Leveraging AI and Automation for Sustainable Synthesis

Expert Insights

Sponsored by:

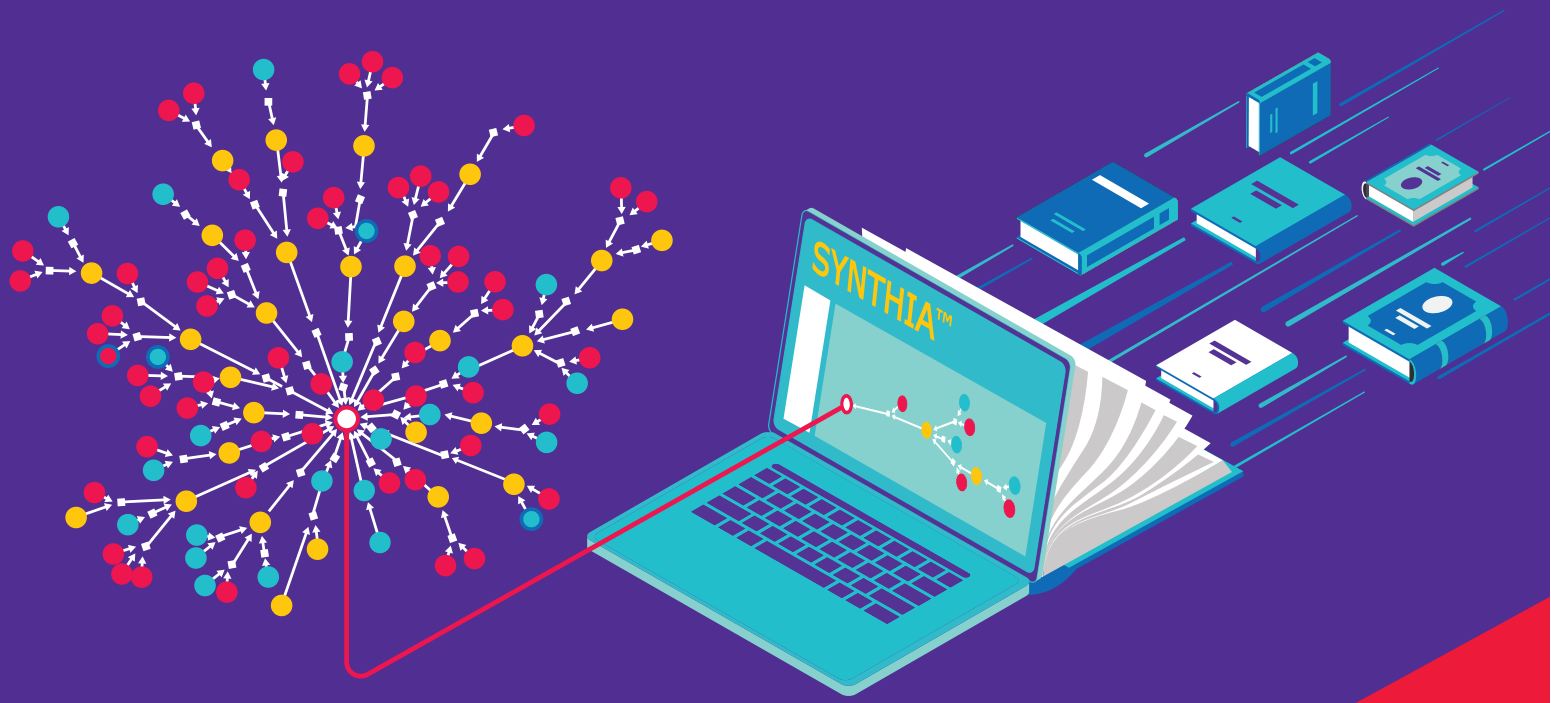
MERCK

WILEY

MERCK

GO BEYOND the literature

SYNTHIA™ Retrosynthesis Software is a unique web-based platform that revolutionizes the way organic chemists design pathways to complex targets. This decision-support tool is the strategic partner of synthetic organic chemists.



Pioneer Your Pathway
www.SynthiaOnline.com

© 2023 Merck KGaA, Darmstadt, Germany and/or its affiliates. All Rights Reserved. Merck, the vibrant M, Sigma-Aldrich and SYNTHIA are trademarks of Merck KGaA, Darmstadt, Germany or its affiliates. All other trademarks are the property of their respective owners. Detailed information on trademarks is available via publicly accessible resources.

MK_AD12813EN Vers. 01 09/2023

The Life Science business of Merck operates as MilliporeSigma in the U.S. and Canada.in the U.S. and Canada.

Sigma-Aldrich®
Lab & Production Materials

Contents

4 Introduction

6 Synthetic Automations: A Revolution from Stone Age to Modern Era
Adapted from Fang, G., Lin, D.-Z. and Liao, K.

9 Computational Analysis of Synthetic Planning: Past and Future
Adapted from Wang, Z., Zhang, W. and Liu, B.

12 Synergy Between Expert and Machine-Learning Approaches Allows for Improved Retrosynthetic Planning
Adapted from Badowski T., *et al.*

15 Computer-Assisted Design of Sustainable Syntheses of Pharmaceuticals and Agrochemicals from Industrial Wastes
Adapted from Le Pogam, P. *et al.*

18 Artificial-Intelligence-Driven Organic Synthesis—En Route towards Autonomous Synthesis?
Adapted from Empel, C. and Koenigs, R. M.

21 Unlocking the Chemistry of Tomorrow with Computer-Aided Synthesis Planning
Interview with Professor Tim Cernak

Imprint

© John Wiley & Sons, Inc.
111 River Street,
Hoboken, NJ 07030-5774
USA
Contact: Customer Service

Associate Editor:
Dr. Christene A. Smith

Senior Account Manager:
Jan Käppler

Merck KGaA, Darmstadt, Germany
www.sigmaaldrich.com/synthia

Cover: © 2023 Adobe Stock

Introduction

In the rapidly evolving landscape of scientific and industrial advancements, the quest for innovation is unending. As industries strive to create novel products and processes, two critical considerations emerge: safeguarding intellectual property (IP) and embracing sustainable practices. Within this dynamic context, the integration of computer-aided synthesis design (CASD) has emerged as a powerful tool for innovation.

Here we include several articles that delve into the synergy between cutting-edge CASD techniques and advancing innovation while fostering environmentally conscious practices. This eBook provides an overview of computer-aided synthesis design, past and present, along with examples of breakthrough developments and applications that align with sustainability and show how CASD, by leveraging advanced algorithms, data-driven insights, and predictive modeling, empowers researchers to streamline the design and optimization of chemical synthesis processes.

We also aim to shed light on the pivotal role CASD plays in promoting green chemistry—a discipline dedicated to minimizing the environmental impact of chemical processes. This eBook underscores how CASD can assist in the identification of environmentally benign reaction pathways, reduction of waste generation, and optimization of resource utilization. By embracing CASD as a green chemistry support tool, industries can not only enhance their sustainability profile but also contribute to a more ecologically balanced world.

Industrial processes often result in significant waste generation, contributing to pollution, landfills, and resource wastage. Traditional methods of waste management involve disposal or incineration, both of which have negative environmental consequences. However, the concept of a circular economy, where waste is considered a resource, is gaining traction. CASD plays a pivotal role in enabling this transition by facilitating the conversion of waste into valuable products, thereby reducing the demand for virgin resources. This is accomplished by combining the power of computer-aided design (CAD), computational chemistry, and sustainability metrics to guide researchers and engineers in designing eco-friendly synthesis processes that reduce waste generation and environmental impact.

In addition, we include several articles that address the evolution of computer-aided synthesis design and consider its future integration with automated synthesis. The future of automated synthesis holds immense promise, driven by advances in technology, artificial intelligence, and sustainability concerns. As we look ahead, several trends and developments are likely to shape the landscape of automated synthesis. The integration of artificial intelligence (AI) and machine learning (ML) techniques will play a pivotal role in automated synthesis. These technologies can analyze vast amounts of data, predict reaction outcomes, optimize reaction conditions, and propose novel synthesis routes. AI-driven systems can learn from both successful and failed reactions, leading to more efficient and innovative processes. Laboratories equipped with robotic systems and automated instruments allow researchers to conduct a large number of experiments simultaneously and will handle various tasks such as sample preparation, reagent addition, and data collection, freeing up researchers' time for more complex and creative tasks.

The future of chemical synthesis is marked by the integration of cutting-edge technologies, sustainability-driven approaches, and increased efficiency. As automation continues to revolutionize the field of chemistry, researchers and industries can anticipate a more streamlined, innovative, and environmentally conscious approach to chemical synthesis. When taking a holistic, ecological approach, Computer-Aided Sustainable Synthesis Design (CASSD) is positioned as a powerful tool that synergizes innovation and sustainability. Its application in transforming industrial waste into valuable products aligns with the principles of a circular economy and paves the way for a greener, more efficient future. As industries continue to adopt CASSD practices, we can anticipate a significant reduction in waste generation, energy consumption, and environmental impact, driving us closer to a truly sustainable industrial landscape.

R. Kopreski, Ph.D.
Digital Chemistry, Merck

BACK TO CONTENTS

Synthetic Automations: A Revolution from Stone Age to Modern Era

Adapted from Fang, G., Lin, D.-Z. and Liao, K.

Article

Traditional organic synthesis has made outstanding progress but manual operation, inconsistent reproducibility, and inadequate efficiency hinder its dependable evolution to intelligent automation. Synthetic chemistry is beginning to embrace Artificial intelligence (AI) software to replace labor-intensive processes such as research for developing potential synthetic pathways, identifying reaction reagents with ranked choices of putative organic synthesis plans and automated synthesis. Herein, Liao and colleagues highlight some of the representative breakthroughs in automated synthesis and present current challenges and future directions in the field.

Background

The Industrial revolution in the 18th, and 19th centuries witnessed the power of automated manufacturing. Organic synthesis has depended on highly trained labor (chemists) to create and perform the molecular assembly process. In the 1960's, Merrifield reported the first automated system in organic chemistry: solid-phase peptide synthesis by attaching the C-terminus to a resin and masking the N-terminus with a protective group. The automation set-up pumps in the relevant reagents and solvents, mixes them with the resin, and removes them in the correct order to achieve deprotection, acylation, separation, and purifications.

In most cases, organic synthesis remains a very time and labor consuming process that provides variable results due to differences in techniques in separate laboratories and facilities. In this Emerging Topic, Fang and colleagues highlight several recent breakthroughs such as AI-driven research, AI-assisted synthesis planning, and AI-integrated robot automation of the actual synthesis process. Challenges in design and implementation of the automated synthetic pathways are also presented.

