PhotoCat: An Artificial Intelligence-Driven Synthesis Planning Platform for Photocatalysis

Jiangcheng Xu,[#,1,2] An Su[#,3,*], Panyi Huang[#,1], Wenbo Yu[2], Kui Du[4], Zhidan Fan[2], Bin Sun[1], Zhicheng Zhong[1], Can Jin[1,*], Weike Su[1,*]

1. Key Laboratory of Pharmaceutical Engineering of Zhejiang Province, Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou, 310014, P. R. China.

2. Hangzhou Vocational & Technical College, Hangzhou, 310014, P. R. China

3. College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, P. R. China

4. School of Chemistry and Chemical Engineering, Shaoxing University, Shaoxing 312000, P. R. China

# These authors contributed equally.

***Corresponding authors:**

Prof. An Su
College of Chemical Engineering, Zhejiang University of Technology, Hangzhou 310014, P. R. China
Email: ansu@zjut.edu.cn

Prof. Can Jin
Key Laboratory of Pharmaceutical Engineering of Zhejiang Province, Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou, 310014, P. R. China
E-mail: canjin@zjut.edu.cn

Prof. Weike Su
Key Laboratory of Pharmaceutical Engineering of Zhejiang Province, Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou, 310014, P. R. China
Email: pharmlab@zjut.edu.cn

ABSTRACT

Photocatalysis is becoming increasingly important in modern chemistry for efficient multicomponent one-pot synthesis. However, predicting the results of photocatalytic reactions using artificial intelligence remains challenging, mostly due to the insufficient number of photocatalytic reactions and the incomplete information on reaction conditions in existing reaction databases. In this study, we curated the Photocatalysis Database (PhotoCatDB), which consists of 6,523 photocatalytic reactions (of which 6,175 are multicomponent) containing reaction condition information such as photocatalysts, bases or acids, additives, and solvents. Before adding reaction conditions to the training data, the attention-based deep learning model PhotoCat pre-trained on USPTO and fine-tuned on PhotoCatDB had a Top-1 accuracy of 78.16%, which was 77.70% higher than the same model trained only on the USPTO database and 14.53% higher than the model fine-tuned by the photocatalytic reactions from Reaxys. After adding reaction conditions to the training data, the Top-1 accuracy of PhotoCat was further increased to 82.25%. In addition, the interpretability of the model was reflected in its attention weights, which can infer the model's understanding of photocatalytic chemistry. Furthermore, five previously unreported photocatalytic reactions predicted by PhotoCat were successfully validated by wet-lab experiments, demonstrating the potential of the model in identifying and verifying novel photocatalysis reactions of real-world significance.

INTRODUCTION

Recently, the rise and rapid development of photocatalytic technology has created exciting opportunities for organic transformations under mild conditions.[1, 2] At the same time, multicomponent reactions (MCRs) have key attributes that are consistent with the principles of green chemistry, such as atom economy and energy efficiency.[3] Thus, combining multicomponent approaches with photocatalysis opens new avenues for synthetic organic chemistry and meets the growing demand for sustainable and efficient reactions. In this context, multicomponent photocatalysis represents a major breakthrough that inspires chemists to explore uncharted territories and discover elusive reaction patterns.[4, 5] However, conducting multicomponent photocatalysis in the laboratory presents considerable challenges - the discovery and optimization of each novel multicomponent photocatalytic reaction require years of effort by chemists.

Deep learning models have made great strides in recent years, largely due to their remarkable ability to extract knowledge from massive amounts of data.[6-10] In the field of organic synthesis, deep learning has brought about a revolution that has impacted several areas, including forward reaction prediction[11-18], retrosynthesis planning[19-22], mechanistic inference[23, 24], inferring experimental procedures[25, 26], and new reaction development[27-29]. Specifically, deep learning models have proven to be effective in predicting specific types of reactions, including enzyme-catalyzed reactions[30, 31], carbohydrate reactions[32], and electrochemical reactions[33]. In addition, the application of deep learning in the field of photocatalysis has garnered significant attention, particularly in the design and optimization of photocatalysts[34, 35].

However, to the best of our knowledge, no deep learning models specifically targeting photocatalytic reactions have been reported to date, not only because of the challenging nature of exploring photocatalytic reactions[4, 5] that leads to the scarcity of photocatalytic reaction data, but also due to the limitation of currently available reaction databases. Although acknowledged chemical reaction databases like Reaxys (https://www.reaxys.com) can provide photocatalytic reaction data through keyword searches, they may have limitations such as incomplete reaction condition information. For example, the water as a reactant is hidden in the word 'wet', which was missed by the Reaxys dataset (Reaction ID: 49168884). Another initiative to address the database format issue is the Open Reaction Database (ORD) framework proposed by Kearnes et al.[36] ORD displays chemical reaction data in a structured format, which provides strong support for machine learning prediction of chemical reactions. However, ORD currently contains fewer than 300 photocatalytic reactions, which is insufficient to train deep learning models effectively.

In this study, a database of multicomponent photocatalytic reactions, named PhotoCatDB, was established through a comprehensive literature search and these reactions were scrutinized by human experts. Our group also added some experimentally recorded reaction data to the database. Most importantly, we added the essential reaction conditions, such as photocatalysts, bases or acids, additives, and solvents, to PhotoCatDB to reflect real-world reaction scenarios. After organizing PhotoCatDB, we developed PhotoCat, an advanced deep learning model based on the Transformer architecture, specifically designed to predict photocatalytic reactions. We used PhotoCat to successfully identify and experimentally validate five previously unreported photocatalytic reactions of practical significance. This study marks a significant progress in predicting multicomponent photocatalytic reactions and provides a powerful tool to accelerate the discovery and validation of novel photocatalytic reactions with diverse applications, thus advancing the field of green chemistry synthesis.

RESULTS

PhotoCatDB, a multicomponent photocatalytic reaction database with reaction conditions. We tried to collect multicomponent photocatalytic reactions as many as possible to increase the uniqueness of PhotoCatDB. Currently, PhotoCatDB contains 6,523 validated photocatalytic reactions, of which 6,175 are multicomponent reactions. The predominant reaction types are three- and four-component reactions, which accounted for 58% and 33% of the database, respectively (Fig. 1a). Various components of reaction conditions are collected, including 59 photocatalysts, 34 acids or bases, 53 additives (of which 37 are ligands), and 42 solvents. We tried to avoid overcollection of similar photocatalytic reactions, which is reflected by a normal distribution of product molecular weights, ranging from 85 to 1251 (Fig. 1b).

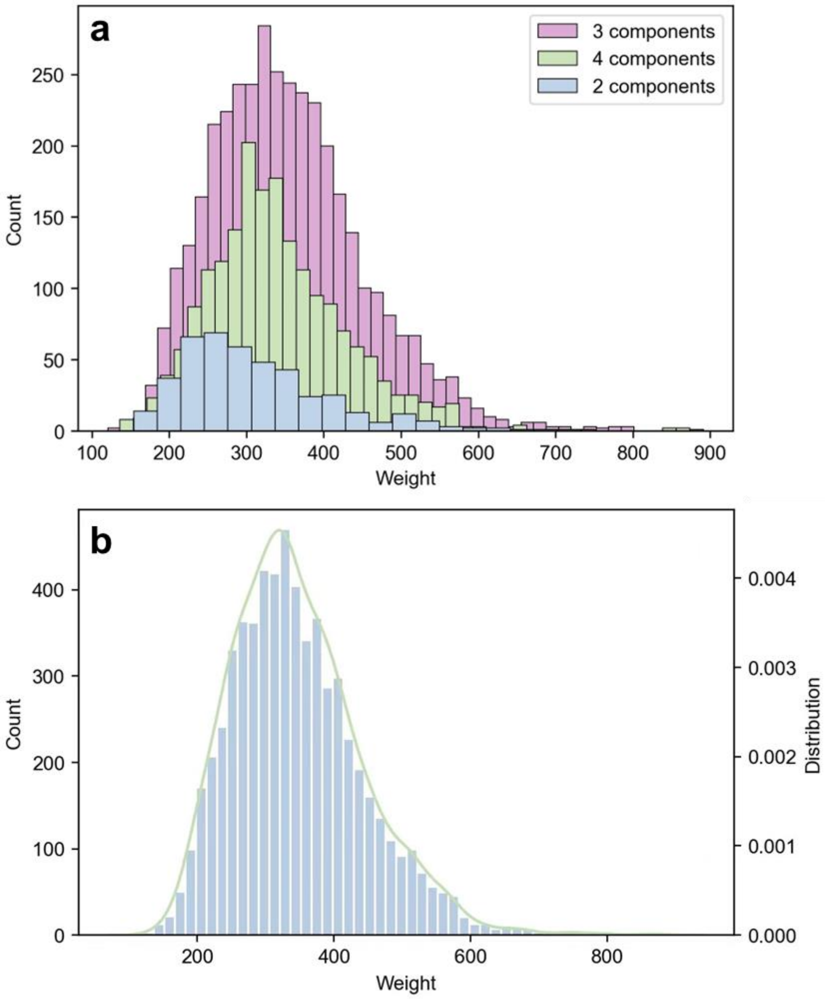


Fig. 1 Data distribution and composition analysis of PhotoCatDB. (a) Multicomponent reactions dominate the dataset, with three- and four-component reactions highlighted in pink and green,

respectively. (b) Molecular weights of reaction products follow a normal distribution.

