nature communications

Article

https://doi.org/10.1038/s41467-024-53204-6

Roboticized AI-assisted microfluidic photocatalytic synthesis and screening up to 10,000 reactions per day

Received: 12 March 2024

Accepted: 4 October 2024

Published online: 12 October 2024

Check for updates

Jia-Min Lu[1,2,7], Hui-Feng Wang[1,2,7], Qi-Hang Guo[1,2,3,7], Jian-Wei Wang[2], Tong-Tong Li[1,3], Ke-Xin Chen[4,5], Meng-Ting Zhang[1], Jian-Bo Chen[1], Qian-Nuan Shi[2], Yi Huang[2], Shao-Wen Shi[2], Guang-Yong Chen[4] ✉, Jian-Zhang Pan[1,2] ✉, Zhan Lu[1,3] ✉ & Qun Fang[1,2,6] ✉

The current throughput of conventional organic chemical synthesis is usually a few experiments for each operator per day. We develop a robotic system for ultra-high-throughput chemical synthesis, online characterization, and large-scale condition screening of photocatalytic reactions, based on the liquid-core waveguide, microfluidic liquid-handling, and artificial intelligence techniques. The system is capable of performing automated reactant mixture preparation, changing, introduction, ultra-fast photocatalytic reactions in seconds, online spectroscopic detection of the reaction product, and screening of different reaction conditions. We apply the system in large-scale screening of 12,000 reaction conditions of a photocatalytic [2 + 2] cycloaddition reaction including multiple continuous and discrete variables, reaching an ultra-high throughput up to 10,000 reaction conditions per day. Based on the data, AI-assisted cross-substrate/photocatalyst prediction is conducted.

How many chemical synthesis reactions can an organic synthesizer complete in a day? In most organic synthesis laboratories around the world, the numbers are usually no more than 10 reactions per day. Chemical synthesis, especially for many organic synthesis reactions, are often time-consuming and take hours or even days to complete. Therefore, the optimization and screening of organic synthesis often requires a significant investment of time and effort by researchers[1]. Recently, various automated and high-throughput organic synthesis and screening techniques have been developed based on automated robotics and microfluidic chemistry techniques. Flow microreactors have advantages of fast reaction speed and high reaction efficiency due to the scale effect of high mass and heat transfer efficiency in microfluidic channels[2–4]. Photocatalytic reactions are a type of reactions frequently-performed in organic synthesis that use light irradiation to excite reactant molecules or catalysts to promote the reaction, converting light energy into chemical energy[5,6]. Compared with the conventional batch photocatalytic reactors, the flow photocatalytic microreactors could reduce the reaction time from several days or a few hours to several hours or even minutes[7,8]. However, in spite of the significant improvement in reaction speed of the current flow photocatalytic systems, their throughputs are still much lower than what is needed for the application of the big data-based artificial intelligence (AI) technique to the field of chemical synthesis[9–14], which has become one of the current hotspots of chemical research as it is considered to have the potential to transformationally improve the efficiency of chemical synthesis research and development[15–18]. The

[1]Department of Chemistry, Zhejiang University, Hangzhou, China. [2]Institute of Intelligent Chemical Manufacturing and iChemFoundry Platform, Engineering Research Center of Functional Materials Intelligent Manufacturing of Zhejiang Province, ZJU-Hangzhou Global Scientific and Technological Innovation Center, Hangzhou, China. [3]Center of Chemistry for Frontier Technologies, Department of Chemistry, Zhejiang University, Hangzhou, China. [4]The Research Center for Life Sciences Computing, Zhejiang Lab, Hangzhou, China. [5]Department of Computer Science and Engineering, The Chinese University of Hong Kong, New Territories, Hong Kong, China. [6]Key Laboratory of Excited-State Materials of Zhejiang Province, Zhejiang University, Hangzhou, China. [7]These authors contributed equally: Jia-Min Lu, Hui-Feng Wang, Qi-Hang Guo. ✉e-mail: gychen@zhejianglab.com; kelvonpan@zju.edu.cn; luzhan@zju.edu.cn; fangqun@zju.edu.cn

lack of a large amount of reliable and high-quality data is also the major factor currently hindering the application of AI technique in chemical synthesis[19–25]. To address this challenge, we developed an automated high-throughput system that can complete ultra-fast photocatalytic reactions in the seconds scale and ultra-large scale screening up to 10,000 reactions per day, using the microfluidic liquid-core waveguide (LCW)[26,27], automated microfluidic liquid-handling and AI techniques.

Results

Ultra-fast photocatalytic synthesis and characterization

We designed and constructed a new type of microfluidic photocatalytic microreactor using the LCW technique to introduce high-intensity laser light into the microfluidic photocatalytic reaction channel to significantly increase the speed of photocatalytic reaction (Fig. 1A). We used four 450-nm lasers as light sources and coupled a total of ca. 4.6 W of light into the reaction area through four optical fibers. The light transmitted by multiple total reflections in the 5-cm-long reaction channel based on the LCW principle, resulting in a laser luminous flux over 3.5×10^4 mW/cm^2 through the channel cross section within a channel length range of 3 cm (Fig. 1B). Such a light irradiation intensity is ca. 10,000 times more than those in conventional batch photocatalytic systems[28,29]. In the microfluidic LCW microreactor, the laser light was reflected multiple times in the photocatalytic microreactor, resulting in a more uniform high-intensity light distribution in the 3-cm-length middle part of the 530-μm-diameter reaction channel, which was 15 times more than the ca. 2 mm irradiation depth in conventional batch photocatalytic systems. However, we observed that the use of the high-intensity laser irradiation caused a significant increase in temperature in the capillary reactor, resulting in an increase in side reactions and a decrease in reaction yield. To address this problem, we designed the microfluidic LCW photocatalytic microreactor as a cannula configuration with an inner reaction capillary and an outer capillary with the circulating cooling solution flowing through to control the temperature of the reaction channel, for instance, 25 ± 2 °C.

We applied the LCW photocatalytic microreactor to a typical organic photocatalytic synthesis, photocatalytic [2 + 2] cycloaddition reaction[30], which has gained extensive attention because cyclobutane obtained through this reaction has been frequently found in a variety of natural products with biological activity[31–33]. Using the LCW photocatalytic microreactor, the photocatalytic [2 + 2] cycloaddition reaction of substrate S-1 could be completely converted with a residence time of only 3.3 s in the microreactor channel, while the reaction yield and diastereomeric ratio (d.r.) were comparable to those of conventional batch photocatalytic systems. As a comparison, the time consumption of this photocatalytic [2 + 2] cycloaddition reaction in a conventional batch photocatalytic reactor was up to 4 h with the same photocatalyst and substrate species, concentration and photocatalyst ratio conditions. Our LCW photocatalytic microreactor could shorten the reaction time by 4300 times with the aid of ultra-high light intensity, uniform and long-length light irradiation, stable reactor temperature control and microfluidic scale effect. To the best of our knowledge, this is the first report of shortening a photocatalytic reaction from several hours to a few seconds, reaching the fastest photocatalytic reaction speed reported in the literature.

In addition to high synthetic reaction speeds, the characterization speed of the reaction products or reactants needs to match the reaction speed to achieve real high-throughput screening. Currently, the routine analysis of organic synthesis products is usually performed by chromatography, mass spectrometry and nuclear magnetic resonance spectroscopy (NMR), most of which are difficult to achieve high-throughput and rapid analysis at the second level[34]. Spectroscopic analysis has the advantages of non-destructive, rapid and easy to implement online detection. Although its weaknesses in qualitative analysis capability and selectivity limit the application for accurate analysis of reaction products, it is well suited in many systems for online monitoring the changes of unsaturated groups in the synthetic reaction process to provide decisive data for the initial high-throughput screening. Before and after the cycloaddition reaction performed in this work, there was a significant change in the UV spectrum of the reaction solution (caused by the consumption of the substrate) in the region of 280 nm–320 nm, thus the reaction solution could be online detected by coupling a capillary flow cell to the LCW photocatalytic microreactor and using a UV-Vis spectrometer with a detection response time of 0.1 s. We also examined the reliability of the UV-Vis detection method in reaction monitoring using the routine characterization methods for organic synthesis reactions—gas chromatography (GC) and NMR.

High-throughput photocatalytic synthesis and screening

In addition to the ability of rapid synthesis and online characterization, a practical high-throughput system must have the ability to automatically change both continuous variables (e.g., reaction temperature, time, pressure, light intensity, and reactant concentration, etc.) and discrete variables (e.g., substrate, photocatalyst and solvent species, etc.) for performing meaningful high-throughput screening of synthesis conditions. In many cases, much attentions are usually paid to achieving rapid organic synthesis reactions under specific conditions, the ability to achieve rapid screening of a large number of different conditions is frequently overlooked. In the present system, we used a liquid handling module consisted of a selective valve and a syringe pump to achieve automated multistep liquid handling operations, including rapid introducing, changing, mixing and driving of different reaction solutions with adjustable flow rates. We uniquely used a combination of a 1-mL and a 5-mL syringes to perform multi-step reciprocal aspirating-dispensing operations for rapidly achieving automated mixing of mL-scale of reactants (substrate, photocatalyst and solvent) within 2 min. We designed a LabVIEW-based program to control the operation of the whole modules of the system (Fig. 1C), allowing the on-demand automated changing of photocatalyst and substrate species, concentration, photocatalyst ratio, laser light intensity and flow rate to achieve the fully-automated operations of the large-scale screening of reaction conditions (>10,000 reactions) without the need of any manual operation and intervention. This photocatalytic synthesis and screening system also served as one of multiple functional islands of the iChemFoundry (IC) platform. The IC platform is a large-scale automated platform for molecular manufacturing which consisted of various functional islands for realizing a series of automated operations of chemical synthesis from reactant preparation, pre-treatment, chemical synthesis reactions, post-treatment, characterization and data analysis. By combining the functional islands of the photocatalytic synthesis and screening system, the reactant preparation system, an orbital robot and an on-island robotic arm of the IC platform, complete automation of all operations could be achieved, including solid reagent weighing, liquid reagent metering, preparation of reagent stock solutions, transferring and aspirating of reagent stocks, preparation and introduction of reactant solutions, on-line flow photocatalytic reactions, UV spectroscopic reaction monitoring, as well as cyclic screening of different conditions (Fig. 1D).