Breakthroughs

Many drugs are small molecules with diverse chemical structures and thus, require customized procedures which consume both money and highly trained labor. Burke developed iterative carbon 2D and 3D cobalt catalysts (C-Csp^2, C-Csp^3) assembly strategy and automated the process to synthesize 14 diverse classes of small molecules. The use of tetramethyl N-methyliminodiacetic acid (TIDA) supported C-Csp^3 bond formation. This synthesis machine coupled with more than 5000 commercial building blocks could support the synthesis of numerous small molecules.

Flow-based synthetic platforms can provide precise control of reaction temperatures, reaction times, and composition, and they can play major roles in automated synthesis. For example, the fast-flow device Tiny

Tides invented by Li and Pentelute *et al.* could efficiently produce cell-penetrating peptides-conjugated peptide nucleic acid [1]. Mo *et al.* accelerated the discovery of novel electroorganic processes with high-throughput experimentation on their automated micro-fluidic single-droplet screening and analysis platform in 2020 [2]. In 2022, Wang *et al.* designed an electrocatalyst testing platform: It performed 942 effective tests on 109 copper-based bimetallic catalysts in 55 hr [3].

Gilmore and coworkers developed an automated multistep synthesizer that stably and reproducibly could provide both linear and convergent synthetic processes by arranging multiple continuous flow modules around a central core [4]. Switching from one module to another did not require manual configuration. The instrument also included inline monitoring with nuclear magnetic resonance (NMR) and infrared (IR) spectroscopy: the monitoring facilitated post-reaction analysis and feedback. They demonstrated the flexibility of the radial flow configuration by synthesizing a library of derivatives of the anti-seizure drug, rufinamide.

Mo's group described an automated platform that collected polarity estimations by inline thin-layer chromatography (TLC) in 2022 [5]. The trained AI-platform could estimate the probability of separation of multiple compounds and aid in proposing purification conditions.

To minimize the need for input from human chemists, Cronin's group (2019) developed the Chemputer that provides methodological instructions into the individual steps and integrates the automation of the platform with bench-scale techniques using a natural language processing (NLP) algorithm [6]. The Chemputer system could extract synthetic procedures from publications, transform the synthesis plan into chemical description language used for procedures, convert the instructions into commands for manipulation of the automated platform, and direct the chemical synthesis. Without human intervention, the Chemputer assembled three high-quality pharmaceuticals with higher yields and purities than those achieved during manual synthetic and purification procedures.

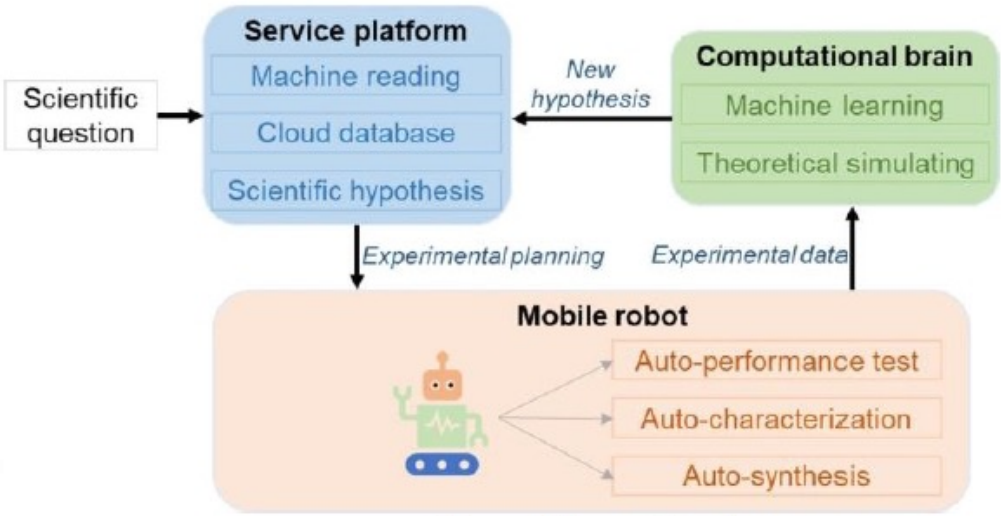


Figure 1. *Workflow of the AI-platform called the AI-Chemist.*

A milestone in automation of organic chemical synthesis was accomplished by Coley *et al.* [7]. Their computer-aided chemical synthesis program involved synthetic planning based on millions of published chemical reactions and in silico simulations to maximize success. The AI-synthesis program directed a modular continuous flow platform that executed the synthesis by automatically reconfiguring the robotic arm. Its power was demonstrated by planning and synthesizing 15 compounds, including several angiotensin-converting enzyme (ACE) inhibitors and NSAID drugs.

Grzybowski and Burke *et al.* described an iterative machine learning system to explore general reaction conditions in the proposed automated synthesis protocol [8]. A simple closed-loop workflow leveraged the machine-learned data-guided matrix to prioritize and select subsequent reactions for testing, and used robotic experimentation to augment precision, throughput, and reproducibility. Its workflow through experimentation and machine learning identified reaction conditions for the hetero(aryl Suzuki-Miyaura coupling reaction which confirmed its utility for multidimensional chemical optimization difficulties.

Cooper's group described the first AI-integrated mobile robot that autonomously ran 688 reactions over eight days to thoroughly test ten variables experimentally [9]. However, the robot did not have the software capacity to capture the existing chemical knowledge, nor machine learning for generating novel scientific hypotheses.

Jiang's group (2022) described an AI-platform, called AI-Chemist, that could perform the essential steps for proposing and ranking synthetic planning, execution of the synthetic steps, monitoring and data collection of the synthesis process through multiple reactions and steps, and machine learning (Fig. 1) [10].

Challenge and Outlook

Much progress in automatic and autonomous AI-driven platforms for organic synthesis have been accomplished. However, widespread adoption would be hastened by addressing the following challenges.

Seamless integration of the automated synthetic platform which contains modules for reagent storage, reaction preparation module, numerous reactor modules, integrated analytical instrument(s) for monitoring reactions, purification system, management system for compound, monitoring unit, and console. Due to space limitations in most laboratories, the size of the instruments and platform ideally would be no larger than a fume hood. The algorithm(s) for comput-

er-aided or AI-driven synthesis planning should be integrated with the computer-aided chemical synthesis (CASP) platform and the monitoring equipment seamlessly.

The physical platform and the software should be easily customizable and reconfigured for future uses.

The full unit should be reasonably (low) priced since many laboratories would hire actual chemists rather than the automation units at similar costs.

The automation platform and software need to be user-friendly: straight-forward in set up and optimization of synthetic planning/ranking and selection and execution of the synthetic process. Using a universal chemical program language for the chemists' input and retrieval of data will help maximize its benefit to human chemists.

As use of automation of the organic chemical synthesis expand, organic chemists will be relieved of repetitive experimentation often used during optimization. Organics chemists will be able to focus more time on answering the questions, "What should we synthesis? Why?" rather than the mechanics of the actual synthesis.

Summary

Multiple groups are advancing AI-driven organic chemistry synthesis, coupling machine learning to propose and experimentally test promising novel synthetic routes automatically with variable needs for human input. Automated synthesis can improve higher yields reliably and liberate chemists from routine manual tasks so they can focus on creative tasks.

References

[1] Li, C. *et al.* (2022). Automated Flow Synthesis of Peptide-PNA Conjugates. *ACS Central Science.* DOI: 10.1021/acscentsci.1c01019.

[2] Mo, Y. *et al.* (2020). A Multifunctional Microfluidic Platform for High-Throughput Experimentation of Electroorganic Chemistry. *Angewandte Chemie - International Edition.* DOI: 10.1002/anie.202009819.

[3] Xie, M. *et al.* (2022). Fast Screening for Copper-Based Bimetallic Electrocatalysts: Efficient Electrocatalytic Reduction of CO2 to C2+ Products on Magnesium-Modified Copper. *Angewandte Chemie - International Edition.* DOI: 10.1002/anie.202213423.

[4] Chatterjee, S. *et al.* (2020). Automated radial synthesis of organic molecules. *Nature.* DOI: 10.1038/s41586-020-2083-5.

[5] Xu, H. *et al.* (2022). High-throughput discovery of chemical structure-polarity relationships combining automation and machine-learning techniques. *Chem.* DOI: 10.1016/j.chempr.2022.08.008.

[6] Steiner, S. *et al.* (2019). Organic synthesis in a modular robotic system driven by a chemical programming language. *Science.* DOI: 10.1126/science.aav2211.

[7] Coley, C.W. *et al.* (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science.* DOI: 10.1126/science.aax1566.

[8] Angello, N.H. et al. (2022). Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. Science. DOI: 10.1126/science.adc8743.

[9] Burger, B. *et al.* (2020). A mobile robotic chemist. *Nature.* DOI: 10.1038/s41586-020-2442-2.

10] Zhu, Q. *et al.* (2022). An all-round AI-Chemist with a scientific mind. *National Science Review.* DOI: 10.1093/nsr/nwac190.

BACK TO CONTENTS

Computational Analysis of Synthetic Planning: Past and Future

Adapted from Wang, Z., Zhang, W. and Liu, B.

Article

Computer-aided synthesis planning (CASP) can play a significant role in organizing and leveraging the flood of novel chemical reactions and expert reaction rules for planning novel and highly efficient synthesis of natural products and drug candidates. This review describes the progress in computational analysis of synthetic planning from the early stage focused on rule-based programs to machine learning and their combined capability.

Introduction

Chemists use retrosynthetic analysis to design a synthetic strategy for a target compound. Briefly, they use their experiences in breaking chemical bonds in the target compound and subsequent precursors in an iterative manner.

Various standardized tools (e.g., CML, SMILES, SMARTS, InChl and ECFP) translate chemical reactions and molecules into machine readable information. More advanced algorithms (e.g., neural networks, reinforcement learning) expand the data processing of chemical reactions.

This review covers three categories of CASP. Two categories use logical deduction from chemists' intuitions and experiences: CASP algorithms based on hand-coded rules or on automatically extracted rules. The third CASP category uses chemical reaction databases for training of machine learning (ML) algorithm(s).

General Structure of CASP System

A typical CASP system has four modules. The reaction template database stores known reactions with rules of bond breaking. The retrosynthetic module aligns known reactions in the template database with structures of input molecules and provides the closest match to commercially available precursors in an iterative manner. The tree guide and evaluation module assess the fit of the candidate precursors to the synthetic routes. The commercially available compound database acts as a stop for the retrosynthetic module.

Hand-coded Rules Combined with Logical Algorithm

Representative CASP systems include LHASA, SECS, IGOR, CHIRON, and Chematica/Synthia™. Both LHASA and SECS CASP systems included a communication module: interfaced writing pad so chemists could evaluate and select the best route from the synthetic tree.

IGOR (Intermediate Generation of Organic Reactions) did not restrict retrosynthetic analysis to empirically derived heuristic rules. IGOR includes all molecules participating in a reaction, requires extensive calculations and can simulate only simple retrosynthetic transformations.

