

AI-driven robotic chemist for autonomous synthesis of organic molecules

Taesin Ha^{1†}, Dongseon Lee^{1†}, Youngchun Kwon¹, Min Sik Park¹, Sangyoonee Lee¹, Jaejun Jang¹, Byungkwon Choi¹, Hyunjeong Jeon¹, Jeonghun Kim¹, Hyundo Choi¹, Hyung-Tae Seo^{1,2}, Wonje Choi¹, Wooram Hong¹, Young Jin Park^{1,3}, Junwon Jang¹, Joonkee Cho¹, Bosung Kim¹, Hyukju Kwon¹, Gahee Kim¹, Won Seok Oh¹, Jin Woo Kim¹, Joonhyuk Choi¹, Minsik Min¹, Aram Jeon¹, Yongsik Jung¹, Eunji Kim^{1,4}, Hyosug Lee^{1,5}, Youn-Suk Choi^{1*}

The automation of organic compound synthesis is pivotal for expediting the development of such compounds. In addition, enhancing development efficiency can be achieved by incorporating autonomous functions alongside automation. To achieve this, we developed an autonomous synthesis robot that harnesses the power of artificial intelligence (AI) and robotic technology to establish optimal synthetic recipes. Given a target molecule, our AI initially plans synthetic pathways and defines reaction conditions. It then iteratively refines these plans using feedback from the experimental robot, gradually optimizing the recipe. The system performance was validated by successfully determining synthetic recipes for three organic compounds, yielding that conversion rates that outperform existing references. Notably, this autonomous system is designed around batch reactors, making it accessible and valuable to chemists in standard laboratory settings, thereby streamlining research endeavors.

INTRODUCTION

The discovery of functional organic materials has led to the emergence of various organic counterparts of electronic devices, such as light-emitting diodes, complementary metal-oxide semiconductor image sensors, and solar cells, with the ongoing challenge of improving their properties. Traditionally, this endeavor has relied on a time-consuming and inefficient trial-and-error approach involving repetitive cycles of molecular design, synthesis, and characterization processes. Recognizing the need for innovation in this methodology, notable efforts spanning decades have aimed to revamp the approach. However, it is time-consuming and inefficient; thus, efforts have been dedicated for decades to innovate this methodology. In the realm of molecular design, the advent of high-throughput computational screening, supported by large-scale first-principles simulations and machine learning, marked a transformative shift aimed at reducing reliance on human knowledge and intuition and minimizing the likelihood of unexpected discoveries (1–4). The drive to streamline laborious experiments gained momentum with the onset of the electronics era, ushering in precise and accessible control over unit operations, such as dispensing, reactions, sample preparation (sample-prep.), work-up, purification, and analysis (5–7). Ultimately, the aspiration for comprehensive laboratory automation initially found its roots in

the life sciences field during the 1980s (8), and substantial progress has been made over the past few decades (9–13). This trend toward automation has also manifested itself in the field of chemistry.

The advancement of artificial intelligence (AI) technologies in the 2010s, coupled with the availability of large-scale datasets, gave rise to the concept of robot chemists, where AI serves as the cognitive brain and the robot acts as the physical body, enabling autonomous chemical research. Challenges have persisted in the development of organic molecules using universal synthetic platforms, particularly in fields such as pharmaceuticals and biology (14, 15). Notably, there has been a recent surge in the adoption of flow-based systems (16–21) due to their cost effectiveness and the ease with which processes can be controlled through configurable fluidic circuits with valves and pumps (8, 22). These systems offer enhanced heat and mass transfer, allow for harsh reaction conditions in terms of temperature and pressure, and facilitate online analytical monitoring. However, flow chemistry faces limitations in handling poorly soluble reagents, lacks dedicated databases for automated synthesis planning, and typically lacks translatability between flow and batch chemistries (8). Innovative hybrid systems have been proposed (23–26) combining round-bottomed flasks for batch reactions and flow systems for chemical transport. Nevertheless, they encounter challenges in handling solid reagents containing metallic elements used in electronics applications. Hence, batch-type synthesis remains practical for chemists, despite its larger footprint and higher cost, due to its status as a standard protocol in mass production and development. While there have been some instances of bio-applications (10, 27), constructing a batch-type automated system by integrating various hardware and software components is complex, resulting in only a limited number of studies with restricted capabilities (12, 28–31).

In pursuit of a versatile and intelligent platform for molecule synthesis, this study introduces an AI-driven robotic chemist, capable of autonomously performing tasks spanning from synthetic

Copyright © 2023
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
License 4.0 (CC BY).

¹Samsung Advanced Institute of Technology, Samsung Electronics Co. Ltd., 130 Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16678, Republic of Korea.

²Department of Mechanical Engineering, Kyonggi University, 154-42, Gwanggyo-san-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, 16227, Republic of Korea. ³School of Mechanical Engineering, Gyeongsang National University, 501, Jinju-daero, Jinju-si, Gyeongsangnam-do, Republic of Korea. ⁴School of Business Administration, Chung-Ang University, 135, Seodal-ro, Dongjak-gu, Seoul 06973, Republic of Korea. ⁵College of Information and Communication Engineering, Sungkyunkwan University (SKKU), 2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do 16419, Republic of Korea.

*Corresponding author. Email: ysuk.choi@samsung.com

†These authors contributed equally to this work.

planning to experiments conducted in batch reactors, capitalizing on the collaborative potential of AI and robots. This platform is aptly named the "Synbot" (synthesis robot). The Synbot comprises three distinct layers: an AI software (S/W) layer, a robot S/W layer, and a robot layer (Fig. 1A). Its primary objective is to synthesize target substances while actively seeking optimal conditions. The AI S/W layer spearheads the synthesis planning process, equipped with the retrosynthesis module, the design of experiments (DoE), and optimization module, and steers the direction of experiments using the decision-making module. This layer adopts a blackboard architecture, enabling individual modules to access a shared database, facilitating communication and collaborative problem solving. Once the synthesis recipe is relayed from the AI S/W layer, the robot S/W layer takes charge, translating it into actionable commands for the robots through the recipe generation module and the translation module. Subsequently, the robot layer operates under the supervision of the online scheduling module (Fig. 1B).

The robot layer modularizes the various functions of the synthetic laboratory and systematically executes the planned recipes, continuously updating the database until the predefined goals are met. The Synbot encompasses essential modules, including pantry, dispensing, reaction, sample preparation, analysis, and transfer-robot modules, with an overall footprint measuring 9.35 m by 6.65 m. This comprehensive integration of AI and robotics represents a significant step toward achieving a versatile and autonomous smart synthesis platform for molecules.