With the present system, we performed a comprehensive screening for the photocatalytic [2 + 2] cycloaddition reaction with a total number of up to 12,000 conditions, including two discrete variables of the photocatalyst and substrate species and four continuous variables as the laser light intensity, concentration, flow rate, and photocatalyst ratios (Fig. 2). Each of the six variables had 4–6 levels or species, resulting in a total orthogonal combination number of

https://doi.org/10.1038/s41467-024-53204-6

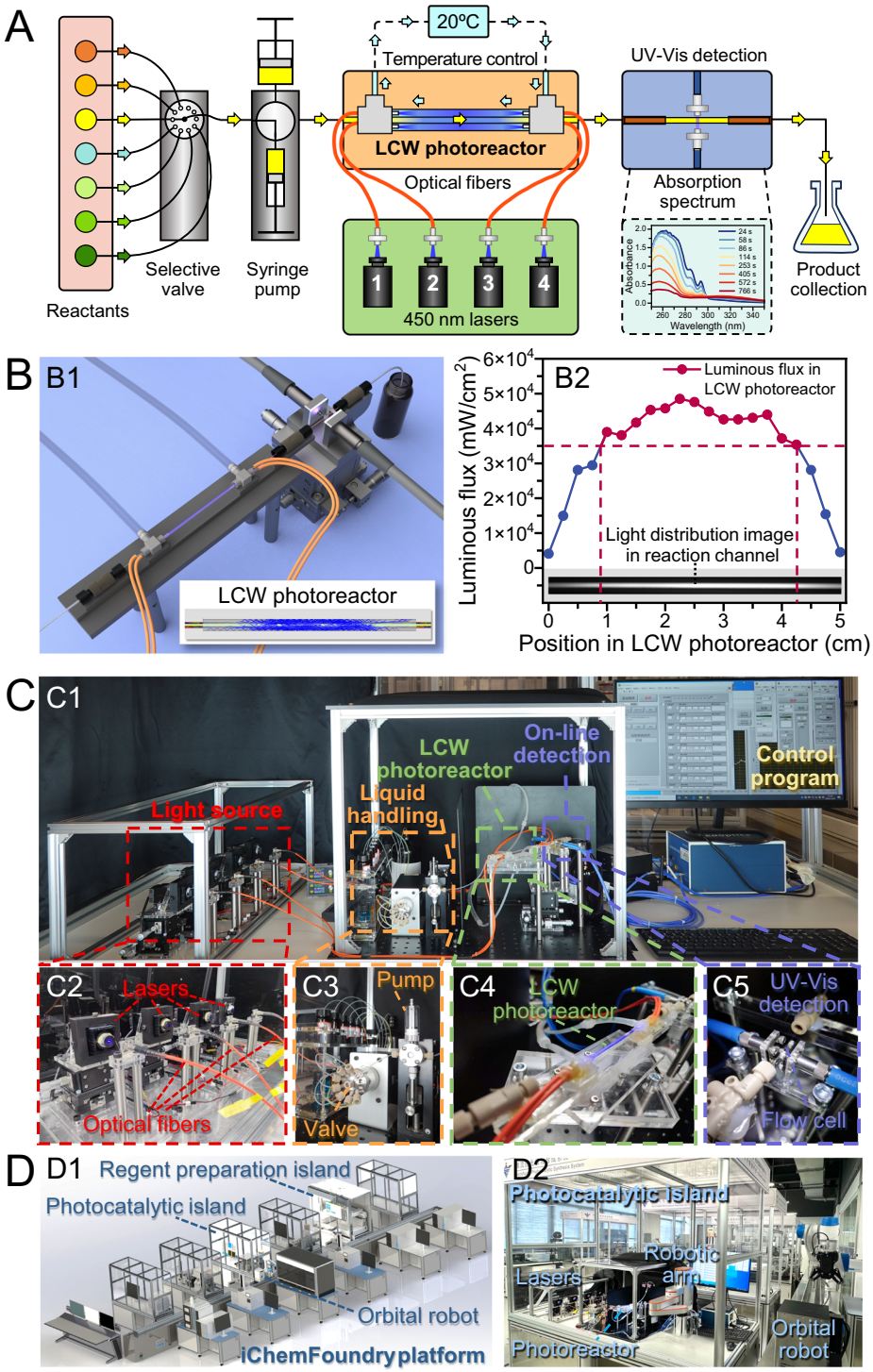


Fig. 1 | Flow manifold and setup of the automated ultra-high throughput photocatalytic synthesis and screening system. A Schematic diagram of the system flow manifold. **B** Structure and light distribution of the LCW photocatalytic microreactor. (B1) Schematic diagram of the LCW photocatalytic microreactor and online UV-Vis absorption spectroscopic detection device; (B2) Cross-sectional light intensity distribution in the reaction channel of the LCW photocatalytic microreactor obtained using TracePro software. **C** Photographs of the system. (C1) Overview of the system. (C2) Light source module, consisting of four 450-nm lasers and four optical fibers. (C3) Liquid handling module with a 10-port selective valve and a syringe pump. (C4) LCW photocatalytic microreactor. (C5) Online UV-Vis absorption spectroscopic detection module. **D** Schematic diagram (D1) and photograph (D2) of the iChemFoundry platform.

12,000 screening conditions (i.e., photocatalytic synthesis and analysis experiments). In our system, the average time required to obtain the data for each photocatalytic [2 + 2] cycloaddition reaction was only 32 s, which enabled the system to reach a screening throughput of 2600 conditions per day.

AI-assisted ultra-high-throughput photocatalytic synthesis and screening

Although the above throughput of 2600 conditions per day with a single flow-reactor format was higher than the highest throughputs (e.g., 1500 reactions per day) in the automated synthesis and screening

https://doi.org/10.1038/s41467-024-53204-6

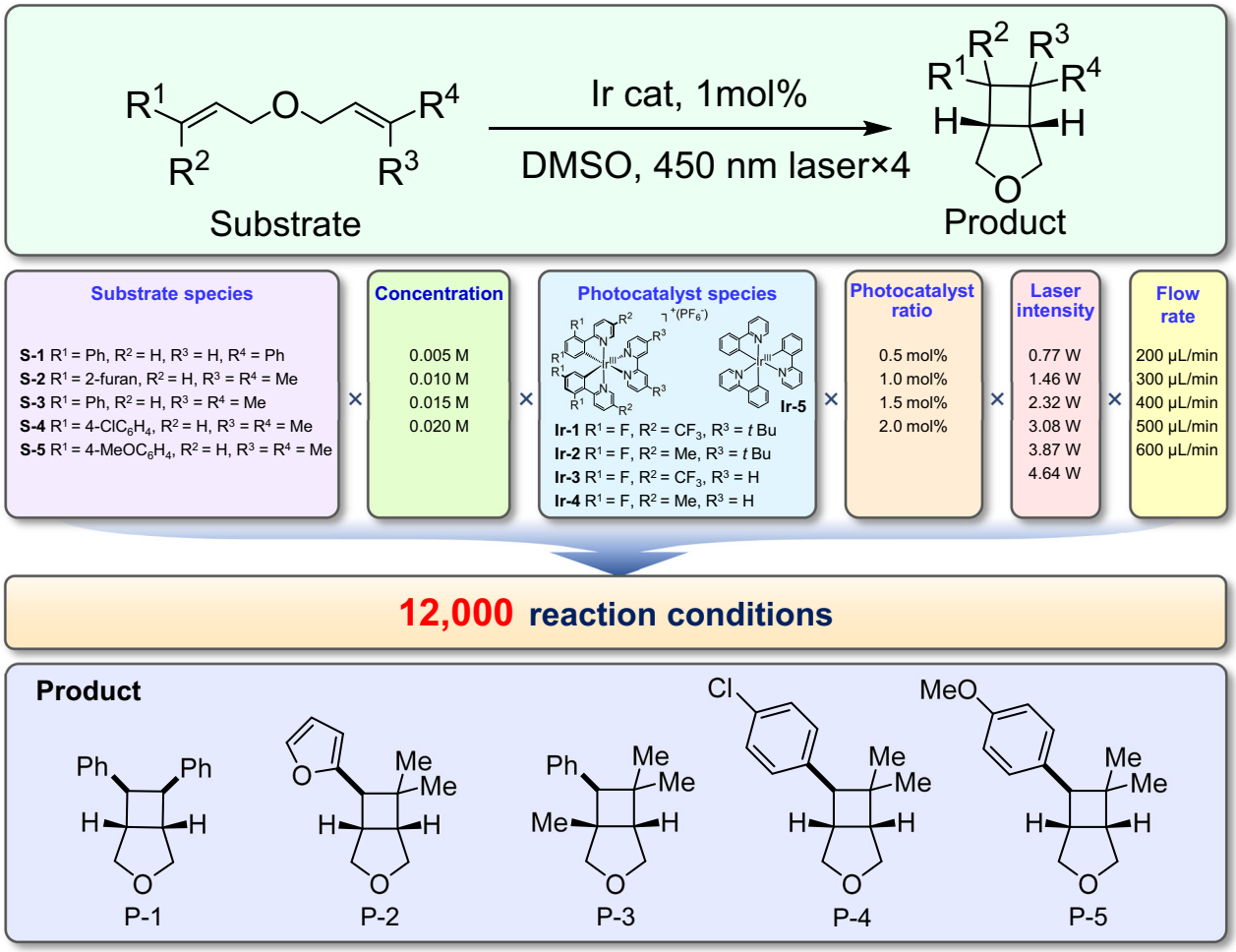


Fig. 2 | High-throughput condition screening for photocatalytic [2 + 2] cycloaddition. List of variables and conditions screened in the photocatalytic [2 + 2] cycloaddition reaction.

systems reported so far[35–37], we noticed that it had the potential to be improved further.