CHIRON can decode complex stereochemistry and functionality which it can correlate to commercially available stere-

ochemistry-enriched precursors. It searches for precursors with closely related skeletons, stereocenters, and functional groups to the target molecule.

Chematica (now called Synthia™) has expanded the Network of Organic Chemistry (NOC) to approx. 10 million compounds and manually added compatibility and context information (e.g., canonical conditions, intolerance of functional groups, regio- and stereoselectivity of specific reactions) using the SMILES/SMART coding method. Its hand-coded reaction rules increased to >100,000 in 2021. Chematica/Synthia™ embedded an intelligent search function and chemical scoring functions allow globally optimal outcomes (e.g., chiral precursor for asymmetric synthesis.)

Chematica/Synthia™ presents the synthetic tree in a dendritic way: each node denotes the retrosynthetic transformation and its associated substrate set (Fig. 1a). Chematica/Synthia™ accelerates the analytic process with a priority queue for the lowest scoring nodes in the search algorithm (Fig. 1b).

Chematica/Synthia™ includes various quantum mechanics and machine learning (ML) methods to optimize the searching algorithm, scoring functions, and stereoselective transformations. Chematica/Synthia™ designed synthetic routes for eight drug-related molecules and several complex natural products. Their syntheses were experimentally accomplished. Synthia™ program designed a more efficient synthetic route for OICR-9429 (Fig. 2). Literature reported a 1% yield of OICR-9429; but Synthia™ route yielded 60%. Furthermore, Synthia™ designed synthetic route simplified its purification from four chromatographic procedures to one recrystallization. Thus, Grzybowski and coworkers clearly demonstrate that Chematica/Synthia™ can solve complex problems in synthetic chemistry.

Manual extraction of reaction templates can broaden the context information of chemical reactions and enhance retrosynthetic analyses. The choice between automatic and manual extraction depends on the consistent description of variables and the desired applications.

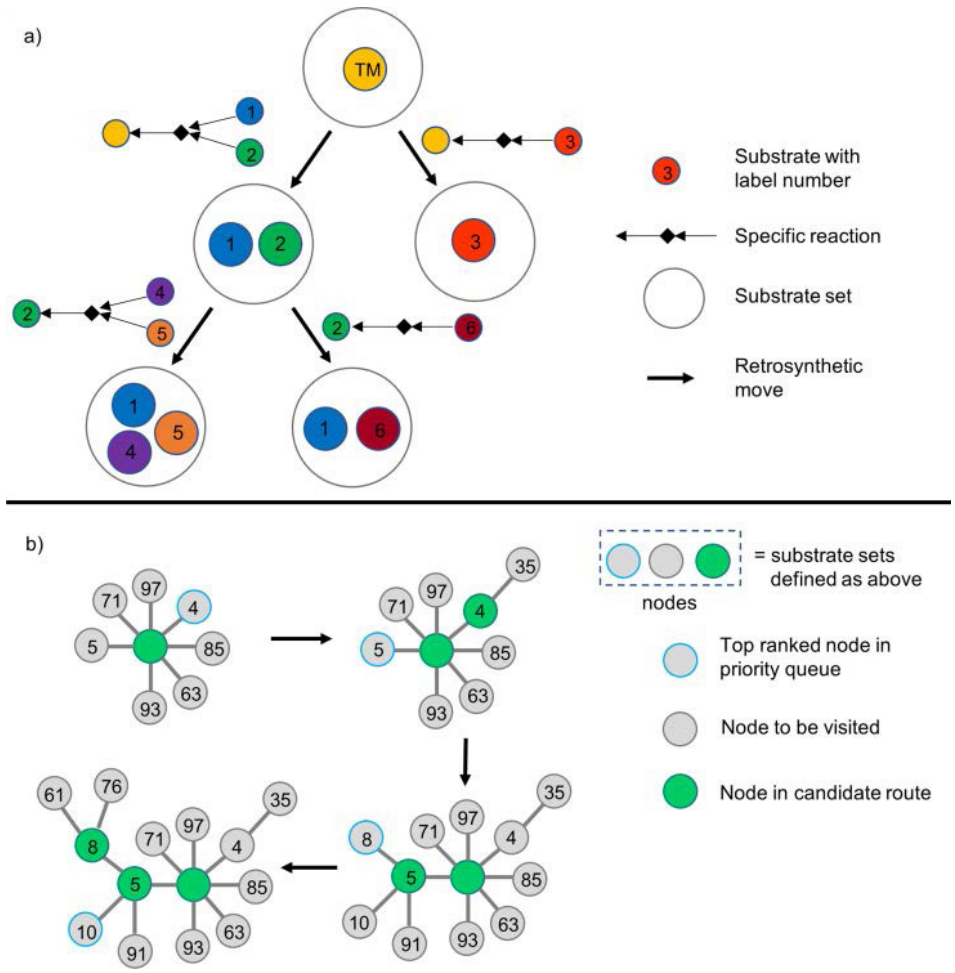


Figure 1. *Graphic representations of Chematica/Synthia™ (a) Graphic depiction of synthetic tree in a dendritic way. (b) The expansion of analytic process at each generation of synthetic tree development. Each node represents a transformation with a substrate set. Low score indicates higher priority for analytic evaluation in next synthetic step.*

Automatically Extracted Rules Combined with Logical Algorithm

Auto-extraction of new chemical reactions and templates daily can efficiently maintain databases but it may miss adjacent functional groups and atoms.

SYNCHEM2 allows both backward and forward synthetic transformations with alternate coding. RETROSYN abstracts the reaction center and builds atomic correlation between products and reactants with a special graph-difference algorithm. RETROSYN searches and sorts the degree of matching with a high to low priority but ignores stereochemistry.

KOSP (Kowledge-base-Oriented System for Synthesis Planning) automatically extracts reaction templates including activating groups/atoms within three bond distances to populate the Reaction Knowledge Base. The new KOSP version enables regio- and stereoselective retrosynthesis analysis and updates have expanded the reaction contents by 10-fold.

ChemPlanner, successor of ARChem, has an exclusive cooperation with American Chemical Abstracts Service and Wiley for SciFinder, a highly accessible database of scientist-curated reaction content. The new ChemPlanner version enables regio- and stereoselective retrosynthesis analysis.

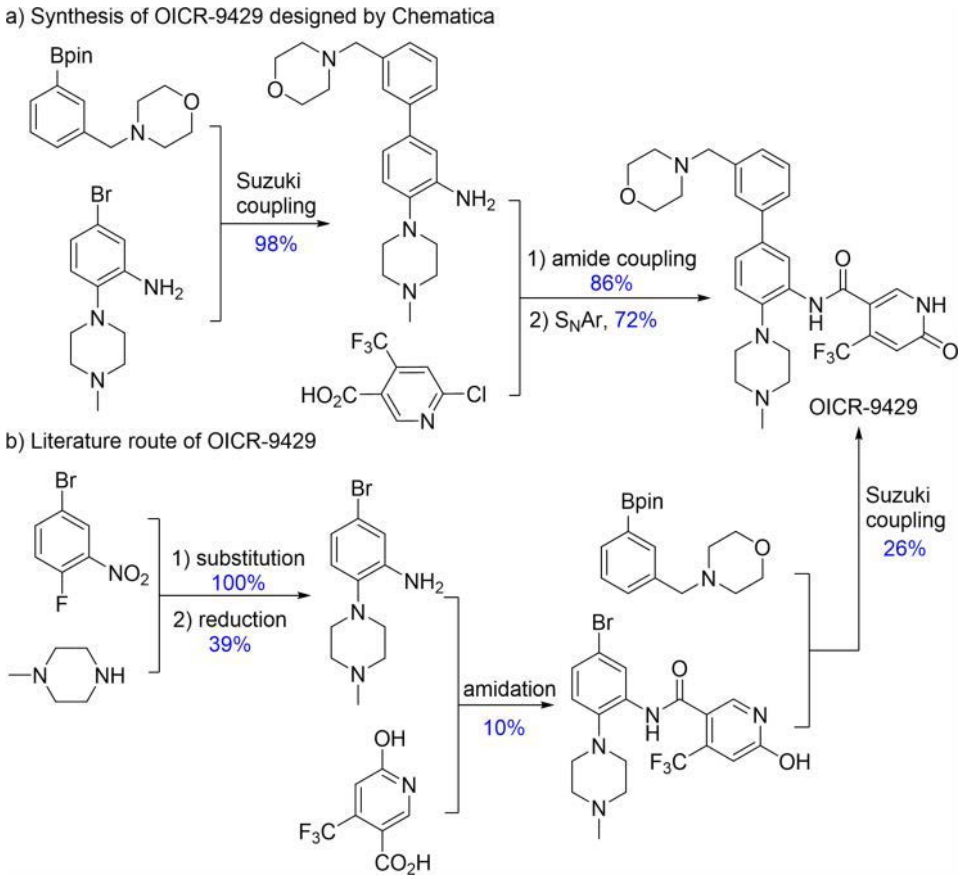


Figure 2. *Two experimentally confirmed synthetic routes of OICR-9429. a) Synthetic route proposed by Chematica/Synthia™. Experimental yield was 60%. b) Literature route for OICR-9429. Reported yield was 1%.*

ICSYNTH represents its reaction knowledge database in graph-basis form. Users can include in-house chemical rules from its confidential reaction database and adapt ICSYNTH for various application scenarios by selecting and editing chemical rules.

ASKCOS calculates the similarity of reaction products with the target molecule to develop a retrosynthetic plan in a stepwise manner. Modules of ASKCOS include One-Step Retrosynthesis, Interactive Path Planning, Tree Builder, and Context Recommendation.

Automatically Extracted Rules Combined with Machine Learning Algorithm

ML algorithms are trained with chemical reaction databases including reactants. Reinforcement learning algorithms continuously interact with the environment which teaches them the optimal strategy via a penalty-reward approach.

The Bishop program combines rule-based retrosynthetic analysis and reinforcement learning. The Chemical Reaction Network compiles the intermediates, connects reactants and products, and has a reinforcement learning module to map a flexibly defined, optimal reaction pathway(s) with potential filters for cost, overall efficiency, and/or environmental impacts.

3N-MCTS (Monte Carlo Tree Search algorithm) uses the artificial neural networks trained by digital sequences of products and relevant precursors from the literature. The ANN-based CASP system reorganizes the specific learned reaction rules which simplifies the calculation process. Each MCTS round consists of Selection, Expansion, Rollout, and Update. Improvements are needed to predict stereoselectivities.

Seq2Seq model with Simplified Molecular Input Line-Entry System (SMILES) translation can process massive dataset and simulate a reaction with global optimal output. AutoSynRoute evaluates synthetic pathways by applying MCTS algorithm with Chematica/Synthia™-inspired heuristic scoring functions. RXN uses two retrosynthetic ML models trained by two databases. RXN can predict suitable reaction conditions for the proposed synthetic route.

