RESEARCH ARTICLE SUMMARY

ORGANIC CHEMISTRY

A robotic platform for flow synthesis of organic compounds informed by AI planning

Connor W. Coley*, Dale A. Thomas III*, Justin A. M. Lummiss*, Jonathan N. Jaworski, Christopher P. Breen, Victor Schultz, Travis Hart, Joshua S. Fishman, Luke Rogers, Hanyu Gao, Robert W. Hicklin, Pieter P. Plehiers, Joshua Byington, John S. Piotti, William H. Green, A. John Hart, Timothy F. Jamison†, Klavs F. Jensen†

INTRODUCTION: The ability to synthesize complex organic molecules is essential to the discovery and manufacture of functional compounds, including small-molecule medicines. Despite advances in laboratory automation, the identification and development of synthetic routes remain a manual process and experimental synthesis platforms must be manually configured to suit the type of chemistry to be performed, requiring time and effort investment from expert chemists. The ideal automated synthesis platform would be capable of planning its own synthetic routes and executing them under conditions that facilitate scale-up to production goals. Individual elements of the chemical development process (design, route development, experimental configuration, and execution) have been streamlined in previous studies, but none has presented a path toward integration of computer-aided synthesis planning (CASP), expert refined chemical recipe generation, and robotically executed chemical synthesis.

RATIONALE: We describe an approach toward automated, scalable synthesis that combines techniques in artificial intelligence (AI) for planning and robotics for execution. Millions of previously published reactions inform the computational design of synthetic routes; expert-refined chemical recipe files (CRFs) are run on a robotic flow chemistry platform for scalable, reproducible synthesis. This development strategy augments a chemist's ability to approach target-oriented flow synthesis while substantially reducing the necessary information gathering and manual effort.

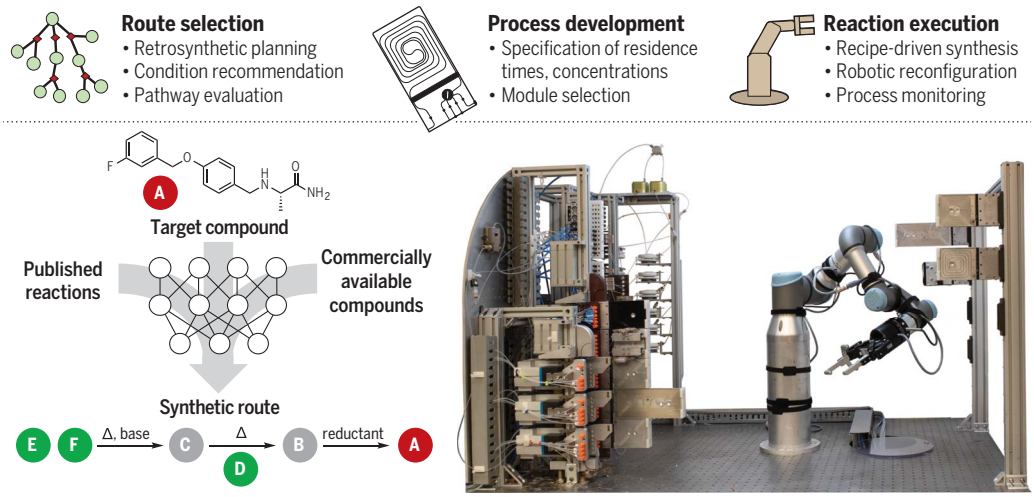
RESULTS: We developed an open source software suite for CASP trained on millions of reactions from the Reaxys database and the U.S. Patent and Trademark Office. The software was designed to generalize known chemical reactions to new substrates by learning to apply retrosynthetic transformations, to identify suitable reaction conditions, and to evaluate whether reactions are likely to be successful when attempted experimentally. Suggested routes partially populate CRFs, which require additional details from chemist users to define residence times, stoichiometries, and concentrations that are compatible with continuous flow. To execute these syntheses, a robotic arm assembles modular process units (reactors and separators) into a continuous flow path according to the desired process configuration defined in the CRF. The robot also connects reagent lines and computer-controlled pumps to reactor inlets through a fluidic switchboard. When that is completed, the system primes the lines and starts the synthesis. After a specified synthesis time, the system flushes the lines with a cleaning solvent, and the robotic arm disconnects reagent lines and removes process modules to their appropriate storage locations.

This paradigm of flow chemistry development was demonstrated for a suite of 15 medicinally relevant small molecules. In order of increasing complexity, we investigated the synthesis of aspirin and secnidazole run back to back; lidocaine and diazepam run back to back to use a common feedstock; (S)-warfarin and safinamide to demonstrate the planning program's stereochemical awareness; and two compound libraries: a family of five ACE inhibitors including quinapril and a family of four nonsteroidal anti-inflammatory drugs including celecoxib. These targets required a total of eight particular retrosynthetic routes and nine specific process configurations.

CONCLUSION: The software and platform herein represent a milestone on the path toward fully autonomous chemical synthesis, where routes still require human input and process development. Over time, the results generated by this and similar automated experimental platforms may reduce our reliance on historical reaction data, particularly in combination with smaller-scale flow-screening platforms. Increased availability of reaction data will further enable robotically realized syntheses based on AI recommendations, relieving expert chemists of manual tasks so that they may focus on new ideas. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aax1566



Planning and execution. A robotically reconfigurable flow chemistry platform performs multistep chemical syntheses planned in part by AI.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: tfj@mit.edu (T.F.J.); kfjensen@mit.edu (K.F.J.)
Cite this article as C. W. Coley *et al.*, *Science* **365**, eaax1566 (2019).
DOI: 10.1126/science.aax1566

Downloaded from http://science.sciencemag.org/ on February 5, 2021

RESEARCH ARTICLE

ORGANIC CHEMISTRY

A robotic platform for flow synthesis of organic compounds informed by AI planning

Connor W. Coley[1]*, Dale A. Thomas III[1,2]*†, Justin A. M. Lummiss[3]*†, Jonathan N. Jaworski[3]‡, Christopher P. Breen[3], Victor Schultz[1], Travis Hart[1], Joshua S. Fishman[2], Luke Rogers[1]§, Hanyu Gao[1], Robert W. Hicklin[3]||, Pieter P. Plehiers[1]¶, Joshua Byington[1]#, John S. Piotti[2], William H. Green[1], A. John Hart[2], Timothy F. Jamison[3]**, Klavs F. Jensen[1]**

The synthesis of complex organic molecules requires several stages, from ideation to execution, that require time and effort investment from expert chemists. Here, we report a step toward a paradigm of chemical synthesis that relieves chemists from routine tasks, combining artificial intelligence–driven synthesis planning and a robotically controlled experimental platform. Synthetic routes are proposed through generalization of millions of published chemical reactions and validated in silico to maximize their likelihood of success. Additional implementation details are determined by expert chemists and recorded in reusable recipe files, which are executed by a modular continuous-flow platform that is automatically reconfigured by a robotic arm to set up the required unit operations and carry out the reaction. This strategy for computer-augmented chemical synthesis is demonstrated for 15 drug or drug-like substances.

The ability to synthesize organic compounds on demand has the potential to transform molecular discovery tasks. Such compounds with typical molecular weights of 50 to 750 g/mol play a central role in a range of disciplines, including specialty polymers, organic photovoltaics, energetics, and medicines. Synthesis is often a bottleneck in small-molecule drug discovery (*1*), where design–synthesize–test iterations have cycle times on the order of weeks and where the scope of a compound library synthesis can determine the accuracy of an empirical structure–activity relationship model (*2*). Materials discovery researchers face similar limitations arising from their inability to synthesize diverse compounds, e.g., candidate organic photovoltaics, and to do so rapidly (*3*).

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [2]Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [3]Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA.
*These authors contributed equally to this work.
†Present address: Mytide Therapeutics, Boston, MA 02111, USA. ‡Present address: Bristol-Myers Squibb, New Brunswick, NJ 08901, USA. §Present address: On Demand Pharmaceuticals, Biotech 8, Richmond, VA 23219, USA. ||Present address: Pfizer, Groton, CT 06340, USA. ¶Present address: Laboratory for Chemical Technology, Ghent University, Ghent, Belgium. #Present address: L3 Technologies Space Systems, Wilmington, MA 01887, USA.
**Corresponding author. Email: tfj@mit.edu (T.F.J.); kfjensen@mit.edu (K.F.J.)

Many chemists and chemical engineers are pursuing the promise of a machine capable of synthesizing large numbers of molecules with little to no human intervention (*4*, *5*). Although major advances in laboratory automation have decreased the manual effort required to perform some classes of chemical reactions (*6*–*8*), the identification and development of synthetic routes to novel molecules remain a largely manual process requiring a time investment from expert chemists. Moreover, current automated synthesis platforms must first be configured to accommodate the necessary sequence of unit operations or be constrained to a subset of otherwise-accessible chemical space. The scope of chemical reactions compatible with current automated platforms tends to be limited by reaction type (*9*), solvent and temperature (*10*), or concentration and time (*11*).

