Article

Chemoenzymatic Synthesis Planning Guided by Reaction Type Score

Hongxiang Li, Xuan Liu, Guangde Jiang, and Huimin Zhao*

Cite This: *J. Chem. Inf. Model.* 2024, 64, 9240–9248

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Thanks to the growing interest in computer-aided synthesis planning (CASP), a wide variety of retrosynthesis and retrobiosynthesis tools have been developed in the past decades. However, synthesis planning tools for multistep chemoenzymatic reactions are still rare despite the widespread use of enzymatic reactions in chemical synthesis. Herein, we report a reaction type score (RTscore)-guided chemoenzymatic synthesis planning (RTS-CESP) strategy. Briefly, the RTscore is trained using a text-based convolutional neural network (TextCNN) to distinguish synthesis reactions from decomposition reactions and evaluate synthesis efficiency. Once multiple chemical synthesis routes are generated by a retrosynthesis tool for a target molecule, RTscore is used to rank them and find the step(s) that can be replaced by enzymatic reactions to improve synthesis efficiency. As proof of concept, RTS-CESP was applied to 10 molecules with known chemoenzymatic synthesis routes in the literature and was able to predict all of them with six being the top-ranked routes. Moreover, RTS-CESP was employed for 1000 molecules in the boutique database and was able to predict the chemoenzymatic synthesis routes for 554 molecules, outperforming ASKCOS, a state-of-the-art chemoenzymatic synthesis planning tool. Finally, RTS-CESP was used to design a new chemoenzymatic synthesis route for the FDA-approved drug Alclofenac, which was shorter than the literature-reported route and has been experimentally validated.
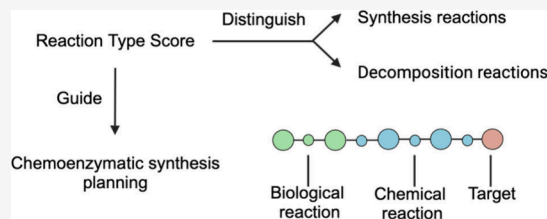
INTRODUCTION

Organic synthesis has been renowned for its long history and regarded as the primary choice in synthesizing target molecules for drugs,[1,2] materials[3] and natural products[4] for years. However, with the rapid development of biocatalysis and directed evolution in the past decade,[5] various new transformations were imported into the traditional organic synthesis space.[6−8] Owing to its high stereoselectivity and regioselectivity, enzymatic catalysis has been increasingly employed to replace chemical reactions in synthesis routes and selectively produce the desired products.[9,10] Since one enzyme might replace multiple chemical reactions, the chemoenzymatic synthesis route can be shorter and with higher yields. For instance, in the synthesis of sitagliptin,[11] the overall hybrid synthesis route was shortened by three steps compared to the original chemical synthesis route and the enantiomeric excess was also increased with enzyme catalysis. Moreover, biocatalysis uses mild conditions and avoids toxic reagents or high pressure, facilitating green chemistry,[12,13] while the one-pot enzymatic cascades help reduce the purification between reactions.[14−17] Not surprisingly, chemoenzymatic synthesis has been applied to multiple practically important small molecules.[18−21]

Other than human efforts and expertise in designing synthesis routes, the computer-aided synthesis planning (CASP) tools have been explored for both organic and enzymatic synthesis.[22−25] CASP is a process of breaking down target molecules step by step, using either machine learning (ML) or rule-based methods, until reaching commercially available molecules. The rule-based CASP tools apply well-defined reaction rules (also called templates) to each target molecule and generate corresponding precursors (ML algorithms can be used for selecting templates) while the ML-based CASP tools predict the precursors using algorithms trained on reaction databases. Synthia[25] is a commercial retrosynthesis tool using chemist curated rules, while Aizynthfinder[26] uses abundant extracted rules from the Reaxys database[27] and employs Monte Carlo tree search (MCTS) to save calculation time. For retrobiosynthesis, novoStoic[28] focuses on enzymes in metabolic engineering and aims to synthesize target molecules from metabolites, while Retro-BioCat[29] specializes in enzymatic cascades for *in vitro* synthesis and uses curated reaction rules for prediction. As the first ML-based retrobiosynthesis tool, the IBM's RXN4Chemistry[30] adapts transformer with ECREACT database[30] and can potentially predict novel reactions.

