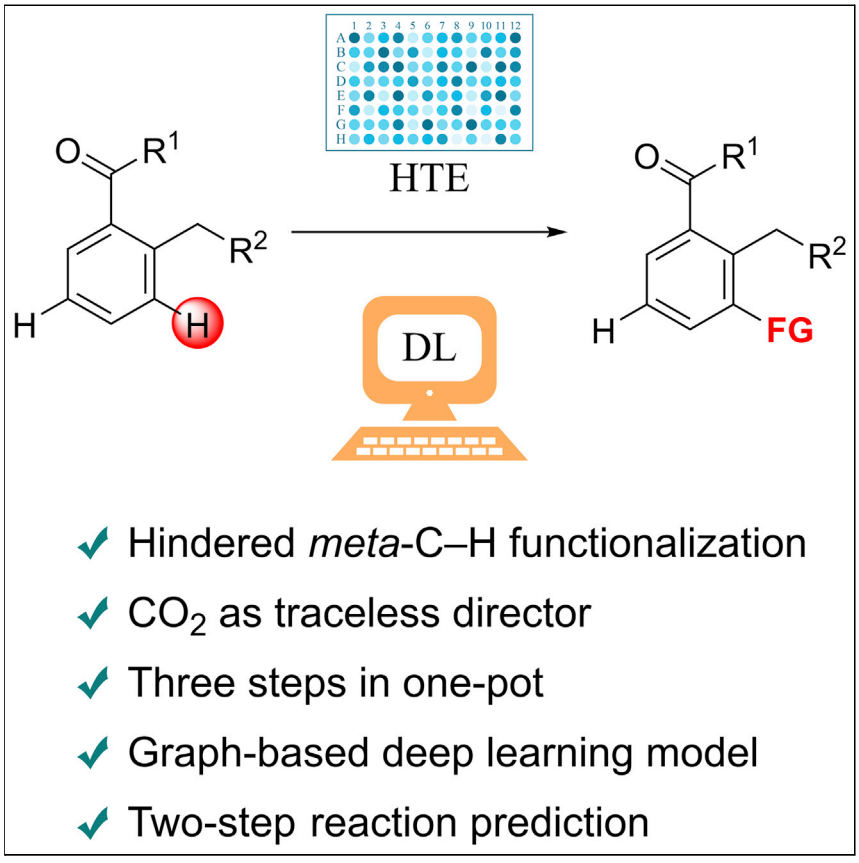
Chem

CellPress

Article

Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning



Modification of C–H bonds is of great interest for chemists, but controlling selectivity is challenging. Differentiating sterically hindered C–H bond over sterically favorable C–H bond is difficult. To develop a reaction for the modification of the hindered aromatic *meta*-C–H bond, automation was used to explore reaction space and artificial intelligence was used to predict reaction outcome. The integration of automation and artificial intelligence will aid the development and application of new synthetic methods.

Jia Qiu, Jiancong Xie, Shimin Su, Yadong Gao, Han Meng, Yuedong Yang, Kuangbiao Liao

yangyd25@mail.sysu.edu.cn (Y.Y.)
liao_kuangbiao@gzlab.ac.cn (K.L.)

Highlights

Hindered *meta*-C–H functionalization

Three steps in one pot (yield up to 76%)

Graph-based prediction model (CMPRY)

Two-step reaction prediction (MAE = 6.6%)

Qiu et al., Chem *8*, 1–13
December 8, 2022
https://doi.org/10.1016/j.chempr.2022.08.015

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

Chem

CellPress

Article

Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning

Jia Qiu,[1,2,5] Jiancong Xie,[3,5] Shimin Su,[1] Yadong Gao,[2] Han Meng,[4] Yuedong Yang,[3,*] and Kuangbiao Liao[1,2,6,*]

SUMMARY

Selective functionalization of the sterically hindered aromatic *meta*-C–H bond is unprecedented and remains to be a major challenge. Promoted by automation-based high-throughput experimentation (HTE) and deep learning (DL), a novel strategy to functionalize the hindered *meta*-C–H bond is disclosed. With carbon dioxide as a traceless director, a one-pot three-step protocol was developed to achieve selective arylation of *o*-alkylaryl ketones at the hindered *meta* position. This novel strategy involved photo-induced C–H carboxylation, carboxyl group-directed Pd-catalyzed C–H functionalization, and microwave-assisted decarboxylation. With HTE and DL, a broad scope of substrates was explored (1,032 reactions) and a DL-based model (CMPRY) for reaction yield prediction was established. Two independent tests with unseen *o*-alkylaryl ketones and/or potassium aryltrifluoroborates were used to evaluate the model. The model gave excellent performances in predicting unseen reactions; mean absolute errors in yield were only 6.6% and 8.4%, suggesting its potential in synthetic application.

INTRODUCTION

In natural products, pharmaceuticals, and functional materials, substituted arenes are ubiquitous and important structural motifs.[1–3] Several selective C–H functionalization of arenes have been reported in past decades.[4–9] Despite the extra steps to install and remove directing groups (DGs), directed C–H functionalization has been demonstrated to achieve excellent selectivity at *ortho*-, *para*-, and unhindered *meta*-position (Figure 1A).[10–17] Yu and co-workers' innovative template strategy enables *meta*-C–H functionalization, but all reacted at unhindered *meta*-position due to the steric preference (Figure 1B).[16–18] Employing carbon dioxide (CO_2) as a traceless relay director for the unhindered *meta*-C–H arylation of phenols was first reported by Larrosa and co-workers in 2014, which involves Kolbe-Schmitt reaction, carboxyl group-directed Pd-catalyzed C–H arylation, and tandem decarboxylation process.[19–21] A similar approach, with norbornene (NBE) as a transient mediator, was introduced by Yu group to achieve unhindered *meta*-C–H alkylation of phenyl acetamide in 2015.[18] The Pd/NBE-relay process (Catellani process)[22] was further exploited for other kinds of unhindered *meta*-C–H functionalization.[23–25]

Reported progresses predominately focused on unhindered *meta*-C–H bond; however, selective functionalization of the hindered *meta*-C–H bond was rarely reported. To circumvent this challenge, we developed a one-pot three-step reaction using CO_2 as a traceless DG (Figure 1C). A photo-induced C(sp^3)–H carboxylation

THE BIGGER PICTURE

Selective C–H functionalization is a key goal in organic chemistry. Although various protocols were reported, distinct electronic or steric properties are generally required for the successful differentiation of one C–H bond from others. For C–H bonds with similar electronic properties, sterically unhindered C–H bonds are strongly preferred under established protocols; therefore, selective functionalization of hindered C–H bonds remains a major challenge. Promoted by automation and artificial intelligence, a different strategy to functionalize the hindered aromatic *meta*-C–H bond was developed. Automation-based high-throughput experimentation was used to explore reaction space and collect reaction data. Artificial intelligence-based models were used to predict reaction yield of unseen substrates. The integration of automation and artificial intelligence in organic synthesis is an emerging field of research in chemical research, which can expedite reaction discovery and application significantly.