Autonomous workflow of the Synbot

The procedure for the autonomous synthesis by the Synbot is illustrated in Fig. 2. According to the target molecule and task given by a user, the AI S/W layer commences synthesis planning (i) and completes the recipe repository with the initial reaction paths and conditions (ii). When the robot S/W layer determines that one of the reactors is available, it requests a new synthesis recipe to the AI S/W layer (iii). The AI S/W layer generates a new synthesis recipe and sends it to the robot S/W layer (iv). The robot S/W layer translates the recipe into commands and sends them to the robot layer (v). The robot layer performs the synthesis according to the commands (vi). Finally, the robot layer sends experimental results back to the AI S/W layer (vii), which updates the database and sends a new synthesis recipe to the robot S/W layer if necessary (viii).

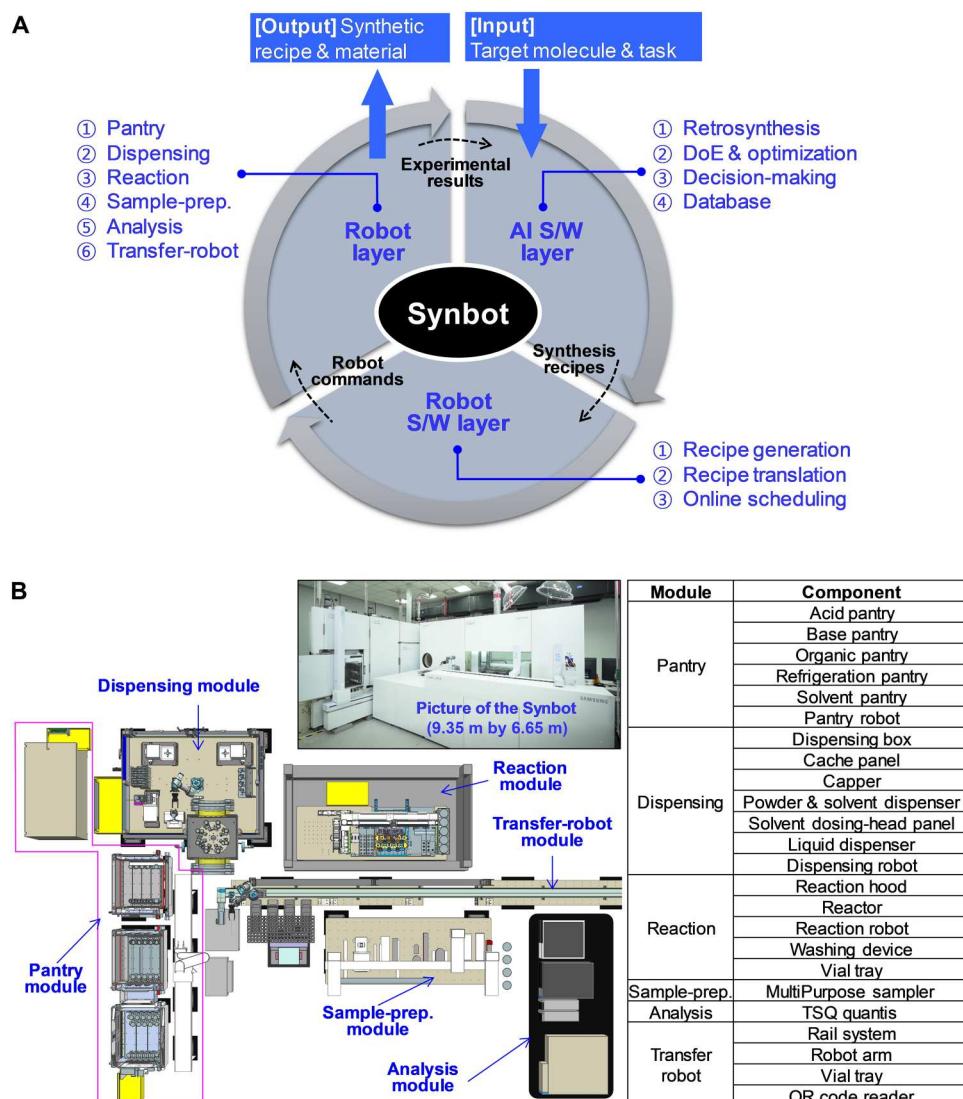


Fig. 1. AI-driven robotic chemist (Synbot). (A) Structure and working concept of the Synbot comprising AI S/W, robot S/W, and robot layers. (B) Layout and configuration of the robot layer comprising six modules: pantry, dispensing, reaction, sample-prep., analysis, and transfer-robot.

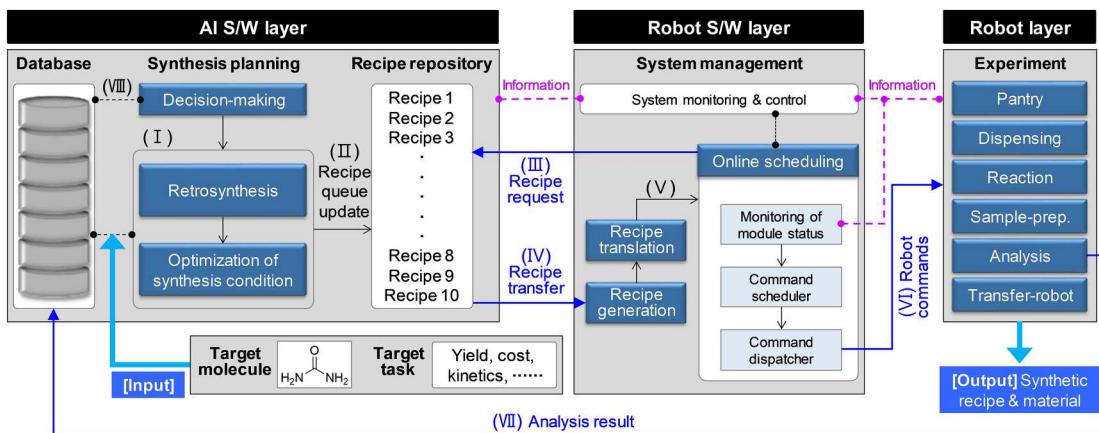


Fig. 2. Workflow of autonomous synthesis for a target molecule and task.

W layer and receives the highest-ranked recipe in the recipe repository (iii and iv). After translating the recipe into detailed robot commands (v), the online scheduler dispatches them to the robot layer (vi) when the relevant robots are prepared for execution. When analyses during the reaction are completed in the robot layer, the results are delivered to the database of the AI S/W layer (vii). The decision-making module determines whether to continue with the current recipe, to try another recipe, or to switch to a new synthetic path. The current recipe continues if the decision-making module determines that the reaction requires more time. If the decision-making module evaluates that the current recipe is not suitable to meet the target, it issues a "Withdraw" signal to the robot S/W layer to halt the current reaction condition and commence a new one. Furthermore, a "Sweep" signal is addressed to the robot S/W to stop all recipes belonging to the current synthetic path when the decision-making module concludes that another synthetic route should be attempted. The DoE and optimization module update its AI model, if the current recipe ends normally, and revise the recipe repository. Thereafter, the entire procedure repeats until the synthetic objective is satisfied.