Steady-state and non-steady-state experimental mode. During the high-throughput screening experiments for different reaction conditions, when the synthesis and characterization of the previous condition experiment were completed in the flow system, a new reaction solution needed to be introduced into the reactor and detection flow-cell channels. Due to the convection and molecular diffusion effects existed in the flow system, the previous reacted solution and the newly-introduced unreacted solution would mix with each other at their junction region, and the absorbance signals detected by the UV-Vis detector exhibited a dynamically-changing format (Fig. 3A1) during the switching process of different solutions. The conventional method is to wait for the newly-introduced unreacted reaction solution to flush all of the previous reacted solution out of the flow system to enable the detector to obtain a stable plateau-type steady-state absorbance signal for data reading. However, such a steady-state experimental mode requires much longer waiting time for obtaining the steady-state signals. In the above photocatalytic screening experiment under the steady-state mode, the lasers kept irradiating the microreactor channel and the system spent most of the time (27 s of the 32 s of one experimental cycle time) in switching the different experimental conditions and waiting for a steady-state detection signal to be obtained (for example, as shown in Fig. 3A2). Such a waiting time far exceeded the actual time (<4 s) for photocatalytic synthesis and characterization for a reaction solution, severely limiting the screening throughput for different reaction conditions. In fact, this is one of the major limiting bottlenecks in the application of current flow chemistry systems to high-throughput screening.

To increase the efficiency of time utilization and screening throughput, we proposed the strategy of non-steady-state experimental mode instead of the steady-state mode by using the laser pulse irradiation method to turn the irradiation laser on and off for achieving the rapid switching between the reacted and unreacted solutions, producing a series of non-steady-state continuous peak-shaped signals as shown in Figure 3A3. Under the non-steady-state mode, the waiting time for reaction solution switching was shortened to 6 s and the average time for each experimental cycle was shortened to 8.5 s, achieving an ultra-high throughput up to 10,000 reaction conditions per day (Fig. 3A3, A4, B).

However, these non-steady-state peak signals included the combined absorbance information of the reactants and products from the previous reacted and the newly-introduced unreacted reaction solutions, which were influenced by multiple factors related to the convection and molecular diffusion effects, such as the reaction solution flow rate, the inner diameters and lengths of the reactor and the detection flow-cell channels, and the reactants and products molecular weights. Therefore, it was a great challenge to acquire the corresponding steady-state absorbance data from the non-steady-state peak signals for evaluating the reaction progress.

https://doi.org/10.1038/s41467-024-53204-6

AI-assisted absorbance prediction. For achieving this complex and challenging task, we developed the AI-assisted absorbance prediction method by using the AI method to analyze the influencing factors related to the convection and molecular diffusion effects and decoupling the non-steady-state data of the adjacent reaction solutions mixed with each other, to predict the corresponding steady-state absorbance data of the respective reaction solutions.

In order to obtain accurate prediction results, we attempted to use 10 regression models based on the principles of linear models, decision tree, neural networks and integrated learning, to process the

Fig. 3 | Recordings of the typical and total absorbance signals obtained in the large-scale screening experiment. A Typical recordings of the steady-state and non-steady-state absorbance signals obtained in the screening for S-5, with a photocatalyst ratio of 2 mol% and a S-5 concentration of 0.01 M. (A1) Typical absorbance recordings of 6 different reaction conditions under the steady-state screening mode. (A2) Enlarged view of the absorbance recordings for the first two conditions of the 6 reaction conditions in (A1). It took an average time of 27 s in each condition cycle to obtain the steady-state absorbance signal, which was calculated from the difference between the reacted steady-state plateau absorbance and the unreacted blank absorbance. (A3) Typical absorbance recordings of 30 different reaction conditions under the non-steady-state screening mode. (A4) Enlarged view of the absorbance recordings for the 17th and 18th conditions of the 30 reaction conditions in (A3). Each condition cycle took an average time of ca. 4 s, with a non-steady-state signal peak containing 40 absorbance data points, obtained using the laser pulse irradiation method. **B** Recordings of the non-steady-state peak signals obtained in the screening experiment of the total 12,000 reaction conditions, which was replicated three times to test the repeatability.

large numbers of the non-steady-state absorbance data and to predict the corresponding steady-state absorbance data under the same reaction condition, from which we searched for the best-performing model. In these models, the non-steady-state absorbance data (i.e., 40 absorbance data points recorded for each non-steady-state signal peak, Fig. 3A4) as well as the all corresponding 8 variables (i.e., flow rate, laser light intensity, wavelength, substrate concentration, photocatalyst ratio, photocatalyst concentration, substrate species, and photocatalyst species) of the present flow photocatalytic system were set as the inputs of the models (Fig. 4A). The corresponding steady-state absorbance data obtained experimentally using the steady-state mode with the same conditions (i.e., real steady-state absorbance data) were set as the targets of the models, and the predicted steady-state absorbance data were set as the outputs of the models. The substrate and photocatalyst species were input to the models in the form of relative molecular weights instead of chemical structures, since they are directly related to the molecular diffusion effect. On the basis of the massive amounts of the output and target data of the 12,000 reaction conditions (Fig. 4B1), we evaluated the performance of the 10 regression models based on the R^2 and RMSE values of the test set. The RMSE values were calculated based on the targets (i.e., real steady-state absorbance data) and the outputs (i.e., predicted steady-state absorbance data) of the models. Among these models, the XGB regression model demonstrated the best prediction performance (Fig. 4B2), which had the smallest RMSE of 0.0140 and the largest R^2 of 0.991 with the 70: 30 proportion of the training and test set. We tried to use less data to predict more data (such as 2.5% of the data as the training set and 97.5% of the data as the test set) to further test the predictive performance of the XGB regression model. Pretty good result was still obtained where 300 conditions data were used to predict the remaining 11,700 conditions with a RMSE of 0.0550 and R^2 of 0.859 (Fig. 4B3).

These results showed that with the use of the non-steady-state mode and the AI-assisted absorbance prediction method, the long-standing challenge limiting the improvement of screening throughput of flow chemical screening systems caused by inefficient and time-consuming condition switching could be solved. Correspondingly, the screening throughput for the photocatalytic [2 + 2] cycloaddition reaction conditions increased from 2600 to 10,000 conditions per day using the non-steady-state experimental mode, which is the highest level reported in the field of organic synthesis so far. If the manual mode of conventional batch photocatalytic organic synthesis experiments is used to complete the same workload, it would require 2000 organic synthesizers to work one day or a synthesizer to work 2000 days, assuming that each person could complete 5 condition screening experiments per day. What's more, we repeated all the non-steady-state experiments 3 times to ensure the repeatability, and all 36,000 peak signals are shown in Fig. 3. In addition, in terms of reagent consumption, only 4.0 mmol of each substrate and 0.05 mmol of each photocatalyst were required to complete the whole screening experiment.

We used 10% of the non-steady-state absorbance data as the training set of the XGB model to predict 12,000 steady-state absorbance data, and a 10-fold cross-validation study was performed and the results are shown in Supplementary Information (Table S9). All the absorbance data were converted into product yields (Fig. 4C1). There are 25 data squares in the heatmap, each of which consists of the data of 1 substrate and 1 photocatalyst condition with different concentrations, photocatalyst ratios, flow rates, and laser light intensities (Fig. 4C2). These data squares show different profiles of the product yield distribution, indicating that the photocatalyst and substrate species have significant impact on the product yields. Among the four variables, although all of them show effects on the product yields, the effects of the flow rate and laser light intensity are more obvious. Therefore, we further present the 12,000 data using multidimensional bubble plots with the x-axis of the flow rate and the y-axis of laser light intensity.

Factors influencing photocatalytic [2 + 2] cycloaddition reaction

In the large majority of experiments, the yield of the reaction products significantly increased with the increase of the laser light intensity (Fig. 5A), indicating that sufficient high light intensity is a necessary prerequisite for the high-efficient photocatalytic [2 + 2] cycloaddition reaction in the LCW photocatalytic microreactor.

As shown in Fig. 5A, the product yield increased as the flow rate of the reaction mixture decreased. A higher flow rate means a higher throughput while a shorter residence time (i.e., light irradiation time) of the reaction solution in the photocatalytic microreactor, which results in a lower reaction yield (Fig. 5A). The present high-efficiency LCW photocatalytic microreactor could overcome the contradiction between the throughput and product yield, with which the photocatalytic reaction speed could be dramatically increased compared to the batch synthesis methods[38–41], while a high level of product yield and selectivity could also be ensured. For example, in the experiment for S-1, a residence time of 3.3 s in the photocatalytic microreactor (corresponding to the flow rate of 200 μL/min) was sufficient to ensure that the substrate S-1 was completely and selectively converted to the target product.

As shown in Fig. 5B, the optimal photocatalyst for each of the five substrates was quite distinct, which reflects the significance of the high-throughput screening. For S-1, S-4 and S-5, Ir-1 was the optimal photocatalyst with the highest yields, while the photocatalyst Ir-5 was the optimal photocatalyst for S-2 and the photocatalyst Ir-3 was the optimal photocatalyst for S-3.

For the effect of the photocatalyst ratio, generally, the higher the photocatalyst ratio, the faster the reaction speed and the higher product yields (Fig. 5A). For meeting the requirements of green chemistry, 1 mol% photocatalyst is an optimal cost-efficiency choice under a sufficient light intensity.

