hartung, director of medicinal chemistry at Merck KGaA. Hartung likens what the algorithm did to using Google Maps to find New York City's Times Square: It did a good job, but the job wasn't a hard one. Hartung says the process's speed is impressive, but he wants to see it tested on harder problems.



AI identified this potential kinase inhibitor drug in 46 days.

CREDIT: INSILICO MEDICINE

Other machine-learning experts are less convinced that Insilico Medicine's 46-day time line is such an achievement. Traditional drug-discovery techniques might have worked just as quickly, says Olexandr Isayev, a computational chemist at the University of North Carolina at Chapel Hill. The researchers don't provide a baseline for comparison. Without that, adds Ash Jogalekar, a medicinal chemist at the AI-oriented biotech firm Strateos, "it's thus impossible to know whether the results attributed to the technique are unique in any way or not."

Alex Zhavoronkov, Insilico Medicine's founder and CEO, says the group has already tested the method on more challenging problems but hasn't made the results public. Zhavoronkov also says the hits found in this research are being tested in disease models other than fibrosis.

He says the current work is a small piece of what needs to be done to make drug discovery by artificial intelligence successful. Identifying the molecules is important, Zhavoronkov says, but molecules are worthless without validation. That's what Insilico Medicine plans to do next.

This article is reprinted with permission from C&EN. A version of this article was published in C&EN on September 9, 2019, on page 11.

ACS
Chemistry for Life®

American Chemical Society

ACS INDUSTRY MEMBER PROGRAMS

help you stay connected to a vibrant community of industry scientists committed to advancing the field of chemistry.

SUBSCRIBE TO THE **ACS INDUSTRY MATTERS NEWSLETTER**

Featuring exclusive ACS interviews with chemical industry executives, career consultants, and other industry leaders who provide valuable insights to help industrial chemists advance their careers.

Preview Industry Matters Content & Subscribe at
acs.org/industrynl

JOIN THE **ACS INNOVATION HUB LINKEDIN GROUP**

Connect, collaborate, and stay informed about the trends leading chemical innovation, with content and conversation curated by ACS.

Join at
bit.ly/ACSinnovationhub

NOMINATE YOUR PEERS FOR THE **ACS HEROES OF CHEMISTRY AWARD**

Recognizing industrial chemical scientists who had an integral role in their company successfully commercializing a product benefitting humankind. Nominations are open Nov. 1, 2019 – Feb. 1, 2020.

Nominate at
acs.org/chemheroes

Visit acs.org/industry for more from ACS Industry Member Programs

READING LIST

Our picks of the hottest new literature on AI in drug discovery

2019

» Zhavoronkov, Alex, Yan A. Ivanenkov, Alex Aliper, Mark S. Veselov, Vladimir A. Aladinskiy, Anastasiya V. Aladinskaya, Victor A. Terentiev, et al. **"Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors."** *Nat. Biotechnol.* 37, no. 9 (Sept. 2019): 1038–40.

» Coley, Connor W., Dale A. Thomas III, Justin A. M. Lummiss, Jonathan N. Jaworski, Christopher P. Breen, Victor Schultz, Travis Hart, et al. **"A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning."** *Science* 365, no. 6453 (Aug. 9, 2019): eaax1566.

» Vamathevan, Jessica, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, et al. **"Applications of Machine Learning in Drug Discovery and Development."** *Nat. Rev. Drug Discov.* 18, no. 6 (June 2019): 463–77.

» Gambini, Luca, Carlo Baggio, Parima Udompholkul, Jennifer Jossart, Ahmed F. Salem, J. Jefferson P. Perry, and Maurizio Pellecchia. **"Covalent Inhibitors of Protein-Protein Interactions Targeting Lysine, Tyrosine, or Histidine Residues."** *J. Med. Chem.* 62, no. 11 (June 13, 2019): 5616–27.

» Brown, Nathan, Marco Fiscato, Marwin H. S. Segler, and Alain C. Vaucher. **"GuacaMol: Benchmarking Models for De Novo Molecular Design."** *J. Chem. Inf. Model.* 59, no. 3 (March 25, 2019): 1096–1108.