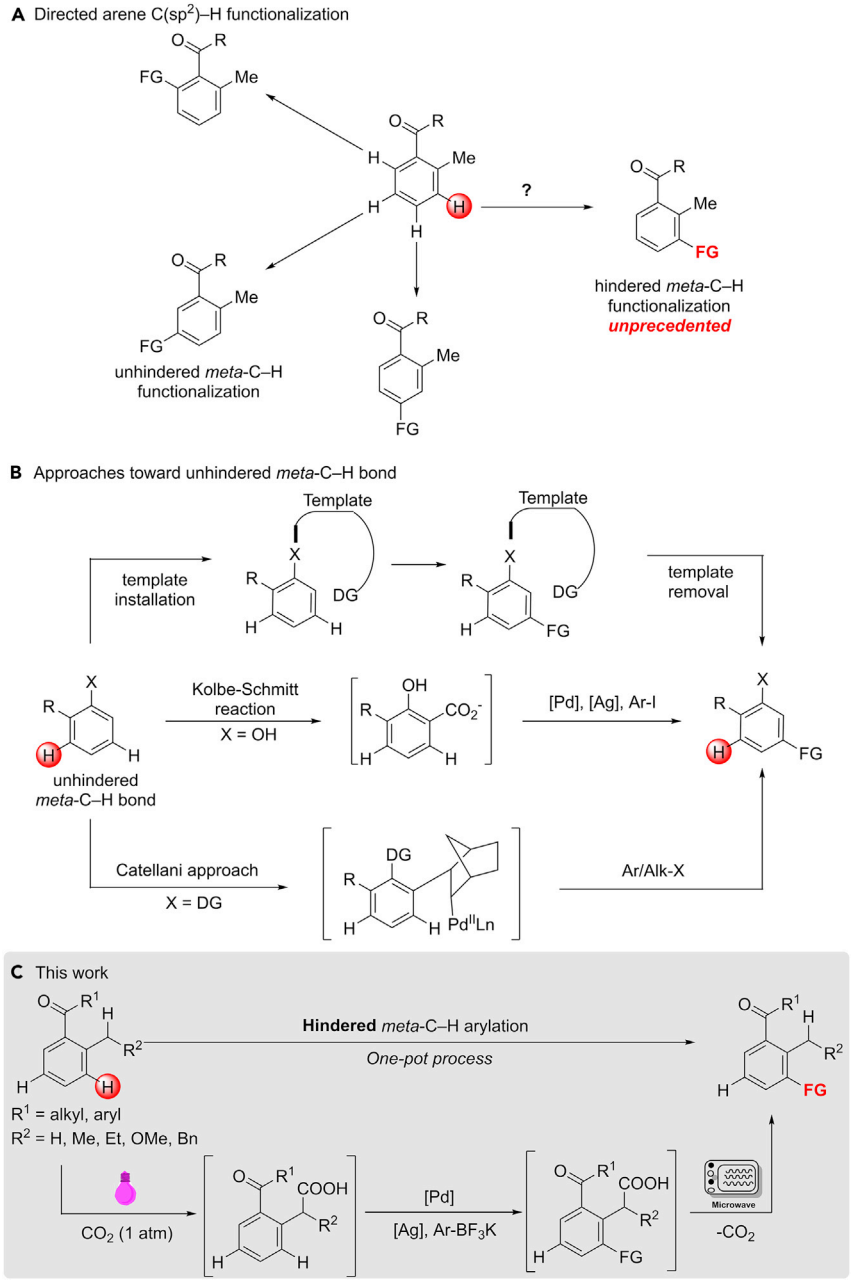


Figure 1. Strategies for regioselective C–H functionalization of substituted arenes

(A) Directed arene C(sp^2)–H functionalization.

(B) Approaches toward unhindered *meta*-C–H bond.

(C) This work: a three-in-one protocol for selective arylation of the hindered *meta*-C–H bond.

reaction[26] was used to install a carboxyl group at the benzylic position, then an arylacetic acid directed Pd-catalyzed reaction was developed to achieve selective functionalization of the hindered *meta*-C–H bond via high-throughput experimentation (HTE); finally, a microwave (MW)-assisted decarboxylation was employed to release CO_2. Notably, HTE[27–31] was the key for success to obtain optimal condition, explore substrates scope, and collect standardized experimental data for deep learning. To broaden the application of this methodology, deep learning models for reaction

[1]Guangzhou Laboratory, Guangzhou 510320, Guangdong, P.R. China

[2]Bioland Laboratory, Guangzhou 510005, Guangdong, P.R. China

[3]School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, P.R. China

[4]Department of Chemistry, Southern University of Science and Technology, Shenzhen 518055, P.R. China

[5]These authors contributed equally

[6]Lead contact

*Correspondence:
yangyd25@mail.sysu.edu.cn (Y.Y.),
liao_kuangbiao@gzlab.ac.cn (K.L.)

https://doi.org/10.1016/j.chempr.2022.08.015

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

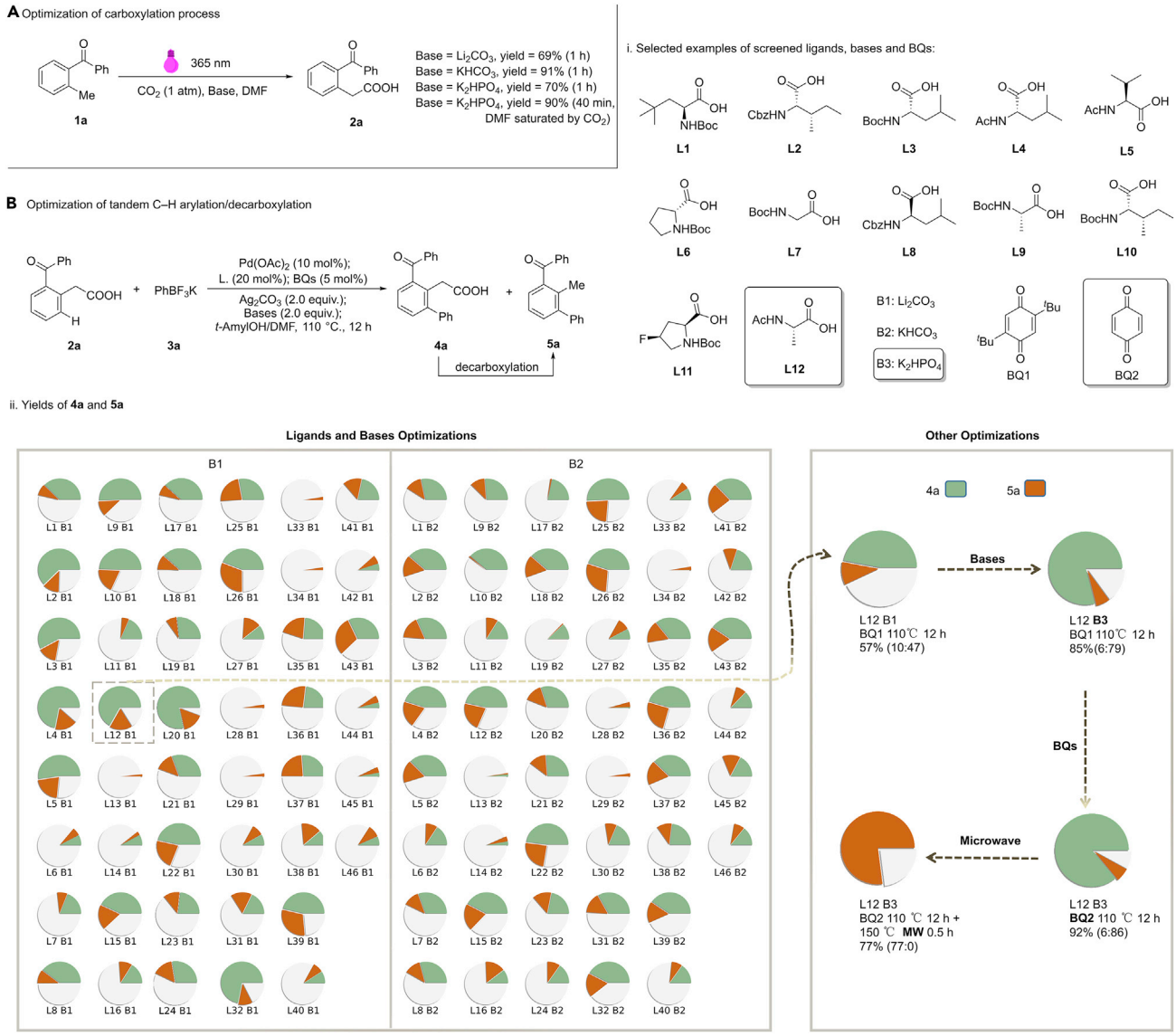


Figure 2. Optimization of C–H carboxylation/C–H arylation/decarboxylation

(A) Optimization of C–H carboxylation process.