Conclusions

Several CASP programs apply heuristic reaction rules and reaction rules from the literature in their algorithms for retrosynthetic chemistry with or without scoring functions and ML (e.g., Chematica/Synthia™) Other CASP programs rely on ML or the combination of ML with heuristic reaction rules and/or literature-based chemical rules. These algorithms have already provided novel synthetic routes that improved yield for complicated molecules. Further improvements can provide novel synthetic routes for complex compounds with additional constraints such as lower cost, lower environmental footprint, and fewer hazardous reagents or solvents.

BACK TO CONTENTS

Synergy Between Expert and Machine-Learning Approaches Allows for Improved Retrosynthetic Planning

Adapted from Badowski T., *et al.*

Article

Grzybowski and colleagues demonstrate that higher synthetic accuracy can be achieved in computer-designed multistep synthetic plans when artificial intelligence (AI) software that combines both expert knowledge and machine-extracted information from large repositories of reactions types.

Introduction

Artificial intelligence (AI) platforms for computer-designed synthetic plans seek commercially available precursor materials, assess individual synthetic steps, and evaluate the vast synthetic possibilities from their resource(s) and training materials. Integral components of AI are scoring functions (SFs) which guide the development of the plans. AI has historically developed chemical synthetic plans based on expert synthesis knowledge or on synthetic pathways reported in the literature such as chemical repositories. However, each dataset has advantages and limitations.

Although heuristic, expert synthetic knowledge usually reflects the successful chemical plans of chemists' intuition, chemists' preferences include central disconnections, reduced numbers of rings and stereocenters, and often multiple steps masking and unmasking of pertinent reactive groups.

In comparison, machine-learning functions based on the literature focus on popular reaction types with sufficient references, and AI uses neural network (NN) algorithms to identify one or more synthetic plans. The SFs of NNs compile information about reactions and final products from a specific database such as the USPTO (US Patent Trademark Office). The SFs output provides probability of specific reactions (identifiers, IDs) but may be overly burdened with popular reactions and miss more efficient reactions known by chemists.

Characteristics of AI training materials for combining machine learning from experts and NN

The NN is trained on analogous product and substrate data from both sources: reactions from the literature and high-quality reaction rules from experts. All analyses utilized approx. 1.6 million reactions reported to synthesize approx. 1.4 million unique products (simple chemicals to complex natural products). Protection and deprotection reactions from either source were not included to avoid their overuse in the synthetic plans. Grzybowski and colleagues required that each reaction included from the literature agreed with an expert's reaction rule(s) from at least one of the 75000 procedures from Chematica The SF-based output may include a synthetic plan involving alternative reaction rules from Chematica, now called Synthia™ and is commercially available.

The analyses provided an average of approx. 60 conflict-free, product-fitting re-

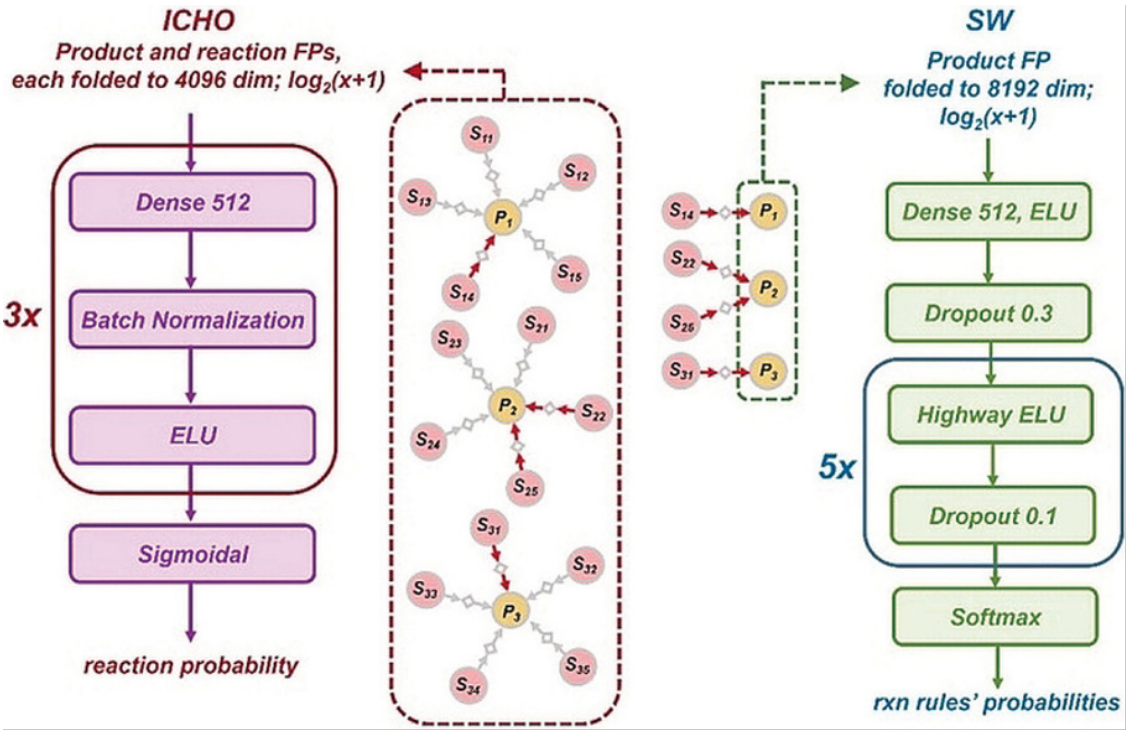


Figure 1. *Architecture of the NN of the ICHO (left) and SW (right) platforms. The NN learn from the information depicted in the middle portion. ICHO learns on published reaction precedents in the literature that match one of the expert coded rules from Chematica (Synthia™) for substrates and a specified product (red arrows). ICHO also evaluates other expert rules that may apply to synthesis of pertinent substrates and the specified product (gray arrows). In contrast, SW learns only from reactions and products in literature precedents. ELU, exponential linear unit.*

actions for a product. In total, Grzybowski and colleagues considered approx. 85 million reactions that were high-chemical quality and conflict-free in the development of synthetic plans for 1.4 million products. The product set was randomly divided into 70% for training, 10% for validation, and 20% for testing.

The authors' program (ICHO) has a NN-based scoring function that contains four layers: three hidden layers that provide possible reactions for producing product 1 (P1), P2, and P3, and an output layer (Fig. 1 left panel). The enhanced program (ICHO+) augmented the NN ICHO architecture with the following expert knowledge of chemically intuitive reactions: number of created or destroyed rings, number of installed or removed stereocenters, selectivity of reaction, sizes of breakdown products (similar vs very disparate), and more. The ICHO+ program thus adjusts the frequency of specific reactions for a particular product in the literature with their frequencies in expert synthetic plans. During the training of the ICHO and ICHO+ training, the program assigns larger probabilities for specific reactions obtained in both literature and expert synthetic plans. In contrast, the program also adjusts the probability lower for a very popular chemist rule that is rarely used for synthesis of a particular product, suggesting the reaction may be tricky, challenging to execute, or inefficient.

Performance of AI platforms

Direct comparison of the NN architecture between ICHO/ICHO+ and the NN-based program by Segler and Walker, denoted as SW, is illustrated in Figure 1 [1,2]. The SW AI platform and other NN-based AI synthetic platforms published by 2019 learn only from reactions in literature precedents. Most AI programs including ICHO and SW use a popular machine-learning activation function called exponential linear unit (ELU). ELU accelerates training and increases performance of the program. The efficiency of the combined ICHO+ program was also compared to an updated heuristic scoring scheme originally called SMILES that assesses the simplicity of the synthesis plan. The updated program called SMALLER advances central disconnections which simulates the organic synthetic intuition and practice of chemists. One advantage of SMALLER is that the frequency of reactions in the literature has minimal influence on the ultimate proposed route.

Within the ICHO and SW programs, the inclusion of learning from the heuristics expert chemical rules (ICHO+, SW+) only marginally improved the efficiency of the synthetic plans. Limiting the SW programs to product-fitting reactions (SW2, SW2+) improved their performance. However, ICHO+ remained the highest ranked pathway, likely due to its additional knowledge of substrates.

The performance of the three types of programs were evaluated on developing synthetic pathways involving both experimentally established reactions and relatively advanced synthetic pathways. Synthetic plans for four complex products developed by the ICHO+, SW2+, and SMALLER programs are compared in Figure 2. ICHO+ ranked highest for the synthetic plans for the four products: the BRD 7/9 inhibitor, the serotonin–norepinephrine reuptake inhibitor (+)-synosutine, the natural product seimatopolide A, and the prostaglandin analogue bimatoprost.

BACK TO CONTENTS

Summary

Grzybowski and colleagues compared their NN-based ICHO+ scoring functions that combine chemical AI with expert knowledge including reaction rules with other NN-based scoring AI programs for development of synthetic plans of complex molecules. Their examples demonstrate a major advantage of combining chemical AI with expert knowledge: the program's ability to propose synthetically powerful reactions that are listed only sparsely in the literature. Chematica is updated and now called Synthia™. It provides AI retrosynthesis software that also can utilize a custom inventory or database (e.g., in-house database of confidential reactions) in addition to several publicly available databases.

References

[1] Segler, M.H.S. *et al.* (2018). Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*. DOI: 10.1038/nature25978.

[2] Segler, M.H.S. and Waller, M.P. (2017). Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chemistry - A European Journal*. DOI: 10.1002/chem.201605499.

