c&en | WHITEPAPERS

THE INFLUENCE OF ARTIFICIAL INTELLIGENCE ON RETROSYNTHESIS

INTRODUCTION

One of the biggest challenges facing researchers today is the increased demand for novel therapeutics. We continue to face new and unprecedented diseases, as most aptly highlighted in the global pandemic caused by the novel coronavirus SARS-CoV-2. It has spurred a flurry of research into developing a COVID-19 vaccine or other prophylactic measures, or increasing the supply of effective treatments. In both cases, retrosynthesis is crucial to bringing effective products to the market.

When faced with synthesizing a specific compound, such as a drug, many chemists start with the end product itself. They then try to identify the individual building blocks and reaction steps needed to make the whole compound. This retrosynthesis process helps chemists develop a synthetic path from available precursors to a target molecule.

“Retrosynthetic planning is a key skill set in any medicinal chemistry lab,” says Jakob Magolan, associate professor of chemistry at McMaster University. “It’s an essential part of making any new compounds.” He compares retrosynthesis to using a map on a road trip: it will help you identify and design a route from your starting point to your final destination.

Yet, it can be a challenge to identify the chemical processes required to synthesize potential therapeutics and other compounds. Chemistry is complicated. Every atom in a chemical reaction can have an impact, so the journey from available starting materials to final compound can seem tedious or even impossible.

At its core, retrosynthesis requires chemists to systematically examine a given compound and work out the process to synthesize it step by step. The first attempt at using computers to aid and accelerate the retrosynthesis process was made in the 1980s and represented a great leap forward in terms of automation.[1] However, computers then had relatively low computing capacity compared with those in laboratories today, and the approach was limited.

Over the years, several computer-supported retrosynthesis resources have become available to help chemists identify synthetic pathways. Ewa Gajewska, a chemistry and software specialist for Merck KGaA, Darmstadt, Germany, says most of these platforms automatically extract their knowledge from reactions that have been published in the literature. The Synthia™ retrosynthesis software, offered by MilliporeSigma, was originally developed in an academic laboratory and uses advanced algorithms to provide synthesis solutions independent of what has appeared in the literature.

Computer-guided retrosynthesis design can save time for researchers by automatically producing synthesis pathways from their desired molecular end product to commercially available building blocks. Other factors—including cost, alternative synthesis strategies, and starting material availability—can be incorporated into the retrosynthesis algorithm (Figure 1).

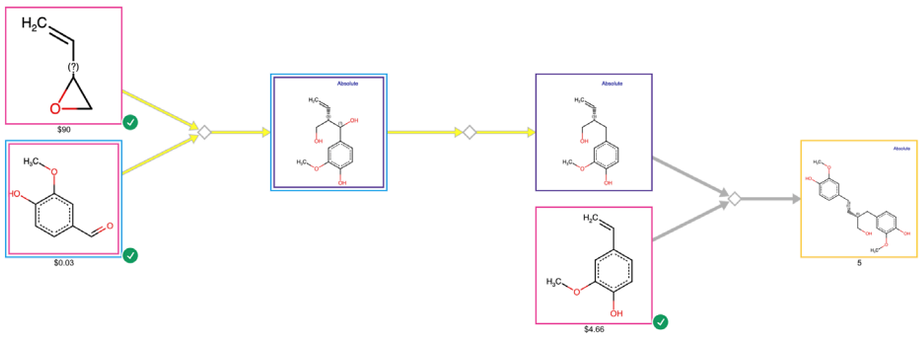

Figure 1. Synthia™ retrosynthesis software uses commercially available building blocks (left, pink boxes) to generate a synthesis route to a desired molecular product (right, yellow box) and includes intermediate reaction steps and expected products. Unknown chemicals are denoted with purple boxes, while blue boxes surrounding the pink and purple boxes indicate that a protecting group is needed for that step in the synthesis.

Image credit: MilliporeSigma

"No matter how well you know the literature, it's not possible for a human to know the ins and outs of every single reaction," says Timothy Cernak, assistant professor of medicinal chemistry and chemistry at the University of Michigan. "We also tend to be biased towards certain types of synthesis solutions, and our biases are dynamic. They change based on the papers we most recently read, for example." Computer-aided retrosynthesis design helps overcome some of these barriers to give researchers a more comprehensive, unbiased look at the synthesis solutions available to them.