PhotoCatDB is specifically designed to summarize the unique features of photocatalytic reactions. It is structured into distinct records, each consisting of three key components: 1) Reaction equations represented using the SMILES[37] (Simplified Molecular Input Line Entry System) notation - a specialized system for converting chemical structures into a computer-readable format. In this notation, the ">>" symbol is the dividing line between reactants and products, while the "." symbol delineates different reactants. 2) Reaction conditions classified into four essential elements: photocatalyst, base or acid, additives, and solvent. Certain reactions, such as those involving quinoxalinone, azobenzene, or EDA complex, do not require an external photocatalyst because the reactants themselves are photosensitizers. These reactions are categorized as "autocatalysis". 3) Additional information including reaction time, yield, and literature sources. An example of a PhotoCatDB data instance is provided in Fig. 2, and additional examples are provided in Tables S1-S12.

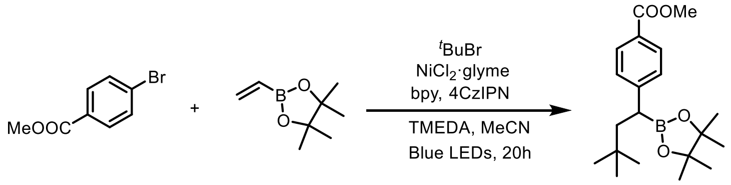
Entry	3251		**Category**	Three-component photocatalysis
Reaction	 Br MeOOC + B O O ᵗBuBr NiCl₂·glyme bpy, 4CzIPN TMEDA, MeCN Blue LEDs, 20h COOMe B O O			
Reaction equation	CC(Br)(C)C.BrC1=CC=C(C(OC)=O)C=C1.C=CB2OC(C)(C)C(C)(C)O2 >>CC(C(C)(C)O3)(C)OB3C(C4=CC=C(C(OC)=O)C=C4)CC(C)(C)C			
Reaction condition	**Photocatalyst**	**Acid or base**	**Additive**	**Solvent**
	4CzIPN	TMEDA	Ni(bpy)Cl_2	MeCN
Additional information	**Reaction time (h)**	**Yield (%)**	**Literature sources**	
	20	73	*Angew. Chem.*, **2020**,*132*, 4400–4404	

Fig. 2 An example of PhotoCatDB data.

For a more insightful analysis of the dataset, TMAP[38], a tree-based unsupervised learning algorithm, was utilized to visualize the chemical reaction mapping of the

PhotoCatDB (Fig. 3a). Despite the wide variety of photocatalysts in the PhotoCatDB, clustering of reactions catalyzed by the same photosensitizer was observed. However, different clustering was observed when reactions were colored according to other types of reaction conditions, such as bases or acids, additives, and solvents (Fig. S2-S4 in ESI), which demonstrates the multivariate effect of reaction conditions and suggests the necessity to include multiple components of reaction conditions in the database. To further demonstrate the advantages of the PhotoCatDB, the photocatalytic reactions collected from Reaxys were also visualized using TMAP[38] (Fig. 3b). PhotoCatDB adopts common names to represent complex photocatalyst structures, making data reading and storage easier. In contrast, photocatalysts in Reaxys were represented with IUPAC names, which makes them more difficult to be classified and understood by AI models. Meanwhile, PhotoCatDB exhibits a higher diversity in reaction types, avoiding the aggregation of similar reactions observed in Reaxys data.

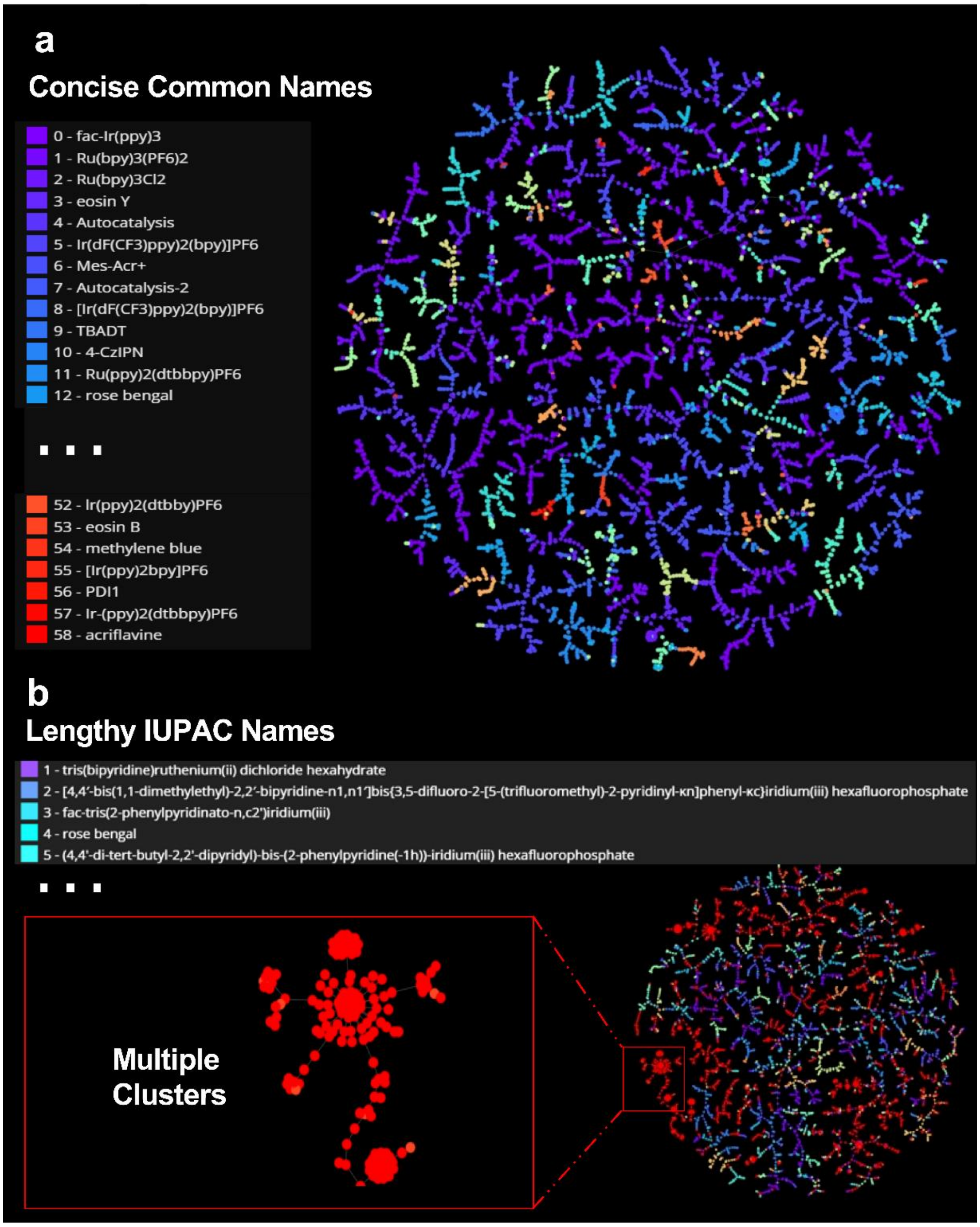
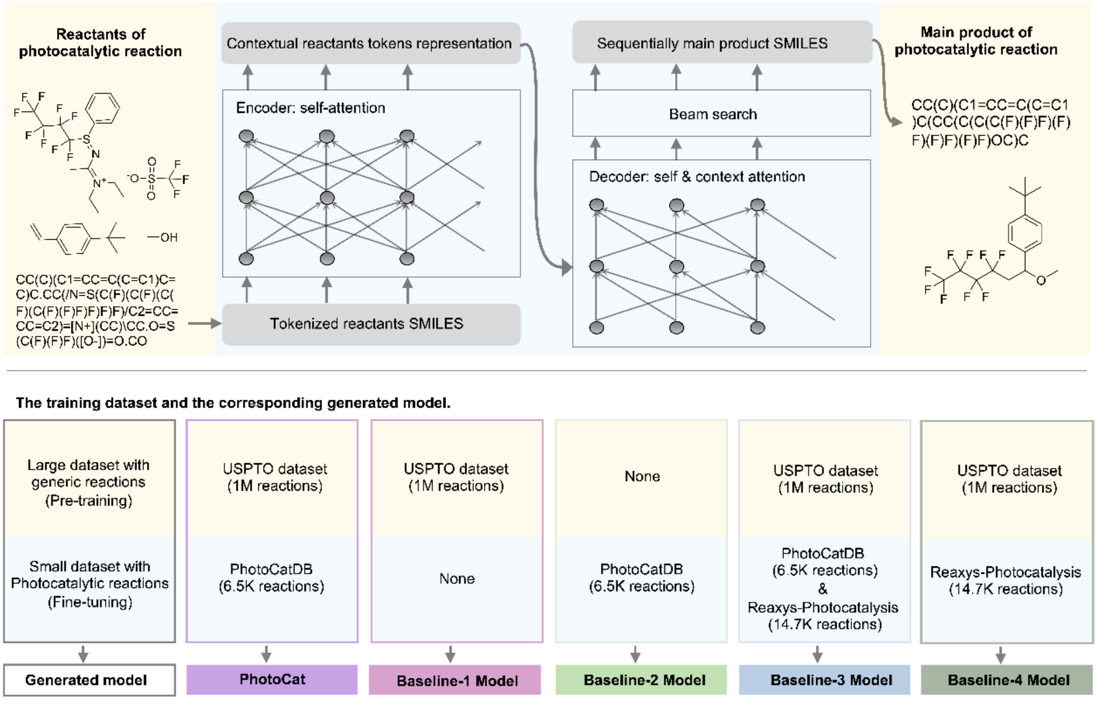


Fig. 3 Comparison of PhotoCatDB and photocatalysis dataset from Reaxys. (a)PhotoCatDB uses common names for photocatalysts, while (b) photocatalysis dataset from Reaxys employs lengthy IUPAC names. Additionally, (a) does not show aggregation of similar reactions as "multiple clusters" as in (b).