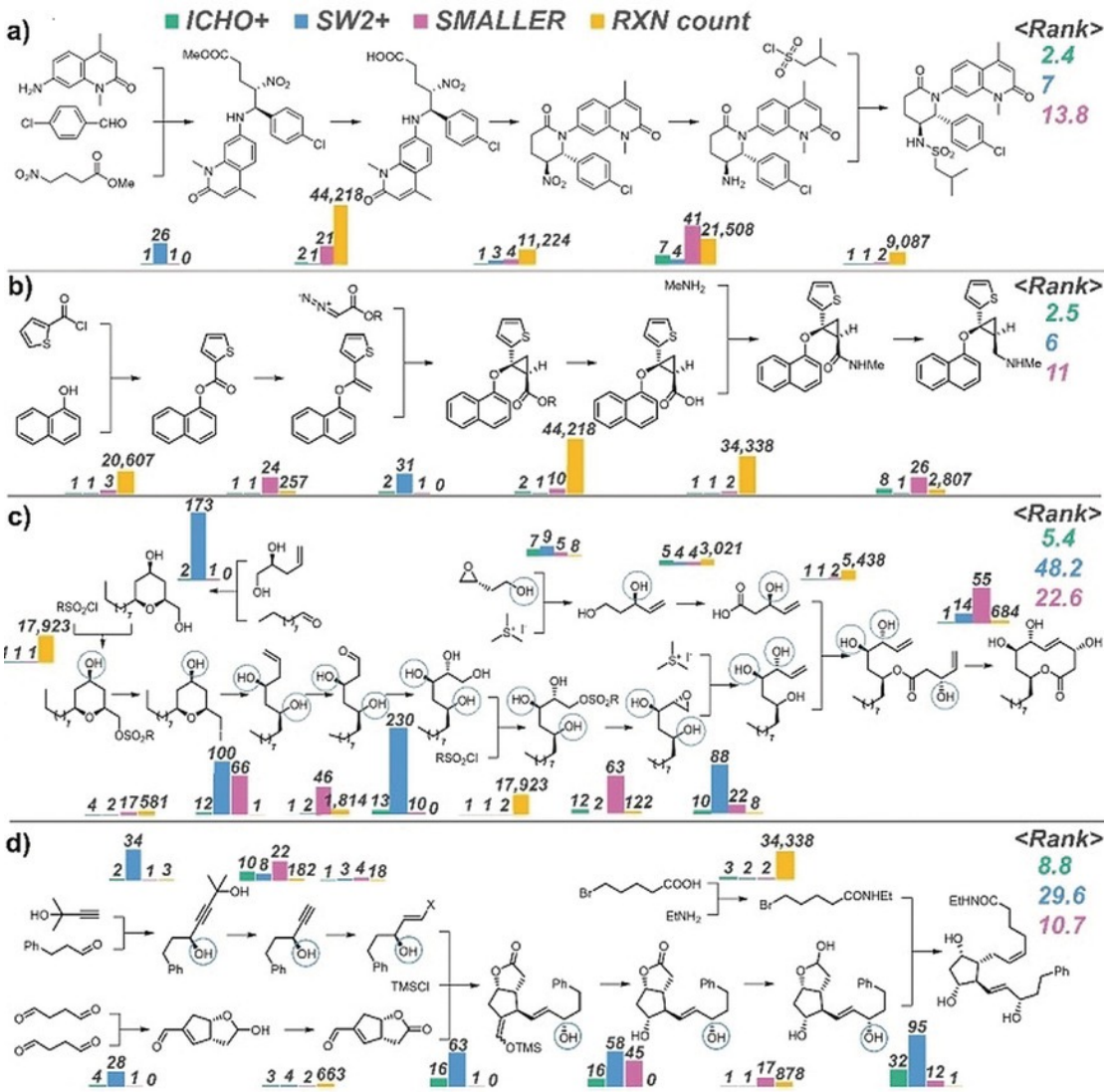


Figure 2. *Comparisons of performance of ICHO+, SW2+, and SMALLER AI programs on developed synthetic pathways for natural products and medicinally-relevant compounds. The histograms located below or above each reaction quantify its ranking: ICHO+ green; SW2+ blue; SMALLER pink. Orange bars depict the number of matches of a specific expert reaction rule. The average rankings for entire pathway are listed on the right side of panel. Synthesis of (a) BRD 7/9 inhibitor, (b) (+)-synosutine, (c) seimatopolide A, and (d) bimatoprost. Blue circles indicate OH protection along pathways to clarify the treatment of protections from Chematica (Synthia™).*

Computer-Assisted Design of Sustainable Syntheses of Pharmaceuticals and Agrochemicals from Industrial Wastes

Adapted from Le Pogam, P. *et al.*

Article

Chemists are seeking to more efficiently use the remaining chemical resources from industrial synthetic pathways, thereby increasing sustainability and overcoming supply chain bottlenecks of precursor molecules. Here, a state-of-the-art computational synthesis-based approach identifies synthesizable molecules of commercial interest that can use one or more common waste products from large-scale industrial processes as substrates or precursors. Its algorithmic ranking of synthetic strategies incorporates metrics of sustainable chemistry.

Introduction

During recent decades, chemists are seeking to improve sustainability of large-scale industrial synthetic processes by repurposing one or more of the waste products. Analogous to computer-based synthetic strategies in analytical chemistry, recent computer algorithms are compiling lists of end-of-pipe industrial waste that may serve as substrates or precursors during novel synthetic pathways of pharmaceutical-relevant products. This process is called the *circular chemistry concept* or *sustainable chemistry*.

Sustainability Characteristics ranked in silico transformation module

Even a small set of substrates can generate synthetic pathways for millions of feasible compounds. Thus, the computer-based programs need to rank the synthetic strategies for simplicity and for sustainability. An *in silico* transformation module learns from 10,000 plus reactions in an in-house collection that includes information on incompatible structural motifs, admissible substrates, suggested conditions, reagents, and typical conditions. Additional information used for developing a synthetic pathway includes suggested solvents, usual temperature range, ability of reactions to be performed in tandem, renewable resources, abundance of commercially available waste product(s) for reagents, geographical origin of industrial chemical waste, and avoidance of problematic solvents according to health or environmental criteria. Figure 1 provides an overview of artificial intelligence (AI) programs and pertinent metadata considered during development of putative synthesizable products that can reuse the industrial waste as building blocks or substrates and increase chemical sustainability.

For example, the AllChemy-based platform also contains an AI algorithm that can predict molecular properties of the synthetic compounds. As a second example, SYNTHIA™ Retrosynthesis Software can efficiently explore the most cost-effective and innovative synthetic routes to produce the target molecules. Chemists can use filter options and art visualization to optimize the synthetic routes of target compounds. SYNTHIA™ allows customization of search

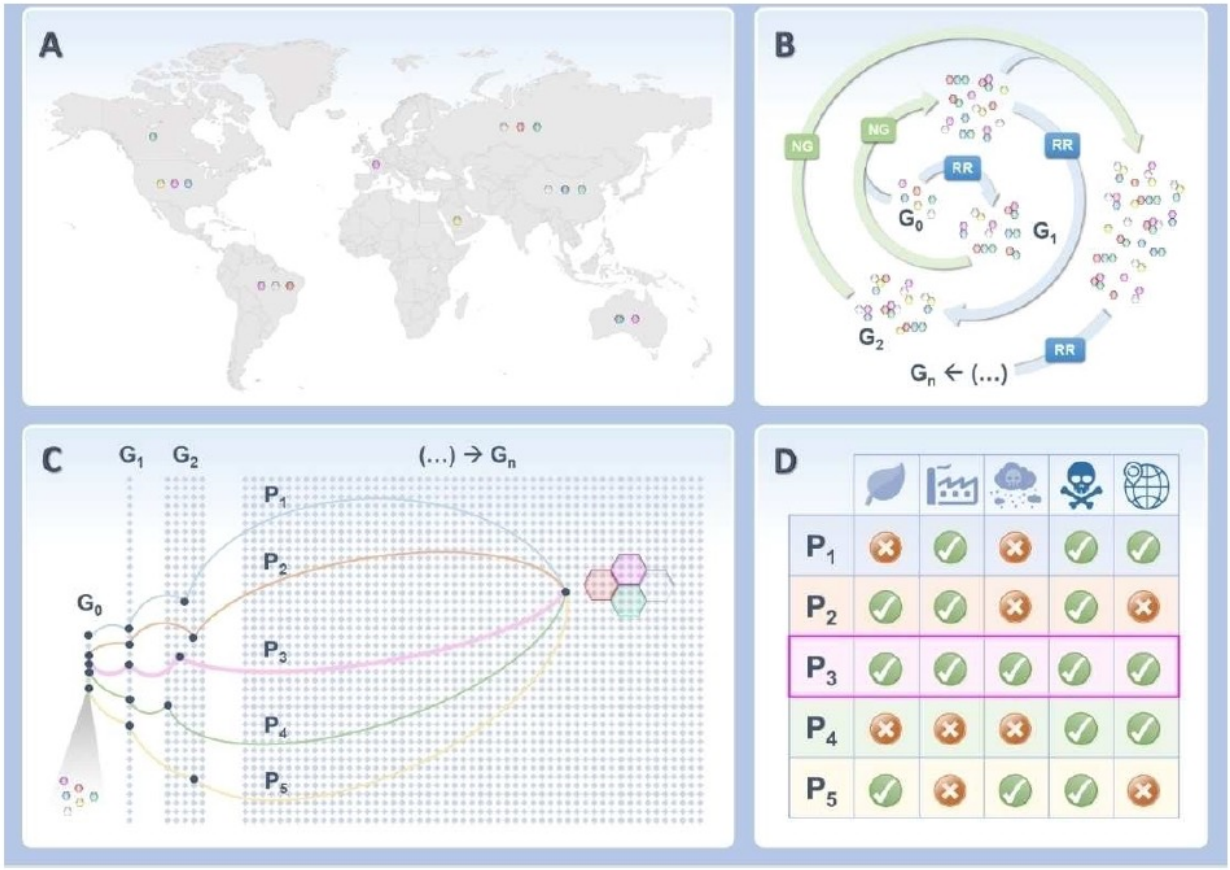


Figure 1. Overview of characteristics, metadata, and steps in AI-based programs for recycling of pharmaceutical wastes of manufacturing process A as precursors or reactants for manufacturing of other medicines. (A) Geographical location of all considered chemical wastes. (B) Illustration showing the numerous analytical generations (G_n) in the iterative, forward-synthesis process of the program. Expert reaction rules (RR) guide the transformation of compounds in the *n* generation to the *n+1* generation. After filtering compounds from both generations for size and similarity to desired product, chemists submit them for the next synthetic AI analysis based on RR. (C) Illustration of the exponential diversification of compounds during a few rounds of transformations. Different arcs represent five distinct synthetic pathways that start with same waste product(s) and lead to the desired product or drug. (D) Depiction of retrieved pathways and their ranking for use of environmental and health-friendly chemicals, low tendency for experimental difficulties (such as intense exo- or endothermic reactions), greenness of overall process (e.g., cumulative Process Mass Intensity, atom economy), avoidance of toxic reagents or by-products, and local geographical origin of waste resources.

parameters to prevent or promote specific reactions, reagents, or classes of molecules, including desired stereoisomers. It can generate a list of commercially available starting materials while guaranteeing the integrity and confidentiality of clients' data with ISO/IEC 27001 information security certification.

Power of AI Program

Le Pogam and colleagues describe Wolos *et al.*'s article [1] that used the Allchemy-based program to identify synthesizable pharmaceutical-relevant products or agrichemicals from 189 chemical wastes. Each synthetic generation of the algorithm compares the synthesized product(s) with the initial substances and those obtained in intermediate generations (Fig. 1B). Since the goal is to facilitate synthesis of high-value compounds from the specific waste in an efficient time frame, researchers can improve efficiency in the next analytic/synthetic generation by retaining small compounds for use as building blocks and molecules with structural similarity to the target synthetic compound.