The ideal automated synthesis platform would be compatible with reaction conditions that can be directly translated from small-scale process development to gram or kilogram manufacturing. Continuous-processing approaches, e.g., synthesis in plug-flow reactors or continuous stirred tank reactors, can offer such scalability and are widely recognized as an enabling technology in many respects, including for process quality improvement (*12*). The smaller length scales relative to batch synthesis enhance heat and mass transfer and are amenable to more precise quantification of the rates thereof before scale-up (*13*). Moreover, flow platforms offer smaller footprints compared with their batch counterparts and provide access to accelerated reaction rates through process intensification (*14*, *15*). Numerous multistep syntheses have been successfully implemented in flow and offer substantial reductions in total reaction time (*14*, *16*–*18*).

The chemical development process for small molecules can be divided into a number of distinct stages, including design (literature search, retrosynthesis, condition selection, feasibility estimation), route development (recipe formulation), experimental configuration (platform reconfiguration), and execution (process execution, scalable synthesis) (Fig. 1A). Previous studies have sought to automate individual aspects of this process but have not presented a path to full automation. Retrosynthesis can be streamlined using Chematica's expert-encoded reaction rules (*19*, *20*) or Segler *et al.*'s algorithmically extracted rules and learned search strategy (*21*), but the former approach is difficult to scale with the growing body of chemical literature and neither explicitly proposes reaction conditions or evaluates feasibility of the forward reaction. Automated chemical synthesis using a predefined instruction set is well proven (*6*, *8*), but has been restricted to batch and thus does not offer a clear path to scaled-up synthesis. Whereas flow chemistry platforms have been developed for automated screening, optimization, and production (*5*, *7*, *22*–*24*), they require manual reconfiguration to the exact flow path required for each process.

Our approach toward automated, scalable synthesis combines techniques in artificial intelligence (AI) for planning and robotics for execution. Specifically, we describe a platform that can design synthetic routes by generalizing millions of previously published reactions (Fig. 1B), including partial specification of reaction conditions and process variables, and then execute human-refined chemical recipe files (CRFs) using a robotically reconfigurable flow chemistry platform (Fig. 1C). Adjustments to the AI-proposed synthetic route required for compatibility with continuous flow are recorded in these reusable recipes for scalable, reproducible synthesis.

This development strategy augments a chemist's ability to approach target-oriented flow synthesis while substantially reducing the necessary information gathering and manual effort. We illustrate this paradigm of chemical development by predicting and automating the synthesis of 15 drug or drug-like molecules.

In this workflow, we rely on expert input from chemist users, minimally including residence times, equivalence ratios, and concentrations, to translate recommendations into practice. In particular, anticipating compatibility with flow necessitates solubility predictions at a level of accuracy currently achievable only empirically; identifying opportunities for telescoped reaction sequences and placement of interstage separations requires quantitative prediction of reaction outcomes and of reagent compatibility. More practically, data describing concentrations, equivalence ratios, and orders of addition are not tabulated in any available reaction databases, precluding data-driven approaches to full process specification, flow or otherwise. In the

Downloaded from http://science.sciencemag.org/ on February 5, 2021

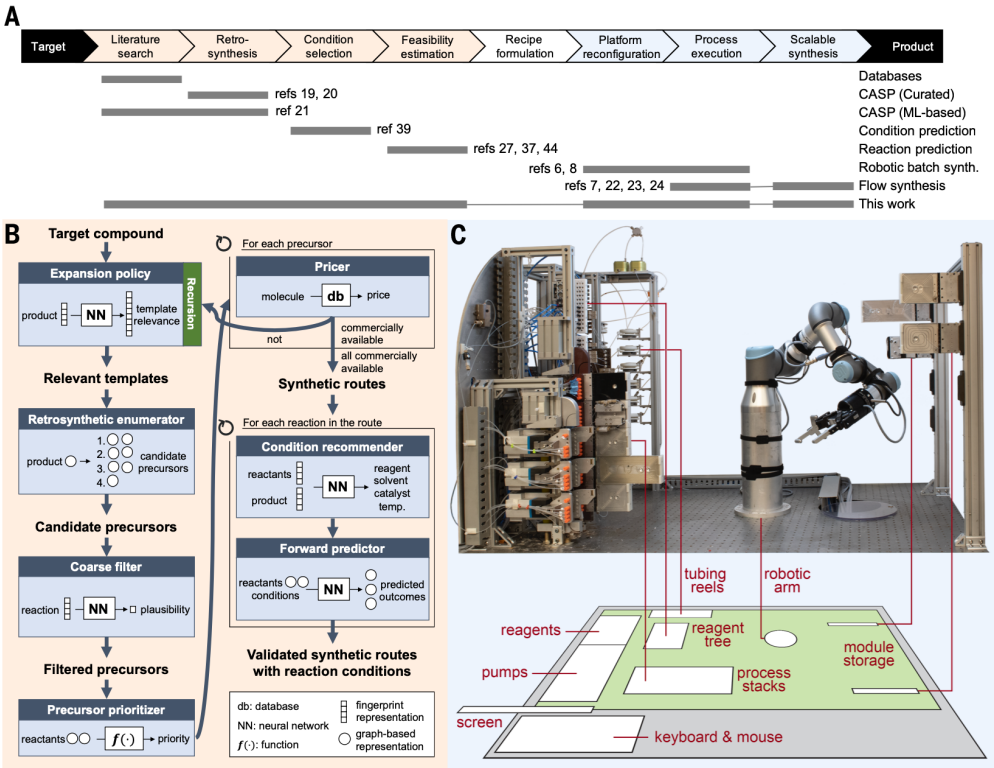


Fig. 1. Overview of the robotically configurable reaction planning and execution platform in the context of the chemical development pipeline. (**A**) Workflow for on-demand synthesis of a targeted organic compound and representative work from Grzybowski (*19*, *20*), Waller (*21*), Jensen (*27*, *37*, *39*), Schwaller (*44*), Godfrey (*6*), Cronin (*8*), Jamison and Jensen (*7*, *22*, *23*), and Ley (*24*) that have streamlined parts of this process; thick gray lines indicate automated steps. (**B**) Software modules combining cheminformatics and machine learning to design and validate synthetic pathways. (**C**) Photograph of the robotic flow chemistry platform, projected floorplan of the 6-foot × 4-foot working table (gray background), and ventilated enclosure (green background). CASP, computer-aided synthesis planning; ML, machine learning.

"Outlook" section, we define challenges that will have to be overcome to automatically generate recipes.

Synthesis-planning module

Computer-aided synthesis planning (CASP) originated as a tool to help chemists identify pathways before executing them in the laboratory (*25*, *26*). Reaction databases (e.g., Reaxys, SciFinder) have streamlined the process of searching for known compounds to find known syntheses and are now routinely used. However, programs that generalize synthesis planning to novel compounds have not achieved widespread adoption, perhaps an indication of their as-yet limited capabilities and costs.

There has been little explicit verification of computer-suggested synthesis plans; one exception is a recent demonstration using the program Chematica to discover new routes based on expert-encoded transformation rules (*19*). A renewed interest in CASP brought about by recent advances in data science and machine learning has led to several recent studies (*27*, *28*–*30*), including Segler *et al.*'s application of a Monte Carlo tree search to expedite the recommendation process (*21*).

This section describes the development of our synthesis-planning program (Fig. 1B). It integrates our previous efforts to generalize known chemistry to new substrates by learning to apply retrosynthetic transformations, to identify suitable reaction conditions, and to evaluate whether reactions are likely to be successful when attempted experimentally into a single open source software framework that we call ASKCOS. This software was trained directly on millions of reactions extracted from the U.S. Patent and Trademark Office (USPTO) (*31*) or tabulated in Reaxys.

We see CASP as a recommendation problem. For a specified target molecule, the program must propose a sequence of chemically viable reaction steps starting from available chemical reactants. If the objective were to identify only known syntheses for known molecules, then this could be treated as a graph search problem. To identify new syntheses, however, candidate reaction steps must be generated on the fly to define the search space.

Reaction templates are subgraph-matching rules that can be algorithmically extracted from literature-precedent reactions and applied to new substrates to recognize structural motifs that lend themselves to retrosynthetic disconnection. From 12.5 million published single-step reactions tabulated in the Reaxys database, we prepared a library of all rules observed ≥10 times and all rules with specified stereochemistry observed ≥5 times, totaling 163,723 rules. Transformations with stereochemistry are inherently more specific and are expected to appear less frequently yet are essential to include to allow the program to predict syntheses of chiral molecules. The program uses RDKit and RDChiral to apply transformations and enforce consistency in handling stereochemistry (*32*, *33*).

