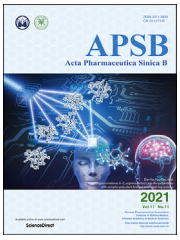
Acta Pharmaceutica Sinica B 2021;11(11):3678–3682

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com



SHORT COMMUNICATION

Preparing anti-SARS-CoV-2 agent EIDD-2801 by a practical and scalable approach, and quick evaluation *via* machine learning

Zhen Qin[a,†], Bin Dong[b,†], Renbing Wang[a], Dechun Huang[b], Jubo Wang[a,*], Xi Feng[a], Jinlei Bian[a,*], Zhiyu Li[a,*]

[a]*Jiangsu Key Laboratory of Drug Design and Optimization, Department of Medicinal Chemistry, School of Pharmacy, China Pharmaceutical University, Nanjing 211100, China*
[b]*Department of Pharmaceutical Engineering, School of Engineering, China Pharmaceutical University, Nanjing 211100, China*

Received 4 October 2021; received in revised form 11 October 2021; accepted 12 October 2021

KEY WORDS

EIDD-2801;
SARS-CoV-2;
Antiviral drug;
Decision tree;
Shapley value

Abstract EIDD-2801 is an orally bioavailable prodrug, which will be applied for emergency use authorization from the U.S. Food and Drug Administration for the treatment of COVID-19. To investigate the optimal parameters, EIDD-2801 was optimized *via* a four-step synthesis with high purity of 99.9%. The hydroxylamination procedure was telescoped in a one-pot and the final step was precisely controlled on reagents, temperature and reaction time. Compared to the original route, the yield of the new route was enhanced from 17% to 58% without column chromatography. The optimized synthesis has been successfully determinated on a decagram scale: the first step at 200 g and the final step at 20 g. Besides, the relationship between yield and temperature, time, and reagents in the deprotection step was investigated *via* Shapley value explanation and machine learning approach-decision tree method. The results revealed that reagents have the greatest impact on yield estimation, followed by the temperature.

*Corresponding authors. Tel.: +86 15151865295 (Jinlei Bian), +86 13951678592 (Zhiyu Li).
E-mail addresses: 1620194588@cpu.edu.cn (Jubo Wang), bianjl@cpu.edu.cn (Jinlei Bian), zhiyuli@cpu.edu.cn (Zhiyu Li).
[†]These authors made equal contributions to this work.
Peer review under responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences

https://doi.org/10.1016/j.apsb.2021.10.011

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19 has severely impacted public health and economy worldwide[1,2]. Currently, some promising repurposed drugs have been approved or studied in large clinical trials for COVID-19 patients[3–5]. EIDD-2801 (**1**, molnupiravir, MK-4482, Fig. 1), an oral prodrug of β-D-*N*4-hydroxycytidine (**2**, NHC, EIDD-1931, Fig. 1), has proved to be effective against SARS-CoV-2 in clinic trials[6,7]. It will be applied for emergency use authorization from the U.S. Food and Drug Administration (FDA) by Merck & Co[6]. It has been demonstrated that **1** has broad-spectrum RNA viruses inhibitory activity by blocking RNA polymerase, especially inhibiting SARS-CoV-2 replication[8,9].

In 2019, Painter and colleagues[10] from Emory University firstly disclosed a synthesis of **1** in a 25 g scale (Scheme 1). The starting material uridine was protected with acetone fork, followed by esterification in a one-pot reaction. Then, the chemical intermediate is activated with the 1,2,4-triazole, which is further transformed into hydroxylamine. After deprotection of the acetone fork, **1** is obtained but no yield has been provided in this step. The low overall yield (not more than 17%) limits the further development. And the intermediates were purified by column chromatography. These challenges of the above route require a simplified and alternative route with improved yield. Thus, we identify an optimized route to afford **1** with simplified procedures and high yields (58% yield) (Scheme 2). Furthermore, process parameters such as reaction concentration, catalyst loading, and stoichiometry has been optimized. Uridine is selected as starting material, followed by acetone fork protection, *O*-acylation, hydroxylamination and deprotection. The 2,4,6-triisopropylbenzenesulfonyl chloride is used to sulfonate 4-carbonyl, followed by reaction with hydroxylamine. As machine learning can evaluate and predict the performance and outcome of a synthetic reaction effectively, we also develop a machine learning model to investigate the importance of solvent, temperature, and reaction time with the yield of **1**[11,12].