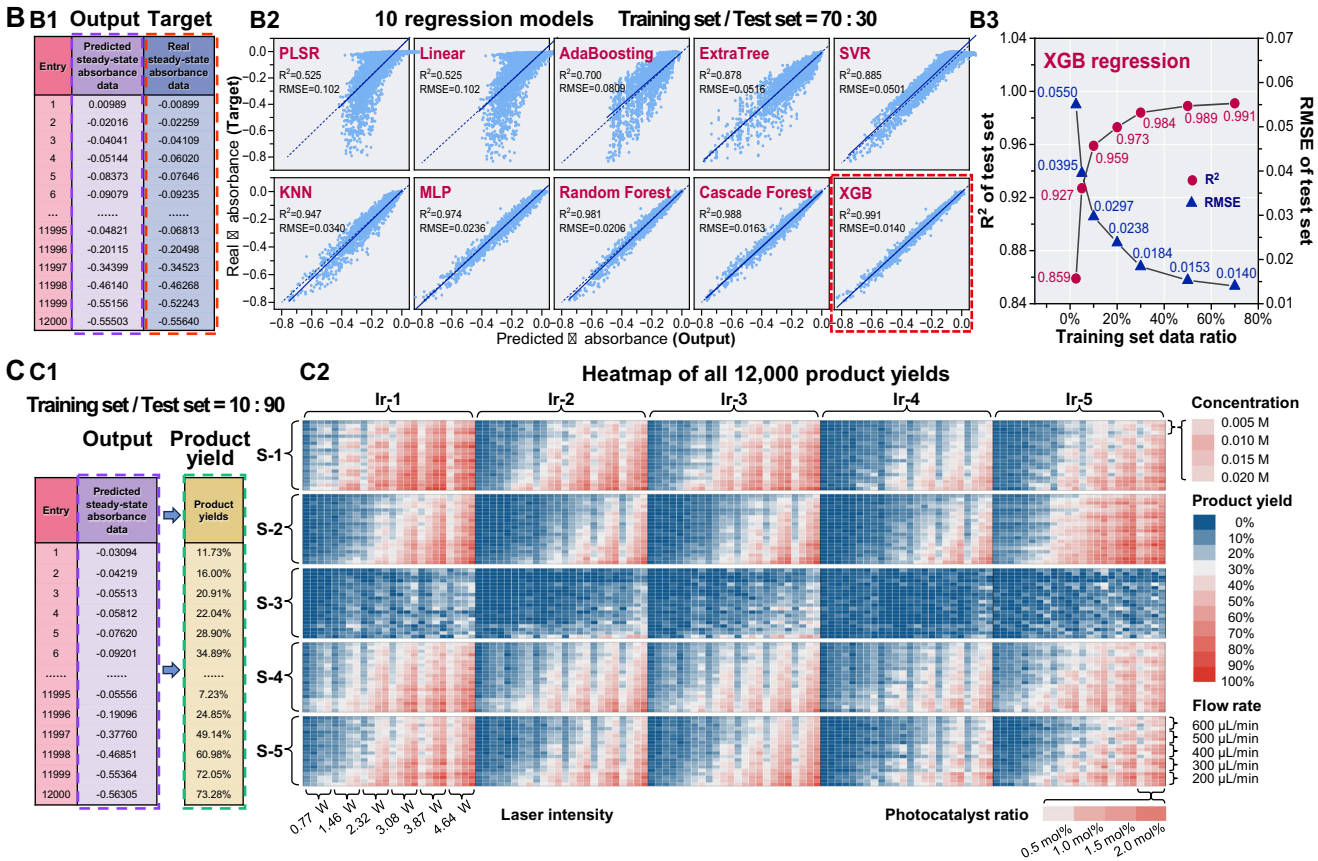
For the effect of the substrate concentration, in most of the plots (Fig. 5C), we did not observe an evident correlation between the product yields and the substrate concentrations of S-1, S-2, S-3, S-4, and S-5 in the tested range of 0.005 M to 0.02 M, which are frequently adopted in conventional photocatalytic reaction experiments. The reason for this phenomenon may be that our system has an ultra-high catalytic capacity which did not show a significant difference in product yield in the tested range of the substrate concentrations.

In the conventional batch photocatalytic systems, how to increase the substrate concentrations while ensuring sufficiently

https://doi.org/10.1038/s41467-024-53204-6

A Input: 8 reaction variables + 40 data points for each non-steady-state signal peak — Target

Entry	Substrate (MW)	Photocatalyst (MW)	Absorbance wavelength (nm)	Concentration (M)	Photocatalyst ratio (mol%)	Photocatalyst concentration (M)	Reaction flow rate (μL/min)	Laser intensity (mW)	40 data points for each non-steady-state signal peak					Real steady-state absorbance data
									x1	x2	...	x39	x40	
1	250.3	1122	293.3	0.005	0.5	0.000025	600	772	0.26216	0.26220	...	0.26235	0.26245	-0.00899
2	250.3	1122	293.3	0.005	0.5	0.000025	600	1459	0.26223	0.26211	...	0.26053	0.26054	-0.02259
3	250.3	1122	293.3	0.005	0.5	0.000025	600	2297	0.26030	0.25999	...	0.25683	0.25657	-0.04109
4	250.3	1122	293.3	0.005	0.5	0.000025	600	3082	0.25605	0.25546	...	0.25258	0.25266	-0.06020
5	250.3	1122	293.3	0.005	0.5	0.000025	600	3868	0.25225	0.25147	...	0.24729	0.24712	-0.07646
6	250.3	1122	293.3	0.005	0.5	0.000025	600	4642	0.24665	0.24572	...	0.24309	0.24463	-0.09235
...	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	...	.....	.....	.....
11995	236.7	655	293.3	0.02	2	0.0004	200	772	1.07372	1.07272	...	1.04654	1.04588	-0.06813
11996	236.7	655	293.3	0.02	2	0.0004	200	1459	1.04423	1.04279	...	0.98814	0.98754	-0.20498
11997	236.7	655	293.3	0.02	2	0.0004	200	2297	0.98596	0.98437	...	0.90900	0.91040	-0.34523
11998	236.7	655	293.3	0.02	2	0.0004	200	3082	0.91079	0.91046	...	0.80804	0.80976	-0.46268
11999	236.7	655	293.3	0.02	2	0.0004	200	3868	0.81055	0.81088	...	0.73706	0.73969	-0.52243
12000	236.7	655	293.3	0.02	2	0.0004	200	4642	0.74216	0.74306	...	0.74551	0.76398	-0.55640



Entry	Predicted steady-state absorbance data	Real steady-state absorbance data
1	0.00989	-0.00899
2	-0.02016	-0.02259
3	-0.04041	-0.04109
4	-0.05144	-0.06020
5	-0.08373	-0.07646
6	-0.09079	-0.09235
...	.....	.....
11995	-0.04821	-0.06813
11996	-0.20115	-0.20498
11997	-0.34399	-0.34523
11998	-0.46140	-0.46268
11999	-0.55156	-0.52243
12000	-0.55503	-0.55640

Entry	Predicted steady-state absorbance data	Product yields
1	-0.03094	11.73%
2	-0.04219	16.00%
3	-0.05513	20.91%
4	-0.05812	22.04%
5	-0.07620	28.90%
6	-0.09201	34.89%
.....	.....	.....
11995	-0.05556	7.23%
11996	-0.19096	24.85%
11997	-0.37760	49.14%
11998	-0.46851	60.98%
11999	-0.55364	72.05%
12000	-0.56305	73.28%

Fig. 4 | AI-assisted prediction of steady-state absorbance data from non-steady-state absorbance data for condition screening of photocatalytic [2 + 2] cycloaddition. A Composition of the large dataset used in the AI-assisted steady-state absorbance prediction, including the 12,000 absorbance data with 48 eigenvalues (8 reaction variables and 40 data points for each non-steady-state signal peak) and 1 target (steady-state absorbance data). **B** AI-assisted prediction models for predicting the steady-state absorbance data from the non-steady-state absorbance data. (B1) Predicted and real steady-state absorbance as the outputs and targets of the models, respectively. (B2) 10 different regression models for the prediction of the steady-state absorbance data from the non-steady-state absorbance data, including partial least squares regression (PLSR), linear regression, adaptive boosting (AdaBoosting) regression, extremely randomized trees (ExtraTree) regression, support vector regression (SVR), k-nearest neighbor (KNN) regression, multi-layer perceptron (MLP) regression, random forest regression, cascade forest regression, and extreme gradient boosting (XGB) regression models. The dashed line is the y = x line, and the solid line is a linear fit curve between the predicted and the true steady-state absorbance values. The two metrics, R^2 and RMSE values, were calculated using the corresponding functions within sklearn.metrics. (B3) Performance of the test set of the XGB regression model. The prediction accuracy of the test set gradually decreases as the proportion of randomly selected training set data decreases from 70% to 2.5%. **C** Screening results of the total 12,000 reaction conditions obtained using the AI-assisted steady-state absorbance prediction method. (C1) Predicted steady-state absorbance data output from the XGB model with training set: test set = 10: 90 were converted to product yields. (C2) Heatmap showing the screening results of a total of 12,000 reaction conditions, including the orthogonal combination of 5 substrate species, 5 photocatalyst species, 4 concentrations, 4 photocatalyst ratios, 5 flow rates, and 6 laser light intensities under the non-steady-state mode.