We trained a feedforward neural network model to predict which of the 163,723 transform rules are most applicable to a target molecule based on its molecular structure [ECFP4 (*34*)] (*28*). Applying only the templates perceived to be most relevant serves two roles: (i) reducing computational cost of exhaustive template subgraph matching and (ii) increasing the likelihood of proposing a chemically feasible reaction. We have focused much of our effort on this latter point: ensuring that reaction suggestions are not false-positives (i.e., recommendations that would not work in the laboratory). This is essential for maximizing the probability of experimental success.

Two modules assess the quality of a single retrosynthetic suggestion. The first is a binary classifier meant to remove only the lowest-quality recommendations based on Segler *et al.*'s "in-scope filter" (*21*). This neural network model is trained on 15 million published positive reaction examples and 115 million artificial negative reaction examples generated by application of algorithmically extracted forward reaction templates, exactly following the protocol in (*27*). The binary classifier is meant to answer the question: Is there any set of conditions for which these reactants will form this product? Reactions passing this filter with a user-tunable threshold are added to the growing search tree.

Retrosynthetic expansion occurs recursively for a predetermined amount of time (no longer than 30 s for all examples described below) or up to a specified depth before all pathways using suitable starting materials are returned. The program uses a root-parallelized Monte Carlo tree search (specifically, an upper confidence–bound method) to balance exploitation of branches thought to be promising and exploration of less frequently visited branches (*35*, *36*,). Our database of buyable chemicals consists of roughly 107,000 compounds available for less than $100/g from eMolecules and Sigma-Aldrich, though the maximum price is a user-tunable parameter in each expansion and

Downloaded from http://science.sciencemag.org/ on February 5, 2021

additional stop criteria are available (see section 1.5 in the supplementary text).

Pathways that consist of reaction steps considered plausible by the binary classifier can be evaluated by the more rigorous forward prediction model, which explicitly generates product molecules based on the provided reactants and reaction conditions (*37*, *38*); the exact model used is described in detail in (*37*). Reaction conditions are provided by a neural network model trained to propose a prioritized list of reagents, solvents, catalysts, and temperature most suitable for that transformation as previously reported (*39*). If the likely outcome according to the forward predictor matches the intended outcome, then we can be more confident that the reaction is chemically viable. The forward predictor model also provides predictions of side products, which assist in impurity identification during process development. Explicit validation of stereoselective reactions is not yet possible because of limitations in three-dimensional molecular representations and inconsistent data reporting.

Even after these stages of validation and filtering, pathway-level concerns might remain. In addition to the flow chemistry considerations stated above, users could also bias pathways toward those with reaction types frequently implemented at large scale (e.g., for process applications), or toward pathways that enable diversification of intermediates (e.g., for discovery applications). Our code framework enables extension to such considerations. A detailed illustration of the code output is shown for an example target in Fig. 2.

The complexity of synthetic routes that we can plan computationally exceeds what makes sense to implement in a telescoped flow synthesis. Moreover, the CASP program can predict routes to targets it has never seen before, such as recently reported modern active pharmaceutical ingredients (APIs) and de novo molecular optimization targets (see section 1.10 in the supplementary text).

Robotic flow chemistry platform

We considered several concepts for designing a self-configurable system, based on previous work on manual plug-and-play unit operations (e.g., reactors and separators) for continuous synthesis (*22*, *40*). As a flexible and process-module–efficient design, we chose to implement a six-axis robotic manipulator to select process modules from storage locations, and then arrange the modules in the sequence required for a particular synthesis (Fig. 1C and figs. S12 to S14). This arrangement also enabled straightforward upgrading of existing process modules as well as the addition of new modules, all using a common fluidic interface. Alternatively, we considered implementing banks of unit operations with two-way valves; however, this arrangement would have created redundancy and greater complexity of the flow path,

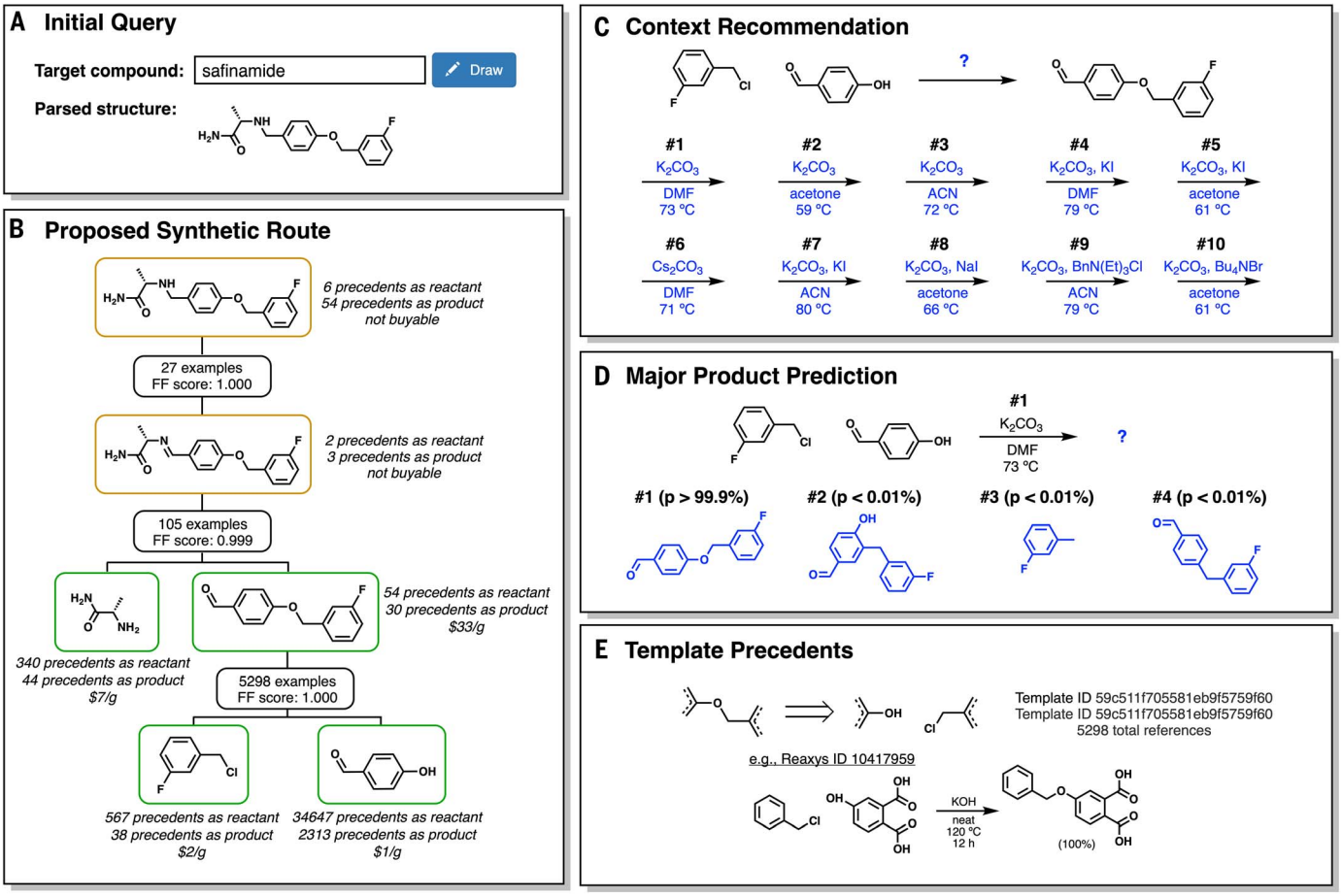


Fig. 2. Typical outputs provided by the software using safinamide as an illustrative target. (**A**) Query format, in which users can specify compounds by common name, drawing, or SMILES string; (**B**) one of many synthetic routes shown to the user, with additional information about commercial availability and number of precedents visible on hover; (**C**) top 10 reaction conditions proposed for the etherification step; (**D**) prediction of the major product under the top recommended conditions, which shows high confidence in the ether product; (**E**) link to summary of literature precedents supporting each reaction template, including the option of exporting a Reaxys query. Website printouts S1 to S34 in the supplementary materials contain the exact display format of results. FF, fast filter used for binary classification of reaction feasibility; DMF, dimethylformamide; ACN, acetonitrile.

Downloaded from http://science.sciencemag.org/ on February 5, 2021

especially with an integrated capability to add fluids (reactants, reagents, or solvents) before each operation. For instance, equipping this alternative design to perform two reactions (with the option of five different reactor types and sizes), one separation with three different configurations, and seven different fluids would require 77 valves, each presenting potential failure points due to leaks and clogging (fig. S11)

In our approach, the robotic manipulator configures the synthesis apparatus by assembling the required unit operations and reagent lines on demand. The process stack is robotically loaded with process modules (e.g., reactor, separator, or packed bed) and then process bays and connections are pneumatically clamped, sealing the fluidic interfaces and forming a continuous flow path. The reagents are plumbed to the process stack through a robotically manipulated "switchboard," analogous to a telephone switchboard yet with flexible tubes rather than wires. In this way, we connect the fluid pumps, outlets, and waste streams to each process module as required by the synthetic schemes (Fig. 3B and fig. S28), thus not requiring complex banks of valves (for demonstrations, see movies S1 to S3).