The target task of the Synbot is currently focused on the maximization of the reaction yield. However, it can be extended to other objectives, such as the minimization of synthetic cost or the optimization of reaction kinetics, if necessary. Furthermore, in addition to the above autonomous mode, it can be operated in a semi-autonomous mode that determines optimal conditions using only the Bayesian optimization (BO) algorithm for areas not covered by deep learning models and in an automation mode that only passively performs user-specified experiments.

AI S/W layer

Competent synthetic planning can save time and cost when obtaining a product by determining suitable combinations of starting materials and reaction conditions. The design of synthetic pathways and determination of suitable reaction conditions for a target molecule are traditionally conducted on the basis of chemists' knowledge and experience. However, advancements in high-performance computing and AI have facilitated computer-assisted synthetic planning. While precision and validity may not yet meet the expectations of researchers, particularly for newly discovered materials, computer-assisted planning reveals implicit information

from a vast body of previous studies and rapidly suggests feasible conditions. Consequently, a computer-assisted approach proves indispensable for an autonomous synthetic platform. For the Synbot, a collaborative retrosynthesis approach is formulated by combining the template-based model (32) and the template-free tied-two-way transformer (33) to increase the viability of the proposed synthetic routes, which increases the top 1 prediction accuracy by 4.5 to 7.0%. When the synthesis path is determined by the retrosynthesis module, suitable reaction conditions are suggested by the DoE and optimization module (34) in the predefined search space (see Supplementary Text). If the target synthesis is within the material database in the AI S/W layer, then message-passing neural networks (MPNNs) (35) can steer the optimization process readily based on previous knowledge. However, if the task is rare, then fresh or peculiar access is crucial for reaching a solution. To address both these cases, a hybrid-type dynamic optimization (HDO) model, which associates MPNNs in conjunction with BO (34), is implemented to coordinate exploitation and exploration harmoniously. Various deep neural network models of the Synbot were built on the basis of the commercial Reaxys DB (Elsevier, Aalborg, Denmark). Details of the AI S/W layer are provided in Supplementary Text.

Robot S/W layer

The synthetic recipes predicted by the AI S/W layers are abstract and cannot drive the robot; thus, they are transformed into more definite robot commands in two steps by the recipe generation and translation modules in the robot S/W layer. First, the recipe generation module produces quantified action sequences that reflect the molecular weight, purity, and concentration of the chemicals. Subsequently, the recipe translation module converts the action sequences into robot commands using concrete parameters for hardware control. The action sequences are independent of H/W configurations and are human-readable; however, the robot commands are specific to the Synbot. The online scheduling module monitors the robots' work status in real-time and executes the commands in order (see Supplementary Text).

Robot layer

The robot layer executes the commands received. The chemical containers of reactants and reagents, which are stored in five types of

pantries (acid, base, organic, refrigeration, and solvent), are transferred to the dispensing module by the pantry robot, and subsequently, the chemicals are dispensed into glass reaction vials, as specified in the recipes. The vials were delivered to the reaction module and subjected to specific temperatures and stirring conditions for the chemical reaction. The reaction status is monitored via repetitive sampling of a small amount of the reaction solution (20 to 25 μm). The sampled solutions are then moved to the sample-prep. module and injected into a liquid chromatography–mass spectrometer (LC-MS; TSQ Quantis; Thermo Fisher Scientific, Waltham, MA). The sample-prep. module is responsible for preprocessing the sampled reaction solutions, such as dilution, mixing, and filtration of solid particles, and the final injection into the LC-MS. Each module, with the exception of the analysis module, has its own robot to handle operations, and the transfer-robot module relays the entire process by transporting the reaction and sample vials between the different modules.

We engineered the system to be robust against variations in the surrounding environment, ensuring stable operation and reliable experimental outcomes. The Synbot laboratory was under the control of a thermo-hygrostat, maintaining a temperature of $\leq 24^\circ\text{C}$ and a relative humidity of $\leq 45\%$, thus ensuring a consistent reaction environment. In addition, the interior of the pantry and dispensing modules was continuously supplied with nitrogen gas to prolong the shelf life of the chemicals. The dispensing module was equipped with several devices for the accurate mixing of the reaction solutions. This module included a capper, dispenser for powders and solvents, dispenser for liquid chemicals, ionizer to remove static electricity, and other supporting devices. In automated systems focused on optimizing synthetic recipes, a significant portion of the experimental time is dedicated to the actual chemical reactions. Therefore, if the other devices remain idle during this phase, then it can lead to reduced overall system utilization. To prevent this, the Synbot's reactor features six reaction slots, allowing simultaneous and independent control of multiple reactions. In addition, to avoid excessive pressure increase and solvent loss during the reaction process, a condensing mechanism and custom-built cap were applied to the reaction vial. Although LC-MS is primarily used to determine the conversion yield, it can also be used to determine the reaction kinetics. This versatility enables the Synbot to be applied to various tasks, including the mitigation of side reactions, elucidating reaction mechanisms, and developing previously unknown synthesis methods. To maintain a contamination-free operation, the Synbot extensively uses disposable glassware and devices. Further details can be found in Supplementary Text.

Reproducibility of the Synbot

Various factors, such as the accuracy of dispensing, consistency of the environment and chemicals, uniformity of the reaction temperature, and mixing, can influence the chemical reaction. If these factors are uncontrollable, then the reliability of the synthesis results may decrease, resulting in inaccurate outcomes. Furthermore, the generated data can negatively affect the chemical database that could otherwise have been used for machine learning. In this regard, the experimental reproducibility of the Synbot in terms of dispensing and conversion yield was examined for three typical aromatic coupling reactions (Suzuki coupling, Buchwald reaction, and Ullmann reaction; Fig. 3 and Supplementary Text).

Identical experiments were conducted 12 times to assess the reproducibility of each reaction scheme. As summarized in table S7, the chemical dispensing is carried out precisely with mean absolute errors $\leq 0.73\text{ mg}$ and coefficients of variance (CVs) $\leq 2.55\%$. In the case of the conversion yield, which reflects the consistency of all process variables including dispensing, reaction, preprocessing, and analysis, the CV values were less than 5% throughout the monitoring time. Moreover, if it is limited only to the latter part of the reaction stage, where the conversion yield converges, the CV values decrease to less than 2.5%. These results validate the performance of the Synbot and can serve as a basis for the Synbot to be used as a common synthesis platform.