The AI-created networks usually provide a high combinatorial repertoire of diverse reactions that can lead to the synthesis of the target product or drug (Fig. 1C). The operator should set a threshold of less than 100,000 compounds to maintain reasonable calculation timeframes. The AI-created networks are queried after each generation for various process variables that directly affect the safety of the synthetic process, its efficiency, cost, and sustainability.

A meta-score compiles the attributes of the synthetic pathway and its ranking reflects also penalties for undesired attributes in at least five categories which are scored from 1 (most desirable) to 10 (least desirable or most harmful) (Fig. 1D). Penalties are provided for use of harmful or problematic reagents or solvents since the use of health and environmentally friendly reagents, substrates, solvents and building blocks are preferred. Penalties are levied for challenging reaction conditions such as extreme

temperatures or experimental difficulties like intense exo- or endo-thermic reactions. Penalties also are levied for producing more by-products and waste, which is referred to as low atom economy and is often measured as process mass intensity (PMI). The preferred reactions produce the least amount of waste or by-products. PMI compiles the mass of all material (reagents, catalysts, reactants, solvents, and workup chemicals) in the synthetic process relative to the yield of the isolated product. Lastly, a local origin of repurposed industrial waste product for building blocks or substrates is preferred to reduce transportation costs and supply chain disruption. Thus, a penalty will be levied for disparate geographical locations of the industrial waste and the putative user.

AI analysis with 189 waste products and seven rounds of chemical transformations provided in silico synthetic pathways for 300 million compounds, including 69 drugs and 98 agrochemicals. Wolos *et al.* [1] identified novel synthetic pathways for several targets that use waste products from the same origin in mostly benign conditions with relatively few steps. Interestingly, the described program can broaden its development of synthetic pathways to several of the most prescribed medicines by adding 1000 basic reagents to the 189 available wastes. Wolos *et al*. [1] have validated several of these computer-predicted pathways experimentally [1].

Summary

The described AI synthesis tools provide opportunities to predict synthetic pathways of diverse compounds from a set of any widely available precursors, including industrial waste products. The current workflow incorporates the expansive knowledge of organic chemical reactions and its associated metadata that can be assessed for its compliance with sustainability and green chemistry metrics at each transformation. These additional checks help ensure that the proposed synthetic pathway for a target compound has a high probability of industrially realistic yields and provides sustainable access to its precursors. This refined AI workflow expands the retrosynthetic chemists' *a priori* network of substrate-based transformations to alternate novel synthetic pathways that may incorporate locally sourced industrial waste products.

References

[1] Wołos, A. *et al.* (2022). Computer-designed repurposing of chemical wastes into drugs. *Nature*. DOI: 10.1038/s41586-022-04503-9.

BACK TO CONTENTS

Artificial-Intelligence-Driven Organic Synthesis— En Route towards Autonomous Synthesis?

Adapted from Empel, C., and Koenigs, R. M.

Article

Organic chemists' creativity and meticulous discipline have enabled organic synthesis of complex molecules for decades and their synthesis is often referred to as "the art of synthesis." Artificial intelligence algorithms are being developed and refined not only to perform literature searches and retrosynthetic analysis but also to identify and rank potential synthesis routes that include the reaction conditions. Embel and Koenigs review a recent Science article that combines artificial intelligence (AI)-driven development of synthetic plans for small molecules with an AI-developed automated synthesis process. They also discuss its limitations and highlight future innovations for enhancing the benefits of the AI-driven organic synthesis output.

Introduction

Traditionally, total synthesis of complex molecules including organic synthesis involves creativity, meticulous assessment of each step for yield of desired product, and reiterative process to modify the reactions for specific biochemical or biological properties (e.g., bioavailability, solubility) and optimize yields. AI is being used to propose reaction methodology such as novel pathways for synthesis of compounds. Empel and Koenigs suggest that the next evolutionary step in AI would be automated multistep synthesis of complex molecules. Although similar, automated and autonomous synthesis differ by their need for human input. Human input is required during automated synthesis to define thresholds, boundaries, reaction parameters, and synthesis protocols in the reaction files. In comparison, autonomous synthesis is a self-governing synthetic process that adjusts to surrounding parameters such as stereoselectivity and reaction yield without human input.

Limitations of Traditional Retrosynthetic Analysis and Automated Synthesis

Without AI support, chemists can easily miss pertinent combinations of novel substrates, reaction optimization, design of improved catalysts and novel reactions in the rapidly expanding body of chemical literature. All disciplines of chemical sciences consider synthesis of small molecules as a bottleneck and automated on-demand synthesis can help overcome this challenge.

Poorly soluble compounds provide extra challenges in automated workflows and often require more human input to improve the process because of clogging the channels. Predictions on the solubility of reactants in the proposed synthesis plan remain limited which hampers the execution

of some synthetic routes. Reactions that require or yield subambient temperatures usually also need advice from an expert chemist to maintain efficient workflow. After the automated synthesis, batch purification will be required for the final compound and may need special equipment such as particular columns.

AI-Planning of Synthesis Pathways and Automated Synthesis

Embel and Koenigs summarize the recent article by Jamison and Jensen and their colleagues [1]. They combined computer-aided retrosynthesis planning and a robotically reconfigurable flow apparatus to provide on-demand synthesis of small molecules (50-750 g/mole), as summarized in Figure 1. The system still requires human input to complement the AI synthetic algorithm with practical considerations (e.g., precise stereochemistry and solvent choices) that help optimize the multistep synthesis process.

AI learns design principles from a literature search of databases that include retrosynthesis reactions and related compound reactions to design one or more synthetic routes. The proposed plans include reaction conditions, precursors, enzymes (as needed), catalysts, substrates, precursors, and by-products at each of the steps. AI also provides an estimation of the feasibility of each proposed synthetic plan so chemist(s) can choose the most appropriate plan for automation, often in microfluidics, with an intention to scale the process. After AI developed an in silico synthesis plan from the reaction and compound databases, chemists use their expert knowledge of synthesis to refine the chemical recipe files, and the experimental and platform configurations. Chemists adjust the recipe files to overcome any inadequacies of the microfluidic flow systems before it is used in a chemical bay in the automated system. Thus, the chemists' input facilitates the robotic implementation of the proposed multistep synthesis, which is a major step toward scalable synthesis.

Jamison, Jensen and colleagues used this strategy to predict de novo synthesis routes of 15 small molecules, modify the relevant recipe files, and automate their synthesis using microfluidic workflows [1]. The 15 compounds include the nonsteroidal anti-inflammatory (NSAID) celecoxib, the blood thinner warfarin, and the ACE inhibitor prodrug, quinapril. Figure 2 (top panel) shows the synthetic process for the nonsteroidal anti-inflammatory (NSAID), celecoxib (structure 1). The first two robot-controlled reaction bays conducted the Claisen condensation of 4-methyl acetophenone (2) with methyl trifluoroacetate (3). Another robot-controlled reaction bay conducted the final condensation of the intermediate with hydrazine (4) and yielded celecoxib.

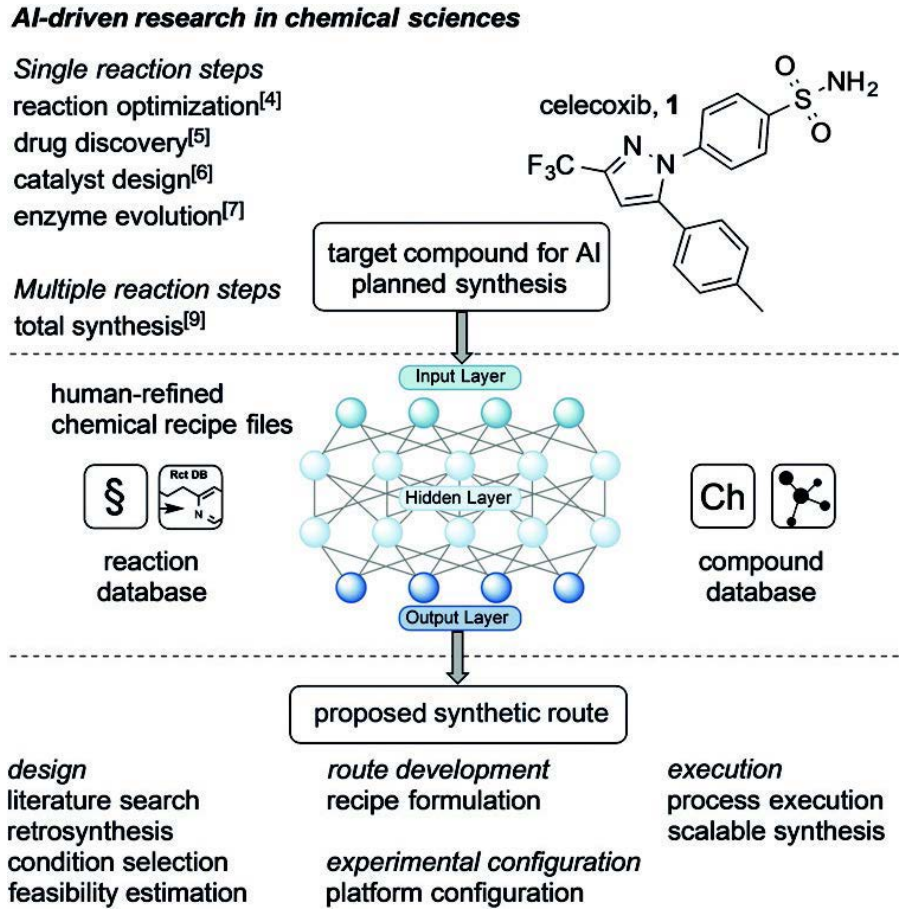


Figure 1. *The impact of AI on chemical sciences, including development of proposed synthetic route for specific compound (e.g., celecoxib) using reaction and compound databases, and potential implementation of automation of compound's multistep synthesis. AI, artificial intelligence*

Jamison, Jensen and colleagues applied the AI-driven technology to develop synthesis plans and synthetic execution of multiple ACE inhibitors and numerous celecoxib analogues [1]. The yield of the small molecules ranged from 342 mg/h to 572 mg/h.

The AI proposed synthesis plan for bezafibrate used a Bargellini reaction involving acetone, chloroform, and phenol (6). However, subambient temperatures occurred during the attempted synthesis of bezafibrate (Fig. 2 compound 5) by microfluidics. Jamison, Jensen and colleagues demonstrated the feasibility of the AI proposed synthetic plan by performing a separate manual synthesis process. Its yield was 76%.

Future Improvements

Multiple groups are working on improving AI-driven analysis and prediction of stereochemistry and methods to favor the desired stereochemical structure in the proposed synthesis. This improvement could reduce the human input needed for the modification(s) and use of the chemical recipe files.

In the future, AI also may be able to propose continuous purification schemes or batch purification systems to provide relatively pure target compound. The addition of on-line reaction analytics could provide important information on reaction progress and allow feedback algorithms to alter reaction parameters in real time and potentially improve efficiency and yield.