The ends of each fluid line are reversibly coupled to the process modules by the robot using magnetically preloaded kinematic couplings, and mechanically controlled actuators seal the flow path, allowing high-pressure operation [up to 250 pounds per square inch gauge (psig)]. To avoid tangling, a constant tension is exerted on each tube using a power-spring–preloaded tubing reel installed on the reagent tree (Fig. 3B and fig. S16). For fluid delivery, the system automatically controls a combination of continuous flow piston pumps. To carry out library syntheses, two pumps are equipped with selector valves to enable selection from up to 24 preloaded feedstocks each.

To support a variety of chemical transformations, we developed a library of process modules, including laminar flow reactors with different volumes (100 μl to 3 ml, Fig. 3D), packed bed reactors (1 to 2 ml, Fig. 3E) capable of operating at temperatures from ambient to 200°C and pressure up to 250 psig, and a membrane separator unit for liquid–liquid extraction (*41*) (Fig. 3H). The process units are placed in one of two process stacks composed of universal process bays (UPBs) (Fig. 3I). UPBs provide sealing and alignment mechanisms for the fluidic, electrical, and pneumatic process connections. Fluid connections needed between adjacent units in the synthesis are achieved by vertically stacking the units in the towers in the required order and, when all units are in place, pneumatically sliding them together to seal the linear fluid path. The two stacks were designed to allow up to either an eight-step linear or a five-step convergent synthesis.

To create versatile, disposable, and chemically compatible fluidic paths, we developed a blow-forming process that integrated millifluidic valves and channels in perfluoroalkoxy alkane (PFA) films with PFA tubes (Fig. 3F and figs. S22 and S23). The inset allows up to four process streams to be connected to each unit. The reactors also have an optional auxiliary port that can be used in the packed bed module to introduce a gas or in the liquid–liquid separator module as a secondary outlet. To enable elevated temperature and pressure operation, the film reactor is enclosed by an

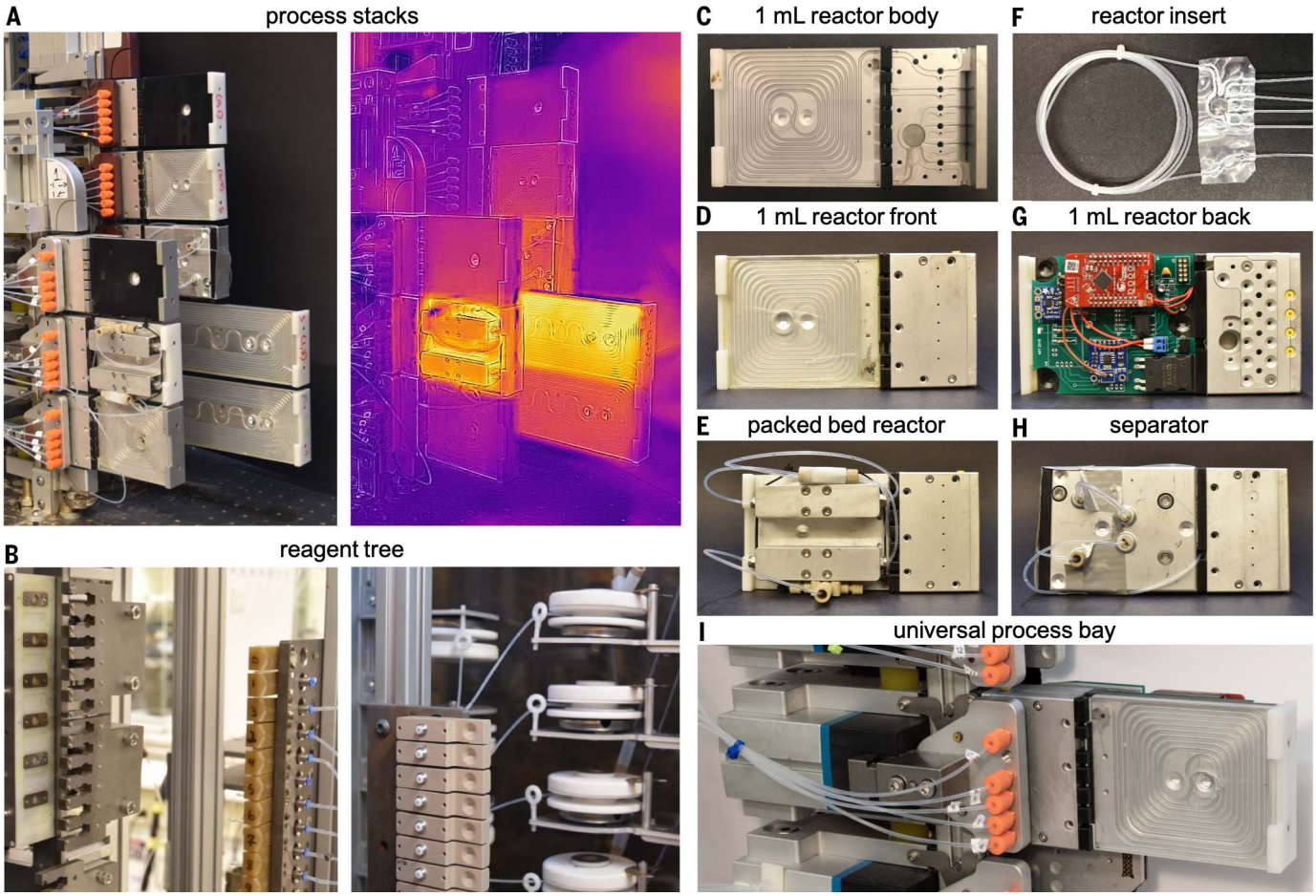


Fig. 3. Process and submodules implemented on robotic flow chemistry platform. (**A**) Process stack where modules interface with UPBs to form a continuous flow path (left) and thermal image showing heated reactors (right); (**B**) front view of the reagent tree and reagent manifold; (**C**) image of front view of reactor body; (**D**) 1.0-ml reactor process module; (**E**) two-column packed bed reactor process module; (**F**) disposable PFA reactor insert; (**G**) integrated electronics on back side of 1.0-ml reactor; (**H**) in-line membrane separator; (**I**) close-up of a UPB holding a 1.0-ml reactor.

Downloaded from http://science.sciencemag.org/ on February 5, 2021

aluminum shell (Fig. 3C). The process modules each have integrated electronics allowing for temperature control (Fig. 3G), as well as the addition of new process modules, without changing or wiring new components to the central control system.

Translation of a chemical synthesis onto the robotic platform occurs through the creation of a CRF. Each CRF specifies the fluidic path that must be constructed; the location of stock solutions; the sequence of process modules that must be moved from storage to the process stack; and the start-up, steady-state, and shutdown flow rates. These files are manipulated and optimized before being run on the automated system based on input from chemist operators. To run a CRF, the platform is loaded with the reagents at their specified locations for synthesis. The user then presses the process start button for the execution of the recipe. The program carries out all of the path planning between the module storage and process stack, and reagent tree and reagent manifold to assemble the flow path of interest. The system then follows the CRF to prime, set flow rates, set pressure and temperature, and wash and disassemble the process.

Predicting and automating the synthesis of 15 small molecules

The primary application of this platform is the on-demand synthesis of target small-molecule organic compounds. An important class of compounds within this context is APIs, which vary greatly in structural and synthetic complexity. We therefore chose a suite of 15 medicinally relevant small molecules, which ultimately required eight particular retrosynthetic routes and nine specific process configurations. Although literature precedents exist for all 15 targets, the synthesis-planning program is prevented from merely recalling any synthetic route from memory as exact matches; all pathways are required to be discovered de novo through abstracted transformation rules and learned patterns of chemical reactivity. A discussion of software capabilities and examples of more complex targets can be found in section 1.10 of the supplementary text. Raw program outputs for the examples shown below are also available in the supplementary materials (website printouts S1 to S34). CRFs for each molecule include additional specification of concentrations, flow rates, and process stack configurations required to achieve the proposed transformations. This information could be predicted in a data-driven manner given a sufficiently detailed database of precedent reactions, or defined by expert-crafted heuristics, but given the insufficient information in current databases, it has been preserved as a manual task in this work.

In order of increasing complexity, we investigated the synthesis of aspirin (**1**) and racemic secnidazole (**4**) run back to back; lidocaine (**7**) and diazepam (**12**) run back to back to use a common feedstock; and (S)-warfarin (**15**) and safinamide (**18**) to demonstrate the planning program's stereochemical awareness (Fig. 4). We also include in the supplementary materials a representative example of bezafibrate as a synthesis that was planned by the software, but after expert evaluation and screening was found to be a poor candidate for translation into continuous flow (see section 2.13.1 in the supplementary text).