(B) Optimization of tandem C–H arylation/decarboxylation. (i) Selected examples of screened MPAA ligands, bases, and BQs; (ii) yields of **4a** (darkseagreen) and **5a** (chocolate).

yield prediction that rely on graph-based descriptors were developed, and the interpretability of the model was also pursued.[32–39]

RESULTS AND DISCUSSION

Reaction development

Our studies began at the optimization of the photo-induced C(sp^3)–H carboxylation (Figure 2A).[26] Owing to the considerable solubility of CO_2 and relatively low boiling point, dimethylformamide (DMF) was selected as solvent instead of dimethyl sulfoxide (DMSO). Then, bases were evaluated to improve the carboxylation yield. Addition of $KHCO_3$, carboxylated compound (**2**) was afforded in excellent yield. With K_2HPO_4 and DMF saturated by CO_2 in advance, C(sp^3)–H carboxylation also proceeded successfully (90% yield in 40 min).

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

Next, we moved on to study C–H arylation/decarboxylation process using the carboxylated compound (**2a**) and potassium phenyltrifluoroborate ($PhBF_3K$, **3a**) as model substrates to make the final product (**5a**) (Figure 2B). As of the arylacetic acid directed C–H arylation, the hindered *meta*-C–H phenylation compound (**4a**) was formed. It was proved that the mono-N-protected amino acids (MPAAs) could accelerate the Pd(II)-catalyzed C(sp^2)–H/arylboron cross-coupling.[4,40–42] Then 46 kinds of MPAAs and 2 bases were selected for our initial condition screening. In total, 92 conditions were quickly evaluated by HTE. Then, further HTE were conducted to optimize MPAAs, bases, benzoquinones (BQs), temperature, and heating method (see details in supplemental information). It is worth mentioning that certain level of decarboxylation product (**5a**) was observed in this process. With L12, K_2HPO_4, and BQ2, the yield was improved to 92% (**5a**:**4a** = 6:86). Finally, hindered *meta*-C–H phenylation final product (**5a**) was obtained in 77% yield with decarboxylation process performed in a MW reactor at 150°C for 0.5 h.

Reaction scope

With the optimized conditions in hand, we began to explore the substrate scope of *o*-alkylaryl ketones and potassium aryltrifluoroborates ($ArBF_3K$) via HTE (Figure 3). Thanks to the unique ^{1}H-NMR shift of the α-H of carboxylic acids, we were able to identify the reaction yield through NMR analysis in a high-throughput manner. Therefore, the HTE enabled substrate scope exploration was focusing on the tandem C–H functionalization, i.e., photo-induced C–H carboxylation/carboxyl group directed Pd-catalyzed C–H functionalization. An array of structurally diverse substrates, consisting of 24 ketones (**1a–1x**) and 43 $ArBF_3K$ (**3a–3zq**), was selected to set up micromole-scale reactions using HTE (Figure 3).

Subsequently, a dataset with 1,032 reaction data was obtained (Figure 4A). It was found that over 70% of the reactions gave desired products, which demonstrated that our strategy was generally applicable for a wide range of substrates. Generally, *o*-tolyl aryl ketones (**1a–u**) afforded the corresponding products in notable yield with most $ArBF_3K$. Substituents such as fluoro (**1c**, **1g**, **1k**, and **1n**), chloro (**1d** and **1h**), and *tert*-butyl (**1f** and **1j**) on the aromatic ring were tolerated. According to the literature, compounds with substitution at the benzylic position were challenging substrates.[26] To our surprise, substrates with Me, Et, Bn, and OMe at the benzylic position (**1q–1t**) were also amenable to this protocol. In the case of *o*-tolyl pyridyl ketone (**1u**), some corresponding products were afforded in moderate yields. It appears that *o*-tolyl *p*-bromophenyl ketones (**1l**) yield traceless or no product across most $ArBF_3K$ probably due to the debromination side reaction; however, 41% overall yield of tandem C(sp^3)–H carboxylation/C(sp^2)–H phenylation was obtained when it reacted with *p*-$OMePhBF_3K$ (**3j**). In the C–H carboxylation step, aryl alkyl ketones (**1v**, **1x**) were generally considered to have lower reactivity than aryl bromophenol ketones; however, in this study, many reactions still gave good overall yield. Phenylation of C(sp^2)–H bonds with two adjacent substituents were known to be extremely challenging due to the huge steric hindrance. In this study, *para*-substituted *o*-tolyl phenyl ketones (**1o**, **1p**) and *para*-substituted *o*-tolyl methyl ketone (**1w**) indeed gave low yield in most cases; however, *p*-$OMePhBF_3K$ (**3j**) still gave 23% overall yield with *para*-chloro *o*-tolyl phenyl ketone (**1p**). In terms of the substrate scope of $ArBF_3K$, electron-donating or electron-withdrawing groups at *ortho*-, *meta*-, or *para*-positions were compatible in this reaction (**3a–3zc**). However, *ortho* substituted $ArBF_3K$ (**3t**-**3zc**) generally gave lower yield due to the steric hindrance. Several reactive functionalities, such as hydroxyl (protected or unprotected) (**3l**, **3s**, and **3zi**), benzodioxole (**3zf**), ketones (**3zg**), and aldehydes (**3zj**),

