In many ways, developing a new drug is like running a marathon – except that it can take more than a decade. But now, scientists are using powerful computers, large quantities of data and intelligent algorithms to accelerate pharmaceutical research. Computer scientists are engaged in every phase of drug development, from the initial idea through to clinical testing.

Computers are playing an increasingly important role for experimental chemists in many areas, such as the search for new cancer drugs. Using computational methods, these scientists can actually design chemical structures on the computer so that the resulting substance should be able to attack a desired molecular target. Rogaratinib, a new cancer drug candidate that is currently in development at Bayer, is the most recent example illustrating the important contributions that computer-aided science can make to successful drug development. “About one-fourth of Bayer’s pre-clinical pipeline comes from lead structures discovered using computer-aided methods,” explains Professor Alexander Hillisch, head of Computational Molecular Design in Wuppertal. By analyzing large quantities of data with intelligent algorithms, computer scientists support all phases of drug development, from the discovery and optimization of new active substances and
BAYER RESEARCH / DECEMBER 2019

Computer-aided drug design: In order to find suitable drugs for disease-causing proteins, Bayer employees such as Dr. Ulrich Lücking and Lara Kuhnke analyze these molecules in 3D.

Identification of a molecular target is the first step of drug discovery

Pharmaceutical research almost always begins with a molecular target protein whose function the scientists aim to modulate with a new drug. This protein is likely to play a key role in the development of the disease, making it a suitable target for controlling, alleviating and or even curing the disease. Many types of cancer are caused by defects in a specific signaling pathway, which is where the new drug to be developed should intervene. Rogaratinib, for example, targets tumor cell activity.

"During the development of this drug, we used computational methods to virtually piece together chemical fragments into a structure that would bind precisely to the disease-promoting protein. This compound and many other modifications were then synthesized in the laboratory, and experiments show that they do actually bind to the target protein in the way we intended – that’s where they block its function,” says Dr. Mario Lobell, a computational chemistry expert at Bayer. Since this strategy requires accurate prediction of the physical interactions between molecules, computational chemists also refer to them as “physics-based methods”.

Another approach is based on large data quantities; researchers evaluate the results of past experiments with substances and try to identify new targets for other drug candidates. For about 15 years now, scientists at Bayer have been working on a platform to predict molecular properties such as uptake, metabolism and toxicity aspects.

Using the company’s own data and artificial intelligence algorithms, and in collaboration with the California-based specialty software company SimulationsPlus, the scientists at Bayer recently achieved a breakthrough in the prediction of the charge states of active ingredients. The charge state of compounds is a key aspect in optimizing molecular prop-
How computer processes accelerate drug development

From the idea to the approved drug, computer-aided data analyses support the entire process. Today, computers generate a steady stream of potential active substances, which Bayer researchers then synthesize and systematically examine in the laboratory.

Molecular target
Researchers use computational methods to analyze the spatial structure of the target molecule, such as a disease-causing protein.

Drug candidate
Digital scientists collaborate with their colleagues in the experimental field to design a completely new suitable active substance using a computer (A) or search for it based on data from the laboratory (B).

Preclinical testing
Researchers translate data from an animal model to humans using a computer, allowing them, among other things, to make proposals for the appropriate dose for a drug candidate. (Small patient group of human volunteers)

Clinical trial
Computational scientists simulate clinical trials to design the actual trials in a manner that delivers meaningful results. (Large patient group)

Approved drug
Using computer-aided methods and protecting patient data

Digital drug researchers generate a huge amount of data, especially in clinical trials. “Protecting this sensitive data is extremely important to us. We go to great lengths with new products to first check whether we are even allowed to use patient data from our own historical clinical trials for this new research,” says Dr. Jörg Lippert, head of Clinical Pharmacometrics in the Pharmaceuticals Division at Bayer. Only after all legal issues have been definitively clarified do the researchers begin their computer-aided analyses.