We additionally planned and executed the synthesis of two compound libraries: one representing a family of angiotensin-converting-enzyme (ACE) inhibitors including quinapril (**24a**), moexipril (**24b**), enalapril (**24c**), ramipril (**24d**), and indolapril (**24e**), and one representing a family of nonsteroidal anti-inflammatory drugs (NSAIDs) including celecoxib (**28aa**) and three analogs (**28ab**, **28ba**, and **28bb**) (Fig. 5). The following sections describe these targets in the context of the platform capabilities that they illustrate. Detailed operator instructions, CRFs, photos, infrared photos, and schematics of each process configuration can be found in the supplementary materials.

Back-to-back synthesis of aspirin and secnidazole

The platform's self-reconfigurability makes it particularly attractive for back-to-back syntheses, which we first demonstrated using aspirin (**1**) and secnidazole (**4**).

The program proposed the well-precedented one-step synthesis of aspirin (**1**) through acetylation of salicylic acid (**2**) with acetic anhydride (**3**), acetyl chloride, or acetic acid. The program is biased toward acetic anhydride because of the template relevance neural network (i.e., because it is more commonly used for similar substrates) and the forward predictor's greater confidence that the anhydride will yield the desired product. The top-recommended reaction conditions used neat reactants (no solvent) at 108°C without additional reagents or catalysts, or with sulfuric acid at a milder temperature. Although the reaction can be and is often run neat in batch, for processability and to mitigate clogging risk, we prepared **2** in ethyl acetate (EtOAc), which also enabled downstream separation to produce aspirin in 91% yield (1.72 g/h).

After several residence times of steady-state operation, the feed streams were rerouted to pure solvent, the temperature set points were lowered for the reactor, and the flow path was depressurized by purging the back-pressure regulator. The robotic arm then replaced the process modules for aspirin with those required for secnidazole, reprimed the system, and began the run.

The proposed one-step synthesis of secnidazole (**4**) was the NaOH-catalyzed opening of propylene oxide (**6**) by 2-methyl-5-nitroimidazole (**5**) at 47°C; lower-ranked suggestions included N-alkylation by 1-bromo-2-propanol or the chloro equivalent under similar conditions. The epoxide opening proceeded readily in flow at 140°C [selected to be just under the atmospheric boiling point of dimethylformamide (DMF)] and excess epoxide (at an unoptimized molar ratio of 3:1 to increase the reaction rate) with triethylamine (TEA), as a standard replacement for NaOH in continuous synthesis, to afford secnidazole in 95% yield (792 mg/h) using one 3.0-ml reactor and one separator module.

Back-to-back synthesis of lidocaine and diazepam with a common feedstock

We continued our exploration of back-to-back syntheses, taking advantage of the robotic reconfigurability, with lidocaine (**7**) and diazepam (**12**). Among the proposed routes were two with precedented reaction steps, shown in Fig. 4A. Whereas the program had a very slight preference for chloroacetyl chloride in the first step toward lidocaine, we chose the pathway using bromoacetyl chloride (**11**) to exploit a common reagent stream (**aj**, corresponding to neat **11**) and because the forward predictor is more confident in bromo substrate's likelihood of completing the ring-closing step. The program proposed suitable conditions (ammonia) for the ring-closing amination from **13** to **12** when using the chloro equivalent of **13**, but proposed a less plausible option of sodium or lithium azide for **13** itself, which might require an additional reduction step.

Commercially available compounds **10** and **11** readily reacted to form intermediate **9**, which in the presence of base, heat, and diethylamine (**8**) afforded **7** in 77% yield (1.09 g/h). After 6 hours of operation, all reagent streams were switched to pure *N*-methylpyrrolidone (NMP) and temperature set points were set to zero. The reaction solvent of NMP was selected as a compromise between suggestions of MeOH, dichloromethane (DCM), and acetonitrile (ACN) for various reaction steps as a multipurpose high-boiling polar aprotic solvent. The system was depressurized and deconstructed by the robot before reconstructing the process stack, exchanging the second reactor from the 3.0- to the 1.0-ml size. Using the same fluid streams **aj** and **ak**, the flow path for diazepam was configured. Diazepam was recovered after the separator in 75% yield (638 mg/hour).

Stereoselective and stereoretentive syntheses of (S)-warfarin and safinamide

To increase the complexity of our targets, we then looked toward (S)-warfarin (**15**), which can be produced in a single step from acetocinnamone (**16**) and 4-hydroxycoumarin (**17**) through an asymmetric Michael addition. The program successfully identified this disconnection, recognizing that this Michael addition can be performed stereoselectively. The context recommender proposed the chiral amine catalyst (8S,9S)-6′-methoxycinchonan-9-amine. Although this catalyst could have been suitable based on prior literature, we substituted the more readily available (1S,2S)-(+)-1,2-diphenylethylenediamine [(S,S)-DPEN)] as its cost exceeded $500/g; whereas the costs of starting materials are considered during planning, the costs of catalysts are not because they are often used in substoichiometric quantities. Translation to flow benefited from premixing **16** and (S,S)-DPEN in the presence of acetone before the Michael addition, which was determined only empirically after manually screening; both steps were run at higher temperatures than recommended (50°C versus 20 to 30°C) to reduce the