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

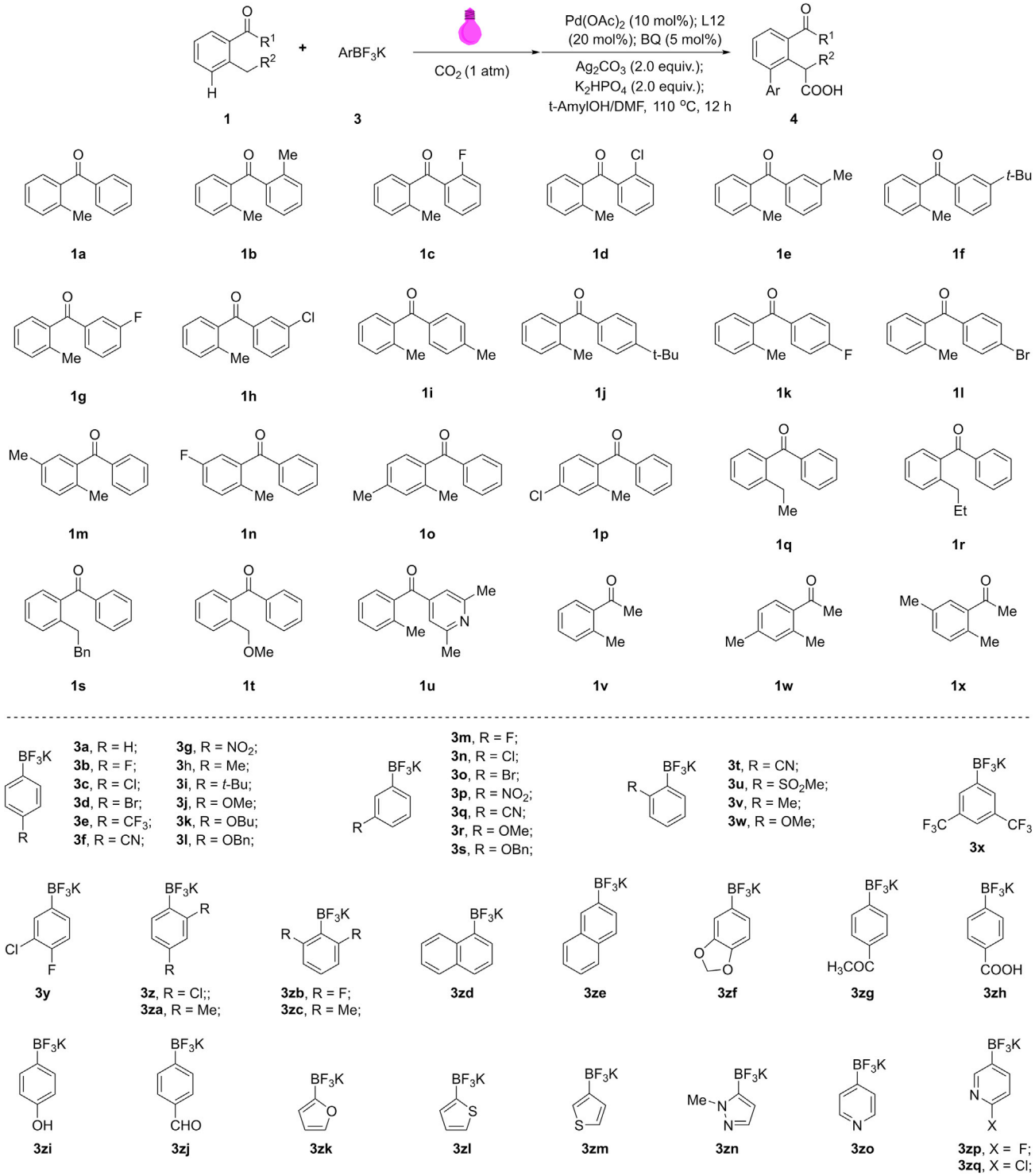


Figure 3. *o*-Alkylaryl ketones and $ArBF_3K$ substrates scope

Yield determined by H-NMR analysis with coumarin as internal standard, see the supplemental information for detailed reaction conditions.

were also suitable for this tandem reaction. However, potassium trifluoroborate with phenyl carboxylic acid (**3zh**), pyrazole (**3zn**), and pyridine (**3zo**) were not compatible in most cases.

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

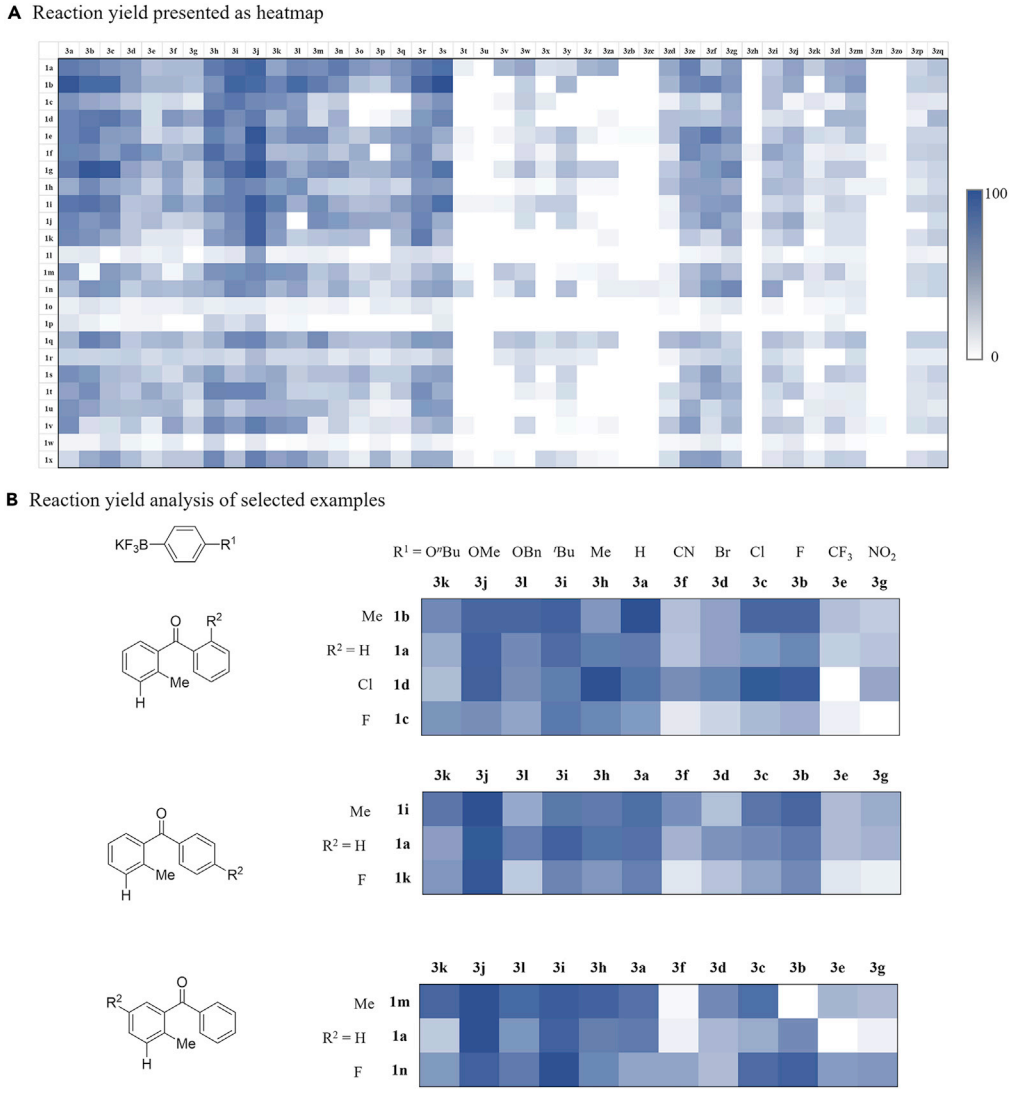


Figure 4. Heatmap of yields of *o*-alkylaryl ketones and $ArBF_3K$ substrates scope

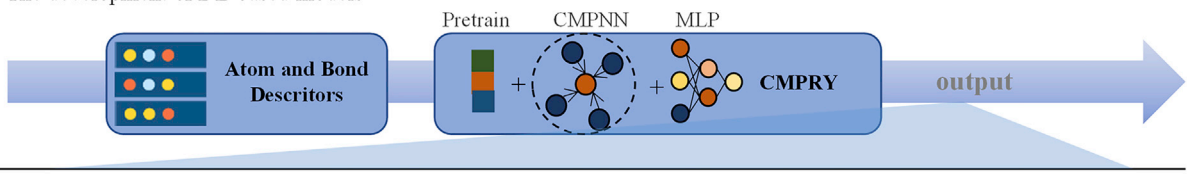
(A) Yields heatmap of 1,032 reactions; (B) selected examples with various substitution. Yields determined by ^{1}H-NMR analysis with coumarin as internal standard.