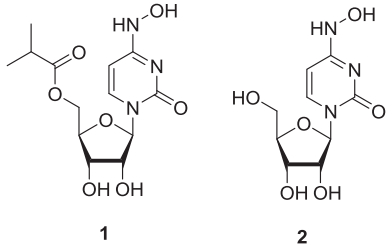


Figure 1 Structures of EIDD-2801 and NHC.

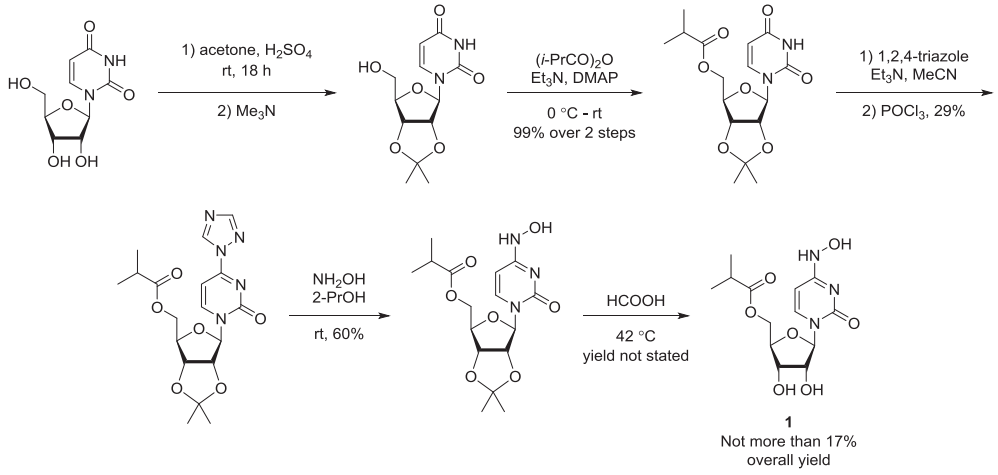
2. Results and discussion

2.1. Chemistry

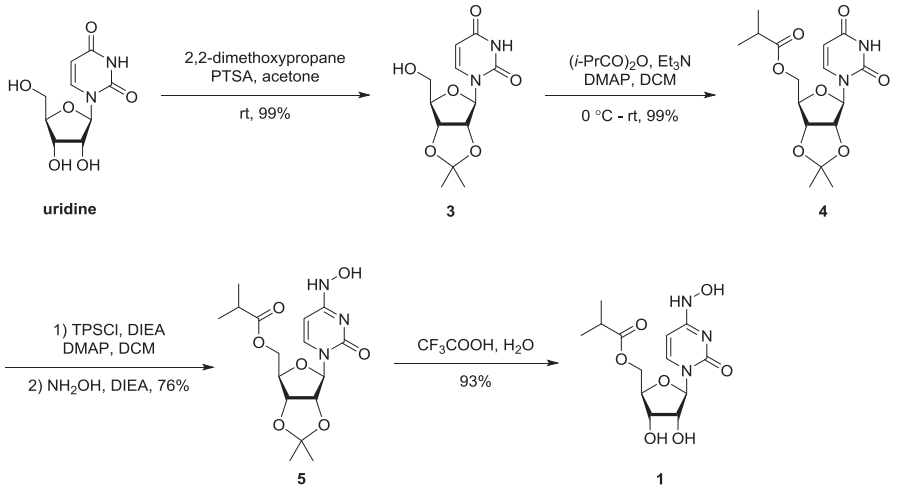
Commercially available uridine was converted to the 2′,3′-*O*-isopropylidine **3** at room temperature. The synthesis of **3** was started with 1.5 equiv. of 2,2-dimethoxypropane and 0.05 equiv. of toluene-4-sulfonic acid monohydrate (PTSA·H_2O). After solvent removed, crude protection product was dissolved in methyl *tert*-butyl ether (MTBE) and filtered to remove PTSA. The compound **3** was obtained in 99% yield.