Summary

Jamison, Jensen and colleagues described their AI-driven on-demand synthesis of small molecules (50-750 g/mole) by combining computer-aided retrosynthesis planning with chemist-refined reaction files that directed a robotically reconfigurable flow apparatus [1]. They used this strategy to predict de novo synthesis routes of 15 small molecules, modify the relevant recipe files, and automate their synthesis using microfluidic workflows. The 15 compounds include the nonsteroidal anti-inflammatory (NSAID) celecoxib, the blood thinner warfarin, and the ACE inhibitor prodrug, quinapril. Embel and Koenigs discussed limitations such as poor solubility of one or more components and target compounds requiring specific stereochemistry. They also suggested expanding the AI-driven proposed plans to include purification scheme(s) and /or in-process reaction analytics to further automate the synthesis process. These improvements would divert the routine synthesis and optimization work to robots so that chemists can spend more effort on curiosity-driven research projects, thorough reaction monitoring and analysis, and serendipitous discoveries.

References

[1] Coley, C.W. *et al.* (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science.* DOI: 10.1126/science.aax1566.

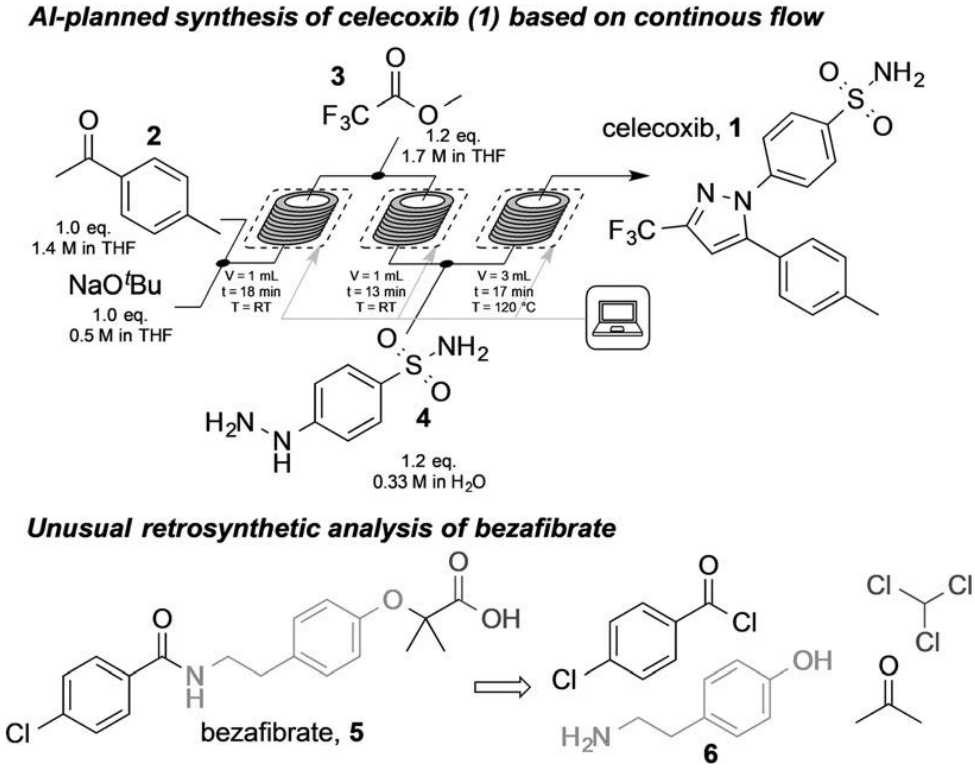


Figure 2. *Two examples of AI-driven analysis that resulted in feasible synthesis plans. Top: AI-driven synthesis of celecoxib (1) used multiple universal reaction bays directed by chemist-refined chemical reaction files in the modules. (2) 4-methyl acetophenone; (3) methyl trifluoroacetate; and (4) hydrazine. Bottom: Unusual retrosynthetic analysis of bezafibrate (5) that proposed the Bargellini reaction (6) involving acetone, chloroform, and phenol.*

BACK TO CONTENTS

Unlocking the Chemistry of Tomorrow with Computer-Aided Synthesis Planning

Interview with Professor Tim Cernak

Prof. Timothy Cernak
Assistant Professor of Medicinal Chemistry and Chemistry

Tim Cernak was born in Montreal, Canada in 1980. He obtained a B.Sc. in Chemistry from the University of British Columbia Okanagan and there studied the aroma profile of Chardonnay wines. Following PhD training in total synthesis with Prof. Jim Gleason at McGill University, Tim was a FQRNT Postdoctoral Fellow with Tristan Lambert at Columbia University. In 2009, Tim joined the Medicinal Chemistry team at Merck Sharp & Dohme in Rahway, New Jersey. There he developed technologies for miniaturized synthesis and late-stage functionalization. In 2013, Tim moved to Merck's Boston site. In 2018, Dr. Cernak joined the Department of Medicinal Chemistry at the University of Michigan in Ann Arbor as an Assistant Professor. The Cernak Lab is exploring an interface of chemical synthesis and data science. Tim is a co-Founder of Entos, Inc.

We had the privilege of interviewing Prof. Tim Cernak, an Assistant Professor of Medicinal Chemistry at the University of Michigan, whose diverse research interests span chemical synthesis, automation, data science, and more. With over a decade of experience, Prof. Cernak has been at the forefront of revolutionizing the field. In this interview, we delve into the world of computer-aided synthesis planning (CASP), a field that leverages automation, computational analysis, and artificial intelligence to streamline chemical syntheses. Prof. Cernak will discuss the pivotal role of CASP, the evolution from rule-based programs to machine learning, and the synergy between human expertise and AI. We'll explore how neural networks and expert-coded reaction rules enhance synthetic accuracy and discuss recent breakthroughs like sustainable synthesis from industrial waste. Moreover, we'll uncover the potential of AI-driven organic synthesis and explore the outlook for the future. Join us in this enlightening conversation with Prof. Tim Cernak as we journey through the realms of CASP, laboratory automation, and the promising future of intelligent automation in synthetic chemistry.

In the field of computer-aided synthesis planning, what role does CASP play in integrating human intuition and computational capabilities?

Computer-Aided Synthesis Planning (CASP) is a rapidly growing field with a rich history. CASP systems are designed to aid chemists in the decision-making process by suggesting synthetic routes that meet specific criteria such as yield, cost, and safety.

Perhaps more than any other field of science, total synthesis embraces and celebrates the art and elegance of its process. The field distills years of developments in strategic planning, systems of developed logic, and innovative and novel experimental reactivity modes into a series of planned reaction steps. The growing corpus of chemical reactions, and their associated rules and mechanisms, cannot possibly be memorized by a human. Interestingly, though, they follow all the rules of a game, and can be encoded into a computer. This logic combines the human decision-making process, expert knowledge, and chemical intuition with the computational power of machine learning models and algorithms to generate efficient synthetic routes.

Can you highlight the progression from rule-based programs to machine learning in the context of computational analysis of synthetic planning?

Sure, in the beginning rule-based programs relied on manually curated sets of chemical reactions and functional group transformations to generate synthetic routes. These approaches involve manual curation of reaction rules, which imparts expert-level context but limits the number of rules that can be considered, especially with the pace that new reaction rules are reported in the primary literature. More recently, machine learning (ML) methods have emerged as a powerful tool for chemical synthesis planning, because they can incorporate large datasets of chemical reactions.

What are the significant milestones and implications of using artificial intelligence algorithms for proposing synthetic routes in small-molecule synthesis?

Modern synthesis algorithms are beginning to push beyond the obvious and into more complex synthesis challenges. Earlier CASD generations could follow the rules of organic chemistry, but the answers weren't too far from the obvious answers. As the corpus of available reactions to consider has grown, more novel suggestions are starting to pop out. Experimental validation of computer-planned routes is finally becoming more popular with some key milestones including fully automated execution on robotic platforms, generation of complex natural products by computer-planned routes that are indistinguishable from human-planned routes, and human-computer partnerships that have arrived at exceptionally brief synthesis recipes to access natural products. The merger high-throughput experimentation techniques are an exciting new area that promises to further accelerate the drug discovery process. As more and more systematically captured and machine-readable reaction data becomes available, machine learning predictions will improve. In the future we will likely see even more physics-based predictions encoded into retrosynthesis calculations.

What is the synergy between expert and machine-learning approaches in improving retrosynthetic planning?

In retrosynthetic planning, the goal is to identify optimal synthetic routes for a target molecule by working backwards from the target molecule to simpler starting materials. Expert and machine learning (ML) approaches can be used together to improve the effectiveness of retrosynthetic planning.

Expert approaches rely on the knowledge and experience of human chemists to identify key steps in a synthesis. These approaches are often based on a set of manually curated chemical rules that have been developed over years of research. While these expert approaches are valuable, they are limited by the scope and complexity of the transformations that can be described using manually curated rules.

ML approaches, on the other hand, can analyze large datasets of known chemical reactions and automatically learn patterns and transformations that are difficult or impossible to capture using expert approaches. The ability of ML algorithms to identify these new patterns and transformations can augment the domain knowledge and intuition of human chemists.

The synergy between expert and ML approaches in retrosynthetic planning can be seen in the development of expert-guided ML algorithms. In these approaches, expert knowledge is used to guide the selection of potential synthetic routes generated by an ML model. This allows for increased accuracy and specificity in the selection of synthetic routes, while still benefiting from the efficiency and scale of ML algorithms.

Overall, the combination of expert and ML approaches in retrosynthetic planning has the potential to significantly improve the speed and effectiveness of this process, leading to more efficient drug discovery and ultimately, the development of new therapies for patients.

Could you elaborate on how neural networks trained on expert-coded reaction rules contribute to achieving higher synthetic accuracy in retrosynthetic planning?

Neural networks that are trained on expert-coded reaction rules can improve the accuracy of synthetic predictions in retrosynthetic planning by incorporating detailed chemical knowledge while remaining flexible enough to address incomplete or novel reactions.

Expert-coded reaction rules are a set of pre-defined chemical rules that describe known chemical transformations. These rules are based on years of research and domain knowledge and can be used as a foundation for training neural networks. By training neural networks on these rules, the networks can learn to recognize patterns of reactions and better predict the outcomes of chemical transformations.

The use of neural networks trained on expert-coded reaction rules can also help address the challenge of incomplete or novel reactions. For example, if a chemical transformation has only been previously observed in a limited set of reactions, the neural network can be trained to predict the outcome of that transformation based on the available data and the expert rules. This is particularly useful in the context of drug discovery, where many of the target molecules have never been synthesized before.

By incorporating detailed chemical knowledge in the form of expert-coded rules, neural networks trained on these rules can achieve higher accuracy in retrosynthetic planning. This not only helps chemists to predict the outcomes of chemical reactions more accurately, but also speeds up the drug discovery process by assisting chemists in identifying optimal synthetic routes for new compounds.