RETROSYNTHESIS PLANNING, STEP BY STEP

Automated retrosynthesis planning starts with a given molecule and identifies synthesis pathways that trace back to commercially available building blocks. The Synthia™ retrosynthesis software can also take into account chemical features like functional groups and stereochemistry when delivering viable pathway solutions. Once the results are in, the user can navigate the proposed pathways and filter the results using a number of criteria, including the similarity of reaction sequences, the number of reaction steps, and the price of starting materials. Users can also upload their own chemical inventory so that the software can plan pathways back to what they already have in the lab (Figure 2). Synthia™ will highlight molecules that need to be protected, and suggest protecting groups that are compatible with the reaction conditions.

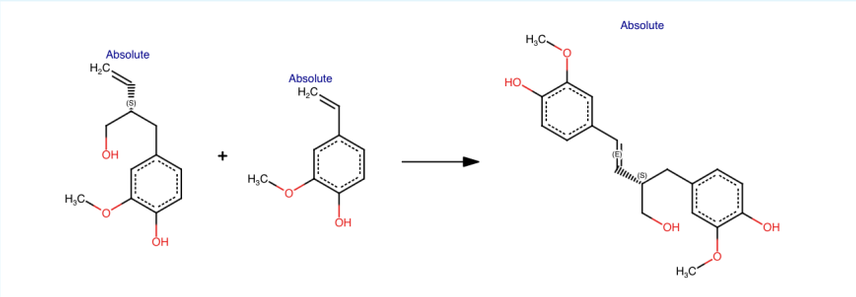

Figure 2. Within any given synthesis route proposed by Synthia™ retrosynthesis software, users can analyze the reaction step details. These include the structures of starting and ending products, existence of any similar reaction chemistries, and the possibility of unwanted or off-target side reactions. This detailed information helps users better identify a suitable retrosynthesis route.

Image credit: MilliporeSigma

Users also have the option to forgo the automated planning process and use Synthia™ retrosynthesis software manually. That gives them the freedom to navigate through any given retrosynthesis pathway, starting with the target molecule and moving iteratively backward based on user preferences. At each step, the software includes reaction details and references. For more information about Synthia™ retrosynthesis software, visit: https://www.sigmaaldrich.com/synthia

ARTIFICIAL INTELLIGENCE IN CHEMISTRY

Magolan and Cernak are among the researchers who have turned to software to speed up and improve their retrosynthesis planning. In the search for synthesis plans, software programs can analyze millions of potential starting materials, intermediates, and reactions, and enable a more thorough analysis of all synthetic possibilities for a given molecule. “Removing some of the tedium of routine literature-searching normally involved in the design of retrosynthesis pathways gives you the ability to ask bigger research questions,” Cernak says.

Software programs may rely on artificial intelligence (AI) to enhance retrosynthesis planning and AI can be applied in different ways. For example, in machine learning, computer algorithms improve as they accrue experience analyzing data sets. In a rule-based system—one of the simplest and oldest forms of AI—algorithms are defined by a set of coded rules.

In the early days of AI, in the 1960s, rule-based systems were applied to simple, restricted domains. For example, the AI systems Dendral and Prospector were designed to analyze the chemical content and geology of rocks and minerals, both of which could be applied to study the soil on Mars.[2,3]

A rule-based AI system mimics the reasoning of a human expert solving a problem; its knowledge comes from a set of rules telling the software what to do or what to conclude in different situations. As technology has advanced, AI has been applied to more complex problems, including chemical synthesis. Synthia™ retrosynthesis software is a rule-based system and derives its knowledge from a database of over 100,000 expert-coded chemical reaction rules.[4]

Sarah Trice, head of cheminformatics technologies for MilliporeSigma, was a member of the original team that vetted Synthia™ retrosynthesis software and helped bring it to the market. Trice says this rule-based approach allows the software to remove pathways containing proposed reactions that wouldn't actually work in the lab, which saves chemists' time.

The power of rule-based systems comes from the fact that all the software's rules are coded with a specific outcome in mind. In the case of retrosynthesis planning, the rules are coded by expert chemists. "The advantage of a software like Synthia™ is the database of hand-coded chemical reactions that doesn't exist anywhere else," Cernak says. "You have a large percentage of the chemistry literature encoded in a searchable format."