Autonomous synthesis of the Synbot

The performance of the autonomous synthesis of the Synbot was investigated using three molecules [4-(2,3-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridine, M1; N-(4-methoxyphenyl)-N-phenylpyrimidin-5-amine, M2; and *N,N*-diphenylquinoxalin-2-amine, M3], which were selected from the literature (36–38) and reported to have isolation yields ranging from 30 to 50%. In advance, the information regarding the target molecules was excluded from the AI training datasets. The reaction conditions reported in the literature were reproduced on the Synbot to obtain the reference conversion yields. The results of the autonomous synthesis of the target products are summarized in Figs. 4 to 6 and tables S14 to S17.

RESULTS

Synthesis of M1

The reference Suzuki coupling reaction for M1 (M1-3 in Fig. 4A) (36) is predicted as the third priority by the retrosynthesis model, while the same synthetic routes as those found in the literature (M2-1 in Fig. 5A and M3-1 in Fig. 6A) (37, 38) are proposed as the first-ranked options for M2 and M3. The reference reaction condition for M1-3 revealed a conversion yield of 86.5% on the Synbot, which is higher than the reported isolation yield of 37.7%. This discrepancy could potentially be attributed to variations in the purification step. However, it is important to acknowledge that even with the same recipe, differences in the experimental apparatus, raw materials, and environmental conditions can lead to distinct outcomes due to variations in mechanical and chemical characteristics. Therefore, a target conversion yield of 91.5%, which is 5% higher than that of the reference, was set for M1 synthesis.

Autonomous synthesis initially follows the reaction scheme M1-1, as described in Fig. 4B and table S14. Although the target yield was as high as 91.5%, a synthetic condition with a conversion yield of 100% was found in the first trial within the search space of 2722 cases. The preference of scheme M1-1 over M1-3 is readily predictable because bromine substituents are generally more reactive than chlorine substituents. However, to confirm this, an experiment for the M1-3 reaction was also conducted, as shown in Fig. 4C and table S15, and a more superior condition than the reference was obtained in the ninth trial. During this process, the Synbot learned that the tetrahydrofuran/water mixture solvent is not favorable and expanded the candidate solvents to include a toluene/ethanol/water mixture and *N,N'*-dimethylformamide. Furthermore, it explored different ligands and catalysts beyond Pd(PPh₃)₄, ultimately achieving perfect conversion using the combination of Pd₂(dba)₃ and BrettPhos combination. After the ninth trial, the search was

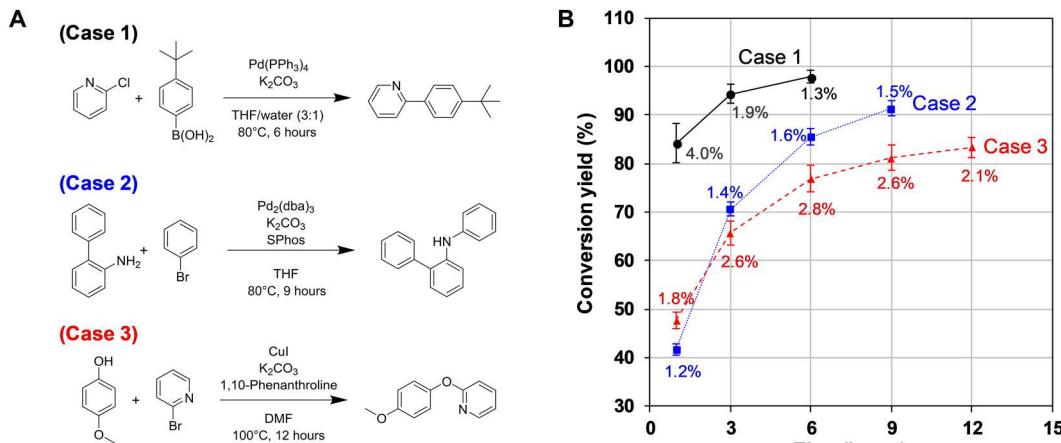


Fig. 3. Experiments to validate reproducibility of the Synbot. (A) Three reaction schemes. (B) Conversion yield variations with time. Each number in the data point indicates the SD obtained from 12 repetitions. THF, tetrahydrofuran; DMF, *N,N'*-dimethylformamide.

continued arbitrarily to further investigate the impact of different reagents, revealing that palladium catalyst sources with dibenzylideneacetone, BrettPhos, nonstrong bases, and toluene/ethanol/water were the optimal conditions for the reaction.

To exemplify the power of the AI model, let us delve into the Suzuki coupling reaction case, denoted as M1-3. In this case, conventional catalyst and base combinations, specifically $\text{Pd}(\text{PPh}_3)_4$ and K_2CO_3 , yielded relatively lower conversion rates within our mild temperature setup. However, under the same temperature conditions, we discovered that the less commonly used reagent combination comprising $\text{Pd}_2(\text{dba})_3$, BrettPhos, and KOAc achieved complete reaction conversion. Notably, historical data in Reaxys DB reveal that the base and catalyst ligand we used in this case are used at only about 1% of the frequency compared to $\text{Pd}(\text{PPh}_3)_4$ or K_2CO_3 .

Synthesis of M2

A common problem encountered in applying AI to molecular synthesis is the scarcity of training data, a limitation driven by the vastness of the chemical space and the high cost associated with experimental data collection. In such scenarios, it becomes crucial to effectively balance both exploitation and exploration strategies. The synthetic task for M2 belongs to this category. A total of 158,609 (19.5%) of the 814,687 data used for the training of the prediction model of reaction conditions are Suzuki coupling-related data, while only 17,705 data (2.2%) belong to Buchwald amination (see Supplementary Text). Consequently, it is anticipated that discovering suitable conditions for Buchwald amination would pose a more significant challenge compared to Suzuki coupling when relying on the HDO model for exploitation.

The conversion yield of the reference M2 recipe (37) was only 15.0% for the Synbot (Fig. 5B and table S16). However, the yield was quantified using LS-MS (as described in eq. S1 in Supplementary Text), which can vary depending on the material's absorbance properties. Therefore, the target conversion yield was set at 70.0%, approximately twice the reported isolation yield. For the M2-1 scheme, most recipes initially exhibited insufficient reactivity. Over time, new recipes were explored, primarily focusing on catalysts and solvents. Eventually, a combination of two types of

palladium dibenzylideneacetone (dba) catalysts, $\text{Pd}(\text{dba})_2$ and $\text{Pd}_2(\text{dba})_3$, APhos ligand, NaOtBu base, and toluene solvent was discovered, resulting in a 100.0% reaction conversion at the 36th and 37th tryouts. Autonomous synthesis continued to elucidate the reaction characteristics in greater detail, leading to more frequent proposals of high-yield reaction conditions. Through these endeavors, it became evident that bulky electron-rich dialkylbiaryl phosphine ligands are less suitable for the reaction compared to simpler monodentate or bidentate ligands such as PtBu₃, APhos, and XantPhos.