high yields has been a critical challenge and has not been well addressed so far. For example, in the photocatalytic [2 + 2] cycloaddition reaction reported by Yoon et al. using a conventional batch photocatalytic reactor, when the substrate concentration increased from 0.01 M to 0.05 M, the product yield reduced from 89% to only 33%[30]. We further increased the concentration of substrate S-1 using 50 μL/min flow rate (i.e., 13.2 s residence time) in the present system, the results showed that the product yield decreased slightly from 91% to 85% when the substrate S-1 concentration increased from 0.01 M to 1.0 M, and further reduced to 67% when the substrate S-1 concentration reached its solubility limit of 2.0 M (Fig. 5D). Compared with that (0.01 M) of the conventional

https://doi.org/10.1038/s41467-024-53204-6

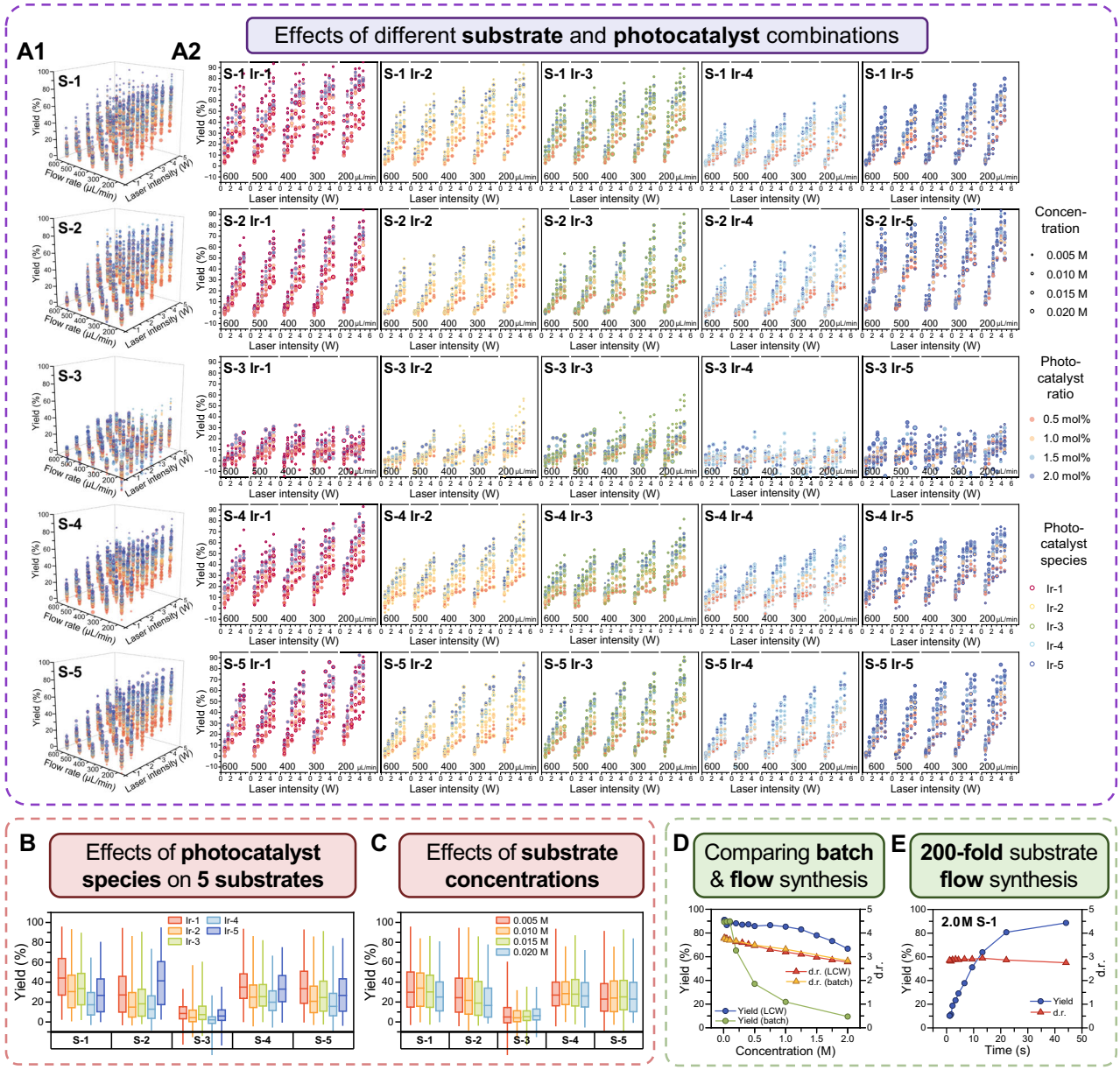


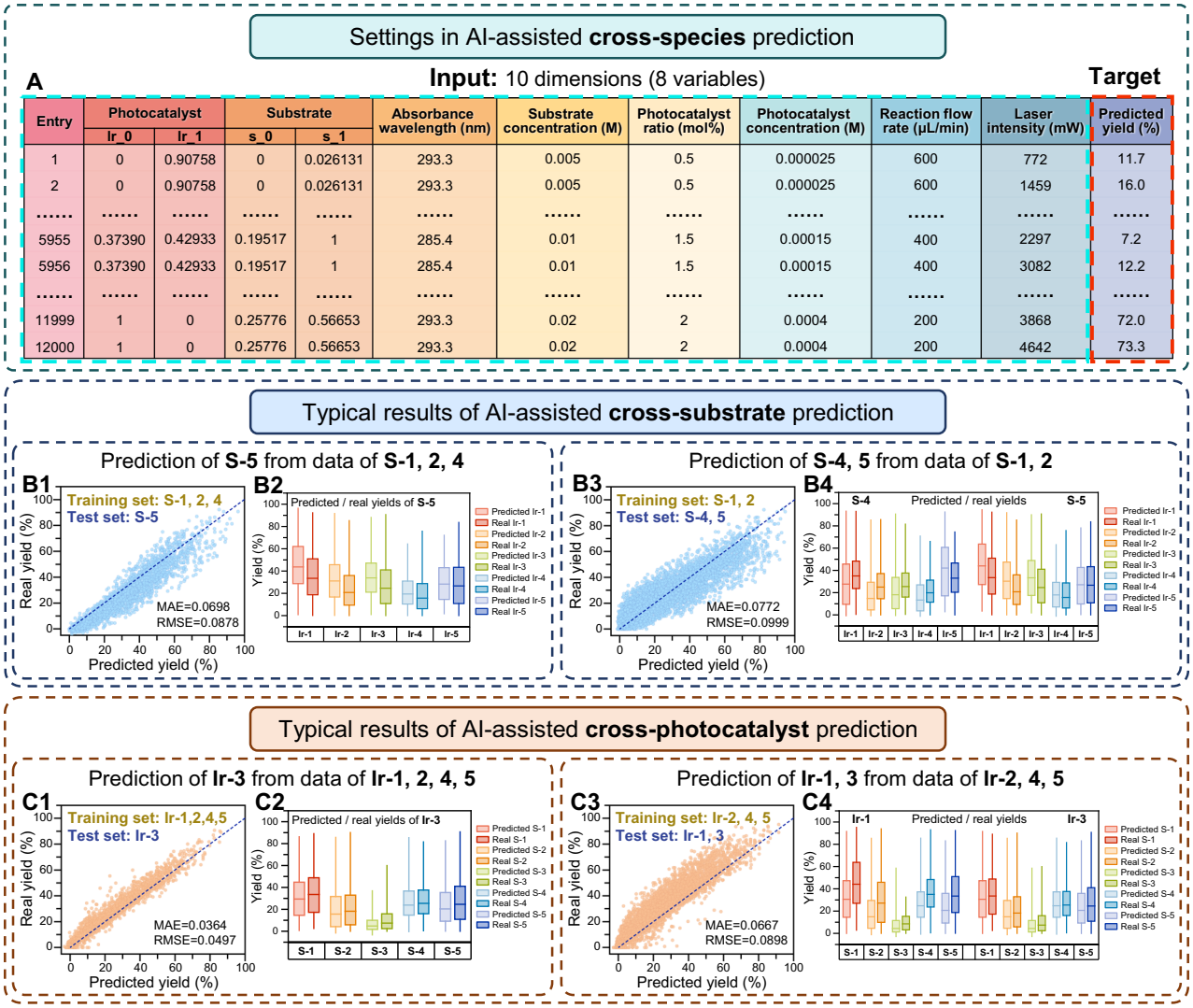
Fig. 5 | Screening results of 12,000 reaction conditions. A Multidimensional bubble plots of the 12,000 data, showing the effects of different variables on the product yield, including the species, concentrations, and ratios of the substrates and photocatalysts, laser light intensity, and flow rate of the reaction solutions. Each 6-dimensional bubble plot in (A1) contains 2400 product yield data of 1 substrate specie and 5 photocatalyst species. Each 5-dimensional bubble plots in (A2) contains 480 product yield data of 1 photocatalyst and 1 substrate specie, corresponding to a data square in the heatmap shown in Figure 4C2. In each 5-dimensional bubble plot, the colors of the contour lines of the bubbles represent different photocatalyst species, the colors filled in the bubbles represent different photocatalyst ratios, and the sizes of the bubbles represent different substrate concentrations. **B** Box plot showing the effect of the photocatalyst species on the product yields of the 5 substrates. The three horizontal lines of each box from top to bottom represent the first, median, and third quartiles of the product yield data, respectively. **C** Box plot showing the effect of the concentrations of the 5 substrates on the product yields. The three horizontal lines of each box from top to bottom represent the first, median, and third quartiles of the product yield data, respectively. **D** Comparisons of the variations of the product yield and d.r. with the increase of the S-1 concentration using the batch and present flow methods. **E** Variations of the product yield and d.r. with the increase of the residence time in the LCW microreactor using S-1 with a high concentration of 2.0 M, which is 200-fold of that in conventional batch systems.

batch system, the substrate concentration could have a 100-fold increase without evidently affecting the reaction yield. We further reduced the flow rate of the 2.0 M substrate reaction solution, the yield could reach 89% at 15 μL/min (i.e., 44 s residence time) (Fig. 5E). The high yields obtained at high substrate concentrations could be attributed to the high light intensity and the fast mass transfer effect in the present photocatalytic microreactor. Such a breakthrough result is of great significance to the application of the photocatalytic cycloaddition reaction in the industrial production of related drugs.