The original route in a patent reported that the compound **3** was acylated using 8.5 equiv. of isobutyric anhydride and 0.39 equiv. of 4-dimethylaminopyridine (DMAP)[10]. Owing to the elimination of redundant reagents, a large number of work was increased. Hence, we investigated the interaction between the equiv. of isobutyric anhydride and conversion (Fig. 2). The 1.0 equiv. and 1.1 equiv. of isobutyric anhydride gave a poor yield. A 1.5 equiv. of isobutyric anhydride provided a complete reaction in a 99% yield of **4**. The 2.0 equiv. and 8.5 equiv. had similar yield to that of 1.5 equiv. It's necessary to remove residual isobutyric anhydride by stirring with petroleum ether to preclude the negative effect of the next step. We also attempted to apply isobutyryl chloride as acylating agent, but it was tended to react with 2- and 4-carbonyl.

Hydroxylamination of **4** to afford **5** was collected by sulfonylation with **4** followed by reacting with hydroxylamine. Compound **4** was sulfonylated at 4-carbonyl using 2,4,6-triisopropylbenzenesulfonyl chloride (TPSCl), *N*,*N*-diisopropylethylamine (DIEA) and a catalytic amount of DMAP in methylene chloride to improve the nucleophilicity of the C-4 carbon that reacted in the hydroxylamination. Sulfonylated intermediate was reacted with hydroxylamine hydrochloride using a stoichiometric amount of DIEA. After quenched, it is worthy to wash off residual **4** with potassium carbonate solution in the layer of



Scheme 1 Initial routes to EIDD-2801.



Scheme 2 Improved route to EIDD-2801.

dichloromethane (DCM). Then, crude product was stirred with isopropyl ether in a 94.8% purity to eliminate impurity that was otherwise found to influence the subsequent deprotection.

To optimize the reaction condition of deprotection of **5**, different reagents, temperature and reaction time were investigated (Supporting Information Table S1). We found that reaction could proceed smoothly in CF_3COOH and $PTSA \cdot H_2O$. The yield of the former could reach 97.6% but the latter could only reach 92.9% (8 h, 28 °C), which were shown in entry 8 and entry 44. Solvents such as hydrochloric acid gave no conversion at 0 °C (entry 16). When increased the temperature (entries 17–24), there were some undetermined impurities. In certain reaction, the product only occupied a small fraction (entry 5 and entry 16). The remaining parts mainly included starting material that was not consumed. Taken purity and operation into consideration, we chose CF_3COOH to convert **5**. This combination of reagents gave **1** with good reproducibility, as well as improved yield and purity.

2.2. *Decision Tree Regressor model*

To forecast the final step based on experimental settings, we constructed a machine learning approach-Decision Tree Regressor (DTR) model to study the interaction between yield and temperature, reaction time, and reagents, and to reveal which factor was the most significant in the determination of yield. The motivation of employing DTR was two folds: (1) the DTR model could be used to predict the yield of EIDD-2801 by using a number of well-designed experimental scenarios with a balanced input distribution. (2) The DTR model had great interpretability and outputted a set of non-overlapping decision rules that could be easily understood. As a result, the type of reagents, temperature and reaction time were employed as independent variable (features) in the DTR model, while the yield as dependent variable (target). While using hyper-parameters tuned by grid search, performance metric of Mean Absolute Error (MAE) = −1.68 was the average results of 20 simulations. Moreover, to avoid overfitting and produce a generalized model, the value of minimum samples per leaf node and the maximum depth of tree was set to 3 and 5, respectively.