PhotoCat, a reaction-prediction model trained with USPTO and PhotoCatDB. PhotoCat uses the Transformer model from the OpenNMT framework[39] as the basis of its model architecture (Fig.4). First, the model was pre-trained using the USPTO database containing 1 million instances of chemical reactions to gain basic knowledge of chemical reactions. After that, the model was fine-tuned using PhotoCatDB. To demonstrate the need for transfer learning, we let the Baseline-1 model be trained only by USPTO and the Baseline-2 model be trained only by PhotoCatDB. In addition, we introduced photocatalytic reaction data from Reaxys (named Reaxys-photocatalysis) to investigate the effect of fine-tuning datasets. We fine-tuned the USPTO-pretrained model using the combined dataset of Reaxys-photocatalysis and PhotoCatDB (Baseline-3) and only the Reaxys-photocatalysis dataset (Baseline-4), respectively. It is important to note that to fairly compare the role of the reactions in USPTO and PhotoCatDB in model training, the PhotoCatDB data for training in this section does not contain any reaction conditions. Training using the complete PhotoCatDB data containing reaction conditions is discussed in the next section.


Reactants of photocatalytic reaction
F
F
F
F
F
F
F
F
S
N
N+
O
O
-O-S-
F
F
F
—OH
CC(C)(C1=CC=C(C=C1)C=C)C.CC(/N=S(C(F)(C(F)(C(F)(C(F)(F)F)F)F)F)/C2=CC=CC=C2)=[N+](CC)\CC.O=S(C(F)(F)F)([O-])=O.CO
Contextual reactants tokens representation
Encoder: self-attention
Tokenized reactants SMILES
Sequentially main product SMILES
Beam search
Decoder: self &amp; context attention
Main product of photocatalytic reaction
CC(C)(C1=CC=C(C=C1)C(CC(C(C(C(F)(F)F)(F)F)(F)F)(F)F)OC)C
F
F
F
F
F
F
F
F
O
The training dataset and the corresponding generated model.

Large dataset with generic reactions (Pre-training)	USPTO dataset (1M reactions)	USPTO dataset (1M reactions)	None	USPTO dataset (1M reactions)	USPTO dataset (1M reactions)
Small dataset with Photocatalytic reactions (Fine-tuning)	PhotoCatDB (6.5K reactions)	None	PhotoCatDB (6.5K reactions)	PhotoCatDB (6.5K reactions) & Reaxys-Photocatalysis (14.7K reactions)	Reaxys-Photocatalysis (14.7K reactions)
Generated model	PhotoCat	Baseline-1 Model	Baseline-2 Model	Baseline-3 Model	Baseline-4 Model

Fig. 4 The Transformer model and training schematic. Sequence-2-sequence prediction of photocatalytic reactions (Upper), and the training schematic of PhotoCat and four baseline models (Lower).

As is shown in Figure 5a, the average Top-1 accuracy of the pre-trained and fine-tuned PhotoCat was 78.16%, and the Top-2 to Top-5 accuracies were all above 84%. In contrast, the prediction accuracy of the Baseline-1 model trained only on USPTO was significantly lower, with an average Top-1 accuracy of only 0.46% and similar low

Top-2 to Top-5 accuracies. In addition, the Baseline-2 model, trained solely on PhotoCatDB, also exhibited lower prediction accuracies than PhotoCat. Fig. 5b provides an alternative perspective on this comparison. Figure 5b depicts the proportion of invalid SMILES predicted by PhotoCat (0.80%), which was lower than that of the Baseline 1 and 2 models. The results above suggest that the USPTO provides the necessary information for the model to understand chemical reactions, while PhotoCatDB provides the model with the ability to target photocatalytic reactions.

Figure 5a shows that the Top-1 accuracy of Baseline-3, which was fine-tuned by combining PhotoCatDB and Reaxys-photocatalysis, is about 3% lower than that of PhotoCat, suggesting that expanding the fine-tuning set does not improve the predictive performance of the model. On the other hand, the Top-1 accuracy of Baseline-4 fine-tuned only with Reaxys-photocatalysis was 14.53% lower than that of PhotoCat. Based on the results of Baseline-3 and Baseline-4, we can conclude that the quality of Reaxys-photocatalysis is lower than that of PhotoCat. This lower quality not only leads to lower prediction accuracy when training with this dataset alone, but also has a negative impact when used in combination with higher-quality datasets.

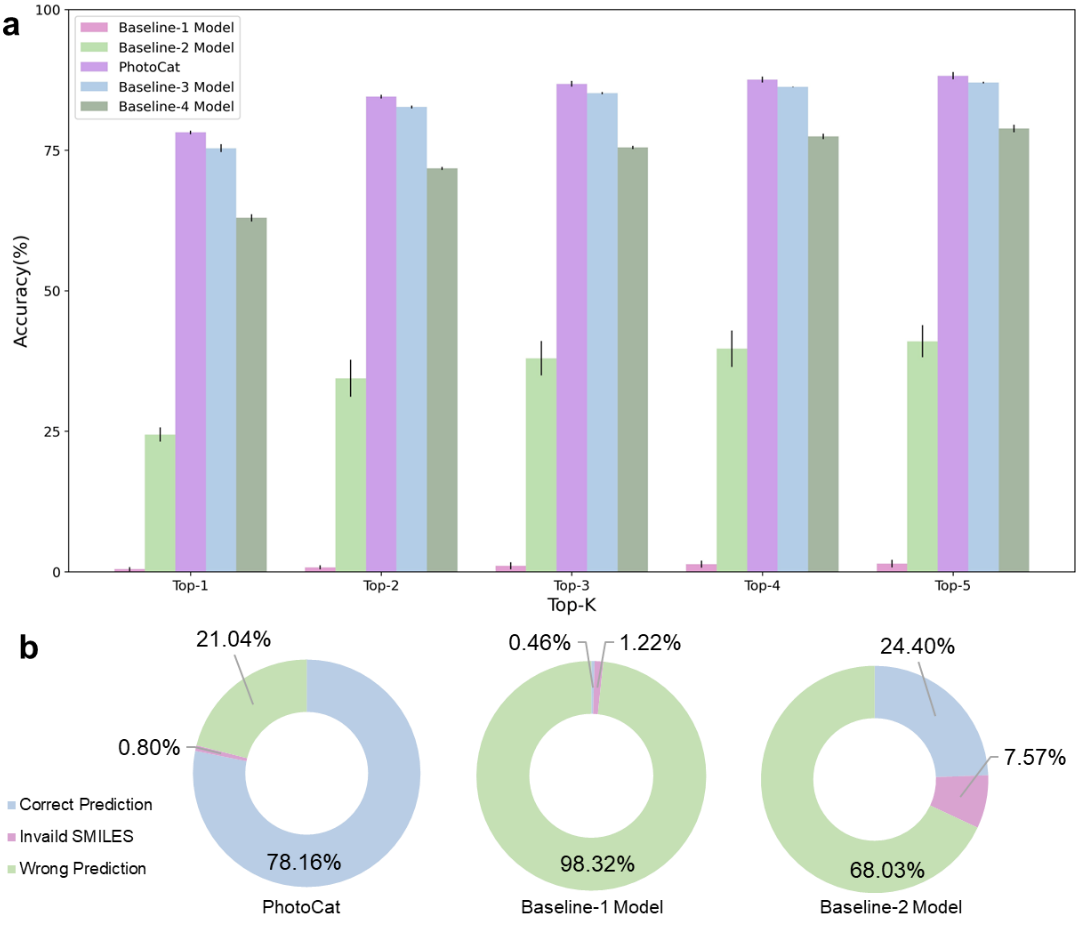


Fig. 5 Comparison of prediction accuracies of PhotoCat and baseline models. (a) The comparison of Top-1 to Top-5 accuracies; (b) Percentage of Invalid SMILES and incorrect predictions in Top-1 accuracies for Baseline-1 and -2 models and PhotoCat.

Reaction conditions improve accuracy and efficiency of reaction prediction. After incorporating detailed chemical reaction conditions into the fine-tuning phase of PhotoCat, the Top-1 accuracy was significantly improved by 4.01% to 82.25% (Fig. 6). We further investigated the effect of the number of reaction conditions on prediction accuracy. The reaction conditions started with only photocatalysts, and then bases or acids, additives, and solvents were added sequentially. As the number of input reaction conditions increased, the accuracy of reaction predictions continued to rise (Fig. 6a). PhotoCat's predicted Top-1 accuracies were 78.1%, 79.8%, and 81.9% when one, two, and three reaction conditions were included in the training, respectively, which were all higher than fine-tuned model that did not include information about the reaction conditions.