Synthesis of M3

The synthesis of M3 was classified as the N-arylation of Buchwald amination, as shown in Fig. 6A. However, the ligand specified in the reference literature, 2-[1,3-bis(dicyclohexylphosphanyl)-1H-inden-2-yl]-N,N-dimethylaniline, was not accessible. Consequently, XPhos was chosen as an alternative since it has been previously reported to induce rapid conversion and excellent yields, similar to the reference ligand in the same literature (38). The conversion yield of this modified reference condition on the Synbot was 50.9%; however, the target yield was 80.0%, considerably higher than the reported isolation yield of 45.0%.

Unlike the commonly used strong bases, such as NaOtBu, the Synbot identified high conversion conditions using the milder base of Cs_2CO_3 (Fig. 6B and table S17). The initial three groups suggested by the MPNN model failed to yield good results; however, the subsequent three recipes from the maximin Latin hypercube sampling (see Supplementary Text) exhibit the possibility of yielding good results. Although no clear improvement was observed until the 33rd run, the frequency of recipes with conversion yields higher than 50% gradually increased as the experiment progressed. Last, the Synbot obtained the target conversion yield in the 42nd trial using $\text{Pd}(\text{OAc})_2$ and XantPhos ligands. A closer observation indicates that a strong base, NaOH, can accelerate kinetics such as NaOtBu in the reference recipe, while Cs_2CO_3 results in a higher yield. Some differences were observed compared to the case of M2. First, the excellent recipes for M2 use the strong base NaOtBu, while that for M3 uses Cs_2CO_3 . In addition, in contrast to M2 synthesis, palladium acetate performs better in M3 synthesis than palladium catalysts prepared with dba. The specificity of these

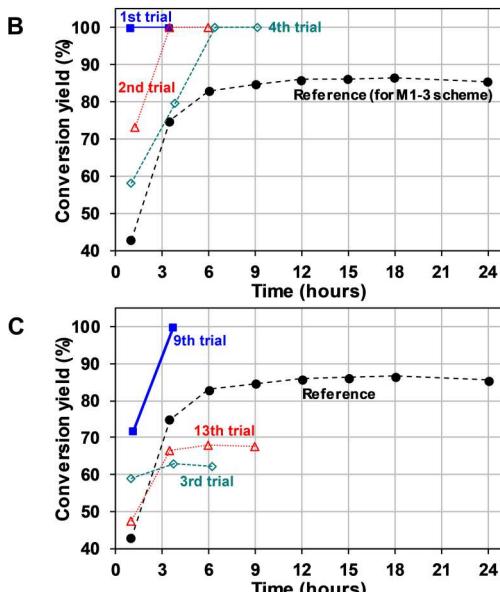
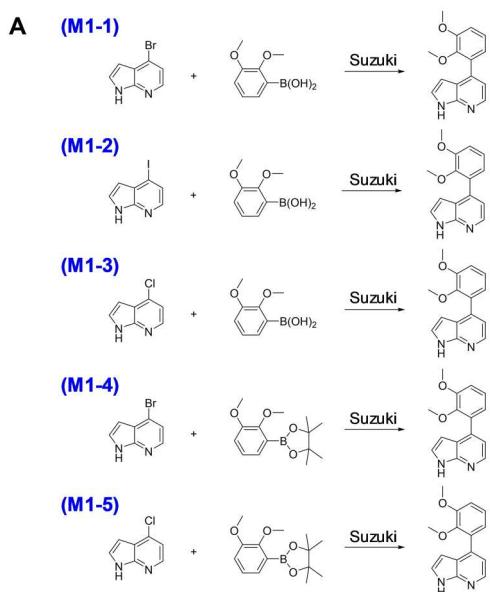


Fig. 4. Autonomous synthesis of M1 [4-(2,3-dimethoxyphenyl)-1H-pyrrolo[2,3-b]pyridine]. (A) Synthetic schemes designed by AI. (B) Conversion yield with time for the reaction scheme M1-1. (C) Conversion yield with time for the reaction scheme M1-3.

reagents may be attributed to the characteristics of the reactants, quinoxaline versus pyrimidine or pure diphenylamine versus methoxy diphenylamine, with slightly different electronic structures. Although further investigation into these nuances falls beyond the scope of this work, it underscores the importance of recipe search in enhancing reaction efficiency and understanding reaction mechanisms. In this context, the utility of the Synbot can be further amplified.

To compare reference results, we conducted additional syntheses of M3 using NaOtBu, as recommended in the reference paper (38), in conjunction with three different ligands: tri-tert-butyl phosphine (PtBu_3), Xphos, and Xantphos as summarized in table S18. Intriguingly, we observed that the reactions halted within just 4 hours, yielding approximately 65 to 70% conversion rates for the Xphos and Xantphos cases and a mere 5% conversion rates for the PtBu_3 cases. While our optimal recipe exhibited a slower reaction rate, it ultimately yielded higher conversion rate.

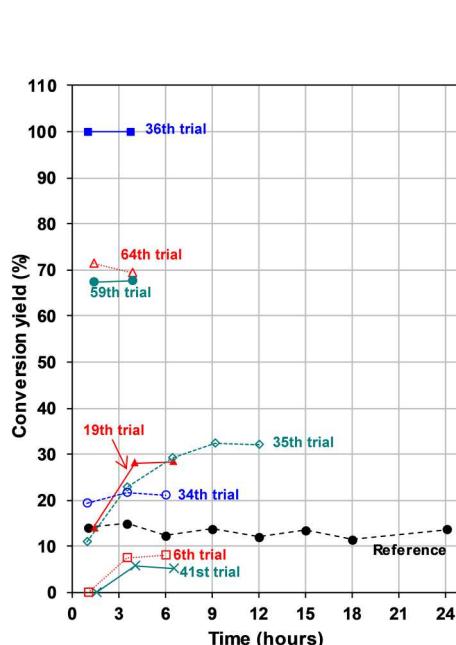
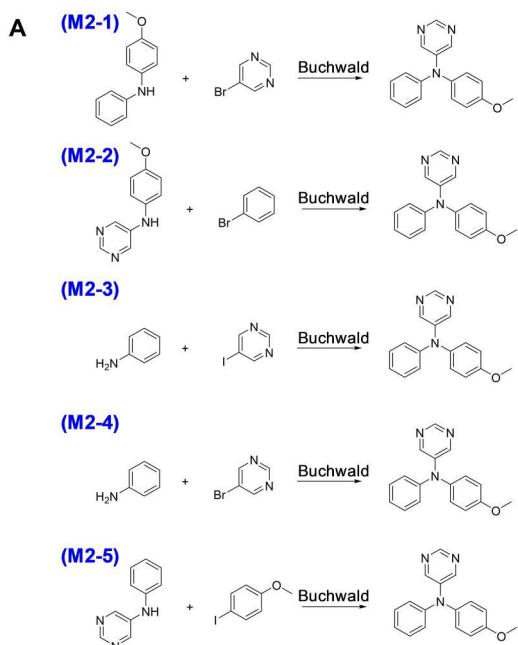


Fig. 5. Autonomous synthesis of M2 [N-(4-methoxyphenyl)-N-phenylpyrimidin-5-amine]. (A) Synthetic schemes designed by AI. (B) Conversion yield with time for the reaction scheme M2-1.