Properties such as how a substance dissolves and how well it penetrates membranes. The leading global role of Bayer and SimulationsPlus in this field is underscored by the results of the SAMPL6 Challenge, a scientific community-wide blind test to predict molecular properties using computational methods. “In two of the three categories, the jointly developed approach was the best out of 32 tested methods at predicting the experimental data; we were clearly the winner of the challenge with this approach,” says Hillisch. “The algorithms for machine learning are evolving and this together with the new data that we constantly receive from our in-house experiments enables us to progressively improve our predictive models.”

This new strategy is transforming the computer into an indispensable tool. As Hillisch remarks, “Computers help us decide whether a chemical structure could be promising. This often means that we also reject some ideas to focus costly laboratory work on the more promising approaches instead.” That is worthwhile in and of itself, since “a wrong decision can sometimes mean that we send a chemist on a six-month wild-goose chase with no end results.”

The researchers view all results produced using computational methods with caution; after all, they are working with models that, as a simplified representation of reality, have fundamental flaws. However, everyone working in this field is well aware of that. The decisions are made together with the experimental chemists who make the molecules in the lab, which means that both partners share the joys of success – and the failures.

Researchers screen billions of virtual molecular structures

The Bayer substance library and commercial databases contain millions of substances for their tests, ensuring that the researchers never lack for ideas. They are often able to find molecules that bind to the target structure, even if they are only weak-binding molecules initially. That is when the computational and experimental chemists begin their optimization work. “It’s important to make a selection. Many compounds cannot be developed into a medicine because, for example, the active substance is poorly absorbed in the body. That’s why we do computer-aided searches for alternatives that fulfill the same biological function, but should have better properties,” explains Hillisch. To conduct these searches, the drug researchers access databases containing millions or even billions of virtual molecular structures that can be easily synthesized. “With our computers, we can screen 100 million molecules and get valuable suggestions for new active ingredient prototypes during the time it takes you to drink a cup of coffee.”

The immense processing power of new computers, sophisticated algorithms and the constantly improving underlying data all facilitate this process. One challenge the researchers are increasingly facing is the flexibility of the proteins they want to target with the new active substances. “Previously, we could only rigidly model target proteins as we were limited by our computing power. Today, our simulations are far more realistic, and we are also able to model the natural
movement of proteins,” explains Dr. Clara Christ, head of Computational Molecular Design Berlin. The researchers are collaborating with academic institutions in this area so that they can quickly integrate findings from basic research into their modeling. For example, Christ and Dr. Katharina Meier, a computational chemistry expert, recently set up an industrial consortium comprising many pharmaceutical companies and a number of leading universities from the United States, with the aim of developing improved computational methods for characterizing interactions between target proteins and active ingredients. As project manager Meier explains, “Quantum chemical calculations on super computers and machine learning are just two of the methods we are using to take the prediction of the binding behavior of active substances to target proteins to the next level.”

Since May 2019, the two Computational Molecular Design departments run by Hillisch and Christ are part of the newly formed Digital Technologies institute, enabling even closer cooperation between adjacent teams. Hillisch and Christ are also working closely with their colleagues in active substance development at Bayer’s Crop Science Division, all using the same IT infrastructure. There is one key difference, however: In contrast to the scientists in the Pharmaceuticals Division, Crop Protection researchers generally do not describe the effect of an active substance in terms of its binding behavior to a target protein; rather, they investigate specific phenomena to determine the behavior, for example looking at the visible changes in the organism of the whole plant. The three-dimensional structures of target proteins are often not yet known in species other than humans, so plant researchers have to use alternative methods to describe structure-activity relationships in these cases.

Drug discovery for humans and plants

What is known as rational drug design is now also being increasingly used in crop protection. This means that a molecule must match the target structure, which is often a protein. The Bayer scientists are therefore confident that in the future these limitations will be overcome with more data and protein structure-based design playing an increasingly significant role in crop science as well.