Upon analyzing these results (Figure 4A), basic rules or conclusions about this tandem reaction were hard to be drawn. For instance, it was difficult to summarize how substituents with different electronic characteristics on *o*-alkylaryl ketones or $ArBF_3K$ would affect the overall yield. Examples in Figure 4B revealed that synergistic effect of all variables instead of any single factor determine the reaction outcome. The electron density on the aromatic ring decreases with the substitution from left to right and from bottom to top. Bis(*o*-tolyl) ketone (**1b**) gave higher yield than *o*-tolyl phenyl ketone (**1a**) when they reacted with $PhBF_3K$ (**3a**) and electron-poor $ArBF_3K$ [Ar as *p*-F, -ClC_6H_4 (**3b–3c**)] and electron-rich $ArBF_3K$ [Ar as *p*-O^{n}Bu, -$OBnC_6H_4$ (**3k** and **3l**)]; however, **1a** gave higher yield than **1b** when it came to *p*-$MeC_6H_4BF_3K$. *o*-Tolyl *o*-chlorophenyl (**1d**) gave higher yield than *o*-tolyl *o*-fluorophenyl (**1c**) when they reacted with $PhBF_3K$ (**3a**) and electron-poor $ArBF_3K$ [Ar as *p*-F, -Cl, -BrC_6H_4 (**3b–3d**)]; however, when it came to electron-rich $ArBF_3K$ [Ar as *p*-Me, -^{t}Bu, -OMe, -O^{n}Bu (**3h–3k**)], **1d** gave higher yield than **1c** for **3h**, but opposite results were obtained

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

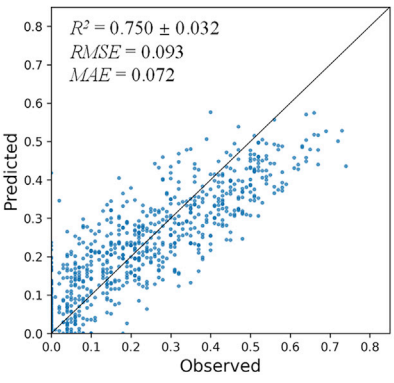
A Dataset split

B The development of DL-based models

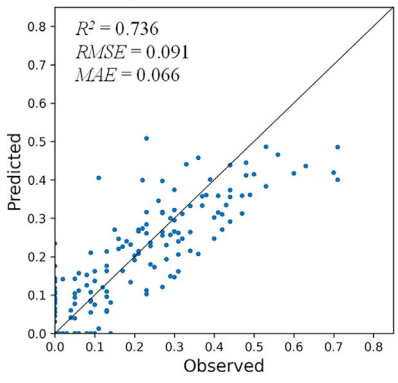


Metrics		CMPRY	CMPNN (CMPRY w/o Pretraining)	GCN	GBM	XGB
Cross Validaton	R^2	0.750±0.03	0.709±0.03	0.677±0.03	0.769±0.02	0.769±0.02
	MAE	0.072±0.005	0.077±0.005	0.082±0.005	0.066±0.004	0.067±0.002
	RMSE	0.093±0.007	0.101±0.008	0.106±0.008	0.089±0.004	0.089±0.001
Independent Test	R^2	**0.736**	0.641	0.638	0.626	0.606
	MAE	**0.066**	0.073	0.080	0.081	0.083
	RMSE	**0.091**	0.106	0.106	0.117	0.111

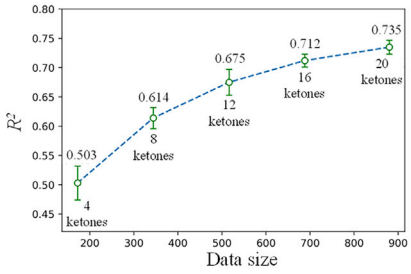
C Cross validation of CMPRY



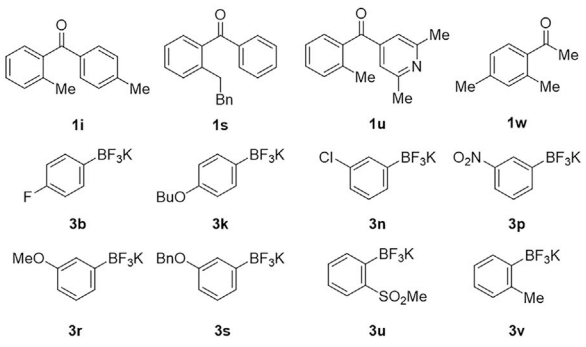
D Independent test of CMPRY



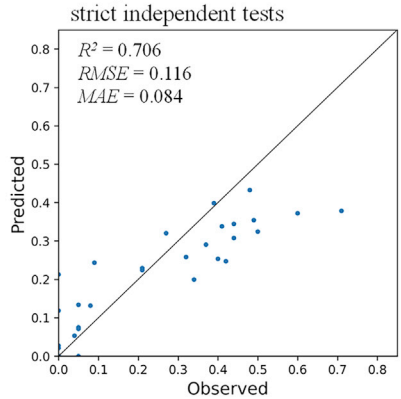
E The performance of CMPRY with different training data sizes.



F 4 ketones and 8 $ArBF_3K$ in strict independent tests



G The performance of CMPRY under strict independent tests



Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

Figure 5. The development of CMPRY
(A) Dataset split.
(B) The development of DL-based models.
(C) 5-fold cross validation result of CMPRY.
(D) Independent test result of CMPRY.
(E) The performance of CMPRY with different training data sizes.
(F) 4 ketones and 8 $ArBF_3K$ in strict independent tests.
(G) The performance of CMPRY under strict independent tests.

for **3i** and **3k**. Similar chaos were observed on other substrates with *para*-substitutions (**1i**, **1k**, and **1m–1n**). There was no obviously privileged or underprivileged electronic characteristics for this reaction because substrates with electron-donating or electron-withdrawing groups could deliver either good or bad results depending on the structure of its reaction partner. In reality, even for a well-established reaction, predicting outcome of an unseen substrate by inference is a challenging task for an experienced chemist. Therefore, developing prediction model for unseen substrates would hold tremendous importance to help chemists to adopt this methodology in their synthetic plan; moreover, it could also significantly lower the bar for non-expert to use this methodology.

CMPRY: DL-based reaction yield prediction

With the experimentally measured 1,032 reaction yield data points, we developed a new model, communicative message passing neural network for reaction yield prediction (CMPRY), to predict reaction yield of this tandem reaction. The CMPRY model was based on the communicative message passing neural network (CMPNN) architecture.[43] First, we split the dataset into training data and external data. We selected 860 reactions, consisting of 20 ketones (**1a–1h**, **1j–1r**, **1t**, **1v**, and **1x**) and 43 $ArBF_3K$ (**3a–3zq**), as the training set. To evaluate the model performance for external data, we used the rest 172 reactions (**1i**, **1s**, **1u**, and **1w** with 43 $ArBF_3K$) as the independent test set (Figure 5A). Due to the small training set, we pretrained the model on a large reaction dataset (USPTO-479k)[44] by following the previous study that preserves the equivalence of molecules with respect to chemical reactions in the embedding space[45] and then fine-tuned it on the training samples. By a random 80/20 split of the training set followed by a 5-fold cross validation, our model achieved R^2 of 0.750 and MAE in yield of 7.2% (Figures 5B and 5C) that are essentially the same as the ones (R^2 of 0.736 and MAE in yield of 6.6%) on the independent test set (Figures 5B and 5D). It is worth mentioning that 4 unseen ketones were included in the independent test set. Achieving such good performance for external data demonstrated that our model was capable of predicting reaction yield in this tandem reaction. We also evaluated the relationship of model performance with regards to the numbers of ketones in the training dataset. As expected, the R^2 was 0.503 with 172 training points (4 ketones) and gradually increased to 0.735 with 688 points (20 ketones). Although R^2 continues growing with the training size, it tends to converge (Figure 5E).

For model comparison, we have employed four commonly used baselines. As shown in Figure 5B, the removal of pretraining caused a big drop with R^2 of 0.643. The change from CMPNN to graph convolutional networks (GCNs) resulted in a small but significant decrease in R^2. Although the Machine Learning-based method extreme gradient boosting (XGB) and gradient boosting machine (GBM) achieved acceptable result in the 5-fold cross validation, they lacked sufficient generalization capability, giving R^2 of 0.606 and 0.626 in the independent test.