Downloaded from http://science.sciencemag.org/ on February 5, 2021

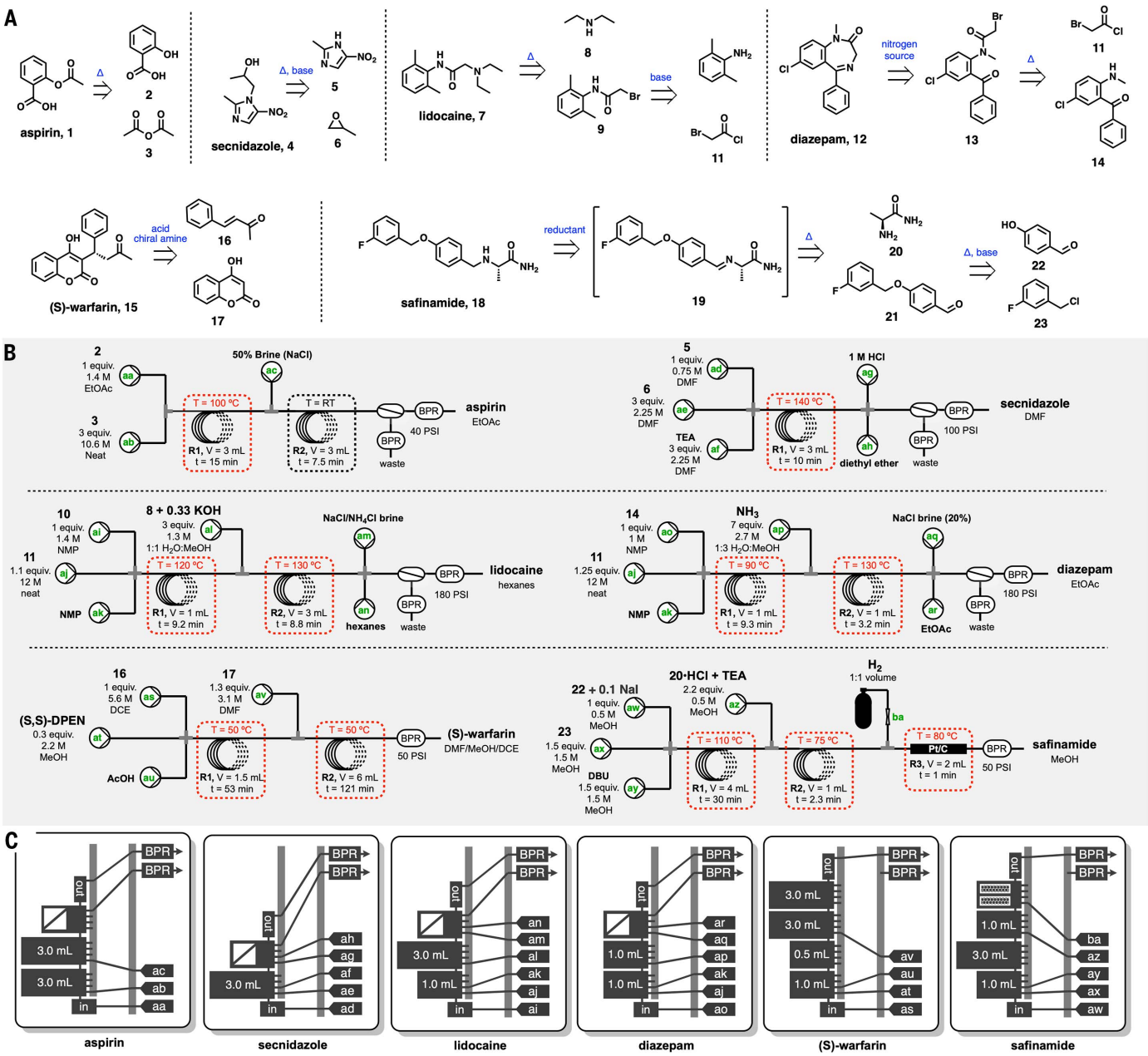


Fig. 4. Synthesis planning and execution for six example drug substances: aspirin 4, secnidazole 4, lidocaine 7, diazepam 12, (S)-warfarin 15, and safinamide 18. (**A**) Synthetic routes proposed by the synthesis-planning program including conditions for the forward reaction (blue); (**B**) continuous flow implementations of the proposed routes; (**C**) robotically configured flow path and unit operations used to execute each of these six syntheses on the modular flow chemistry platform. Secnidazole and aspirin were run back to back. Lidocaine and diazepam were similarly run back to back, taking additional advantage of the common precursor **11** and the corresponding reagent stream **aj**. (S)-Warfarin and safinamide were run to demonstrate the successful identification of stereoselective and stereoretentive routes. Specific fluid streams are labeled alphabetically (green). DMF,dimethylformamide; TEA, triethylamine; NMP, *N*-methylpyrrolidone; (S,S)-DPEN, (1S,2S)-(+)-1,2-diphenylethylenediamine; DBU, 1,8-diazabicyclo[5.4.0]undec-7-ene.

residence time at the expense of enantioselectivity. The final route achieved 78% yield for a throughput of 730 mg/hour, 4.1:1 enantiomeric ratio.

A stereoretentive route to the chiral drug substance safinamide (**18**) was identified by the synthesis-planning program using L-alaninamide (**20**). The reductive amination of aldehyde **21** with **20** was proposed as a single-step reaction or the equivalent two-step reaction with corresponding imine **19** as an explicit intermediate. Although **21** is in our database of buyable compounds for \$33/g, we imposed a maximum price of \$10/g to force the recursive expansion further back to chloride **23** and phenol **22** as simpler, cheaper starting materials. The suggested conditions for this etherification and subsequent reductive amination were modified for flow compatibility [DBU (1,8-diazabicyclo{5.4.0}undec-7-ene) is a common organic, soluble replacement for K_2CO_3 that does not produce gaseous by-products]. The imine reduction was completed in methanol with H_2 in a heated Pt/C packed-bed reactor module (the sixth-ranked set of conditions, rather than using sodium borohydride or cyanoborohydride as they

Downloaded from http://science.sciencemag.org/ on February 5, 2021

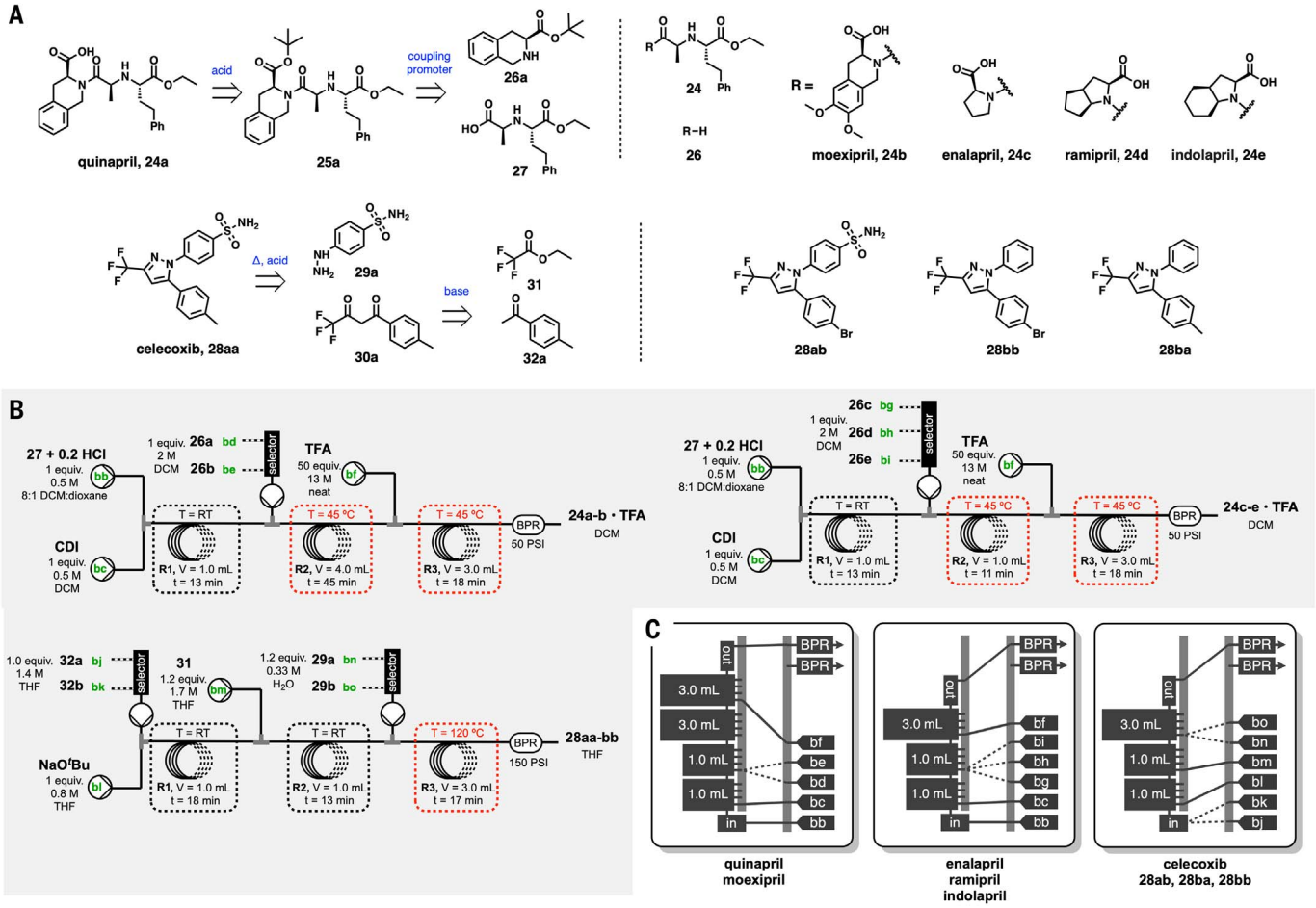


Fig. 5. Synthesis planning and execution for two compound libraries based around quinapril (24a) and celecoxib (28aa). (**A**) Synthetic routes proposed by the synthesis-planning program including conditions for the forward reaction (blue); (**B**) continuous flow implementations of the proposed routes using selector valves for on-the-fly changing of reagent streams; (**C**) robotically configured flow path and unit operations used to execute these syntheses on the modular flow chemistry platform. Specific fluid streams are labeled alphabetically (green). DCM, dichloromethane; CDI, 1,1′-carbonyldiimidazole; TFA, trifluoroacetic acid; THF, tetrahydrofuran.

are highly hygroscopic and prone to quenching) to produce **18** in 32% yield (265 mg/h).

Synthesis planning and execution of an ACE inhibitor library

Automatic rerouting of fluidic connections and two 24-way selector valves enables the use of dozens of different feedstocks. To showcase this capability, we prepared a library of five ACE inhibitors based around quinapril (**24a**) and derivatives accessible by variation of the secondary amine reactant **26**.

Quinapril (**24a**) features three tetrahedral centers that the program identified in purchasable precursors. Activation with 1,1′-carbonyldiimidazole (CDI) enabled the coupling of amino acid derivative **27** with **26a** to produce ^{t}Bu-protected quinapril (**25a**); diethyl cyanophosphonate and *N*,*N*′-dicyclohexylcarbodiimide (DCC) were recommended as coupling reagents but were avoided because of concerns over acute toxicity. As recommended by the program, the ester was readily cleaved in the presence of trifluoroacetic acid (TFA). The program also proposed a single-step amide coupling with the carboxylic acid equivalent of **26a**; however, this pathway was flagged as potentially problematic by the forward predictor because of the competing amidation at the secondary amine in **27**, although the desired product was predicted to be dominant.

Initial screening revealed that quinapril and moexipril (**24b**) benefited from longer residence times to achieve the initial coupling between **26** and **27** before the deprotection; preparation of enalapril (**24c**), ramipril (**24d**), and indolapril (**24e**) proceeded more rapidly, enabling the use of a smaller reactor.

We prepared two CRFs encompassing routes to this pair and trio of targets. A small 1.0-ml reactor at room temperature was installed to allow sufficient time for activation of **27** with CDI before introducing **26** into the subsequent 1.0- or 4.0-ml heated reaction module. Fluidic lines were switched between precursors via 24-port selector valves. Crude quinapril, moexipril, enalapril, ramipril, and indolapril were collected as their TFA salts and isolated offline in 70%, 50%, 58%, 59%, and 66% yield (459, 369, 342, 390, and 420 mg/h, respectively). The full library was run for a total of 68 hours.