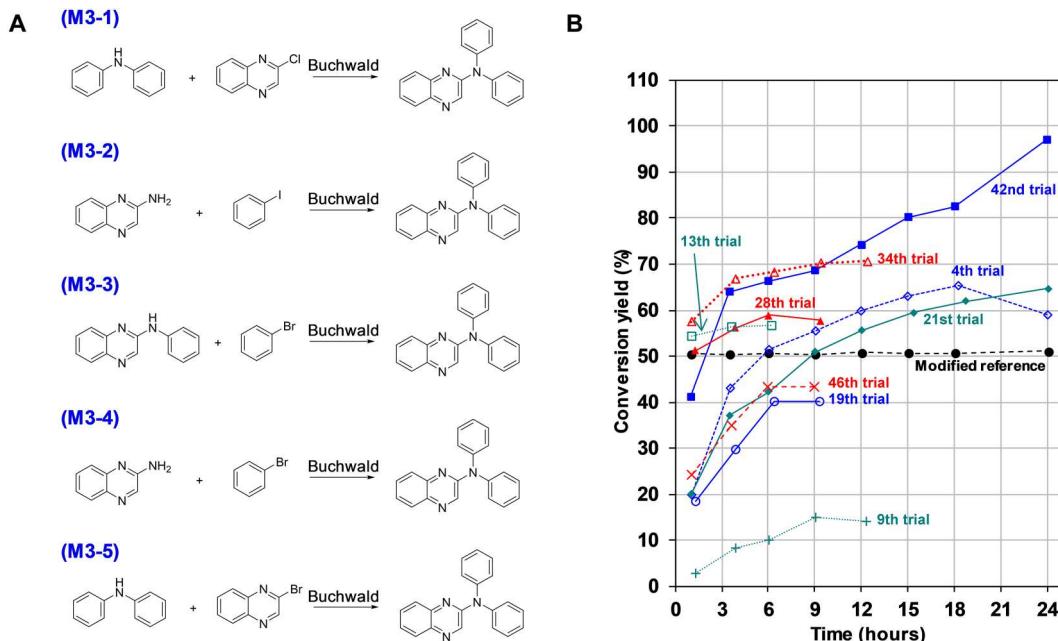


Fig. 6. Autonomous synthesis of M3 (*N,N*-diphenylquinoxalin-2-amine). (A) Synthetic schemes designed by AI. (B) Conversion yield with time for the reaction scheme M3-1.

DISCUSSION

The customized Synbot exhibited its exceptional capabilities by consistently delivering competitive synthetic recipes with yields on par with or surpassing known references. This achievement was made possible through a closed-loop feedback mechanism between the robotic system and AI. The MPNN model effectively determined solutions for well-established Suzuki coupling reactions (M1) in a relatively straightforward, data-driven manner. Conversely, for M2 and M3, the MPNNs faced challenges in individually identifying favorable conditions but succeeded in finding solutions through collaboration with BO. The goals were achieved in all cases, with fewer than 1% of trials from the total search space, highlighting the efficiency of HDO in chemical research compared to traditional methods reliant on human expertise and knowledge. The Synbot uses not only its high-throughput experimentation capabilities but also its real-time recipe design strategy guided by AI models. This stands as a testament to the Synbot's effectiveness in accelerating the discovery and optimization of chemical processes.

The Synbot's ability to monitor kinetics during synthesis has the potential to enhance synthesis quality while reducing research costs. In manual experiments, the periodic inspection of reaction progress can be labor intensive, leading to reactions often proceeding for excessive durations, resulting in yield losses due to side reactions or unnecessary time wastage. The automatic analysis capabilities of the Synbot naturally address this issue. Although LC-MS provides precise quantification, its relatively lengthy and complex preprocessing is a drawback. Therefore, integrating simpler yet somewhat qualitative techniques, such as thin-layer chromatography, could enhance overall efficiency.

Depending on the total reaction time, the Synbot can conduct an average of 12 reactions within 24 hours, encompassing dispensing and analysis. Assuming a researcher can perform two experiments of this type per day, the Synbot exhibits at least sixfold increase in

efficiency compared to human counterparts. This efficiency is further amplified when considering automatic synthetic planning and optimization. While the Synbot currently requires periodic human intervention to replenish chemicals, consumables like vials and filters, and dispose of waste, these challenges can be addressed by expanding pantry capacity, introducing automatic feeding robots, and implementing continuous waste-discharging mechanisms.

Efficiently assessing the properties and synthetic feasibility of materials in the early stages of development is crucial for screening potential candidates and identifying underlying issues. In this regard, the Synbot offers multiple contributions. Automated synthetic planning and decision-making guide robots to explore chemical spaces efficiently with minimal resources, enabling research even for individuals lacking extensive chemical knowledge. Accurate robot operation produces reliable experimental results, forming the basis for a high-quality DB that can be used in future studies. In addition, the Synbot provides access to numerous negative data, which are often challenging to find in typical research papers, and rich metadata for detailed causal analysis. Moreover, the batch-type reaction format aligns well with conventional synthesis practices, making it highly practical for chemists. The Synbot can accelerate the time to market for novel materials, granting researchers more time to focus on creative research activities beyond the realm of AI and robotics.

Existing chemistry DBs suffer from insufficient data and imbalance data distribution compared within the vast chemical space, leading to subpar AI performance compared to general machine learning applications like language translation or image recognition. This issue can be overcome by accelerating data accumulation through an automation platform such as the Synbot. However, ensuring compatibility of experimental results across different systems is paramount. Even batch-type reactors may exhibit variations in

heating, cooling, and mixing characteristics, potentially causing discrepancies in experimental outcomes. In this respect, global standardization of experimental devices becomes imperative. Now, the Synbot is undergoing upgrades to transform to a multistep synthesis platform, including work-up and purification steps, aiming to serve as a versatile, general-purpose platform.