AI-assisted cross-species prediction

To further utilize the above 12,000 data and preliminarily explore the potential possibility of applying AI technique to intelligent chemical synthesis screening, we used the XGB algorithm to perform AI-assisted prediction of product yields cross-substrate and cross-photocatalyst. For cross-species prediction of product yields, the inputs of the models were the reaction conditions (i.e., the 8 reaction variables), the targets were the product yields obtained experimentally, and the outputs were the predicted product yields (Fig. 6A). Differing from the

https://doi.org/10.1038/s41467-024-53204-6



Entry	Photocatalyst		Substrate		Absorbance wavelength (nm)	Substrate concentration (M)	Photocatalyst ratio (mol%)	Photocatalyst concentration (M)	Reaction flow rate (μL/min)	Laser intensity (mW)	Predicted yield (%)
	Ir_0	Ir_1	s_0	s_1							
1	0	0.90758	0	0.026131	293.3	0.005	0.5	0.000025	600	772	11.7
2	0	0.90758	0	0.026131	293.3	0.005	0.5	0.000025	600	1459	16.0
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
5955	0.37390	0.42933	0.19517	1	285.4	0.01	1.5	0.00015	400	2297	7.2
5956	0.37390	0.42933	0.19517	1	285.4	0.01	1.5	0.00015	400	3082	12.2
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
11999	1	0	0.25776	0.56653	293.3	0.02	2	0.0004	200	3868	72.0
12000	1	0	0.25776	0.56653	293.3	0.02	2	0.0004	200	4642	73.3

Fig. 6 | AI-assisted cross-species prediction. A Settings used for AI-assisted cross-species prediction, including inputs and targets of the models. **B** Typical results of the AI-assisted cross-substrate prediction, including prediction of S-5 from data of S-1, S-2, and S-4 (B1), and prediction of S-4, S-5 from data of S-1 and S-2 (B3) with box plots showing product yields results (B2, B4), respectively. **C** Typical results of the AI-assisted cross-photocatalyst prediction, including prediction of Ir-3 from data of Ir-1, Ir-2, Ir-4 and Ir-5 (C1), and prediction of Ir-1, Ir-3 from data of Ir-2, Ir-4 and Ir-5 (C3) with box plots showing product yields results (C2, C4), respectively. The model performance metrics for the cross-validation studies include the MAE and RMSE values. The entire results of the cross-species prediction are shown in Fig. S11, Table S10, S11.

AI-assisted absorbance prediction, the cross-species prediction required detailed chemical structure information of the substrate and photocatalyst species, which were described, digitized, and input to the models using the SMILES strings to generate Modred descriptors (Table S1). The Mordred descriptor dimensions of each substrate and photocatalyst species were reduced to 2 in order to match the dimensions of the other variables and facilitate visualization in subsequent study, as described in Methods. The inputs consisted of 10 dimensions, with 2 representing the substrate species, 2 representing the photocatalyst species, and the remaining 6 representing the other variables (Fig. 6A).

Both the results of the cross-species prediction with different training set ratios and the cross-validation studies are described in the Supplementary Information (Fig. S11, Table S10, S11). The 12,000 experimental data were divided into training and test sets with different ratios for AI-assisted cross-species prediction. As a typical result of the cross-substrate prediction, we used the data of S-1, S-2, and S-4 as the training set to predict the yields of S-5, achieving MAE = 0.0698 and RMSE = 0.0878 (Fig. 6B1). The distinct effects of the 5 photocatalyst species on S-5 were accurately predicted and Ir-1 was predicted as the optimal photocatalyst, which is consistent with the experimental results (Fig. 6B2). With a smaller training set of S-1 and S-2, the prediction for S-4 and S-5 achieved with MAE = 0.0772 and RMSE = 0.0999 (Fig. 6B3, B4). For the cross-photocatalyst prediction, the data of Ir-1, Ir-2, Ir-4 and Ir-5 could be used to predict the yields of Ir-3, with MAE = 0.0364 and RMSE = 0.0497, which presented similar results to the real product yields (Fig. 6C1, C2). When the training set was reduced to include three photocatalysts of Ir-2, Ir-4 and Ir-5, pretty good prediction for Ir-1 and Ir-3 could still be obtained with MAE = 0.0667 and RMSE = 0.0898 (Fig. 6C3, C4). These surprising results exemplarily demonstrated the attractive possibility of utilizing AI algorithms for cross-substrate and cross-photocatalyst prediction of organic synthesis reactions.

Discussion

In summary, we developed a fully-automated and integrated system for high-throughput chemical synthesis, online characterization, and large-scale of condition screening of photocatalytic reactions. With the LCW, microfluidic liquid-handling, and AI techniques, the present

system achieved fast photocatalytic cycloaddition reaction in only a few seconds, and reached a high-throughput screening of 10,000 reaction conditions per day that has never been reached before. It can produce massive amount of high-quality data with extremely-low time and reagent consumption, which is difficult to accomplish using conventional systems, providing a powerful platform to facilitate the application of AI techniques in chemical synthesis and to help researchers explore the unknown chemical space more efficiently. As we initially demonstrated in AI-assisted cross-species prediction, such a large amount of data from the same experimental system could provide a solid data base for AI applications. In the future, it would be meaningful to make full use of the 12,000 data and further incorporate AI techniques, such as Bayesian optimization method, for rapid optimization of new products.

In addition, the system can be further improved to meet the needs of different types of photocatalytic reactions, for instance, changing the light source with other wavelengths or trying other types of reactions such as photocatalytic cross-coupling reactions to extend its application scope. For the scale-up of photocatalytic reactions for industrial production, the present system showed a unique potential for increasing the concentration of photocatalytic substrates. In addition to organic synthesis, the application of the system can also be extended to other synthesis fields such as photoinduced material synthesis or biological molecule synthesis in the future.

Methods

Building of the robotic system

The robotic system consisted of a LCW photocatalytic microreactor module, a liquid handling module, and an online UV-Vis absorption spectroscopy module, which is capable of achieving rapid synthesis and high-throughput condition screening of photocatalytic synthesis reactions. We designed a new photocatalytic microreactor based on the LCW principle, which could significantly increase the irradiation light intensity inside the flow microreactor and realize rapid photocatalytic synthesis. In the LCW photocatalytic microreactor module, light from 4 lasers was coupled into flow microreactor with 4 optical fibers to provide high and uniform light irradiation, while the reaction mixture temperature was under control of the cannula-shaped circulating-water temperature control device in order to ensure the selectivity of target products. The inlet and outlet of the LCW photocatalytic microreactor module were connected to the liquid handling module and the online UV-Vis absorption spectroscopy module by capillaries, respectively, allowing the reaction mixture to flow through. The liquid handling module utilized a syringe pump and a 10-port selective valve for liquid handling in the system, including automated introducing of different reactants, preparing reaction mixtures and delivering them through the LCW photocatalytic microreactor and UV-Vis detection module at set flow rates. In order to realize high-throughput detection that could match the efficiency of high-throughput synthesis, the absorption spectra of the reaction mixture after synthesis were recorded in real time by the online UV-Vis absorption spectroscopy module. The operation of all the modules of the robotic system was automatically controlled by a self-written program without human intervention.

LCW photocatalytic microreactor module

In order to realize the LCW-based light illumination and temperature control of the reaction mixture, the LCW photocatalytic microreactor was designed as a cannula configuration with an inner fused-silica capillary (530 μm i.d. and 690 μm o.d., Refined Chromatography Co., Yongnian, China) and an outer glass capillary (1.5 mm i.d. and 2.5 mm o.d., Jingke, Guangzhou, China). The inner fused-silica capillary served as the reaction channel with the reaction mixture solution flowing through, and the outer glass capillary acted as the temperature control channel with the circulating cooling solution flowing through (Fig. S1A). The inner and the outer capillary were installed with two polypropylene tees (3.2 mm, 1/8, Jieliante Technology Co., Beijing, China). The inlet and outlet of the reaction channel were connected with a PFA capillary (1512 L, IDEX, Northbrook, USA) and the detection capillary (250 μm i.d. and 375 μm o.d., Refined Chromatography Co., Yongnian, China), respectively, using unions (P-702, IDEX, Northbrook, USA) or tees (P-712, IDEX, Northbrook, USA).

To provide sufficiently high laser light intensity, the light source consisted of 4 lasers (5.5 W, λmax = 450 nm, Oxlasers, Shanghai, China), from which 4 laser beams were coupled and conducted in the LCW photocatalytic microreactor through 4 optical fibers (400-μm diameter, Lanwin Technology, Zhongshan, China), respectively (Fig. S1B). Each laser beam with a focused spot diameter of ca. 400 μm could be fine-tuned by an x-y-z translation stage to couple as much light as possible into the optical fiber. The 1-meter-long optical fibers consisted of a core/cladding/coating three-layer structure made of quartz/quartz/acrylic with a diameter of 400/440/700 μm and a refractive index of 1.446/1.43/1.51, respectively. With one end of the optical fiber connected with the laser for optical coupling, the other end of the optical fiber was placed between the inner and outer channels to introduce the laser light, and there were two optical fibers both on the left and right sides of the LCW photocatalytic microreactor for providing uniform light illumination (Fig. S1C). The light emitted from the optical fibers spread out naturally according to the numerical aperture (NA) of 0.22, and could be uniformly distributed in the microreactor by multiple total reflections, based on the LCW principle. With the aid of the TracePro software (7.4.3, Lambda Research Co., Littleton, USA), we simulated the light intensity distribution in the cross-section of the reaction channel of the LCW photocatalytic microreactor, which showed that the main section of the reaction channel was under the illumination of strong and uniform laser light (Fig. S1D).

In order to deflect the unwanted heat caused by the high-intensity laser light and to perform the synthesis at set temperature, the cooling solution was circulated in the outer capillary using a diaphragm pump (KVP004, Kamoer Fluid Technology Co., Shanghai, China), and the temperature of the circulating cooling solution was monitored in real time by a digital temperature controller (XH-W3002, Xinghe Electronic Technology Co., Suqian, China). The on and off of the semiconductor chilling plate (TEC1-12706, Xinghe Electronic Technology Co., Suqian, China) was controlled by the digital temperature controller to maintain the temperature of the reaction mixture (Fig. S1E). When the temperature of the cooling solution was set at 20–21 °C, the actual temperature of the reaction mixture in the reaction channel could be maintained in the range of 22.4–27.2 °C measured by an ultrafine thermocouple.