Synthesis planning and execution of an NSAID library

We next prepared a two-dimensional compound library of NSAIDs based around celecoxib (**28aa**). Celecoxib features an indole core that the software proposed to synthesize through a condensation of hydrazine **29a** and diketone **30a**, prepared from acetophenone **32a** and ethyl trifluoroacetate (**31**). The suggested use of a base followed by acid and heating in the forward direction is consistent with literature precedents. By exchanging the hydrazine with phenylhydrazine (**29b**) and/or 4-methylacetophenone with 4-bromoacetophenone (**32b**), three additional celecoxib analogs (**28ab**, **28bb**, **28ba**) were prepared.

The CRF for celecoxib was written to use two room-temperature reactors (each 1.0 ml) to achieve the proper order of addition (i.e., generation of the

Downloaded from http://science.sciencemag.org/ on February 5, 2021

acetal from **32a** before reaction with **31**). The four targets were synthesized sequentially by switching the reagent line from **32a** to **32b**, then from **29a** to **29b**, and **32b** back to **32a**. Celecoxib and **28ab**, **28bb**, and **28ba** were obtained in 91%, 59%, 79%, and 86% yield (572, 417 478, and 432 mg/h, respectively). The full library was run for a total of 28 hours.

Outlook

By integrating CASP and a robotic flow chemistry platform, we have enabled the streamlined execution of AI-planned synthetic routes to small-molecule targets. CRFs serve as an intermediate between the two to record the full process details as specified by expert chemists.

Approximate conditions for batch synthesis can be generated based on the literature, as we have done in this study, but their direct implementation in flow is challenging. The desire for process intensification (e.g., to decrease reaction times), the need to mitigate solids formation to avoid clogging, and the importance of telescoping multiple unit operations requires deviation from batch conditions and a level of confidence of predictions that flow chemistry has not yet achieved. Computational prediction of solubilities to within even a factor of 2 in nonaqueous solvents and at nonambient temperatures remains elusive. Predicting suitable purification procedures is a general challenge, not just for flow chemistry, particularly when using nonchromatographic methods. To develop routes to new target molecules, the identification and quantification of species in the crude product will need to be automated as well. Coupling the platform to automated optimization routines would be straightforward for single- and dual-step reactions given a reasonable set of starting conditions as demonstrated previously using other flow chemistry platforms (*7*, *23*). However, optimization of multistep reactions will be more complex owing to the propagation of parameter changes from early steps to later reactions.

Although the routes currently benefit from offline process development, this is largely a limitation of the data that exist in the public domain and our interest in leveraging the benefits of flow chemistry. Over time, the results generated by this and similar automated experimental platforms may obviate our reliance on historical reaction data, particularly in combination with smaller-scale flow-screening platforms (*11*, *42*). Increased availability of reaction data would further enable the robotically realized syntheses based on AI recommendations, relieving expert chemists of manual tasks so that they may focus on new ideas.

Methods summary

A more detailed description of the synthesis-planning software, the robotic platform, stock solution preparation, and chemical analysis are provided in the supplementary materials. A summary of the user workflow follows.

Recipe generation

Suggested synthetic routes and reaction conditions were obtained from the web-based graphical user interface of the ASKCOS program for each target molecule. A preference was set for shorter synthetic routes, as requiring many reaction steps complicates implementation in flow. Reaction conditions were adapted from the batch-generated recommendations through noncodified flow chemistry intuition. Changes are described in the text for each target and typically included the use of soluble organic bases rather than inorganic bases, common solvents, and elevated temperatures to increase reaction rates. Conditions were manually screened to determine appropriate concentrations and approximate residence times; future work should aim to automate this step. A spreadsheet file was populated ("Recipe Planning File S1" in the supplementary materials) and programmatically expanded into the full CRF ("Recipe Planning Script S1" in the supplementary materials) for each target ("Recipes S1 to S7" in the supplementary materials).

Platform operation

Stock solutions were prepared by operators according to CRF-defined concentrations and loaded onto the platform at the specified reagent line numbers. Reagent solutions and pure solvent duplicates (as rinse solutions) were placed under a 5 psig argon atmosphere. The CRF was loaded into the platform control graphical user interface and executed. The robotic arm placed the required unit operations in the process stack and made the necessary fluidic connections in the reagent tree. During automated priming and pressurization with rinse solutions, the system was inspected for leaks as a safety measure. Fraction collection was started when fluidic lines were automatically changed from rinse solutions to reagent solutions and temperature set points were reached. After operating for a duration specified in the CRF or after an aliquot of the product stream showed reasonable conversion and/or yield, fluidic lines were changed back to rinse solutions for the system to begin cleaning and depressurization. After a cleaning cycle, the robot arm automatically returned modules to storage and disconnected fluidic lines at the reagent tree.

REFERENCES AND NOTES

1. D. C. Blakemore *et al.*, Organic synthesis provides opportunities to transform drug discovery. *Nat. Chem.* **10**, 383–394 (2018). doi: 10.1038/s41557-018-0021-z; pmid: 29568051
2. G. Schneider, Automating drug discovery. *Nat. Rev. Drug Discov.* **17**, 97–113 (2018). doi: 10.1038/nrd.2017.232; pmid: 29242609
3. D. P. Tabor *et al.*, Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018). doi: 10.1038/s41578-018-0005-z
4. M. Peplow, Organic synthesis: The robo-chemist. *Nature* **512**, 20–22 (2014). doi: 10.1038/512020a; pmid: 25100466
5. S. V. Ley, D. E. Fitzpatrick, R. J. Ingham, R. M. Myers, Organic synthesis: March of the machines. *Angew. Chem. Int. Ed. Engl.* **54**, 3449–3464 (2015). doi: 10.1002/anie.201410744; pmid: 25586940
6. A. G. Godfrey, T. Masquelin, H. Hemmerle, A remote-controlled adaptive medchem lab: An innovative approach to enable drug discovery in the 21st Century. *Drug Discov. Today* **18**, 795–802 (2013). doi: 10.1016/j.drudis.2013.03.001; pmid: 23523957
7. A.-C. Bédard *et al.*, Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018). doi: 10.1126/science.aat0650; pmid: 30237351
8. S. Steiner *et al.*, Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, (2019). doi: 10.1126/science.aav2211
9. J. Li *et al.*, Synthesis of many different types of organic small molecules using one automated process. *Science* **347**, 1221–1226 (2015). doi: 10.1126/science.aaa5414; pmid: 25766227
10. A. Buitrago Santanilla *et al.*, Organic chemistry. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49–53 (2015). doi: 10.1126/science.1259203; pmid: 25554781
11. D. Perera *et al.*, A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **359**, 429–434 (2018). doi: 10.1126/science.aap9112; pmid: 29371464
12. Food and Drug Administration, Strategic plan for preventing and mitigating drug shortages (2013); https://www.fda.gov/downloads/Drugs/DrugSafety/DrugShortages/UCM372566.pdf.
13. R. L. Hartman, J. P. McMullen, K. F. Jensen, Deciding whether to go with the flow: Evaluating the merits of flow reactors for synthesis. *Angew. Chem. Int. Ed. Engl.* **50**, 7502–7519 (2011). doi: 10.1002/anie.201004637; pmid: 21710673
14. M. B. Plutschack, B. Pieber, K. Gilmore, P. H. Seeberger, The Hitchhiker's Guide to Flow Chemistry. *Chem. Rev.* **117**, 11796–11893 (2017). doi: 10.1021/acs.chemrev.7b00183; pmid: 28570059
15. K. F. Jensen, Flow chemistry-Microreaction technology comes of age. *AIChE J.* **63**, 858–869 (2017). doi: 10.1002/aic.15642
16. B. Gutmann, D. Cantillo, C. O. Kappe, Continuous-flow technology—A tool for the safe manufacturing of active pharmaceutical ingredients. *Angew. Chem. Int. Ed. Engl.* **54**, 6688–6728 (2015). doi: 10.1002/anie.201409318; pmid: 25989203
17. C. Wiles, P. Watts, Continuous flow reactors: A perspective. *Green Chem.* **14**, 38–54 (2012). doi: 10.1039/C1GC16022B
18. D. R. Snead, T. F. Jamison, A three-minute synthesis and purification of ibuprofen: Pushing the limits of continuous-flow processing. *Angew. Chem. Int. Ed. Engl.* **54**, 983–987 (2015). doi: 10.1002/anie.201409093; pmid: 25470733
19. T. Klucznik *et al.*, Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem* **4**, 522–532 (2018). doi: 10.1016/j.chempr.2018.02.002
20. S. Szymkuć *et al.*, Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed. Engl.* **55**, 5904–5937 (2016). doi: 10.1002/anie.201506101; pmid: 27062365
21. M. H. S. Segler, M. Preuss, M. P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018). doi: 10.1038/nature25978; pmid: 29595767
22. A. Adamo *et al.*, On-demand continuous-flow production of pharmaceuticals in a compact, reconfigurable system. *Science* **352**, 61–67 (2016). doi: 10.1126/science.aaf1337; pmid: 27034366
23. B. J. Reizman, K. F. Jensen, Feedback in Flow for Accelerated Reaction Development. *Acc. Chem. Res.* **49**, 1786–1796 (2016). doi: 10.1021/acs.accounts.6b00261; pmid: 27525813
24. D. E. Fitzpatrick, C. Battilocchio, S. V. Ley, A Novel Internet-Based Reaction Monitoring, Control and Autonomous Self-Optimization Platform for Chemical Synthesis. *Org. Process Res. Dev.* **20**, 386–394 (2016). doi: 10.1021/acs.oprd.5b00313
25. E. J. Corey, W. L. Jorgensen, Computer-assisted synthetic analysis. Synthetic strategies based on appendages and the use of reconnective transforms. *J. Am. Chem. Soc.* **98**, 189–203 (1976). doi: 10.1021/ja00417a030
26. A. Cook *et al.*, Computer-aided synthesis design: 40 years on. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**, 79–107 (2012). doi: 10.1002/wcms.61
27. C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **3**, 434–443 (2017). doi: 10.1021/acscentsci.7b00064; pmid: 28573205
28. M. H. S. Segler, M. P. Waller, Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chemistry* **23**, 5966–5971 (2017). doi: 10.1002/chem.201605499; pmid: 28134452
29. B. Liu *et al.*, Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models. *ACS Cent. Sci.* **3**, 1103–1113 (2017). doi: 10.1021/acscentsci.7b00303; pmid: 29104927
30. C. W. Coley, W. H. Green, K. F. Jensen, Machine Learning in Computer-Aided Synthesis Planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018). doi: 10.1021/acs.accounts.8b00087; pmid: 29715002