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

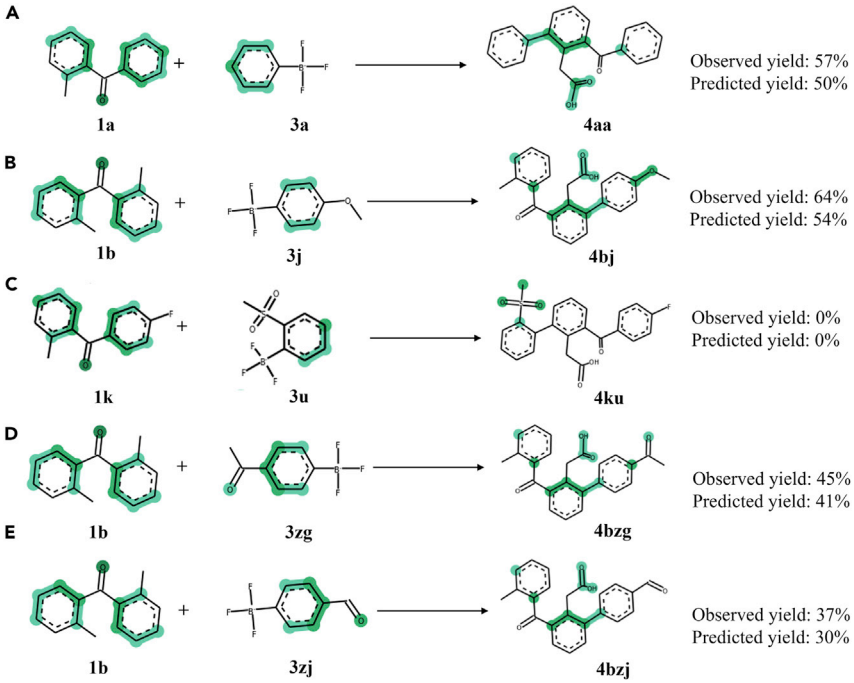


Figure 6. Important structural fragments derived by CMPRY using CAM-GRAD

(A–E) Selected examples for the model interpretability calculated by CAM-GRAD. The atoms and bonds marked in green indicate their importance in the reaction.

For a stricter test, we attempted to make an independent test set with both ketones and $ArBF_3K$ unseen. We selected 4 ketones (**1i**, **1s**, **1u**, and **1w**) and 8 $ArBF_3K$ (**3b**, **3k**, **3n**, **3p**, **3r**, **3s**, **3u**, and **3v**) and removed their related reaction data from the training set (Figures 5A and 5F). We re-trained the model using the shrunk training set and tested it on the strict independent test set. The model achieved a reasonable R^2 value of 0.706 and MAE in yield of 8.4%, indicating the generality and the robustness of CMPRY (Figure 5G). The performance could be regarded superior because of the complexity of predicting tandem reactions, in regard with that most prediction models reported in the literatures[46–51] were dealing with one-step reaction.

To interpret our DL-based model, we have employed the gradient-weighted class activation mapping (CAM-GRAD)[52,53] to identify critical atoms for predictions. As shown in Figure 6, CMPRY highlighted the new $C(sp^2)$–$C(sp^2)$ bond formation, indicating that the model learned the basic reaction information. For the reaction 6C that gave no product, CMPRY highlighted the *ortho*-substitution, methylsulfonyl group, agreeing with the rule that the steric hindrance is harmful for the C–H bond activation. Interestingly, CMPRY found that the carboxylic acid group was important at all successful reactions, although such information was not input into the predictions. Mechanistically, carboxyl group is needed to direct palladium catalyst for C–H activation. Last but not the least, there are some peculiar structural fragments highlighted in reactions 6D and 6E. For example, both oxygen atoms in the aldehyde and acetyl groups were considered important atoms, but only oxygen in the acetyl group was highlighted as an important atom in products.

Reaction utilities

To improve the synthetic utilities of this methodology and avoid the tedious separation of C–H carboxylation and C–H arylation products, a one-pot three-step reaction

Please cite this article in press as: Qiu et al., Selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones promoted by automation and deep learning, Chem (2022), https://doi.org/10.1016/j.chempr.2022.08.015

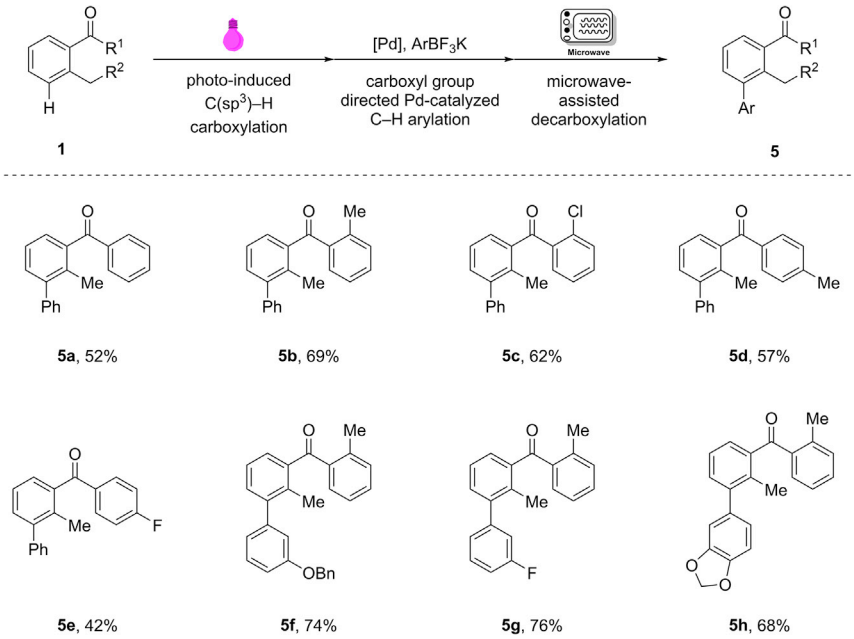


Figure 7. One-pot process for hindered *meta*-C–H arylation with CO_2 as traceless directing group
Reaction conditions unless specified otherwise: **1** (0.1 mmol), K_2HPO_4 (2.0 equiv), 365 nm UV light, 40 min; then, **3** (0.3 mmol), $Pd(OAc)_2$ (10 mol %), L12 (20 mmol %), Ag_2CO_3 (2.0 equiv), and K_2HPO_4 (1.0 equiv), 110°C, 12 h; finally, microwave reactor, 150°C, 0.5 h. Isolated yields.

for selective functionalization of the hindered *meta*-C–H bond was pursued (Figure 7). Following the photo-induced C–H carboxylation, the unpurified intermediate was subjected to carboxyl group directed Pd-catalyzed C–H arylation with $ArBF_3K$, and then, the vial was transferred to an MW reactor for decarboxylation to deliver the final product. Gratifyingly, this one-pot procedure furnished the desired hindered *meta*-C–H functionalization products in good overall yields (up to 76% for three steps, meaning up to 91% yield on average for each step). The synthetic applicability and efficacy transformation were further demonstrated by these results. More importantly, CO_2 was acted as a traceless relay director in this protocol.

In summary, this methodology successfully achieved selective functionalization of hindered *meta*-C–H bond of *o*-alkylaryl ketones. A tandem transformation, involving photo-induced C–H carboxylation, carboxyl group-directed Pd-catalyzed C–H functionalization, and MW-assisted decarboxylation, was developed to streamline the access of multiply substituted arenes that would otherwise require extra steps. Further application of this strategy is under investigation in our laboratory. Moreover, several DL models, which were capable of predicting reaction yield over a range of unseen substrates for this one-pot reaction, have been developed. Overall, we hope that the use of HTE-DL paradigm in the development of hindered *meta*-C–H arylation will stimulate its further application in the development of other synthetic methods.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kuangbiao Liao (liao_kuangbiao@gzlab.ac.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All datasets, codes, and workflows for modeling are available at https://github.com/biomed-AI/CMPRY.