Liquid handling module

In the liquid handling module, the reaction mixture was prepared using a syringe pump (20738325 Cavro XCalibur Pump, Tecan, Mannedorf, Switzerland) and a 10-port selective valve (C5-3000EUTA, VICI, Schenkkan, Switzerland). In order to ensure the data quality of the condition screening, the stock solutions of different reactants need to be freshly prepared every day, which were made in 20 mL vials as 0.2 M substrates and 0.0035 M photocatalysts in DMSO. For performing automated liquid handling, 7 positions of the 10-port selective valve were utilized to select the stock solutions of different reactants for condition screening, with position 4 connected to the solvent, position 3 connected to the substrate, and the remaining 5 positions (i.e., position 1, 2, 5, 6, 7) connected to different photocatalyst species. For the syringe pump, the valve position 1 was connected to the 10-port selective valve for introducing the reactants, and the valve position 2 was connected to a 5 mL syringe that worked in conjunction with another 1 mL syringe for automated dispensing and rapid mixing (Fig. S2A). With the innovative use of these two syringes of different volumes, the reactants could be mixed thoroughly by quickly aspirating and mixing at a rate of more than 50 mL/min for several times according to the mixture volume. Subsequently, the reaction mixture

was introduced into the reaction channel of the LCW photocatalytic microreactor through valve position 3 of the syringe pump.

During the reagent sampling process, the solvent DMSO was sampled in the first step and was used to rinse the reaction channel in the last step. The stock solutions of the substrates and photocatalysts were sampled in the intermediate steps to facilitate the homogeneous mixing of the reaction mixture. In order to save time in preparation and ensure high throughput, the volume of reaction mixture prepared each time should be sufficient to complete the screening for a set of 30 conditions with different flow rates and laser light intensities ($5 \times 6 = 30$). In the screening for each set of 30 steady-state conditions, at least 15 reciprocal aspirating-dispensing operations were required to achieve the complete mixing of 5 mL of reaction mixture solution, which took 2 min 30 s. In the screening for each set of 30 non-steady-state conditions, at least 10 reciprocal aspiration-dispensing operations were required to achieve the complete mixing of 3 mL of reaction mixture solution, which took 1 min 15 s. After screening conditions of 6 different laser light intensities at the same flow rate, the reaction channel was rinsed with the unreacted mixture solution to allow the absorbance signal back to the baseline before changing the flow rate. When changes in substrate concentration, photocatalyst ratio, photocatalyst or substrate species were required, it was necessary to thoroughly clean the reaction channel and the detection flow-cell channel with solvent DMSO to avoid cross-contamination. In addition, we tested the performance of the system by alternatively performing condition screenings on reaction mixture of S-1 (0.01 M) catalyzed by Ir-1 (1 mol%) and reaction mixture of S-1 (0.01 M) catalyzed by Ir-4 (1 mol%), using operations of changing the stock solutions with the 10-port selective valve. Each set of experiments was repeated 3 times. The results showed good reproducibility with RSD below 5% (Fig. S2B).

Online UV-Vis absorption spectroscopy module

In the online UV-Vis absorption spectroscopy module, a 6 cm-long fused-silica capillary (250 μm i.d. and 375 μm o.d., Refined Chromatography Co., Yongnian, China) was connected to the LCW photocatalytic microreactor module as a detection flow-cell with an effective detection path length of 240 μm, and the protective layer was removed to allow UV-Vis light to pass through. UV-Vis spectra were recorded on a spectrometer (QEPro, Ocean Optics Co., Dunedin, USA) using a light resource (250-2500 nm, DT-MINI2-GS, Ocean Optics Co., Dunedin, USA) and 2 optical fibers placed on both sides of the detection capillary for light output (OP600-2-UV-VIS, Ocean Optics Co., Dunedin, USA) and acquisition of the absorbance signal (OOS-003946-13, Ocean Optics Co., Dunedin, USA), respectively.

In order to verify the feasibility of this online UV-Vis absorption spectroscopy method, the absorption spectra of carmine pigment as a standard sample were compared in a conventional 10 mm cuvette and in the present 250 μm detection capillary, and the result showed a good fit between the two ones at the wavelengths longer than 260 nm (Fig. S3A). However, when the wavelength was less than 260 nm, the absorbance measured in the detection capillary was slightly shifted compared to those measured in the 10 mm cuvette, which may be related to the relatively weak intensity of the light source in the deep UV wavelength range. Considering that most of the organic synthesis mixture solutions have significant absorption at the wavelengths longer than 260 nm, it is feasible to use this kind of capillary as a detection cell for online UV-Vis absorption spectroscopy detection.

Subsequently, we measured the absorption spectra of each component of the photocatalytic [2 + 2] cycloaddition reaction with the detection module, and there was obvious absorption change of the substrate compared to the products (Fig. S3B). Taking substrate S-1 with the corresponding products P-1 as an example, the products had almost no absorption at the measured wavelength of 293.3 nm. Since the amount of photocatalysts would not consume as the reaction proceeds, theoretically the absorption of the photocatalyst should be constant (Fig. S3C). Catalyzed by photocatalyst Ir-1, the absorbance of the reaction mixture at 293.3 nm gradually decreased as the flow rate decreased, which could indicate the consumption of substrate S-1 and the generation of the products P-1 (Fig. S3D). In the actual experiments, all the UV-Vis absorption spectra information at the wavelength range of 200–1100 nm was recorded to retain as much information as possible, and all absorption spectra were smoothed five times to ensure data quality. For the S-2, S-3, S-4 and S-5 systems, the appropriate wavelengths for absorbance observation were also selected, with 293.3 nm for S-2 and S-5, 285.4 nm for S-3 and 310.7 nm for S-4, respectively.

Basic operation and synthesis procedure of the robotic system

Based on the conditions to be screened, an automated program was written with LabView (8.0, National Instruments, Austin, USA) to control the operation of the 10-port selective valve, syringe pump, adjustable power supply for controlling the laser light intensity and the operation of the on-line UV-Vis spectrometer. In the liquid handling module, different substrates and photocatalyst species were selected, and the desired volumes of the stock solutions were aspirated using the 10-port selective valve and the syringe pump to prepare the reaction mixtures according to the set conditions. After fast and homogeneous mixing, the reaction mixture was delivered into the reaction channel of the LCW photocatalytic microreactor module for rapid synthesis. Unless otherwise mentioned, the photocatalytic [2 + 2] cycloaddition reaction was performed in the LCW photocatalytic microreactor at ambient temperature, with automatic control of the reaction temperature (e.g., 25 ± 2 °C) and uniform light irradiation from the 450 nm lasers. After synthesis, the product yields were evaluated based on the absorbance data obtained by the online UV-Vis absorption spectroscopy module. For the representative conditions, and the GC chromatogram data of the reaction mixture after synthesis was measured with a GC system (7890 A, HP-5 column, Agilent Technologies, Santa Clara, USA) with a flame ionization detector (FID). The detailed procedures followed for each compound and characterization are shown in Supplementary Information (Fig. S8).

AI-assisted prediction of the steady-state absorbance data from the non-steady-state absorbance data

We attempted to use AI technique to process the large numbers of the non-steady-state peak signals and predict the corresponding steady-state absorbance data. First, the non-steady-state absorbance data were aggregated into a large dataset with 12,000 samples, each of which contained 48 eigenvalues (8 reaction variables and 40 absorbance values of a non-steady-state signal peek) and 1 standard output value (steady-state absorbance) (Fig. 4A3).

We tried 3 different principal component analysis (PCA) to downscale the original eigenvalues, and normalized the eigenvalues to improve the model performance. In PCA scheme 1, the 40 non-steady-state absorbance values were downscaled to 1 parameter and then combined with the remaining 8 reaction variables to form new 9 eigenvalues. In PCA scheme 2, we directly downscaled the 48 eigenvalues, while in PCA scheme 3, the 48 eigenvalues were directly used without dimensionality reduction. Taking the XGB model as an example, the model trained with eigenvalues form PCA scheme 1 had the best performance (Fig. S4A).

On the basis of the massive amounts of the data of the 12,000 reaction conditions, using the principles of linear models, decision tree, neural networks and integrated learning, we evaluated 10 regression models including partial least squares regression (PLSR), linear regression, adaptive boosting (AdaBoosting) regression, extremely randomized trees (ExtraTree) regression, support vector regression (SVR), k-nearest neighbor (KNN) regression, multi-layer perceptron (MLP) regression, random forest regression, cascade forest regression, and extreme gradient boosting (XGB) regression (Fig. 3A2). Python (Version 3.6.8) and Microsoft Visual Studio Professional 2017 were used as the

development environments. The Python packages including sklearn (Version 0.22), xgboost (Version 1.5.2), and deepforest (Version 0.1.5) were used to train the regression models and to predict the results of the test set. To compare the model performance, the dataset was randomly divided into training and test sets with a proportion of 70% and 30%, respectively. The training set was subjected to k-fold cross-validation, and the hyperparameter combinations of the models were optimized by grid search to improve the model performance. The model performance evaluation metrics RMSE and R^2 were calculated using the corresponding built-in functions within sklearn.metrics.

Using the XGB model which showed the best performance among all 10 AI regression models, we respectively used 10% and 30% of non-steady-state data as the training sets to predict the steady-state absorbance data. Then, all the 12,000 absorbance data (including 5 substrate species, 5 photocatalyst species, 4 substrate concentrations, 4 photocatalyst ratios, 5 reaction flow rates, and 6 laser light intensities) were converted to product yields, and the absolute error between the non-steady-state and steady-state product yields are shown in the heatmaps (10% in Fig. S4B, 30% in Fig. S4C).

Descriptor and algorithms used in AI-assisted cross-species prediction

In order to conduct the cross-substrate and cross-photocatalyst prediction, Mordred descriptor was employed to represent different substrates and photocatalysts, which is a popular choice of molecular descriptor calculation software in cheminformatics studies. Mordred descriptor is a combination of known structural and functional properties, such as topological indices and adjacency matrix, and its robustness and comprehensive descriptor calculation capabilities make it well-suited for our prediction task.

With the SMILES string of each molecule as part of the inputs of the models (Table S1), Mordred could generate both 2D and 3D descriptors, encompassing a list of more than 1800 descriptors. However, such a dimension of the Mordred descriptors was very high compared with other numerical variables (i.e., laser light intensity, flow rate, substrate concentration, photocatalyst ratio, photocatalyst concentration, and absorbance wavelength), which may hamper the predictive capabilities of the model. Given the specificity of our problem, where the number of Mordred descriptors far exceeded those of other numerical variables, we did not perform feature selection. Instead, we reduced the dimension of Mordred descriptor of each substrate and photocatalyst species from more than 1800 to 2 with T-distributed stochastic neighbor embedding (TSNE) algorithm, which is a widely used unsupervised dimension reduction technique owing to its advantage in capturing local data characteristics and revealing subtle data structures. Given the large number of Mordred descriptors, the TSNE algorithm is preferable due to its ability to manage high-dimensional data. The reduced Mordred descriptors were then concatenated with other 6 numerical variables to construct the final reaction fingerprints.