Integrating reaction conditions also speeds up the training process. Fig. 6 shows that the training curve grows faster for the models with more reaction conditions provided in their training. With only 0 reaction conditions (blue), the accuracy reaches about 76% after 40,000 training steps. In contrast, when all four reaction conditions are considered (purple), the training accuracy grows faster, reaching over 80% after only 30,000 steps. Meanwhile, in both the validation set (Fig. S11) and test set (Fig. S12-13), the Top-K accuracy (with K values ranging from 2 to 5) grows in a similar trend as the Top-1 accuracy. These observations demonstrated the importance of incorporating detailed and comprehensive reaction conditions during the training and prediction phases, which promotes superior model performance and faster convergence to peak prediction accuracy.

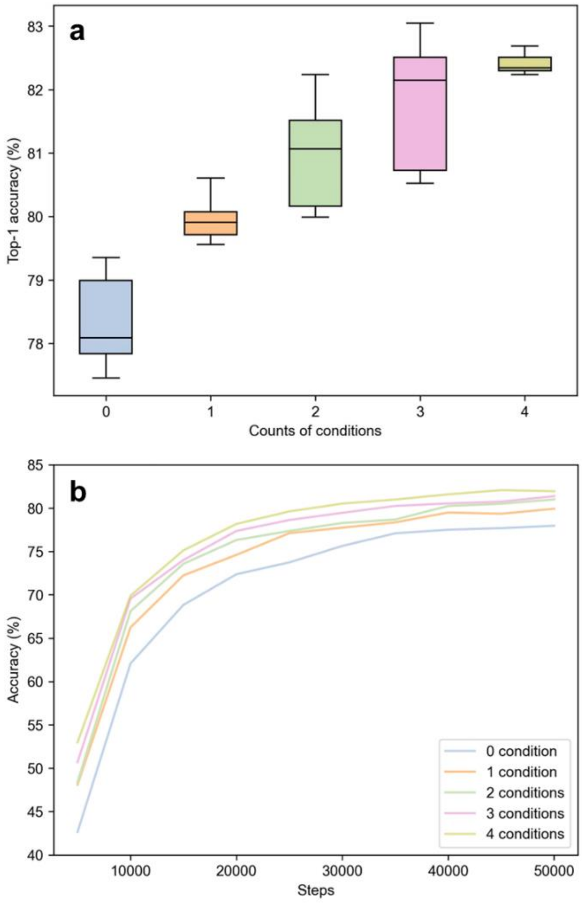
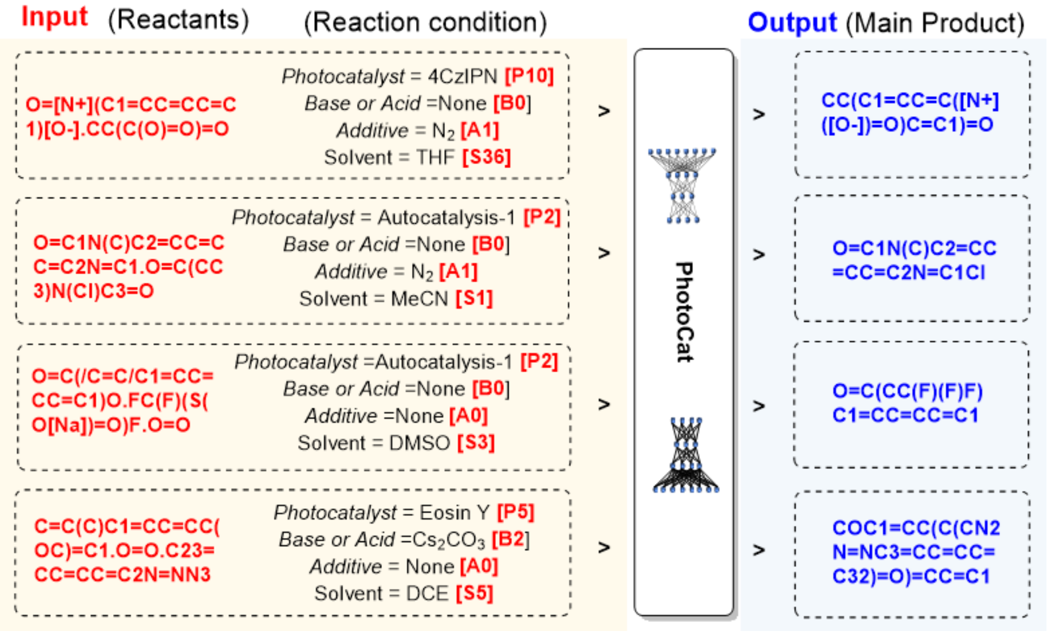


Fig. 6 Effects of reaction condition inputs on model predictions and training Efficiency. (a) As the quantity of input reaction conditions grows, there's a consistent ascent in prediction accuracy. (b) The adaptive inclusion of various reaction condition inputs significantly amplifies the model's training agility.

Photocatalytic synthesis planning with PhotoCat. In this section, we demonstrate how PhotoCat can assist the synthetic planning of photocatalytic reactions. The first step is to use PhotoCat to screen the human-designed reactions through a dry experiment, where the reactants and reaction conditions of these reactions were input to PhotoCat (Fig. 7). Human experts in our group designed six photocatalytic reactions (Fig. 8), which had not reported in literature or validated by wet-lab experiments. Of the main products predicted by PhotoCat, reactions **a-d** were consistent with those anticipated by human experts, reaction **e** was inconsistent, and the main product of reaction **f** had an invalid SMILES.



Input (Reactants) (Reaction condition)
Output (Main Product)
O=[N+](C1=CC=CC=C1)[O-].CC(C(O)=O)=O
Photocatalyst = 4CzIPN [P10]
Base or Acid =None [B0]
Additive = N2 [A1]
Solvent = THF [S36]
O=C1N(C)C2=CC=CC=C2N=C1.O=C(CC3)N(Cl)C3=O
Photocatalyst = Autocatalysis-1 [P2]
Base or Acid =None [B0]
Additive = N2 [A1]
Solvent = MeCN [S1]
O=C(/C=C/C1=CC=CC=C1)O.FC(F)(S(O[Na])=O)F.O=O
Photocatalyst =Autocatalysis-1 [P2]
Base or Acid =None [B0]
Additive =None [A0]
Solvent = DMSO [S3]
C=C(C)C1=CC=CC(OC)=C1.O=O.C23=CC=CC=C2N=NN3
Photocatalyst = Eosin Y [P5]
Base or Acid =Cs2CO3 [B2]
Additive = None [A0]
Solvent = DCE [S5]
>
PhotoCat
CC(C1=CC=C([N+]([O-])=O)C=C1)=O
O=C1N(C)C2=CC=CC=C2N=C1Cl
O=C(CC(F)(F)F)C1=CC=CC=C1
COC1=CC(C(CN2N=NC3=CC=CC=C32)=O)=CC=C1

Fig. 7 The process of conducting dry experiments using PhotoCat

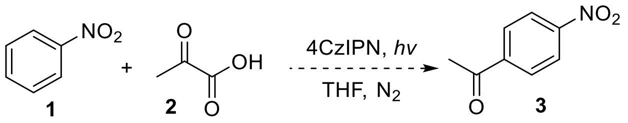
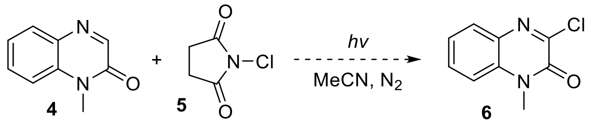
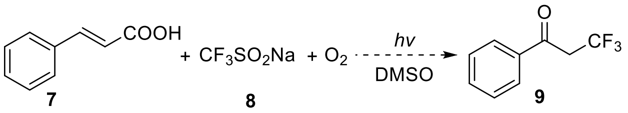
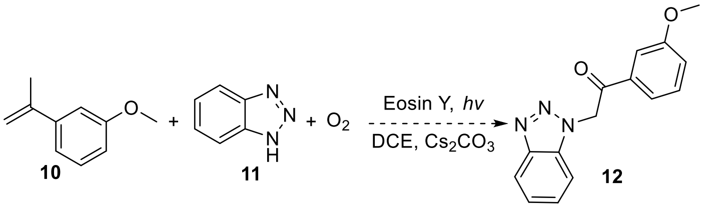
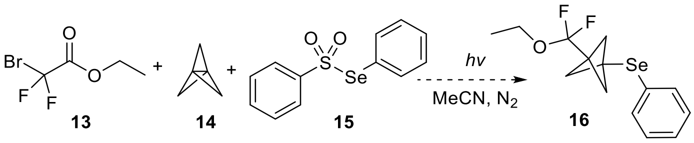
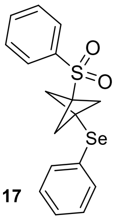
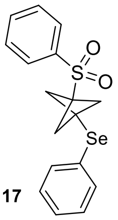
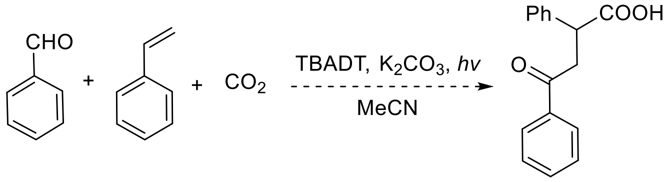
Entry	Chemists' designed photocatalytic reactions and predicted main product structures	Main product structure	
		Predicted by PhotoCat	Repeatable by wet-lab experiment
a	 NO2 O OH 1 2 O + 4CzIPN, *hv* THF, N2 NO2 O 3	Correct	Correct
b	 N N O 4 + O N−Cl O 5 *hv* MeCN, N2 N Cl N O 6	Correct	Correct
c	 COOH 7 + CF3SO2Na 8 + O2 *hv* DMSO O CF3 9	Correct	Correct
d	 O 10 + N N N H 11 + O2 Eosin Y, *hv* DCE, Cs2CO3 O O N N N 12	Correct	Correct
e	 O Br O F F 13 + 14 + O O S Se 15 *hv* MeCN, N2 O F F Se 16		
f	 CHO + + CO2 TBADT, K2CO3, *hv* MeCN Ph COOH O	Invalid SMILES	N.D.