» Sieg, Jochen, Florian Flachsenberg, and Matthias Rarey. **"In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening."** *J. Chem. Inf. Model.* 59, no. 3 (March 25, 2019): 947–61.

» Topol, Eric. ***Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.*** New York: Basic Books, 2019.

» Sharma, Yugal. **"Data Quality: The Not-So Secret Sauce for AI and Machine Learning."** *CAS: Where Science and Strategy Converge* (blog), March 8, 2019.

» Wang, Qianwen, Yao Ming, Zhihua Jin, Qiaomu Shen, Dongyu Liu, Micah J. Smith, Kalyan Veeramachaneni, and Huamin Qu. **"ATMSeer: Increasing Transparency and Controllability in Automated Machine Learning."** Preprint, submitted Feb. 13, 2019.

» Esteva, Andre, Alexandre Robicquet, Bharath Ramsundar, Volodymyr Kuleshov, Mark DePristo, Katherine Chou, Claire Cui, Greg Corrado, Sebastian Thrun, and Jeff Dean. **"A Guide to Deep Learning in Healthcare."** *Nat. Med.* 25, no. 1 (Jan. 2019): 24–29.

2018

» Jordan, Allan M. **"Artificial Intelligence in Drug Design—the Storm before the Calm?"** *ACS Med. Chem. Lett.* 9, no. 12 (Dec. 13, 2018): 1150–52.

» Gao, Hanyu, Thomas J. Struble, Connor W. Coley, Yuran Wang, William H. Green, and Klavs F. Jensen. **"Using Machine Learning to Predict Suitable Conditions for Organic Reactions."** *ACS Cent. Sci.* 4, no. 11 (Nov. 28, 2018): 1465–76.

» Smith. Justin S., Adrian E. Roitberg, and Olexandr Isayev. **"Transforming Computational Drug Discovery with Machine Learning and AI."** *ACS Med. Chem. Lett.* 9, no. 11 (Nov. 8, 2018): 1065–69.

» Donner, Yoni, Stéphane Kazmierczak, and Kristen Fortney. **"Drug Repurposing Using Deep Embeddings of Gene Expression Profiles."** *Mol. Pharmaceutics* 15, no. 10 (Oct. 1, 2018): 4314–25.

» Sellwood, Matthew A., Mohamed Ahmed, Marwin H. S. Segler, and Nathan Brown. "**Artificial Intelligence in Drug Discovery."** *Future Med. Chem.* 10, no. 17 (Sept. 2018): 2025–28.

» Popova, Mariya, Olexandr Isayev, and Alexander Tropsha. **"Deep Reinforcement Learning for De Novo Drug Design."** *Sci Adv.* 4, no. 7 (July 2018): eaap7885.

» Rifaioglu, Ahmet Sureyya, Heval Atas, Maria Jesus Martin, Rengul Cetin-Atalay, Volkan Atalay, and Tunca Doğan. **"Recent Applications of Deep Learning and Machine Intelligence on In Silico Drug Discovery: Methods, Tools and Databases."** *Brief. Bioinform.* (July 31, 2018): bby061.

» Coley, Connor W., William H. Green, and Klavs F. Jensen. **"Machine Learning in Computer-Aided Synthesis Planning."** *Acc. Chem. Res.* 51, no. 5 (May 15, 2018): 1281–89.

» Ching, Travers, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, et al. "**Opportunities and Obstacles for Deep Learning in Biology and Medicine."** *J. R. Soc., Interface.* 15, no. 141 (April 30, 2018): 20170387.

» Segler, Marwin H. S., Mike Preuss, and Mark P. Waller. **"Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI."** *Nature* 555, no. 7698 (March 29, 2018): 604–10.

» Gómez-Bombarelli, Rafael, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. **"Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules."** *ACS Cent. Sci.* 4, no. 2 (Feb. 28, 2018).

Note: This list was chosen by scientists that work in the field, editors of ACS journals, and C&EN editorial staff.

Download a copy of this and future reports at www.acs.org/discoveryreports