MATERIALS AND METHODS

Preparation of reagents

All reagents and starting materials were purchased from Sigma-Aldrich (Burlington, MA, USA), Tokyo Chemical Industry (Tokyo, Japan), and Daejung Chemicals (Siheung, Republic of Korea) and were meticulously prepared before storage in our laboratory's pantries. Solid chemicals exceeding a size of 1 mm were initially subjected to grinding and sieving through a 500- μm -aperture metal sieve (TS-F0500; Glenammer, Ayrshire, UK). Subsequently, these materials were securely stored in designated chemical containers (QH010-CNMW; Mettler-Toledo, Greifensee, Switzerland), equipped with powder dispensers for convenient access. The various reaction solvents were transferred to 1-liter bottles, while nonsolvent liquid materials were carefully housed within in-house syringes. All these containers, both for chemicals and solvents, were systematically arranged in designated slots within our pantry.

Reaction condition for autonomous synthesis

The autonomous synthesis process was executed to derive optimal reaction recipes achieving the desired target yields. This process involved navigating a four-dimensional space defined by the catalyst, ligand, base, and solvent parameters. The reaction temperature was set as a constant, contingent on the specific solvent type [for detailed temperature settings, refer to Supplementary Text (35, 39–44)]. In all cases, the equivalent ratios of reactant 2, catalyst, and base to reactant 1 were determined from relevant literature (36–38). However, it is worth noting that due to the unavailability of equivalent ratios for the ligands in M1-1 and M1-3, ligands with twice the equivalent ratios of the catalysts were used. Concentrations for each reaction were established following established literature protocols. For comprehensive information regarding the Synbot system and the experimental procedures, please consult the Supplementary Materials.

Supplementary Materials

This PDF file includes:

Technical descriptions of the Synbot

Procedure and results of experiments

Figs. S1 to S20

Tables S1 to S18

Legend for movie S1

Legend for data S1

Other Supplementary Material for this manuscript includes the following:

Movie S1

Data S1

REFERENCES AND NOTES

- R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel, D. Duvenaud, D. Maclaurin, M. A. Blood-Forsythe, H. S. Chae, M. Einzinger, D.-G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W. Huang, S. I. Hong, M. Baldo, R. P. Adams, A. Aspuru-Guzik, Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **15**, 1120–1127 (2016).
- K. Kim, S. Kang, J. Yoo, Y. Kwon, Y. Nam, D. Lee, I. Kim, Y.-S. Choi, Y. Jung, S. Kim, W.-J. Son, J. Son, H. S. Lee, S. Kim, J. Shin, S. Hwang, Deep-learning-based inverse design model for intelligent discovery of organic molecules. *npj Comput. Mater.* **4**, 67 (2018).
- B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).
- F. Häse, L. M. Roch, P. Friederich, A. Aspuru-Guzik, Designing and understanding light-harvesting devices with machine learning. *Nat. Commun.* **11**, 4587 (2020).
- K. Olsen, The first 110 years of laboratory automation: Technologies, applications, and the creative scientist. *J. Lab. Autom.* **17**, 469–480 (2012).
- 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).
- G. R. D. Prabhu, P. L. Urban, The dawn of unmanned analytical laboratories. *TrAC Trends Anal. Chem.* **88**, 41–52 (2017).
- J. A. Selekman, J. Qiu, K. Tran, J. Stevens, V. Rosso, E. Simmons, Y. Xiao, J. Janey, High-throughput automation in chemical process development. *Annu. Rev. Chem. Biomol. Eng.* **8**, 525–547 (2017).
- R. D. King, J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova, A. Sparkes, K. E. Whelan, A. Clare, The automation of science. *Science* **324**, 85–89 (2009).
- K. Williams, E. Bilsland, A. Sparkes, W. Aubrey, M. Young, L. N. Soldatova, K. De Grave, J. Ramon, M. de Clare, W. Sirawaraporn, S. G. Oliver, R. D. King, Cheaper faster drug development validated by the repositioning of drugs against neglected tropical diseases. *J. R. Soc. Interface* **12**, 20141289 (2015).
- V. Rosso, J. Albrecht, F. Roberts, J. M. Janey, Uniting laboratory automation, DoE data, and modeling techniques to accelerate chemical process development. *React. Chem. Eng.* **4**, 1646–1657 (2019).
- B. Burger, P. M. Maffettone, V. V. Gusev, C. M. Aitchison, Y. Bai, X. Wang, X. Li, B. M. Alston, B. Li, R. Clowes, N. Rankin, B. Harris, R. S. Sprick, A. I. Cooper, A mobile robotic chemist. *Nature* **583**, 237–241 (2020).
- Y. Jiang, D. Salley, A. Sharma, G. Keenan, M. Mullin, L. Cronin, An artificial intelligence enabled chemical synthesis robot for exploration and optimization of nanomaterials. *Sci. Adv.* **8**, eab02626 (2020).
- C. W. Coley, N. S. Eyke, K. F. Jensen, Autonomous discovery in the chemical sciences Part I: Progress. *Angew. Chem. Int. Ed. Engl.* **59**, 22858–22893 (2020).
- C. W. Coley, N. S. Eyke, K. F. Jensen, Autonomous discovery in the chemical sciences Part II: Outlook. *Angew. Chem. Int. Ed. Engl.* **59**, 23414–23436 (2020).
- A. Adamo, R. L. Beingessner, M. Behnam, J. Chen, T. F. Jamison, K. F. Jensen, J.-C. M. Monbalui, A. S. Myerson, E. M. Revalor, D. R. Snead, T. Stelzer, N. Weeranoppanant, S. Y. Wong, P. Zhang, On-demand continuous-flow production of pharmaceuticals in a compact, reconfigurable system. *Science* **352**, 61–67 (2016).
- A.-C. Bédard, A. Adamo, K. C. Aroh, M. G. Russell, A. A. Bedermann, J. Torosian, B. Yue, K. F. Jensen, T. F. Jamison, Reconfigurable system for automated optimization of diverse chemical reactions. *Science* **361**, 1220–1225 (2018).
- N. Cherkasov, Y. Bai, A. J. Expósito, E. V. Rebrov, OpenFlowChem – a platform for quick, robust and flexible automation and self-optimisation of flow chemistry. *React. Chem. Eng.* **3**, 769–780 (2018).
- A. M. Schweidtmann, A. D. Clayton, N. Holmes, E. Bradford, R. A. Bourne, A. A. Lapkin, Machine learning meets continuous flow chemistry: Automated optimization towards the Pareto front of multiple objectives. *Chem. Eng. J.* **352**, 277–282 (2018).
- C. W. Coley, D. A. Thomas, J. A. M. Lummiss, J. N. Jaworski, C. P. Breen, V. Schultz, T. Hart, J. S. Fishman, L. Rogers, H. Gao, R. W. Hicklin, P. P. Plehiers, J. Byington, J. S. Piotti, W. H. Green, A. J. Hart, T. F. Jamison, K. F. Jensen, A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).
- A. M. K. Nambiar, C. P. Breen, T. Hart, T. Kulesza, T. F. Jamison, K. F. Jensen, Bayesian optimization of computer-proposed multistep synthetic routes on an automated robotic flow platform. *ACS Cent. Sci.* **8**, 825–836 (2022).
- M. Shevlin, Practical high-throughput experimentation for chemists. *ACS Med. Chem. Lett.* **8**, 601–607 (2017).
- V. Dragone, V. Sans, A. B. Henson, J. M. Granda, L. Cronin, An autonomous organic reaction search engine for chemical reactivity. *Nat. Commun.* **8**, 15733 (2017).