The best fitted decision tree was shown in Supporting Information Fig. S1, with each node visualising how decision nodes split up the feature space. The yield of each experiment was depicted as scatters, with their average values displayed as dotted lines. The DTR fits the data well, with the exception of nodes 4 and 10, due to the minimum sample size per node was set to 3 to prevent overfitting. Here, a run from the root node (node 0) down to leaf nodes 3 and 4 was used to explain the decision tree. Starting at the root node, experiments were split by the type of reagents. Experiments with the solvents HCl/MeOH and $HCOOH/H_2O$ were placed in node 1, whereas those with CF_3COOH/H_2O and $PTSA \cdot H_2O$/THF were placed in node 8. The plot on node 0 demonstrates that, on average, when using CF_3COOH/H_2O and $PTSA \cdot H_2O$/THF as solvent yield over 90%; while, using HCl/MeOH and $HCOOH/H_2O$ yield around 60%. The experiments in node 1 were then further divided by temperature. Experiments with temperatures less than 28 °C inclusive were grouped in node 2, while those with temperatures greater than 50 °C inclusive were grouped in node 5. Consequently, experiments involving an HCl/MeOH or $HCOOH/H_2O$ and a temperature less than 28 °C inclusive were classified into node 2. The final decision was determined based on the experiment's duration. If the period was less than 1 h inclusive, the yield would be 23.07% on average of 8 samples that fell into node 3. If this was not the case, the yield would be 57.84%.

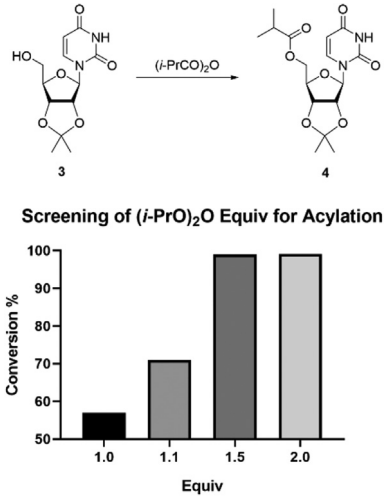


Figure 2 Equiv. of isobutyric anhydride screened for the acylation of **3** (conditions: 1.0 equiv. of **3**, 2.0 equiv. of Et_3N, 0.1 equiv. of DMAP).