Once the reaction rules are established, they don't need to be modified again. However, with retrosynthesis software, new chemistries that may be discovered can quickly be added to the system to expand its knowledge.

Another benefit of a rule-based approach is that it is not limited to the existing literature. Gajewska, who was part of the team that developed the Synthia™ retrosynthesis software,[4] says rule-based systems can generate feasible synthetic pathways that haven't been previously published.

A comprehensive look at possible retrosynthesis pathways can lead to novel solutions for chemical process development, according to Chaomin Li, a principal scientist of chemical process development at Biogen. These solutions are more difficult to arrive at when researchers are combing through the literature themselves or are concentrating on a particular subset of reaction chemistry.

ACCELERATING MEDICINAL CHEMISTRY

The COVID-19 pandemic has highlighted the need for quick and efficient drug development and manufacturing capabilities. The process of synthesizing potential therapeutic candidates for clinical testing has historically been a significant bottleneck, partly because of the time it takes to identify viable synthesis pathways and starting materials.

To further complicate matters, successful manufacture of a drug in high demand relies on adequate and sufficient production of all the component parts involved in the synthesis pathway. Depletion of one of these key buildings blocks could result in supply shortages. Automated retrosynthesis design—including software like Synthia™—helps researchers navigate both these considerations.

• *Speeding up drug discovery*

Magolan says many therapeutic candidates can now be designed computationally, before being synthesized and experimentally validated. Identifying which of these compounds can be easily and rapidly synthesized is crucial to accelerating the drug discovery pipeline. Narrowing down lists of potential drug candidates is particularly helpful for medicinal chemists like Magolan, who runs a synthetic organic chemistry laboratory focused on drug discovery and preclinical development.

After the COVID-19 outbreak, Magolan began working with colleagues to leverage computational strategies to design hundreds of new potential antiviral drug candidates for SARS-CoV-2. The researchers then have to decide which compounds to move to the next steps of synthesis and experimental validation.

"It would be time consuming to manually evaluate the potential synthesis pathways for each unique proposed compound," Magolan says. To avoid a logjam, Magolan and colleagues used Synthia™ retrosynthesis software. "Every compound that comes through this computational drug discovery project goes through Synthia,™" he says. "It quickly becomes clear which compounds may be the easiest to make."

Automated retrosynthesis planning gives researchers the ability to test and validate compounds faster. The acceleration helps scientists better respond to public health crises that demand novel therapeutics.

• *Improving process chemistry*

Beyond drug discovery, automated retrosynthesis planning can aid in improving the manufacturing of therapeutics. In the context of COVID-19, any drug that proves to be an effective treatment will likely face demand that outweighs supply. Moreover, many drug synthesis processes are patented or kept confidential to avoid competition. However, during a global pandemic, public health workers need multiple supply lines of the same drug.

The Cernak Lab recently worked on a project to determine how automated retrosynthesis planning could help the design of 12 different SARS-CoV-2 therapeutic compounds.[5] Synthesis protocols for each already exist but rely on established starting material supply chains. Recognizing that a pandemic could strain these supply chains, Cernak wanted to see if Synthia™ retrosynthesis software could generate new protocols that used alternative starting products (Figure 3).

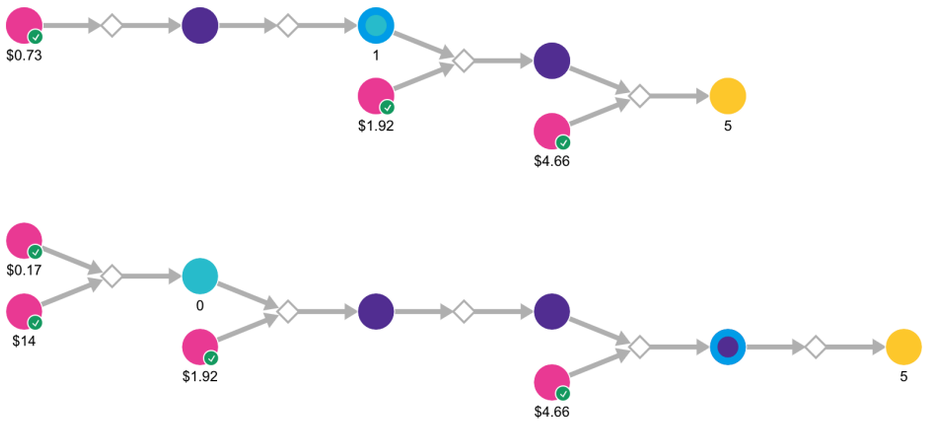

Figure 3. Synthia™ retrosynthesis software can generate multiple synthesis pathways to the same target molecule. The number of reactions steps, cost, and commercial availability of the starting products and intermediates can be assessed at each point along the synthesis route. This allows researchers to assess multiple synthesis routes to identify which fit best within their desired parameters.