Fig. 8 The dry experimental outcomes of 6 photocatalytic reactions (including 4 multi-component photocatalytic reactions) that remain unverified through practical experimentation, as determined using the PhotoCat.

Wet-lab experiments were conducted to further investigate these reactions (Fig. 8). The results show that reactions a-d, unanimously recognized by PhotoCat and human experts, generated the expected products with high yields in the laboratory. Reaction **e** generated the main product **17**, which was consistent with PhotoCat's prediction, rather than product **16** expected by human experts. Reaction **f** that failed in the PhotoCat remained unsuccessful despite extensive exploration by the three chemists for a month, unable to obtain the desired products.

In addition to validating PhotoCat's predictions, the five successfully completed reactions are of great scientific importance to the field of photocatalysis. In reaction **a**, pyruvate **1** and nitrobenzene **2** combine under photocatalytic conditions to produce aromatic ketone **3**. This reaction is reminiscent of the Friedel-Crafts acylation reaction. The introduction of strong electron-withdrawing groups, like the nitro group, desensitizes the benzene ring, making the execution of the Friedel-Crafts acylation reaction more challenging.[40] Furthermore, a significant advantage of reaction **a** is its ability to sidestep the customary reliance on large amounts of Lewis acids, such as $AlCl_3$, typically required in traditional Friedel-Crafts acylation procedures. In reaction **b**, using N-methyl quinoxalinone **4** and NCS **5**, the 3-chlorinated product **6** is obtained without the need for an external photocatalyst. While recent literature has described photocatalytic synthesis methods for the production of compound **6**, reaction **b** introduces innovative approaches, expanding the chemist's repertoire with additional options.[41] In reaction **c**, the photocatalytic synthesis of α-trifluoromethyl-substituted ketone **7** is achieved using cinnamic acid **8** and CF_3SO_2Na **9**, without the necessity of metals, external oxidants, or photocatalysts.[42] To the best of our knowledge, this represents the inaugural documentation of an oxidative decarboxylative trifluoromethylation of α,β-unsaturated carboxylic acid employing cost-effective CF_3SO_2Na via any photocatalytic approach.[43] In reaction **d**, a photo-triggered oxo-amination of an inactivated alkene **10** is developed, leading to the synthesis of α-amino ketones **12.**[44] This strategy showcases a vicinal heterodifunctionalization of widely accessible olefin feedstock, enabling the construction of the target product in a single step. In reaction **e**, bicyclo[1.1.1]pentane (BCP), a nonclassical bioisosteric replacement of aryl[45] and internal alkynes **17** was achieved in good yields in the absence of any additives. Detailed experimental procedures and mechanism derivation can be found in the ESI, Fig. S18 to S25.

DISCUSSION

Unlike previous studies that considered only a few commonly used reagents[11-16] or ignored reaction conditions altogether[17,18], this study shows that the inclusion of reaction conditions can improve the predictive accuracy and training efficiency of chemical reaction prediction models. There are two main reasons for this improvement. First, in previous studies that considered reaction conditions, the reactants were input simultaneously as mixtures, which may increase the challenge of the model in interpreting the reaction conditions. In this study, the photocatalytic reaction conditions were methodically grouped into four different types, which significantly reduced the complexity of the model in grasping the intricate chemical reaction conditions. In addition, the terminology of the reaction conditions was simplified by using concise common names or identifiers, which establishes a direct correspondence between the structure of the reagents and their names (Fig. 9) and simplifies model learning.

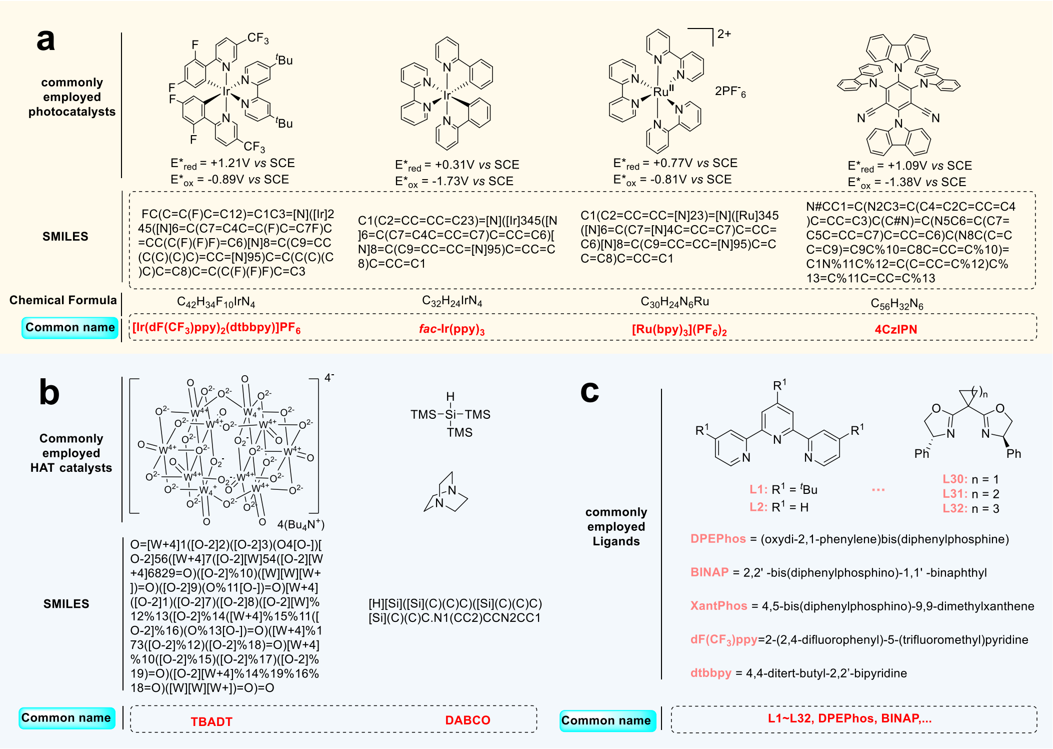

a
commonly employed photocatalysts
E*red = +1.21V vs SCE
E*ox = -0.89V vs SCE
E*red = +0.31V vs SCE
E*ox = -1.73V vs SCE
E*red = +0.77V vs SCE
E*ox = -0.81V vs SCE
2PF-6
E*red = +1.09V vs SCE
E*ox = -1.38V vs SCE
SMILES
FC(C=C(F)C=C12)=C1C3=[N]([Ir]245([N]6=C(C7=C4C=C(F)C=C7F)C=CC(C(F)(F)F)=C6)[N]8=C(C9=CC(C(C)(C)C)=CC=[N]95)C=C(C(C)(C)C)C=C8)C=C(C(F)(F)F)C=C3
C1(C2=CC=CC=C23)=[N]([Ir]345([N]6=C(C7=C4C=CC=C7)C=CC=C6)[N]8=C(C9=CC=CC=[N]95)C=CC=C8)C=CC=C1
C1(C2=CC=CC=[N]23)=[N]([Ru]345([N]6=C(C7=[N]4C=CC=C7)C=CC=C6)[N]8=C(C9=CC=CC=[N]95)C=CC=C8)C=CC=C1
N#CC1=C(N2C3=C(C4=C2C=CC=C4)C=CC=C3)C(C#N)=C(N5C6=C(C7=C5C=CC=C7)C=CC=C6)C(N8C(C=CC=C9)=C9C%10=C8C=CC=C%10)=C1N%11C%12=C(C=CC=C%12)C%13=C%11C=CC=C%13
Chemical Formula
C42H34F10IrN4
C32H24IrN4
C30H24N6Ru
C56H32N6
Common name
[Ir(dF(CF3)ppy)2(dtbbpy)]PF6
fac-Ir(ppy)3
[Ru(bpy)3](PF6)2
4CzIPN
b
Commonly employed HAT catalysts
4-
4(Bu4N+)
TMS−Si−TMS
SMILES
O=[W+4]1([O-2]2)([O-2]3)(O4[O-])[O-2]56([W+4]7([O-2][W]54([O-2][W+4]6829=O)([O-2]%10)([W][W][W+])=O)([O-2]9)(O%11[O-])=O)[W+4]([O-2]1)([O-2]7)([O-2]8)([O-2][W]%12%13([O-2]%14([W+4]%15%11([O-2]%16)(O%13[O-])=O)([W+4]%173([O-2]%12)([O-2]%18)=O)[W+4]%10([O-2]%15)([O-2]%17)([O-2]%19)=O)([O-2][W+4]%14%19%16%18=O)([W][W][W+])=O)=O
[H][Si]([Si](C)(C)C)([Si](C)(C)C)[Si](C)(C)C.N1(CC2)CCN2CC1
Common name
TBADT
DABCO
c
commonly employed Ligands
L1: R1 = tBu
L2: R1 = H
L30: n = 1
L31: n = 2
L32: n = 3
DPEPhos = (oxydi-2,1-phenylene)bis(diphenylphosphine)
BINAP = 2,2' -bis(diphenylphosphino)-1,1' -binaphthyl
XantPhos = 4,5-bis(diphenylphosphino)-9,9-dimethylxanthene
dF(CF3)ppy=2-(2,4-difluorophenyl)-5-(trifluoromethyl)pyridine
dtbbpy = 4,4-ditert-butyl-2,2'-bipyridine
Common name
L1~L32, DPEPhos, BINAP,...