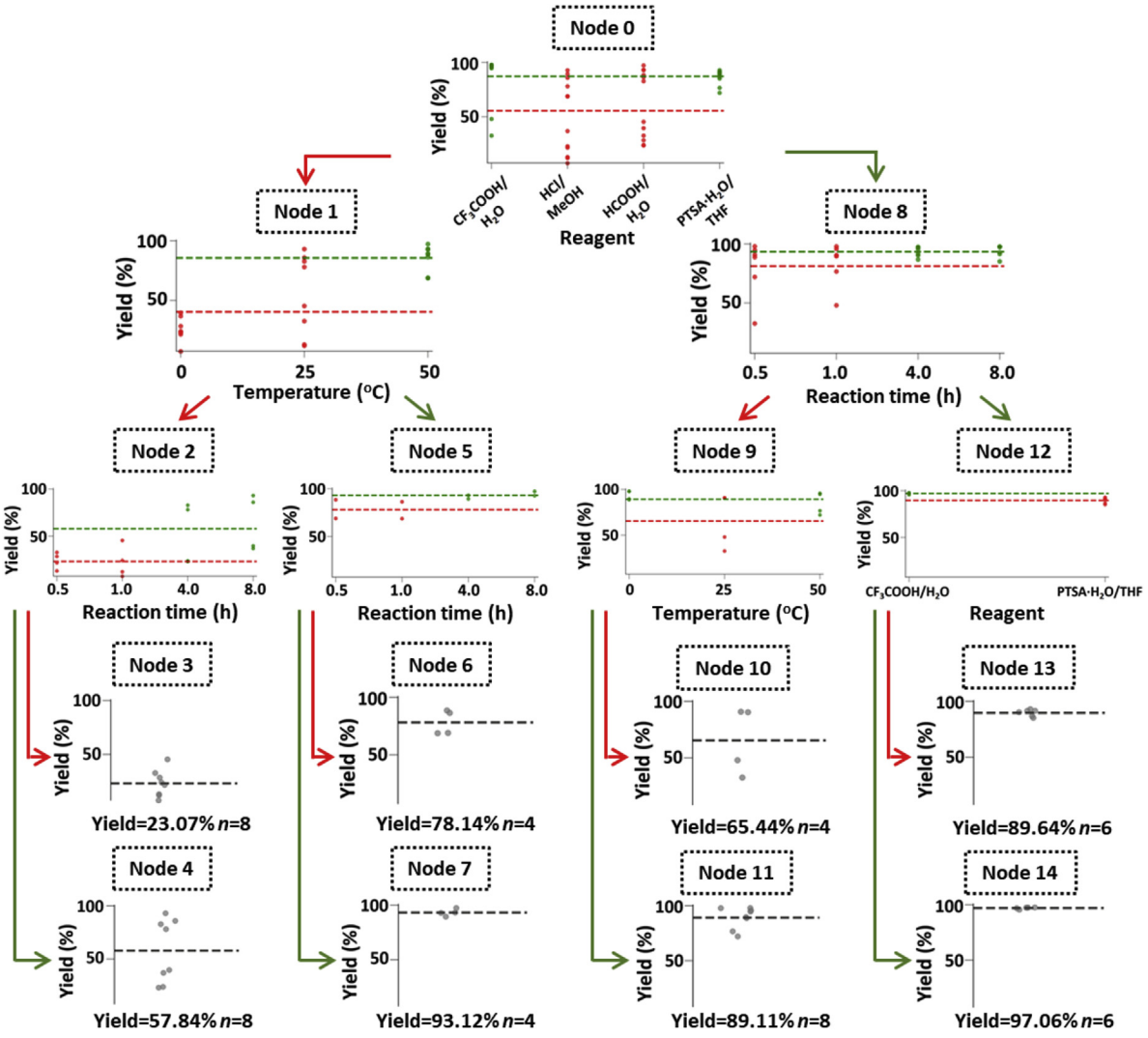


Figure 3 The best fitted Decision Tree Regressor (DTR) to study the relationship between yield and temperature, reaction time, and reagents (the nodes are used to split up the feature space).

2.3. *Shapley value explanation*

Furthermore, Shapley value explanation was applied in this study to evaluate the importance of temperature, time, and solvents as features. Shapley value has been widely used as local explanations of predictions from machine learning models, and computed by inputting different features into a conditional expectation function of the model's output[13]. Fig. 4 shows the proportion of relative Shapley importance derived by dividing each unique absolute Shapley value by the sum of all absolute Shapley values. The findings revealed that reagents had the greatest impact on yield estimation (56.7%), followed by the temperature (36.6%) and reaction time (6.7%). Shapley value was used instead of information gain (*e.g.*, Gini or MSE) approach which could also be used to explain feature importance for DTR, due to the gain method was biased to attribute more importance to lower splits. In other words, since trees were constructed greedily, it was expected features near the root of the tree to be more important than features split on near the leaves. This bias potentially leads to inconsistency using the information gain-based method to determine feature importance. For the decision tree in Fig. 3, feature solvent was applied to divide the tree not just on node 0 but also on node 12, so aggravating the bias. The Shapley importance, on the other hand, reduced this bias because it was mathematically equivalent to average differences in predictions over all possible orderings of the features, rather than just the ordering specified by their position in the tree.

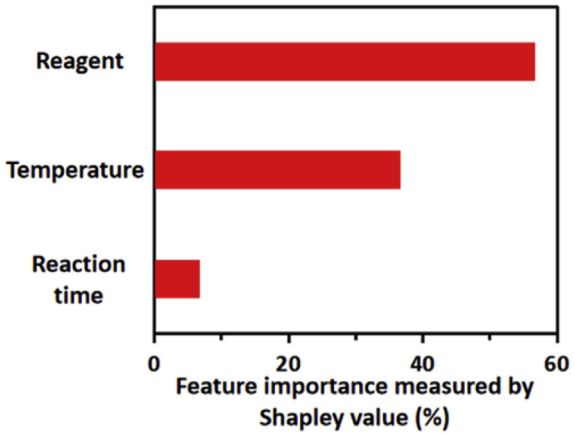


Figure 4 Feature importance of reagent, temperature and time measured by SHAP value.

3. Conclusions

In summary, an optimized and practical method for the preparation of oral drug EIDD-2801 (**1**) to treat COVID-19 has been developed. The method from available uridine to EIDD-2801 exhibits easy process with overall yield of 58% and high purity of 99.9%. A notable improvement is that the hydroxylamination was telescoped in one step, comparing to originally reported two steps in a patent. Furthermore, a problematic deprotection of the 2′,3′-*O*-isopropylidine intermediate **5** by acidolysis was carried out, which was precisely controlled on reagents (CF_3COOH/H_2O, HCl/MeOH, $HCOOH/H_2O$, and $PTSA\cdot H_2O$/THF), temperature (0, 28 and 50 °C) and reaction time (0.5, 1, 4 and 8 h). Furthermore, the relationship among yield, temperature, time, and reagents was also investigated *via* Shapley value explanation and DTR model to understand the critical role of different factors on the yield of the key step for the synthesis of EIDD-2801. Although we don't use machine learning method to optimize the experiment in this paper, this study will allow us to construct experiments more efficiently and further improve the yield of EIDD-2801 in our future work for large-scale preparation.