In pharmaceutical research, it usually takes more than five years of thorough in vitro and in vivo testing before a new active substance can be tested in humans. “We translate the results from animal experiments concerning tolerability and efficacy to predict the scenario in the human organism,” says Jörg Lippert, a clinical scientist at Bayer who works on pharmacometric computer models. He is researching how the human body processes drugs. As part of this, Lippert and his team incorporate numerous contributing factors, which makes their analyses complex and computationally intensive.

Powerful computers and server centers facilitate the researchers’ work. “We carry out complex calculations in the cloud,” explains Lippert, which allows today’s computational pharmacologists to solve complex problems with a simple laptop and good internet connection. The researchers intend to use intelligent algorithms and statistical methods to gain new insights from old data. Concepts such as artificial intelligence, machine learning and neural networks play a decisive role in their daily work (see text box 2).

Does this mean that digital pharmacology will soon be advanced enough that laboratory tests and clinical trials with patients are no longer necessary? Lippert doubts that this will be the case. “I think in 20 years we’ll still be testing a great deal in the clinical setting.” A reason for this is that clinical data sets are often poorly structured and extremely diverse, ranging from tables with blood results to x-rays. That makes it difficult to automate and standardize analyses.

Computer models facilitate the pathway from the laboratory to the clinical setting

Nonetheless, pharmacokinetic computer models played a role in all Bayer drugs that are currently undergoing clinical trials. “Such models can be based on simple formulas,” adds Lippert. Take, for example, the dosage of a drug candidate when translating from an animal model to a clinical trial in humans. The appropriate

How a computer learns

Artificial intelligence: Artificial intelligence refers to the capability of a machine to perform intellectual tasks such as learning or problem-solving. Well-known fields of application for artificial intelligence include driverless vehicles and banks where intelligent computer algorithms detect unusual money movement.

Machine learning: In this area of artificial intelligence, computers learn from data, meaning the computer uses statistical methods to identify patterns. This allows computer scientists to make predictions about future events based on available data.

Artificial neural networks: The base unit of this network is an artificial neuron, which sends a signal to the next neuron downstream when the strength thereof exceeds a certain level. Several base units are connected in a network, which convert an input signal to an output signal. Artificial neurons are dynamic, meaning they can change their activity over time, which allows the network to “learn”.

"I think in 20 years we’ll still be testing a great deal in the clinical setting.” A reason for this is that clinical data sets are often poorly structured and extremely diverse, ranging from tables with blood results to x-rays. That makes it difficult to automate and standardize analyses.

Computer models facilitate the pathway from the laboratory to the clinical setting

Nonetheless, pharmacokinetic computer models played a role in all Bayer drugs that are currently undergoing clinical trials. “Such models can be based on simple formulas,” adds Lippert. Take, for example, the dosage of a drug candidate when translating from an animal model to a clinical trial in humans. The appropriate

How a computer learns

Artificial intelligence: Artificial intelligence refers to the capability of a machine to perform intellectual tasks such as learning or problem-solving. Well-known fields of application for artificial intelligence include driverless vehicles and banks where intelligent computer algorithms detect unusual money movement.

Machine learning: In this area of artificial intelligence, computers learn from data, meaning the computer uses statistical methods to identify patterns. This allows computer scientists to make predictions about future events based on available data.

Artificial neural networks: The base unit of this network is an artificial neuron, which sends a signal to the next neuron downstream when the strength thereof exceeds a certain level. Several base units are connected in a network, which convert an input signal to an output signal. Artificial neurons are dynamic, meaning they can change their activity over time, which allows the network to “learn”.

"I think in 20 years we’ll still be testing a great deal in the clinical setting.” A reason for this is that clinical data sets are often poorly structured and extremely diverse, ranging from tables with blood results to x-rays. That makes it difficult to automate and standardize analyses.