Downloaded from http://science.sciencemag.org/ on February 5, 2021

31. D. Lowe, Chemical reactions from US patents (1976-Sep2016) (2017); https://figshare.com/articles/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873.
32. G. Landrum, RDKit: Open-source cheminformatics; http://www.rdkit.org.
33. C. W. Coley, W. H. Green, K. F. Jensen, RDChiral: An RDKit Wrapper for Handling Stereochemistry in Retrosynthetic Template Extraction and Application. *J. Chem. Inf. Model.* **59**, 2529–2537 (2019). doi: 10.1021/acs.jcim.9b00286; pmid: 31190540
34. D. Rogers, M. Hahn, Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50**, 742–754 (2010). doi: 10.1021/ci100050t; pmid: 20426451
35. G. M. J.-B. Chaslot, M. H. M. Winands, H. J. van den Herik, in *Computers and Games*, H. J. van den Herik, X. Xu, Z. Ma, M. H. M. Winands, Eds. (Springer, 2008), pp. 60–71.
36. D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, D. Hassabis, Mastering chess and shogi by self-play with a general reinforcement learning algorithm. arXiv:1712.01815 (2017).
37. C. W. Coley *et al*., A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **10**, 370–377 (2018). doi: 10.1039/C8SC04228D; pmid: 30746086
38. W. Jin, C. Coley, R. Barzilay, T. Jaakkola, Predicting organic reaction outcomes with Weisfeiler-Lehman network, in *Advances in Neural Information Processing Systems* (NeurIPS 2017), pp. 2604–2613.
39. H. Gao *et al*., Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Cent. Sci.* **4**, 1465–1476 (2018). doi: 10.1021/acscentsci.8b00357; pmid: 30555898
40. P. Zhang *et al*., Advanced Continuous Flow Platform for On-Demand Pharmaceutical Manufacturing. *Chemistry* **24**, 2776–2784 (2018). doi: 10.1002/chem.201706004; pmid: 29385292
41. A. Adamo, P. L. Heider, N. Weeranoppanant, K. F. Jensen, Membrane-Based, Liquid–Liquid Separator with Integrated Pressure Control. *Ind. Eng. Chem. Res.* **52**, 10802–10808 (2013). doi: 10.1021/ie401180t
42. Y.-J. Hwang *et al*., A segmented flow platform for on-demand medicinal chemistry and compound synthesis in oscillating droplets. *Chem. Commun. (Camb.)* **53**, 6649–6652 (2017). doi: 10.1039/C7CC03584E; pmid: 28585652
43. C. Coley, ASKCOS Version 0.2.9, Zenodo (2019); doi: 10.5281/zenodo.3261361
44. P. Schwaller, T. Gaudin, D. Lányi, C. Bekas, T. Laino, "Found in Translation": Predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chem. Sci.* **9**, 6091–6098 (2018). doi: 10.1039/C8SC02339E; pmid: 30090297

ACKNOWLEDGMENTS

The authors acknowledge M. Fortunato and our Computer Science colleagues R. Barzilay, T. Jaakkola, W. Jin, and Y. Qian for help with aspects of the code. We thank S. O'Meara for help with the robotic platform and P. Morse for preliminary work on developing the CRF for safinamide. We also thank Elsevier for use of the Reaxys API and F. Frankel for help photographing the robotic platform. **Funding**: This work was supported by the DARPA Make-It program under contract ARO W911NF-16-2-0023. **Author contributions**: C.W.C. designed, developed, and implemented the synthesis-planning software; D.A.T. designed, developed, and supervised construction of the robotic platform; J.A.M.L., J.N.J., C.P.B., V.S., and L.R. developed chemistry, performed experiments with the robotic platform, and interpreted results; T.H., J.S.F., J.B., and J.S.P. assisted in the development, maintenance, and assembly of the robotic platform; H.G. and P.P.P. assisted in the development of synthesis-planning code; R.W.H., W.H.G., and K.F.J. advised on the development of the synthesis-planning software; A.J.H. and K.F.J. advised on the development of the robotic platform; T.F.J. supervised chemistry development; C.W.C., D.A.T., and J.A.M.L. prepared the manuscript; T.F.J., K.F.J., and A.J.H. edited the manuscript; and K.F.J. supervised the project and secured funding. **Competing interests**: T.F.J. is a cofounder of Snapdragon Chemistry, Inc., and a scientific adviser for Zaiput Flow Technologies, Continuus Pharmaceuticals, Paraza Pharmaceuticals, and Asymchem; some of these companies develop continuous processes and technologies but there is no direct connection between their activities and the results described herein. MIT has filed a patent (WO 2018/200236) with D.A.T. and K.F.J. as inventors of robotic handling of fluid connections. **Data and materials availability**: All experimental data are available in the main text or the supplementary materials. All code and trained models used in the synthesis-planning program are available through GitHub and Git LFS at https://github.com/connorcoley/ASKCOS (*43*); the original Reaxys data upon which our models are trained are the intellectual property of Elsevier and can only be accessed through a direct request to the publisher.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/365/6454/eaax1566/suppl/DC1
Supplementary Text
Figs. S1 to S95
Tables S1 to S3
References (*45–49*)
Movies S1 to S3
Recipe Files, Batch Sheets, and Website Printouts

1 March 2019; accepted 2 July 2019
10.1126/science.aax1566

Downloaded from http://science.sciencemag.org/ on February 5, 2021

Science

A robotic platform for flow synthesis of organic compounds informed by AI planning

Connor W. Coley, Dale A. Thomas III, Justin A. M. Lummiss, Jonathan N. Jaworski, Christopher P. Breen, Victor Schultz, Travis Hart, Joshua S. Fishman, Luke Rogers, Hanyu Gao, Robert W. Hicklin, Pieter P. Plehiers, Joshua Byington, John S. Piotti, William H. Green, A. John Hart, Timothy F. Jamison and Klavs F. Jensen

Science **365** (6453), eaax1566.
DOI: 10.1126/science.aax1566

Pairing prediction and robotic synthesis

Progress in automated synthesis of organic compounds has been proceeding along parallel tracks. One goal is algorithmic prediction of viable routes to a desired compound; the other is implementation of a known reaction sequence on a platform that needs little to no human intervention. Coley *et al.* now report preliminary integration of these two protocols. They paired a retrosynthesis prediction algorithm with a robotically reconfigurable flow apparatus. Human intervention was still required to supplement the predictor with practical considerations such as solvent choice and precise stoichiometry, although predictions should improve as accessible data accumulate for training.

Science, this issue p. eaax1566

ARTICLE TOOLS http://science.sciencemag.org/content/365/6453/eaax1566

SUPPLEMENTARY MATERIALS http://science.sciencemag.org/content/suppl/2019/08/07/365.6453.eaax1566.DC1

REFERENCES This article cites 39 articles, 6 of which you can access for free
http://science.sciencemag.org/content/365/6453/eaax1566#BIBL

PERMISSIONS http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Downloaded from http://science.sciencemag.org/ on February 5, 2021