24. P. J. Kitson, G. Marie, J.-P. Francoia, S. S. Zalesskiy, R. C. Sigerson, J. S. Mathieson, L. Cronin, Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals. *Science* **359**, 314–319 (2018).
25. S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin, Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
26. S. Rohrbach, M. Šiaūulis, G. Chisholm, P.-A. Pirvan, M. Saleeb, S. H. M. Mehr, E. Trushina, A. I. Leonov, G. Keenan, A. Khan, A. Hammer, L. Cronin, Digitization and validation of a chemical synthesis literature database in the ChemPU. *Science* **377**, 172–180 (2022).
27. R. D. King, K. E. Whelan, F. M. Jones, P. G. K. Reiser, C. H. Bryant, S. H. Muggleton, D. B. Kell, S. G. Oliver, Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252 (2004).
28. A. Weber, E. von Roedern, H. U. Stilz, SynCar: An approach to automated synthesis. *J. Comb. Chem.* **7**, 178–184 (2005).
29. 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).
30. C. A. Nicolaou, C. Humblet, H. Hu, E. M. Martin, F. C. Dorsey, T. M. Castle, K. I. Burton, H. Hu, J. Hendle, M. J. Hickey, J. Duerksen, J. Wang, J. A. Erickson, Idea2Data: Toward a new paradigm for drug discovery. *ACS Med. Chem. Lett.* **10**, 278–286 (2019).
31. N. H. Angello, V. Rathore, W. Beker, A. Wolos, E. R. Jira, R. Roszak, T. C. Wu, C. M. Schroeder, A. Aspuru-Guzik, B. A. Grzybowski, M. D. Burke, Closed-loop optimization of general reaction conditions for heteroaryl Suzuki-Miyaura coupling. *Science* **378**, 399–405 (2022).
32. M. H. S. Segler, M. P. Waller, Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chemistry* **23**, 5966–5971 (2017).
33. E. Kim, D. Lee, Y. Kwon, M. S. Park, Y.-S. Choi, Valid, plausible, and diverse retrosynthesis using tied two-way transformers with latent variables. *J. Chem. Inf. Model.* **61**, 123–133 (2021).
34. Y. Kwon, D. Lee, J. W. Kim, Y.-S. Choi, S. Kim, Exploring optimal reaction conditions guided by graph neural networks and bayesian optimization. *ACS Omega* **7**, 44939–44950 (2022).
35. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl, Neural Message Passing for Quantum Chemistry. arXiv:1704.01212 [cs.LG] (2017).
36. L. Pieterse, L. J. Legoabe, R. M. Beteck, B. Josselin, S. Bach, S. Ruchaud, Synthesis and biological evaluation of selected 7-azaindole derivatives as CDK9/Cyclin T and Haspin inhibitors. *Med. Chem. Res.* **29**, 1449–1462 (2020).
37. A. C. Hernandez-Perez, A. Caron, S. K. Collins, Photochemical synthesis of complex carbazoles: Evaluation of electronic effects in both UV- and visible-light methods in continuous flow. *Chem. A Eur. J.* **21**, 16673–16678 (2015).
38. M.-Q. Yan, J. Yuan, Y.-X. Pi, J.-H. Liang, Y. Liu, Q.-G. Wu, X. Luo, S.-H. Liu, J. Chen, X.-L. Zhu, G.-A. Yu, Pd-indenyl-diphosphine: An effective catalyst for the preparation of triarylamines. *Org. Biomol. Chem.* **14**, 451–454 (2016).
39. 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).
40. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG] (2014).
41. M. H. S. Segler, M. Preuss, M. P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
42. G. Klein, Y. Kim, Y. Deng, J. Senellart, A. Rush, "OpenNMT: Open-Source Toolkit for Neural Machine Translation" in *Proceedings of ACL 2017, System Demonstrations* (Vancouver, Canada, 2017), pp. 67–72.
43. B. J. Shields, J. Stevens, J. Li, M. Parasram, F. Damani, J. I. M. Alvarado, J. M. Janey, R. P. Adams, A. G. Doyle, Bayesian reaction optimization as a tool for chemical synthesis. *Nature* **590**, 89–96 (2021).
44. A. M. J. Olsson, G. E. Sandberg, Latin hypercube sampling for stochastic finite element analysis. *J. Eng. Mech.* **128**, 121–125 (2002).

Acknowledgments: We thank I. Han and G. Jin for help with the Synbot (synthesis robot) development and H. Park for detailed comments on the manuscript. **Funding:** This work was supported by Samsung Electronics Co. Ltd. **Author contributions:** Conceptualization: Y.-S.C. and H.L. AI S/W layer: D.L., Y.K., M.S.P., J.W.K., J. Choi, and E.K. Robot S/W layer: T.H., S.L., Jaejun Jang, and B.C. Robot layer: J.K., H.C., H.-T.S., W.C., W.H., Y.J.P., Junwon Jang, J. Cho, and B.K. Chemical Analysis: H.K., G.K., and W.S.O. Experimental Design: H.J., H.K., M.M., A.J., and Y.J. Supervision: Y.-S.C. and H.L. Writing—original draft: T.H. and D.L. Writing—review and editing: T. H., D.L., Y.K., and Y.-S.C. **Competing interests:** All authors are inventors of a patent application related to this work filed by Samsung Electronics Co. Ltd. (US20210125060A1, 29 May 2020; US20210180884A1, 29 July 2020; US20210174910A1, 8 December 2020; US20220331772A1, 5 November 2021; US20220036182A1, 12 July 2021; US20220319642A1, 21 October 2021; US20210170597A1, 17 September 2020). The authors declare that they have no other competing interests. **Data and materials availability:** The datasets of the Reaxys DB used for the training of our AI models are not allowed to be disclosed according to Elsevier's policy. However, the Reaxys DB can be provided under individual licenses since it is commercially available. Requests for the license should be submitted to <https://beta.elsevier.com/products/reaxys/contact-us>. All other data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 3 June 2023

Accepted 27 September 2023

Published 1 November 2023

10.1126/sciadv.adj0461