How do you leverage a library of chemical reactions and metadata to design sustainable syntheses?

This is an important aspect of future work. The beauty of computational retrosynthesis is that you can demerit protocols that use ecological insensitive reagents, for instance reactions requiring dichloromethane as solvent or that produce a large amount of metal waste. Meanwhile, you can reward protocols that leverage more environmentally friendly options.

One way to do this is to incorporate information on the environmental impact of specific chemical reactions into the library. A variety of factors such as the amount of waste generated, energy required, and toxicity of reagents, can be encoded. This enables chemists to identify sustainable synthetic routes for target molecules, by selecting reactions from the library that meet the required sustainability criteria.

In addition to information on the environmental impact of specific chemical reactions, metadata about chemical reactions can also be used to design sustainable syntheses. This metadata can include information such as reaction yields, solvents used, and energy requirements. By analyzing this metadata, chemists can identify more efficient and sustainable synthetic routes.

By leveraging a library of chemical reactions and metadata to design sustainable syntheses, chemists can reduce the environmental impact of synthetic processes, leading to a more sustainable and less wasteful chemical industry. Furthermore, by selecting reactions from a library of sustainable reactions, chemists can also improve the speed and cost-effectiveness of the drug discovery process, leading to more accessible and affordable treatments for patients.

In the context of artificial intelligence-driven organic synthesis, how can AI algorithms and robotic platforms be coupled together?

The coupling of AI algorithms and robotic platforms in artificial intelligence-driven organic synthesis could significantly improve the speed and efficiency of the discovery process. AI driven organic synthesis is automating the logic of synthesis, while robotic platforms are automating the hands-on lab work so it's a powerful combination.

AI algorithms can be used to predict the optimal synthetic routes for a given target molecule, leveraging vast databases of chemical reactions and related data. Robotic platforms can then be used to synthesize the target molecule based on the predicted routes. The use of these platforms allows for rapid experimentation and high-throughput synthesis of a large number of compounds.

The combination of AI algorithms and robotic platforms can be further enhanced by using feedback loop mechanisms to optimize the synthetic process in real-time. For example, an AI algorithm could monitor the synthetic process and provide feedback on reaction conditions, leading to further optimization of the synthetic route and ultimately better outcomes.

Another way in which AI algorithms and robotic microfluidic platforms can be coupled together is through the use of machine learning algorithms to improve the performance of the microfluidic platform. Through continuous monitoring and feedback, machine learning algorithms can learn to predict optimal reaction conditions and improve the efficiency and accuracy of the microfluidic platform.

Ultimately, by combining the power of AI algorithms and robotic microfluidic platforms, the drug discovery process can be accelerated, allowing for the rapid identification of novel drug candidates. This has the potential to significantly improve patient outcomes by providing faster and more effective treatments for a wide range of diseases.

The exciting thing is that laboratory robotics are becoming more and more accessible to all, through commercial vendors or open-source democratized hardware platforms. The field is sure to accelerate even further as lab automation is playing an increasing role in the education of undergraduate and graduate programs, preparing the next generation of researchers for this future of work.

How has laboratory automation transformed traditional synthesis, and what are the key advantages it offers over manual operation?

Laboratory automation has transformed traditional synthesis by allowing for high-throughput experimentation and faster, more precise synthesis of complex organic compounds. Some of the key advantages it offers over manual operation include increased efficiency, higher precision, improved reproducibility, ability to handle hazardous materials and integration with artificial intelligence and machine learning algorithms.

Automation allows for the rapid and reproducible synthesis of large numbers of compounds, reducing the time and effort required for routine chemical tasks and freeing up scientists to focus on more complex research. Automated equipment can dispense precise volumes of reagents, leading to less variability in reaction outcomes and improved accuracy of results and can handle hazardous materials and reactions safely, protecting researchers from potentially harmful chemicals.

Automation ensures that experiments are performed under consistent conditions, reducing the likelihood of human error and increasing the reproducibility of results. It can be integrated with artificial intelligence and machine learning algorithms to further optimize laboratory operations and accelerate the development of new compounds.

Overall, laboratory automation has the potential to significantly improve the efficiency and precision of traditional synthesis, leading to faster drug discovery and ultimately more effective treatments for patients.

What is the outlook of the future of synthetic automation, and how do you envision the progression from labor-intensive processes to intelligent automation?

The future of synthetic automation is expected to be marked by continued advances in robotics, machine learning, and artificial intelligence, leading to increasingly sophisticated and intelligent automation systems.

One of the key areas of focus in future synthetic automation will be the development of autonomous laboratories that can operate 24/7 without human intervention. These systems will be equipped with robotics and machine learning algorithms that are capable of performing routine tasks, monitoring experiments, and making autonomous decisions based on the data generated.

Another area of focus will be the further integration of artificial intelligence into the drug discovery process. This will involve the development of AI algorithms that can analyze vast amounts of chemical data, predict the outcomes of chemical reactions, and optimize experimental conditions in real-time.

Over time, it is expected that synthetic automation will become more and more intelligent, with the automation systems taking on increasingly complex tasks and generating new insights that can accelerate drug discovery. These highly automated systems will allow researchers to perform a wide range of experiments with minimal labor and human oversight, leading to faster and more efficient drug development.

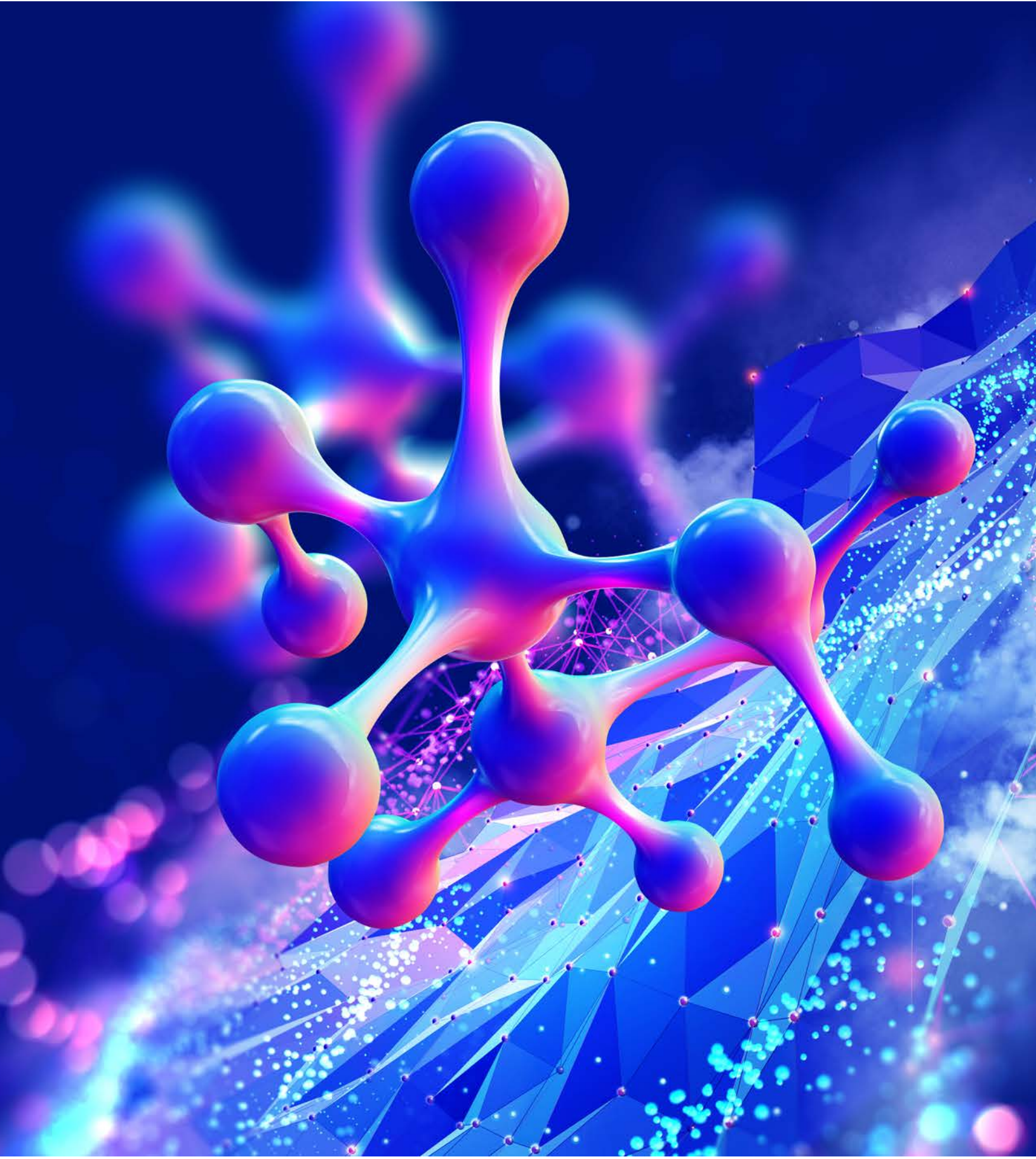
Overall, the trajectory of synthetic automation is toward more intelligent and sophisticated systems that can handle complex experiments and rapidly generate large amounts of data that can be used to further optimize the drug discovery process. In the future, we can expect to see a progression from labor-intensive processes to highly automated, intelligent systems that can revolutionize drug discovery and lead to new treatments for a wide range of diseases.

Considering the insights from these papers and the advancements in computer-aided retrosynthesis, how do you see the continued integration of artificial intelligence and automation shaping the landscape of chemical synthesis in the years to come?

The continued integration of artificial intelligence and automation is expected to dramatically shape the landscape of chemical synthesis in the future. AI and automation are expected to speed up drug discovery by allowing chemists to synthesize compounds more quickly and efficiently. This will enable researchers to test a greater number of molecules and explore a wider range of chemical space. The integration of AI and automation may also help reduce the environmental impact of chemical synthesis by enabling researchers to identify more sustainable chemical routes.

AI algorithms with automated synthesis can lead to reductions in human error and result in more precise synthesis, thereby yielding higher quality products. It will accelerate the exploration of chemical space in drug discovery and lead to the identification of new compounds that would otherwise be difficult to discover through traditional methods. The integration of AI and automation into chemical synthesis has the potential to transform the field in profound ways, greatly accelerating the speed and efficiency of drug discovery while simultaneously improving our ability to identify sustainable and effective drug molecules. The future of chemical synthesis looks bright with the continued integration of these technologies.

Disclosure: The Cernak Lab has received research funding or in-kind donations from MilliporeSigma, Burlington, MA, an affiliate of Merck, Relay Therapeutics, Janssen Therapeutics, SPT Labtech, and MSD Inc. T.C. holds equity in Scorpion Therapeutics and is a cofounder of and equity holder in Iambic Therapeutics.



MERCK

WILEY

MK_BR13106EN