Fig. 9 Simplification strategy for describing reaction conditions in PhotoCatDB. (a) Given the limited dataset of 59 photosensitizers, extensive SMILES notations are circumvented in favor of deep learning. Common names are employed to effectively denote photocatalysts, capturing intricate conjugate structures with central metals. (b) Molecular structures of widely-used hydrogen atom transfer (HAT) catalysts can often be more intricate than those of photocatalysts. To achieve

concise representation, common names are utilized. (c) Since ligands frequently feature chiral configurations, SMILES notations are not ideal. Instead, common names are favored for depicting ligand structures. Some ligands are abbreviated with labels (L1 to L32; refer to Fig S6 for comprehensive details), and prevalent ligands retain their common names.

Fig. 10 illustrates the improvement of model interpretability by adding reaction conditions to model training.[46] When the reactants (styrene **18**, DMSO **19**, NHP ester **20**) are subjected to a trivalent ruthenium photocatalyst and light, ketone **21** is the main product in the absence of acid. However, when a strong acid (e.g., trifluoromethanesulfonic acid) is present, the reaction shifts to alkene **22** as the main product (Fig. 10a). A high concentration of protons prevents the deprotonation of the alkoxysulfonium intermediate **23**, thus hindering the formation of ketone **21** (Fig. 10b). Based on complete reaction conditions, PhotoCat precisely predicts the main product (Fig. 10c). In this case, [B0] indicates the absence of “base or acid”, while [B31] represents the inclusion of trifluoromethanesulfonic acid in the reaction conditions. In the attention heatmap, the thickness of the lines represents the attention weight between the input and output. Fig. 10d-e show that PhotoCat clearly focuses on the input reaction conditions (highlighted in purple) when predicting the main product. Notably, PhotoCat gives higher attention to [B0] when outputting the key ketone carbonyl group “C=O” (Fig. 10d), and [B31] when projecting the critical alkene group “C=C” (Fig. 10e). This is consistent with human experts’ approach in determining the main products of photocatalytic reactions based on reaction conditions.

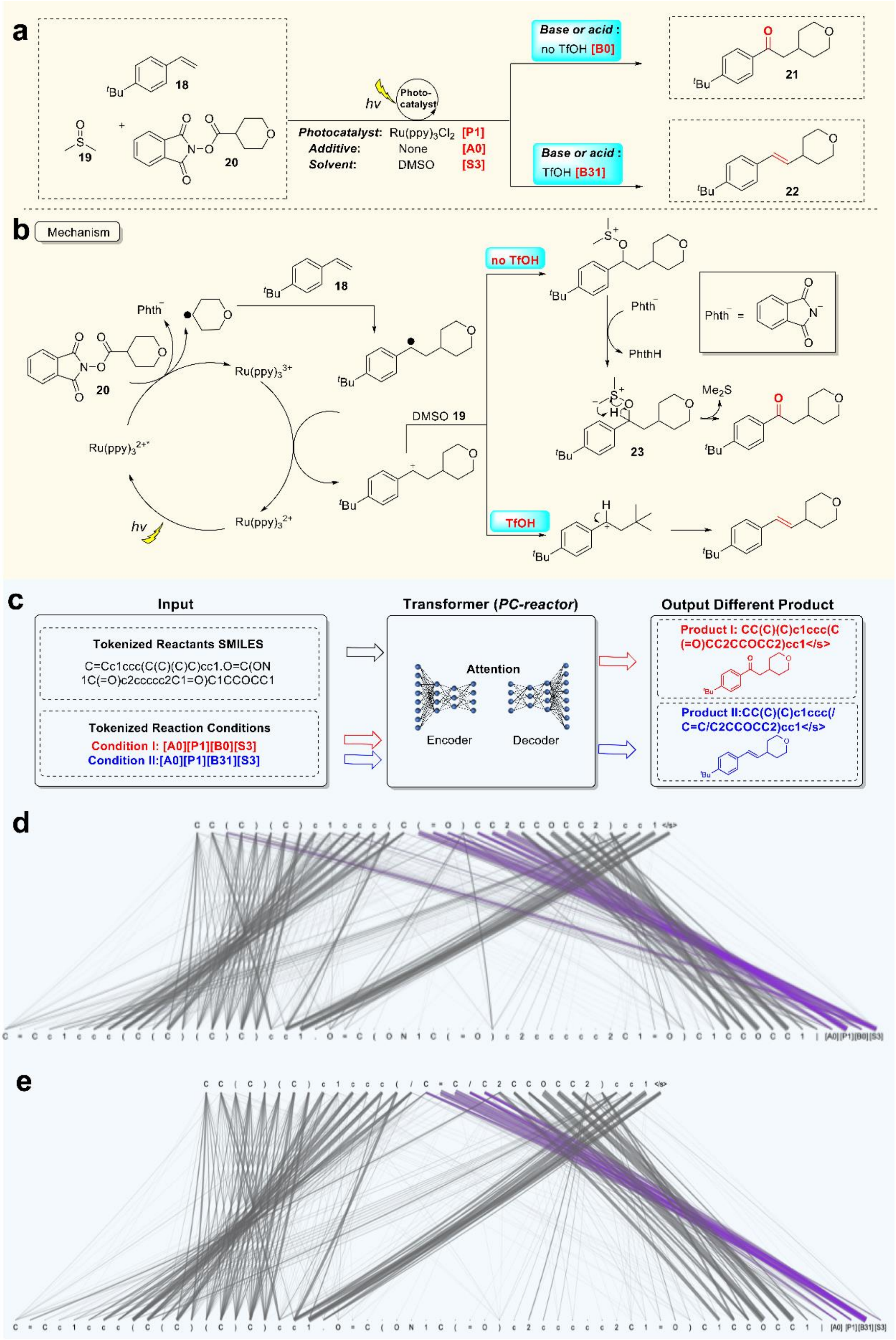

a
Base or acid : no TfOH [B0]
Base or acid : TfOH [B31]
18
19
20
21
22
hv
Photo-catalyst
Photocatalyst: Ru(ppy)3Cl2 [P1]
Additive: None [A0]
Solvent: DMSO [S3]
b
Mechanism
no TfOH
TfOH
Phth⁻
PhthH
Phth⁻ =
Ru(ppy)3 3+
Ru(ppy)3 2+*
Ru(ppy)3 2+
DMSO 19
Me2S
23
c
Input
Tokenized Reactants SMILES
C=Cc1ccc(C(C)(C)C)cc1.O=C(ON1C(=O)c2ccccc2C1=O)C1CCOCC1
Tokenized Reaction Conditions
Condition I: [A0][P1][B0][S3]
Condition II:[A0][P1][B31][S3]
Transformer (PC-reactor)
Attention
Encoder
Decoder
Output Different Product
Product I: CC(C)(C)c1ccc(C(=O)CC2CCOCC2)cc1</s>
Product II:CC(C)(C)c1ccc(/C=C/C2CCOCC2)cc1</s>
d
e

Fig. 10 Interpretability and attentional analysis of PhotoCat. (a) Presence or absence of trifluoromethanesulfonic acid in the reaction system dictates the main product outcome[46], producing aldehyde **21** or alkene **22**. A plausible reaction mechanism is presented in (b). (c) PhotoCat accurately predicts main products when provided with corresponding reaction conditions ([B0] and

[B31] respectively refer to “base or acid” being “none” and “TfOH”). In the attention heatmaps (d) and (e), the upper strings represent the SMILES of the main product from the photocatalytic reaction (output), while the lower strings denote the SMILES of the reactants and the input of the four reaction conditions. The output of the product's SMILES places a particular emphasis on the input of the four reaction conditions (highlighted in purple). Notably, when the key functional group of ketone carbonyl “C(=O)” (d) or alkene “C=C” (e) is outputted, PhotoCat pays special attention to the corresponding inputs of [B0] and [B31], respectively.

CONCLUSION

This study introduces a novel Transformer-based deep learning model, PhotoCat, which is designed to predict photocatalytic reactions. The model was trained by PhotoCatDB, a new photocatalysis database curated by our group, and achieved a top-1 accuracy of 78.16% in predicting complex multicomponent photocatalytic reactions. This accuracy was further improved to 82.25% by introducing the reaction conditions that were properly simplified and classified in PhotoCatDB. In addition, the analysis of the model's attention weights is consistent with chemists’ assessments, demonstrating the interpretability of the model. Most importantly, five previously unreported photocatalytic reactions were successfully predicted by PhotoCat and subsequently validated by wet-lab experiments. This study highlights the impact of reaction conditions on the reaction prediction using deep learning models. PhotoCat provides chemists with a predictive tool for photocatalytic reactions, facilitating the application of photocatalysis as a green chemical synthesis technique. Meanwhile, this research broadens the scope of deep learning applications in chemical synthesis, providing insightful discussions for scientists engaged in the development of scientific databases.