Full experimental procedures are provided in the supplemental information.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chempr.2022.08.015

ACKNOWLEDGMENTS

We are grateful for financial support from Guangzhou Laboratory, Bioland Laboratory, and the National Natural Science Foundation of China (22071249).

AUTHOR CONTRIBUTIONS

J.Q. and K.L. conceived the study; J.Q. carried out the reaction and analyzed the data, J.Q., S.S., Y.G., J.X., Y.Y., and K.L. prepared the manuscript and supplemental information; J.Q., S.S., and J.X. developed the prediction model; Y.G. and H.M. performed the physical descriptors calculation. All authors discussed the results and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 11, 2022
Revised: August 8, 2022
Accepted: August 18, 2022
Published: September 19, 2022

REFERENCES

1. Horton, D.A., Bourne, G.T., and Smythe, M.L. (2003). The combinatorial synthesis of bicyclic privileged structures or privileged substructures. Chem. Rev. *103*, 893–930. https://doi.org/10.1021/cr020033s.

2. Hudgens, D.P., Taylor, C., Batts, T.W., Patel, M.K., and Brown, M.L. (2006). Discovery of diphenyl amine based sodium channel blockers, effective against hNav1.2. Bioorg. Med. Chem. *14*, 8366–8378. https://doi.org/10.1016/j.bmc.2006.09.010.

3. Ackermann, L. (2009). Modern Arylation Methods (Wiley-VCH Verlag).

4. Engle, K.M., Mei, T.-S., Wasa, M., and Yu, J.-Q. (2012). Weak coordination as a powerful means for developing broadly useful C–H functionalization reactions. Acc. Chem. Res. *45*, 788–802. https://doi.org/10.1021/ar200185g.

5. Sambiagio, C., Schönbauer, D., Blieck, R., Dao-Huy, T., Pototschnig, G., Schaaf, P., Wiesinger, T., Zia, M.F., Wencel-Delord, J., Besset, T., et al. (2018). A comprehensive overview of directing groups applied in metal-catalysed C–H functionalisation chemistry. Chem. Soc. Rev. *47*, 6603–6743. https://doi.org/10.1039/C8CS00201K.

6. Meng, G., Lam, N.Y.S., Lucas, E.L., Saint-Denis, T.G., Verma, P., Chekshin, N., and Yu, J.-Q. (2020). Achieving site-selectivity for C–H activation processes based on distance and geometry: a carpenter's approach. J. Am. Chem. Soc. *142*, 10571–10591. https://doi.org/10.1021/jacs.0c04074.

7. Rej, S., Ano, Y., and Chatani, N. (2020). Bidentate directing groups: an efficient tool in C–H bond functionalization chemistry for the expedient construction of C-C Bonds. Chem. Rev. *120*, 1788–1887. https://doi.org/10.1021/acs.chemrev.9b00495.

8. Dutta, U., Maiti, S., Bhattacharya, T., and Maiti, D. (2021). Arene diversification through distal C(sp^2)-H functionalization. Science *372*, eabd599. https://doi.org/10.1126/science.abd5992.

9. Rogge, T., Kaplaneris, N., Chatani, N., Kim, J., Chang, S., Punji, B., Schafer, L.L., Musaev, D.G., Wencel-Delord, J., Roberts, C.A., et al. (2021). C–H activation. Nat. Rev. Methods Primers *1*. https://doi.org/10.1038/s43586-021-00041-2.

10. Kakiuchi, F., Matsuura, Y., Kan, S., and Chatani, N. (2005). A $RuH_2(CO)(PPh_3)_3$-catalyzed regioselective arylation of aromatic ketones with arylboronates via carbon–hydrogen bond cleavage. J. Am. Chem. Soc. *127*, 5936–5945. https://doi.org/10.1021/ja043334n.

11. Giri, R., Maugel, N., Li, J.J., Wang, D.H., Breazzano, S.P., Saunders, L.B., and Yu, J.-Q. (2007). Palladium-catalyzed methylation and arylation of sp^2 and sp^3 C–H bonds in simple carboxylic acids. J. Am. Chem. Soc. *129*, 3510–3511. https://doi.org/10.1021/ja0701614.

12. Zhuang, Z., and Yu, J.Q. (2020). Lactonization as a general route to beta-C(sp(3))-H functionalization. Nature *577*, 656–659. https://doi.org/10.1038/s41586-019-1859-y.

13. Zhang, Z., Tanaka, K., and Yu, J.Q. (2017). Remote site-selective C–H activation directed by a catalytic bifunctional template. Nature *543*, 538–542. https://doi.org/10.1038/nature21418.

14. Jin, Z., Chu, L., Chen, Y.Q., and Yu, J.Q. (2018). Pd-catalyzed remote meta-C–H functionalization of phenylacetic acids using a pyridine template. Org. Lett. *20*, 425–428. https://doi.org/10.1021/acs.orglett.7b03336.

15. Park, H., Verma, P., Hong, K., and Yu, J.Q. (2018). Controlling Pd(IV) reductive elimination pathways enables Pd(II)-catalysed enantioselective C(sp^3)-H fluorination. Nat. Chem. *10*, 755–762. https://doi.org/10.1038/s41557-018-0048-1.

16. Shi, H., Herron, A.N., Shao, Y., Shao, Q., and Yu, J.Q. (2018). Enantioselective remote meta-C–H arylation and alkylation via a chiral transient mediator. Nature *558*, 581–585. https://doi.org/10.1038/s41586-018-0220-1.

17. Xu, J., Chen, J., Gao, F., Xie, S., Xu, X., Jin, Z., and Yu, J.Q. (2019). Sequential functionalization of meta-C–H and ipso-C-O bonds of phenols. J. Am. Chem. Soc. *141*, 1903–1907. https://doi.org/10.1021/jacs.8b13403.

18. Wang, X.C., Gong, W., Fang, L.-Z., Zhu, R.-Y., Li, S., Engle, K.M., and Yu, J.-Q. (2015). Ligand-enabled meta-C–H activation using a transient mediator. Nature *519*, 334–338. https://doi.org/10.1038/nature14214.

19. Luo, J., Preciado, S., and Larrosa, I. (2014). Overriding ortho-para selectivity via a traceless directing group relay strategy: the meta-selective arylation of phenols. J. Am. Chem. Soc. *136*, 4109–4112. https://doi.org/10.1021/ja500457s.

20. Font, M., Spencer, A.R.A., and Larrosa, I. (2018). meta-C–H arylation of fluoroarenes via traceless directing group relay strategy. Chem. Sci. *9*, 7133–7137. https://doi.org/10.1039/c8sc02417k.

21. Spencer, A.R.A., Korde, R., Font, M., and Larrosa, I. (2020). meta-Selective olefination of fluoroarenes with alkynes using CO_2 as a traceless directing group. Chem. Sci. *11*, 4204–4208. https://doi.org/10.1039/d0sc01138j.

22. Catellani, M., Motti, E., and Della Ca', N. (2008). Catalytic sequential reactions involving palladacycle-directed aryl coupling steps. Acc. Chem. Res. *41*, 1512–1522. https://doi.org/10.1021/ar800040u.

23. Wang, P., Farmer, M.E., Huo, X., Jain, P., Shen, P.-X., Ishoey, M., Bradner, J.E., Wisniewski, S.R., Eastgate, M.D., and Yu, J.-Q. (2016). Ligand-promoted meta-C–H arylation of anilines, phenols, and heterocycles. J. Am. Chem. Soc. *138*, 9269–9276. https://doi.org/10.1021/jacs.6b04966.

24. Wang, J., and Dong, G. (2019). Palladium/norbornene cooperative catalysis. Chem. Rev. *119*, 7478–7528. https://doi.org/10.1021/acs.chemrev.9b00079.