Computer models facilitate the pathway from the laboratory to the clinical setting

Nonetheless, pharmacokinetic computer models played a role in all Bayer drugs that are currently undergoing clinical trials. “Such models can be based on simple formulas,” adds Lippert. Take, for example, the dosage of a drug candidate when translating from an animal model to a clinical trial in humans. The appropriate

How a computer learns

Artificial intelligence: Artificial intelligence refers to the capability of a machine to perform intellectual tasks such as learning or problem-solving. Well-known fields of application for artificial intelligence include driverless vehicles and banks where intelligent computer algorithms detect unusual money movement.

Machine learning: In this area of artificial intelligence, computers learn from data, meaning the computer uses statistical methods to identify patterns. This allows computer scientists to make predictions about future events based on available data.

Artificial neural networks: The base unit of this network is an artificial neuron, which sends a signal to the next neuron downstream when the strength thereof exceeds a certain level. Several base units are connected in a network, which convert an input signal to an output signal. Artificial neurons are dynamic, meaning they can change their activity over time, which allows the network to “learn”. 
ate drug dose is often determined using what is known as the allometric principle, which is based on the fact that regardless of whether a single cell or blue whale, the metabolism of an organism does not increase relative to weight, but rather with a lower potency that can also be derived from theoretical considerations.

Digital drug discoverers are now developing these formulas and empirical values into extremely complex models. For example, they can simulate blood coagulation using a computer model. “We include many parameters to do this, such as the reactions of the different coagulation factors with each other, the role of the cells in blood and the surrounding tissue.” Scientists can even model deposits in blood vessels and the impact of drugs, for example acetylsalicylic acid, the active ingredient of Aspirin™, on the vascular system. The computer model can help to predict, for example, the optimal way of switching a patient’s treatment from one blood-thinning substance to another product. “That is important since different medications do not have the same speed of onset of action and do not all remain effective for the same length of time. We therefore have to adjust the starting dose so that at any given time, there is not too much or too little anticoagulation,” explains Lippert.

Before a new active substance reaches the home stretch and is granted approval, researchers investigate its efficacy and side effects in large groups of patients. That necessitates different methods. “We then look at a complete population of patients,” says Lippert. Researchers also initially conduct a complete computer simulation of these large clinical trials, which has several advantages. "We can calculate the effect-specific characteristics of a patient population have on the result – for example, if we expect that a clinical trial population will have mainly overweight or underweight patients, or patients with reduced kidney function.

**Computer modeling optimizes the planning of clinical trials**

In addition, scientists examine the potential effects of other factors such as if clinical trial participants forget to take the study drug. Ultimately, the computer models influence the design of the real clinical study. For instance, the researchers determine when a blood test should be conducted and what blood levels are to be expected at a given time.

Such considerations are particularly important in clinical trials with children or other highly vulnerable patients so that they are preferably put under minimal stress and not exposed to any risk when, for example, they have blood drawn. As Lippert explains, “It is our ethical duty to do everything we can to eliminate risks, while reducing the burden for the clinical trial participants to a minimum. Computer simulations with their quantitative predictions help to ensure that we succeed in living up to this responsibility.”

Another aspect is the natural variation of medical measurements. The larger the deviations expected by drug developers, the larger the patient group has to be. Only with a larger patient group can the researchers make a statistically sound statement. “We have to choose the number of clinical trial participants in a manner that ensures the results are not simply left to chance,” says Lippert. Scientists refer to this as a “power analysis”. Only valid clinical trials ultimately lead to the approval of a new drug.

In the case of rogaratinib, everything is going according to plan thus far. As Hillisch explains, “The efficacy of this active substance is now being tested on patients in a Phase 2 clinical trial.” That will soon show whether the initial idea can make it all the way to becoming a new drug product.

> With our computers, we can screen 100 million molecules and find suitable lead structures during the time it takes you to drink a cup of coffee.

Professor Alexander Hillisch