Image credit: MilliporeSigma

Using computer-aided retrosynthesis analysis, Cernak and his team sifted through the almost infinite number of possible synthesis pathways to their drug compounds, seeking routes that were cost-competitive with commercial routes but initiated from alternate starting materials. "Pulling this off 12 times over for each of the compounds was a really big data challenge," Cernak says. "We had to customize our search parameters within Synthia™, but we were able to get results for all of the compounds." For example, Cernak was able to limit the search results to include only those synthesis routes that were similar in the number of steps and the cost to existing synthesis reactions.

In at least one case, the proposed synthetic pathway used lower-cost starting materials than those currently in use.[5] From a drug manufacturing perspective, Cernak's research suggests companies could mitigate supply chain issues that may arise during crises, such as a pandemic, using AI-assisted retrosynthesis. The work also notes ways to decrease manufacturing risk by providing alternate synthesis routes if the supply of any one product or starting material is disrupted.

Synthesis and manufacturing are necessary components of any drug development pipeline—during a pandemic or not. At Biogen, Li relies on retrosynthetic analysis to develop efficient manufacturing processes for active pharmaceutical ingredients. Developing these processes involves many considerations, including cost, safety, environmental impact, robustness, and quality control.

Li uses Synthia™ retrosynthesis software to look for pathways or starting materials that aren't obvious. Many of the considerations mentioned above can be taken into account when setting up the initial search criteria in the platform and allow researchers to narrow in on synthetic solutions that work within their parameters. "Innovative ideas inspired by Synthia™ can serve as a novel starting point for chemical process development," Li says.

EVOLUTION AND EXPANSION

Automated retrosynthesis software will undoubtedly continue to improve through the years as computer scientists and chemists apply AI strategies. One benefit of retrosynthesis software with large, hand-coded databases such as Synthia™ is the ability to go beyond traditional drug discovery and expand the chemical space available with synthetic exploration, according to Cernak. "Synthia™ will give you answers to really complex synthesis processes," he says. "It has the database and computational power to navigate an extensive number of reaction steps."

In the future, AI and computational chemistry could open new doors to chemical design and allow researchers to ask deeper, more fundamental questions about the molecules that make up the world.

REFERENCES

1. E. J. Corey, A. K. Long, and S. D. Rubenstein, "Computer-Assisted Analysis in Organic Synthesis," *Science* 228, no. 4698 (April 26, 1985): 408–19, https://doi.org/10.1126/science.3838594.
2. Buchanan, Bruce G., Sutherland, Georgia L., and Feigenbaum, Edward A. "Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry." Chapter 12 in Volume 4 of Machine Intelligence, edited by B. B. Meltzer and D. Michie. Edinburgh University Press (1969): 209-254.
3. Richard Duda, John Gaschnig, and Peter Hart, "Model Design in the PROSPECTOR Consultant System for Mineral Exploration," *Exp. Sys. Microelec. Age*., 1979, 153–68, DOI: 10.1016/B978-0-934613-03-3.50028-3 *Readings in Artificial Intelligence* (1981): 334–48, https://doi.org/10.1016/B978-0-934613-03-3.50028-3.
4. Ewa Gajewska et al., "Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses," *Chem*. 6, no. 1 (Jan. 9, 2020): 280–93, https://doi.org/10.1016/j.chempr.2019.11.016.
5. Yingfu Lin et al., "Reinforcing the Supply Chain of COVID-19 Therapeutics with Expert-Coded Retrosynthetic Software," *ChemRxiv* (preprint), https://doi.org/10.26434/chemrxiv.12765410.v1.

The life science business of Merck KGaA, Darmstadt, Germany operates as MilliporeSigma in the U.S. and Canada.