Acknowledgments

We gratefully acknowledge financial support from the National Natural Science Foundation of China (no. 81703347, 21672260, and 21372260), the Natural Science Foundation of Jiangsu Province of China (no. BK20170743 and BK20171393), the "Double First-class" University Project (CPU2018GY07, China), the State Key Laboratory of Drug Research (SIMM1903KF-03, China), and the Undergraduate Innovation and Entrepreneurship Training Program (202110316002Z, China).

Author contributions

Jinlei Bian and Zhiyu Li designed this project. Jinlei Bian, Jubo Wang, Dechun Huang and Zhiyu Li supervised this project. Zhen Qin, Jubo Wang, Renbing Wang and Xi Feng performed the chemical synthesis. Bin Dong and Dechun Huang performed the machine learning. All authors analyzed the data and wrote the manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supporting information

Supporting information to this article can be found online at https://doi.org/10.1016/j.apsb.2021.10.011.

References

1. Wahl A, Gralinski LE, Johnson CE, Yao W, Kovarova M, Dinnon 3rd KH, et al. SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature* 2021;**591**:451−7.
2. Gil C, Ginex T, Maestro I, Nozal V, Barrado-Gil L, Cuesta-Geijo MA, et al. COVID-19: drug targets and potential treatments. *J Med Chem* 2020;**63**:12359−86.
3. De Savi C, Hughes DL, Kvaerno L. Quest for a COVID-19 cure by repurposing small-molecule drugs: mechanism of action, clinical development, synthesis at scale, and outlook for supply. *Org Process Res Dev* 2020;**24**:940−76.
4. Hughes DL. Quest for a cure: potential small-molecule treatments for COVID-19, part 2. *Org Process Res Dev* 2021;**25**:1089−111.
5. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-CoV and naming it SARS-CoV-2. *Nat Microbiol* 2020;**5**:536−44.
6. Toots M, Yoon JJ, Cox RM, Hart M, Sticher ZM, Makhsous N, et al. Characterization of orally efficacious influenza drug with high resistance barrier in ferrets and human airway epithelia. *Sci Transl Med* 2019;**11**:eaax5866.
7. Merck Sharp & Dohme Corp. Study of MK-4482 for prevention of coronavirus disease 2019 (COVID-19) in adults (MK-4482-013) (MOVe-AHEAD). ClinicalTrials.gov identifier: NCT04939428. Available from: https://clinicaltrials.gov/ct2/show/study/NCT04939428. Accessed July 2021.
8. Toots M, Yoon JJ, Hart M, Natchus MG, Painter GR, Plemper RK. Quantitative efficacy paradigms of the influenza clinical drug candidate EIDD-2801 in the ferret model. *Transl Res* 2020;**218**:16−28.
9. Yoon JJ, Toots M, Lee S, Lee ME, Ludeke B, Luczo JM, et al. Orally efficacious broad-spectrum ribonucleoside analog inhibitor of influenza and respiratory syncytial viruses. *Antimicrob Agents Chemother* 2018;**62**:e00766.
10. Painter GR, Bluemling GR, Natchus MG, Guthrie D, Inventors. Emory University, assignee. *N*4-hydroxycytidine and derivatives and antiviral uses related thereto. WO2019113462A1, 2019.
11. Ahneman DT. Predicting reaction performance in C–N cross-coupling using machine learning. *Science* 2018;**360**:613.
12. Jethava KP, Fine J, Chen Y, Hossain A, Chopra G. Accelerated reactivity mechanism and interpretable machine learning model of n-sulfonylimines toward fast multicomponent reactions. *Org Lett* 2020;**22**:8480−6.
13. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intel* 2020;**2**:56−67.