METHODS

PhotoCatDB. The PhotoCatDB is based on a review of multicomponent photocatalytic reactions and our group's experience in photocatalytic reactions[47-51]. All the records included in the dataset are sourced exclusively from published literature. Data from pre-printed versions of papers or papers deemed subjectively unreasonable were deliberately excluded. We assembled a team of 15 data collectors who extracted reactions from literature sources. By analyzing the mechanisms, they organized the information into three main categories: reaction equations (expressed using SMILES, standardization of all reactions was performed using RDKit[52].), reaction conditions, and additional details. After cross-checking, the collected data was incorporated into PhotoCatDB.

Reaxys-Photocatalysis. The dataset was compiled from photocatalysis reactions obtained from Reaxys[53] by directly querying the keywords “visible light induced” in the search field. This process resulted in a dataset of approximately 87K photocatalytic reactions. To ensure the reliability of the dataset for model testing, a restriction was imposed, including only reactions with “irradiation” specified in the reaction conditions, and duplicates and erroneous reactions were removed. This refining process led to the creation of a dataset contains approximately 14.7K photocatalytic reactions.

USPTO dataset. The reactions were originally from Lowe’s dataset, extracted from patents filed in the United States Patent and Trademark Office (USPTO). The dataset was preprocessed by removing reagents, solvent, temperature, and other reaction conditions, and subsequently filtered to eliminate duplicate, incorrect, and incomplete reactions.

Transformer model. Transformer model proposed by Schwaller et al.[14] was used. The model's backend deep learning language of choice was PyTorch. Hyperparameters are as follows:

Both encoder and decoder with 6 layers.

Word vectors, and RNN of size = 512.

the gradient was accumulated 8 times (maximum vector norm of 0.0).

optimizer = adam (β1 = 0.9, β2 = 0.998).

Batch size = 4096, the batch type and the gradient normalization method were tokens,

Learning rate = 2.0.

decay method = noam.

Dropout and label smoothing (ε) = 0.1.

Parameter initialization was disabled, and position encoding was enabled.

Transfer learning. Multi-task transfer learning was implemented using a convex weighting scheme for the USPTO and fine-tuning dataset, with weights of 9 and 1, respectively, as described by Pesciullesi[32].

Cross-validation. Cross-validation, a widely adopted machine learning evaluation technique, assesses a model's performance and generalization ability by dividing the dataset into exclusive subsets for training and validation. It effectively tackles overfitting and underfitting concerns while offering a comprehensive understanding of the model's real-world performance. In this paper, all models were constructed and evaluated using 5-fold cross-validation, ensuring the robustness of the results.

Chemical synthesis. Reaction **a**: A mixture of nitrobenzene **1** (0.2 mmol), pyruvic acid **2** (0.4 mmol), and THF (2 mL) were added to a reaction tube. The tube was evacuated and backfilled with N_2 for three times. The mixture was then irradiated by 360–365 nm (10 w) for 24 h. After completion of the reaction, the resulting mixture was extracted with CH_2Cl_2, and the organic phase was then removed under vacuum. The residue was purified by column chromatography using a mixture of petroleum ether and ethyl acetate as eluent to give the desired product **3** with 70% yield.

Reaction **b**: A mixture of 1-methylquinoxalin-2(1H)-one **4** (0.2 mmol), NCS **5** (0.4 mmol), and MeCN (2 mL) were added to a reaction tube. The tube was evacuated and backfilled with N_2 for three times. The mixture was then irradiated by blue light for 24 h. After completion of the reaction, the resulting mixture was extracted with CH_2Cl_2, and the organic phase was then removed under vacuum. The residue was purified by column chromatography using a mixture of petroleum ether and ethyl acetate as eluent to give the desired product **6** with 71% yield.

Reaction **c**: A mixture of cinnamic acid **7** (0.2 mmol), CF_3SO_2Na **8** (0.4 mmol), and DMSO (2 mL) were added to a reaction tube. The reaction mixture was opened to the air and stirred at room temperature under the irradiation of purple light for 5 h. After completion of the reaction, the resulting mixture was extracted with CH_2Cl_2, and the organic phase was then removed under vacuum. The residue was purified by column chromatography using a mixture of petroleum ether and ethyl acetate as eluent to give the desired product **9** with 75% yield.

Reaction **d**: In an oven-dried reaction tube equipped with a magnetic stirrer bar was charged with α-methylstyrene **10** (0.9 mmol), benzotriazole **11** (0.3 mmol), caesium carbonate (0.9 mmol), Eosin Y (3.0 mol %) and DCE (2.0 mL). The tube was then exposed to blue LEDs irradiation at room temperature under O_2 atmosphere with stirring for 36 h. After completion of the reaction, the resulting mixture was extracted with CH_2Cl_2, and the organic phase was then removed under vacuum. The residue was purified by column chromatography using a mixture of petroleum ether and ethyl acetate as eluent to give the desired product **12** with a 63% yield.

Reaction **e**: A mixture of ethyl bromodifluoroacetate **13** (0.2 mmol), bicyclo[1.1.1]pentane **14** (0.4 mmol), selenosulfonate **15** (0.2 mmol) and MeCN (2 mL) were added to a reaction tube. The tube was evacuated and backfilled with N_2 for three times. The mixture was then irradiated by 400-405 nm for 24 h. After completion of the reaction, the resulting mixture was extracted with CH_2Cl_2, and the organic phase was then removed under vacuum. The residue was purified by column chromatography using a mixture of petroleum ether and ethyl acetate as eluent to give the product **17** with 67% yield.

Unless otherwise specified, all reagents and solvents were obtained from commercial suppliers and used without further purification. The NMR spectra were recorded on a Bruker Avance 400 or 500 spectrometer at 400 or 500 MHz in $CDCl_3$ with tetramethylsilane as the internal standard. Chemical shifts (δ) are reported in parts per million (*ppm*) and coupling constants (*J*) are reported in hertz (*Hz*). Melting points were determined using a Büchi B-540 capillary melting point apparatus. High-resolution mass spectra were obtained with a Bruker Impact II UHR-QTOF by electrospray ionization (ESI) on a time-of-flight (TOF) mass analyzer. Steady-state and time-resolved emission spectroscopy were conducted using an Edinburgh FLS1000. Column chromatography was performed on silica gel (200–300 mesh).

Conflicts of interest

There are no conflicts to declare.

Data availability

The PhotoCatDB and supplementary datasets used in this study are available at https://github.com/su-group/PhotoCat

Code availability

The source code of PhotoCat and associated data preparation python v3.7 scripts are available at https://github.com/su-group/PhotoCat

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant No. 22108252 and the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China under Grant No. LHDMZ23B060001.

Many thanks to Lu Jiejia, Hu Bing, Zhang Zhenyang, Zhuang Yu, Liu Rongwu, Zhou Na, Zhou Shengqi, Huang Xiao, Lei Rong, Peng Jiehai, Shen Shoutao, Xu Zhengcheng, Zhang Xinyi, Li Yue, and Fan Zhidan in Hangzhou Vocational & Technical College for their contributions to the data collection and verification in PhotoCatDB.

Notes and references

1. Melchiorre, P., Introduction: photochemical catalytic processes. *Chem. Rev.* **122**, 1483-1484. (2022).
2. Huang H., Steiniger, K. A., Lambert, T. H., Electrophotocatalysis: Combining light and electricity to catalyze reactions. *J. Am. Chem. Soc.* **144**, 12567-12583 (2022).
3. Rotstein B. H., Zaretsky S., Rai V., Yudin A. K., Small heterocycles in multicomponent reactions. *Chem. Rev.* **114**, 8323-59 (2014).
4. S. Garbarino, D. Ravelli, S. Protti, A. Basso, Photoinduced multicomponent reactions. *Angew. Chem.* **55**, 15476-15484 (2016).
5. G. A. Coppola, S. Pillitteri, E. V. Eycken, Multicomponent reactions and photo/electrochemistry join forces: atom economy meets energy efficiency. *Chem. Soc. Rev.* **51**, 2313-2382 (2022).
6. Goh G. B., Hodas, N. O., Vishnu, A., Deep learning for computational chemistry. *J. Comput. Chem.*

38, 1291-1307 (2017).

7. Keith J. A., Vassilev-Galindo V., Cheng B., Chmiela S.; Gastegger M., Muller K. R., Tkatchenko A., Combining machine learning and computational chemistry for predictive insights into chemical systems. *Chem. Rev.* **121**, 9816–9872 (2021).

8. Coley C. W., Green W. H., Jensen K. F., Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **51**, 1281-1289 (2018).

9. Mater A. C., Coote M. L., Deep learning in chemistry. *J. Chem. Inf. Model.* **59**, 2545-2559 (**2019)**.

10. Ozturk H., Ozgur A., Schwaller P., Laino T., Ozkirimli E., Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Discov. Today* **25**, 689-705 (2020).

11. Coley, C. W., Barzilay R., Jaakkola T. S., Green W. H., Jensen K. F., Prediction of organic reaction outcomes using machine learning. *ACS Cent. Sci.* **3**, 434-443 (2017).

12. Gale E. M., Durand D. J., Improving reaction prediction. *Nat. Chem.* **12**, 509-510 (2020).

13. Wei J. N., Duvenaud D., Aspuru-Guzik A., Neural networks for the prediction of organic chemistry reactions. *ACS Cent. Sci.* **2**, 725-732 (2016).