We used the XGB algorithm for cross-substrate and cross-photocatalyst prediction, which was a highly efficient and flexible machine learning algorithm based on the gradient boosting framework. It is renowned for its outstanding performance and high efficiency, which are optimized through parallel processing and tree-pruning. Additionally, the XGB's regularization method could prevent overfitting, thereby improving the model's generalizability.

Data availability

Source data are provided with this paper. All data are available from the corresponding author upon request.

Code availability

The source code and dataset are available at GitHub https://github.com/LJM-1997/NCOMMS-24-15301-T and Zenodo https://doi.org/10.5281/zenodo.13738699[42].

References

1. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, 144–144 (2019).
2. Elliott, L. D. et al. Batch versus flow photochemistry: a revealing comparison of yield and productivity. *Chem Eur. J.* **20**, 15226–15232 (2014).
3. Sagmeister, P. et al. Advanced real-time process analytics for multistep synthesis in continuous flow. *Angew. Chem. Int. Ed.* **60**, 8139–8148 (2021).
4. Harenberg, J. H., Weidmann, N., Karaghiosoff, K. & Knochel, P. Continuous flow sodiation of substituted acrylonitriles, alkenyl sulfides and acrylates. *Angew. Chem. Int. Ed.* **60**, 731–735 (2021).
5. Douglas, J. J., Sevrin, M. J. & Stephenson, C. R. J. Visible light photocatalysis: applications and new disconnections in the synthesis of pharmaceutical agents. *Org. Process Res. Dev.* **20**, 1134–1147 (2016).
6. Tucker, J. W., Zhang, Y., Jamison, T. F. & Stephenson, C. R. J. Visible-light photoredox catalysis in flow. *Angew. Chem. Int. Ed.* **51**, 4144–4147 (2012).
7. Coley, C. W., Abolhasani, M., Lin, H. & Jensen, K. F. Material-efficient microfluidic platform for exploratory studies of visible-light photoredox catalysis. *Angew. Chem. Int. Ed.* **56**, 9847–9850 (2017).
8. Nguyen, J. D., Reiss, B., Dai, C. & Stephenson, C. R. J. Batch to flow deoxygenation using visible light photoredox catalysis. *Chem. Commun.* **49**, 4352–4354 (2013).
9. Davies, I. W. The digitization of organic synthesis. *Nature* **570**, 175–181 (2019).
10. Wilbraham, L., Mehr, S. H. M. & Cronin, L. Digitizing chemistry using the chemical processing unit: from synthesis to discovery. *Acc. Chem. Res.* **54**, 253–262 (2021).
11. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
12. Chan, H. C. S., Shan, H., Dahoun, T., Vogel, H. & Yuan, S. Advancing drug discovery via artificial intelligence. *Trends Pharmacol. Sci.* **40**, 592–604 (2019).
13. Paul, D. et al. Artificial intelligence in drug discovery and development. *Drug Discov. Today* **26**, 80–93 (2020).
14. Jimenez-Luna, J., Grisoni, F., Weskamp, N. & Schneider, G. Artificial intelligence in drug discovery: recent advances and future perspectives. *Expert Opin. Drug Discov.* **16**, 949–959 (2021).
15. Mehr, S. H. M., Craven, M., Leonov, A. I., Keenan, G. & Cronin, L. A universal system for digitization and automatic execution of the chemical synthesis literature. *Science* **370**, 101–108 (2020).
16. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
17. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: generative models for matter engineering. *Science* **361**, 360–365 (2018).
18. Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377-–3381 (2018).
19. Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *Aiche J.* **65**, 466–478 (2019).
20. Bergman, R. G. & Danheiser, R. L. Reproducibility in chemical research. *Angew. Chem. Int. Ed.* **55**, 12548–12549 (2016).
21. Baker, M. Is there a reproducibility crisis? *Nature* **533**, 452–454 (2016).
22. Tshitoyan, V. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
23. Ley, S. V., Fitzpatrick, D. E., Ingham, R. J. & Myers, R. M. Organic synthesis: march of the machines. *Angew. Chem. Int. Ed.* **54**, 3449–3464 (2015).

24. Trobe, M. & Burke, M. D. The molecular industrial revolution: automated synthesis of small molecules. *Angew. Chem. Int. Ed.* **57**, 4192–4214 (2018).
25. Schmink, J. R., Bellomo, A. & Berritt, S. Scientist-led high-throughput experimentation (HTE) and its utility in academia and industry. *Aldrichimica Acta* **46**, 71–80 (2013).
26. Duggan, M. P., McCreedy, T. & Aylott, J. W. A non-invasive analysis method for on-chip spectrophotometric detection using liquid-core waveguiding within a 3D architecture. *Analyst* **128**, 1336–1340 (2003).
27. Pan, J.-Z., Yao, B. & Fang, Q. Hand-held photometer based on liquid-core waveguide absorption detection for nanoliter-scale samples. *Anal. Chem.* **82**, 3394–3398 (2010).
28. Corcoran, E. B., McMullen, J. P., Levesque, F., Wismer, M. K. & Naber, J. R. Photon equivalents as a parameter for scaling photoredox reactions in flow: translation of photocatalytic C-N cross-coupling from lab scale to multikilogram scale. *Angew. Chem. Int. Ed.* **59**, 11964–11968 (2020).
29. Plutschack, M. B., Pieber, B., Gilmore, K. & Seeberger, P. H. The hitchhiker's guide to flow chemistry (II). *Chem. Rev.* **117**, 11796–11893 (2017).
30. Lu, Z. & Yoon, T. P. Visible light photocatalysis of 2+2 styrene cycloadditions by energy transfer. *Angew. Chem. Int. Ed.* **51**, 10329–10332 (2012).
31. Du, J., Skubi, K. L., Schultz, D. M. & Yoon, T. P. A dual-catalysis approach to enantioselective 2+2 photocycloadditions using visible light. *Science* **344**, 392–396 (2014).
32. Kleinmans, R. et al. Intermolecular [2π+2σ]-photocycloaddition enabled by triplet energy transfer. *Nature* **605**, 477–482 (2022).
33. Poplata, S., Troester, A., Zou, Y.-Q. & Bach, T. Recent advances in the synthesis of cyclobutanes by olefin 2+2 photocycloaddition reactions. *Chem. Rev.* **116**, 9748–9815 (2016).
34. Sans, V. & Cronin, L. Towards dial-a-molecule by integrating continuous flow, analytics and self-optimisation. *Chem. Soc. Rev.* **45**, 2032–2043 (2016).
35. Perera, D. et al. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **359**, 429–434 (2018).
36. Gesmundo, N. J. et al. Nanoscale synthesis and affinity ranking. *Nature* **557**, 228–232 (2018).
37. Santanilla, A. B. et al. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. *Science* **347**, 49–53 (2015).
38. Lin, S. et al. Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS. *Science* **361**, eaar6236 (2018).
39. González-Esguevillas, M. et al. Rapid optimization of photoredox reactions for continuous-flow systems using microscale batch technology. *ACS Cent. Sci.* **7**, 1126–1134 (2021).
40. Ruffoni, A. et al. Practical and regioselective amination of arenes using alkyl amines. *Nat. Chem.* **11**, 426–433 (2019).
41. Nicastri, M. C., Lehnherr, D., Lam, Y., DiRocco, D. A. & Rovis, T. Synthesis of sterically hindered primary amines by concurrent tandem photoredox catalysis. *J. Am. Chem. Soc.* **142**, 987–998 (2020).
42. Lu, J.-M. et al. Roboticized AI-assisted microfluidic photocatalytic synthesis and screening up to 10,000 reactions per day. *Zenodo* https://doi.org/10.5281/zenodo.13738699 (2024).

Acknowledgements

The authors acknowledge the financial support provided by the National Ministry of Science and Technology (Grants 2021YFA1301601 to Q.F., 2021YFA1500200 to Z.L. and 2021YFF0701600 to Z.L.), the National Natural Science Foundation of China (Grants 22234007 to Q.F. and 22271249 to Z.L.), the Fundamental Research Funds for the Central Universities (Grants 226-2022-00224 and 226-2023-00115 to Z.L.), the Zhejiang Provincial Natural Science Foundation of China (Grant LDQ24B020001 to Z.L.) and R&D Program of (Grant 2024C03144 to Q.F.), and the iChem-Foundry Platform Cooperation Fund of ZJU-Hangzhou Global Scientific and Technological Innovation Center (to Z.L., J.-Z.P., and Q.F.).

Author contributions

J.-M.L. developed the workflow, designed and carried out experiments, analyzed the data and wrote the manuscript. H.-F.W. contributed the idea of photoreactor and rapid mixing method, wrote and developed the control software. J.-W.W. performed AI prediction methods of absorbance data and developed the absorbance data processing software. Q.-H.G. and T.-T.L. synthesized, isolated, and characterized the compounds, and performed NMR analysis. K.-X.C. performed AI algorithms and methods for cross-species prediction. M.-T.Z. helped to perform the light distribution simulations. J.-B.C. constructed the three-dimensional figure of the photoreactor. Q.-N.S., Y.H., and S.-W.S. helped with GC analysis. G.-Y.C. supervised the AI-assisted cross-species prediction. J.-Z.P. helped to develop the control software and supervised the flow manifold design. Z.L. supervised the photocatalysis synthesis work and secured funding. Q.F. conceived the idea, supervised the project, revised the manuscript, and secured funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53204-6.

Correspondence and requests for materials should be addressed to Guang-Yong Chen, Jian-Zhang Pan, Zhan Lu or Qun Fang.

Peer review information *Nature Communications* thanks Mehrad Ansari and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.