25. Shi, H., Lu, Y., Weng, J., Bay, K.L., Chen, X., Tanaka, K., Verma, P., Houk, K.N., and Yu, J.-Q. (2020). Differentiation and functionalization of remote C–H bonds in adjacent positions. Nat. Chem. *12*, 399–404. https://doi.org/10.1038/s41557-020-0424-5.

26. Masuda, Y., Ishida, N., and Murakami, M. (2015). Light-driven carboxylation of o-alkylphenyl ketones with CO_2. J. Am. Chem. Soc. *137*, 14063–14066. https://doi.org/10.1021/jacs.5b10032.

27. Buitrago Santanilla, A.B., Regalado, E.L., Pereira, T., Shevlin, M., Bateman, K., Campeau, L.-C., Schneeweis, J., Berritt, S., Shi, Z.-C., Nantermet, P., et al. (2015). Organic chemistry. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. Science *347*, 49–53. https://doi.org/10.1126/science.1259203.

28. Krska, S.W., DiRocco, D.A., Dreher, S.D., and Shevlin, M. (2017). The evolution of chemical High-throughput experimentation to address challenging problems in pharmaceutical synthesis. Acc. Chem. Res. *50*, 2976–2985. https://doi.org/10.1021/acs.accounts.7b00428.

29. Gaunt, M.J., Janey, J.M., Schultz, D.M., and Cernak, T. (2021). Myths of high-throughput experimentation and automation in chemistry. Chem *7*, 2259–2260. https://doi.org/10.1016/j.chempr.2021.08.012.

30. González-Esguevillas, M., Fernández, D.F., Rincón, J.A., Barberis, M., de Frutos, O., Mateos, C., García-Cerrada, S., Agejas, J., and MacMillan, D.W.C. (2021). Rapid optimization of photoredox reactions for continuous-flow systems using microscale batch technology. ACS Cent. Sci. *7*, 1126–1134. https://doi.org/10.1021/acscentsci.1c00303.

31. Kang, K., Loud, N.L., DiBenedetto, T.A., and Weix, D.J. (2021). A general, multimetallic cross-Ullmann biheteroaryl synthesis from heteroaryl halides and heteroaryl triflates. J. Am. Chem. Soc. *143*, 21484–21491. https://doi.org/10.1021/jacs.1c10907.

32. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature *559*, 547–555. https://doi.org/10.1038/s41586-018-0337-2.

33. Sanchez-Lengeling, B., and Aspuru-Guzik, A. (2018). Inverse molecular design using machine learning: generative models for matter engineering. Science *361*, 360–365. https://doi.org/10.1126/science.aat2663.

34. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., and Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nat. Rev. Drug Discov. *18*, 463–477. https://doi.org/10.1038/s41573-019-0024-5.

35. Sandfort, F., Strieth-Kalthoff, F., Kühnemund, M., Beecks, C., and Glorius, F. (2020). A structure-based platform for predicting chemical reactivity. Chem *6*, 1379–1390. https://doi.org/10.1016/j.chempr.2020.02.017.

36. Strieth-Kalthoff, F., Sandfort, F., Segler, M.H.S., and Glorius, F. (2020). Machine learning the ropes: principles, applications and directions in synthetic chemistry. Chem. Soc. Rev. *49*, 6154–6168. https://doi.org/10.1039/C9CS00786E.

37. Artrith, N., Butler, K.T., Coudert, F.X., Han, S., Isayev, O., Jain, A., and Walsh, A. (2021). Best practices in machine learning for chemistry. Nat. Chem. *13*, 505–508. https://doi.org/10.1038/s41557-021-00716-z.

38. Keith, J.A., Vassilev-Galindo, V., Cheng, B., Chmiela, S., Gastegger, M., Müller, K.-R., and Tkatchenko, A. (2021). Combining machine learning and computational chemistry for predictive insights Into chemical systems. Chem. Rev. *121*, 9816–9872. https://doi.org/10.1021/acs.chemrev.1c00107.

39. Jorner, K., Tomberg, A., Bauer, C., Sköld, C., and Norrby, P.-O. (2021). Organic reactivity from mechanism to machine learning. Nat. Rev. Chem. *5*, 240–255. https://doi.org/10.1038/s41570-021-00260-x.

40. Engle, K.M., Thuy-Boun, P.S., Dang, M., and Yu, J.Q. (2011). Ligand-accelerated cross-coupling of C(sp^2)–H bonds with arylboron reagents. J. Am. Chem. Soc. *133*, 18183–18193. https://doi.org/10.1021/ja203978r.

41. Thuy-Boun, P.S., Villa, G., Dang, D., Richardson, P., Su, S., and Yu, J.Q. (2013). Ligand-accelerated Ortho-C–H alkylation of arylcarboxylic acids using alkyl boron reagents. J. Am. Chem. Soc. *135*, 17508–17513. https://doi.org/10.1021/ja409014v.

42. Salazar, C.A., Flesch, K.N., Haines, B.E., Zhou, P.S., Musaev, D.G., and Stahl, S.S. (2020). Tailored quinones support high-turnover Pd catalysts for oxidative C-H arylation with O_2. Science *370*, 1454–1460. https://doi.org/10.1126/science.abd1085.

43. Song, Y., Zheng, S., Niu, Z., Fu, Z.H., Lu, Y., and Yang, Y. (2020). Communicative representation learning on attributed molecular graphs. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, pp. 2831–2838. https://doi.org/10.24963/ijcai.2020/392.

44. Zheng, S., Rao, J., Zhang, Z., Xu, J., and Yang, Y. (2020). Predicting retrosynthetic reactions using self-corrected transformer neural networks. J. Chem. Inf. Model. *60*, 47–55. https://doi.org/10.1021/acs.jcim.9b00949.

45. Wang, H., Li, W., Jin, X., Cho, K., Ji, H., Han, J., and Burke, M.D. (2022). Chemical-reaction-aware molecule representation learning. In International Conference on Learning Representations. https://doi.org/10.48550/arXiv.2109.09888.

46. Ahneman, D.T., Estrada, J.G., Lin, S., Dreher, S.D., and Doyle, A.G. (2018). Predicting reaction performance in C-N cross-coupling using machine learning. Science *360*, 186–190. https://doi.org/10.1126/science.aar5169.

47. Reid, J.P., and Sigman, M.S. (2019). Holistic prediction of enantioselectivity in asymmetric catalysis. Nature *571*, 343–348. https://doi.org/10.1038/s41586-019-1384-z.

48. Zahrt, A.F., Henle, J.J., Rose, B.T., Wang, Y., Darrow, W.T., and Denmark, S.E. (2019). Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. Science *363*, eaau5631. https://doi.org/10.1126/science.aau5631.

49. Haghighatlari, M., Li, J., Heidar-Zadeh, F., Liu, Y., Guan, X., and Head-Gordon, T. (2020). Learning to make chemical predictions: the interplay of feature representation, data, and machine learning methods. Chem *6*, 1527–1542. https://doi.org/10.1016/j.chempr.2020.05.014.

50. Shields, B.J., Stevens, J., Li, J., Parasram, M., Damani, F., Alvarado, J.I.M., Janey, J.M.,

Adams, R.P., and Doyle, A.G. (2021). Bayesian reaction optimization as a tool for chemical synthesis. Nature *590*, 89–96. https://doi.org/10.1038/s41586-021-03213-y.

51. Żurański, A.M., Martinez Alvarado, J.I., Shields, B.J., and Doyle, A.G. (2021). Predicting reaction yields via supervised learning. Acc. Chem. Res. *54*, 1856–1865. https://doi.org/10.1021/acs.accounts.0c00770.

52. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626.

53. Rao, J., Zheng, S., and Yang, Y. (2021). Quantitative evaluation of explainable graph neural networks for molecular property prediction. Preprint at arXiv. https://doi.org/10.48550/arXiv.2107.04119.