14. Schwaller P., Laino T., Gaudin T., Bolgar P., Hunter C. A., Bekas C., Lee A. A., Molecular Transformer: A model for uncertainty-calibrated chemical reaction prediction. *ACS Cent. Sci.* **5**, 1572-1583 (2019).

15. Zhang Y., Wang L., Wang X., Zhang C., Ge J, Tang J., Su A., Duan H., Data augmentation and transfer learning strategies for reaction prediction in low chemical data regimes, *Org. Chem. Front.*, 8, 1415-1423 (2021).

16. Schwaller P., Gaudin T., Lanyi D., Bekas C., Laino T., "Found in Translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chem. Sci.* **9**, 6091-6098 (2018).

17. Wang L., Zhang C., Bai R., Li J., Duan H., Heck reaction prediction using a transformer model based on a transfer learning strategy. *Chem. Commun.* **56**, 9368-9371 (2020).

18. Wu Y., Zhang C., Wang L., Duan H., A graph-convolutional neural network for addressing small-scale reaction prediction. *Chem. Commun.* **57**, 4114-4117 (2021).

19. Liu B., Ramsundar B., Kawthekar P., Shi J., Gomes J., Luu N. Q., Ho S., Sloane J., Wender P., Pande V., Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Cent. Sci.* **3**, 1103-1113 (2017).

20. Schwaller P., Petraglia R., Zullo V., Nair V. H., Haeuselmann R. A., Pisoni R., Bekas C., Iuliano A., Laino T., Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.* **11**, 3316-3325 (2020).

21. Wang X., Li Y., Qiu J., Chen G., Liu H., Liao B., Hsieh C. Y., Yao X., RetroPrime: A Diverse, plausible and Transformer-based method for Single-Step retrosynthesis predictions. *Chem. Eng. J.* **420**, 129845 (2021).

22. Tetko I. V., Karpov P., Van D. R., Godin G., State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat. Commun.* **11**, 5575 (2020).

23. Xu J.; Zhang Y., Han J., Su A., Qiao H., Zhang C., Tang J., Shen X., Sun B., Yu W., Zhai S., Wang X., Wu Y., Su W., Duan H., Providing direction for mechanistic inferences in radical cascade cyclization using a Transformer model. *Org. Chem. Front.* **9**, 2498-2508 (2022).

24. Meuwly M., Machine learning for chemical reactions. *Chem. Rev.* **121**, 10218-10239 (2021).

25. Vaucher A. C., Schwaller P., Geluykens J., Nair V. H., Iuliano A., Laino T., Inferring experimental procedures from text-based representations of chemical reactions. *Nat. Commun.*, **12**, 2573 (2021).

26. Tu Z., Stuyver T., Coley C. W., Predictive chemistry: machine learning for reaction deployment, reaction development, and reaction discovery. *Chem. Sci.* **14**, 226-244 (2023).

27. de Almeida A. F., Moreira R., Rodrigues T., Synthetic organic chemistry driven by artificial intelligence. *Nat. Rev. Chem.* **3**, 589-604 (2019).

28. Caramelli D., Granda J. M., Mehr S. H. M., Cambie D., Henson A. B., Cronin L., Discovering new chemistry with an autonomous robotic platform driven by a Reactivity-Seeking neural network. *ACS Cent. Sci.* **7**, 1821-1830 (2021).

29. Su A., Wang, X., Wang L., Zhang C., Wu Y., Wu X., Zhao Q., Duan H., Reproducing the invention of a named reaction: zero-shot prediction of unseen chemical reactions. *Phys. Chem. Chem. Phys.* **24**, 10280-10291 (2022).

30. Probst D., Manica M., Nana Teukam Y. G., Castrogiovanni A., Paratore F., Laino T., Biocatalysed synthesis planning using data-driven learning. *Nat. Commun.*, **13**, 964 (2022).

31. Kreutter D., Schwaller P., Reymond J. L., Predicting enzymatic reactions with a molecular transformer. *Chem. Sci.* **12**, 8648-8659 (2021).

32. Pesciullesi G., Schwaller P., Laino T., Reymond J. L., Transfer learning enables the molecular transformer to predict regio- and stereoselective reactions on carbohydrates. *Nat. Commun.* **11**, 4874 (2020).

33. Zahrt A. F., Mo Y., Nandiwale K. Y., Shprints R., Heid E., Jensen K. F., Machine-Learning-Guided discovery of electrochemical reactions. *J. Am. Chem. Soc.* **144**, 22599-22610 (2022).

34. Mai H., Le T. C., Chen D., Winkler D. A., Caruso R. A., Machine learning for electrocatalyst and photocatalyst design and discovery. *Chem. Rev.* **122**, 13478-13515 (2022).

35. Su A., Zhang X., Zhang C., Ding D., Yang Y., Wang K., She Y., Deep transfer learning for predicting frontier orbital energies of organic materials using small data and its application to porphyrin photocatalysts, *Phys. Chem. Chem. Phys.* **25**, 10536-10549 (2023).

36. Kearnes S. M., Maser M. R., Wleklinski M., Kast A., Doyle A. G., Dreher S. D., Hawkins J. M., Jensen K. F., Coley C. W., The Open reaction database. *J. Am. Chem. Soc.* **143**, 18820-18826 (2021).

37. David W., SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).

38. TMAP. https://github.com/reymond-group/tmap, Accessed 10 Sep 2023.

39. Klein G., Kim Y., Deng Y., Senellart J., Rush A., OpenNMT: Open-source toolkit for neural machine translation. in Proceedings of ACL 2017, System Demonstrations. Vancouver, Canada: Association for Computational Linguistics. 67–72. https://www.aclweb.org/anthology/P17-4012 (2017).

40. Maggi, G. S. R., Use of Solid Catalysts in Friedel−Crafts Acylation Reactions. *Chem. Rev.* **106**, 1077–1104 (2006).

41. Wu M. C., Li M. Z., Chen J. Y., Xiao J. A., Xiang H. Y., Chen K., Yang H., Photoredox-catalysed chlorination of quinoxalin-2(1H)-ones enabled by using $CHCl_3$ as a chlorine source. *Chem. Commun.* **58**, 11591-11594 (2022).

42. Deb A., Manna S., Modak A., Patra T., Maity S., Maiti D., Oxidative trifluoromethylation of unactivated olefins: an efficient and practical synthesis of alpha-trifluoromethyl-substituted ketones. *Angew. Chem.* **52**, 9747-50 (2013).

43. Muralirajan K., Kancherla R., Bau J. A., Taksande M. R., Qureshi M., Takanabe K., Rueping M., Exploring the structure and performance of Cd–Chalcogenide photocatalysts in selective trifluoromethylation. *ACS Catal.* **11**, 14772-14780 (2021).

44. Nguyen Q. H., Hwang H. S., Cho E. J., Shin S., Energy transfer photolysis of N-

Enoxybenzotriazoles into benzotriazolyl and α-carbonyl radicals. *ACS Catal.* **12**, 8833-8840 (2022).

45 Zhu J., Guo Y., Zhang Y., Li W., Zhang P., Xu J., Visible-light-induced direct perfluoroalkylation/heteroarylation of [1.1.1]propellane to diverse bicyclo[1.1.1]pentanes (BCPs) under metal and photocatalyst-free conditions. *Green Chem.* **25**, 986−992 (2023).

46. Xia Z. H., Zhang C. L., Gao Z. H., Ye S., Switchable decarboxylative Heck-type reaction and Oxo-alkylation of styrenes with N-Hydroxyphthalimide esters under photocatalysis. *Org. Lett.* **20**, 3496-3499 (2018).

47. Zhuang X., Ling L., Wang Y., Li B., Sun B., Su W., Jin C., Photoinduced cascade C-N/C=O bond formation from bromodifluoroalkyl reagents, amines, and H_2O via a Triple-Cleavage process. *Org. Lett.* **24**, 1668-1672 (2022).

48. Zhuang X., Shi X., Zhu R., Sun B., Su W., Jin C., Photocatalytic intramolecular radical cyclization involved synergistic SET and HAT: synthesis of 3,3-difluoro-γ-lactams. *Org. Chem. Front.,* **8**, 736-742 (2021).

49. Yan Z., Sun B., Huang P., Zhao H., Ding H., Su W., Jin C., Visible-light-promoted radical alkylation/cyclization of allylic amide with N-hydroxyphthalimide ester: Synthesis of oxazolines. *Chin. Chem. Lett.* **33**, 1997-2000 (2022).

50. Yan Z., Sun B., Zhang X., Zhuang X., Yang J., Su W., Jin C., Construction of C(sp^2)-C(sp^3) bond between quinoxalin-2(1H)-ones and N-hydroxyphthalimide esters via photocatalytic decarboxylative coupling. *Chem. Asian. J.* **14**, 3344-3349 (2019).

51. Huang P., Yan Z., Ling J., Li P., Wang J., Li J., Sun B., Jin C., Catalyst-free intramolecular radical cyclization cascades initiated by the direct homolysis of Csp3–Br under visible light. *Green Chem.*, **25**, 3989–3994 (2023).

52. Landrum G. et al, RDKit: Open-Source Cheminformatics Software, Release 2019_03_4. https://doi.org/10.5281/zenodo.3366468. Accessed 10 Sep 2023.

53. Reaxys database. https://www.reaxys.com. Accessed 10 Sep 2023.

54. Lowe, D. M. Extraction of Chemical Structures and Reactions from the Literature (University of Cambridge, 2012).