However, CASP tools for chemoenzymatic synthesis planning remain rare despite that many chemoenzymatic
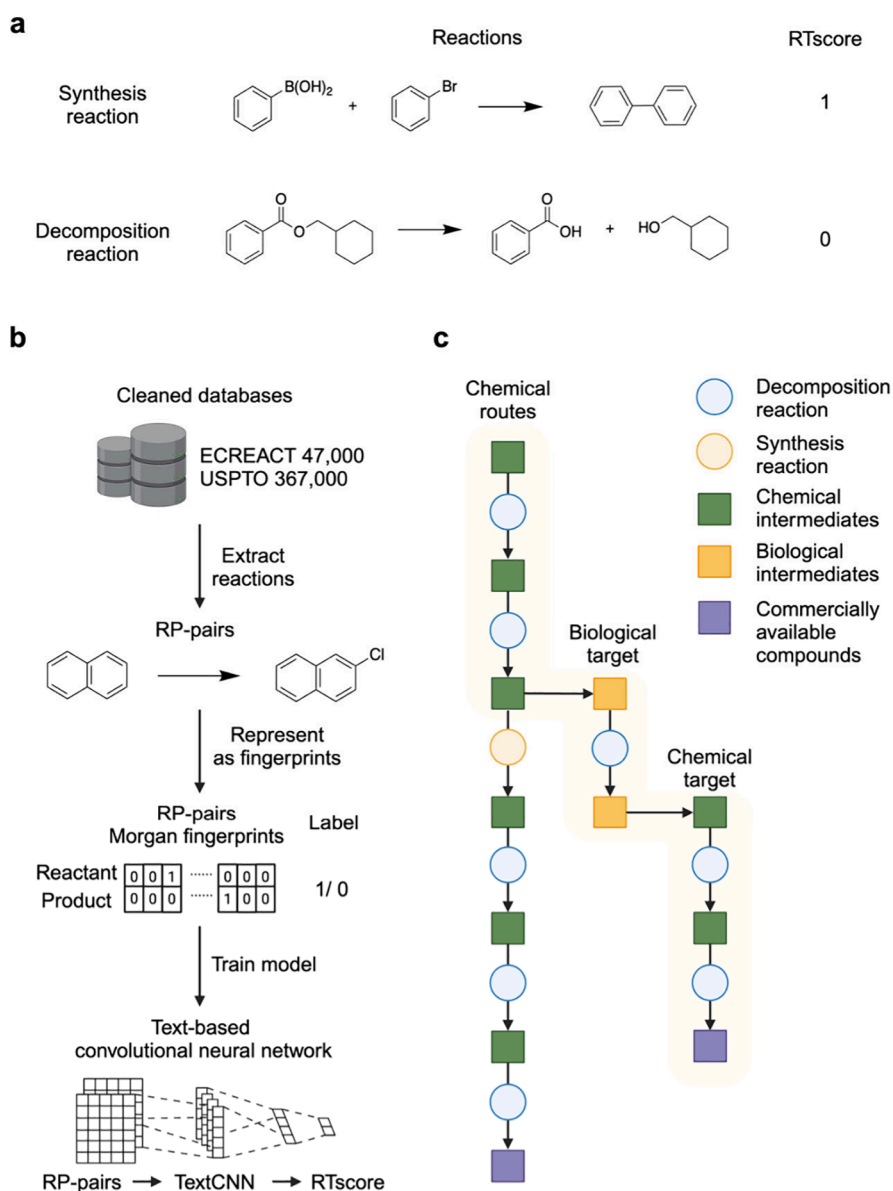
Figure 1. RTscore training process and searching algorithm. (a) Examples of synthesis reactions and decomposition reactions. In the process of breaking down the target molecules in a retrosynthesis manner, the decomposition reactions are preferred. (b) Procedure for training the RTscore model. The USPTO and ECREACT databases were first cleaned to remove cofactors in reactions; then, the major reactant-product pairs (RP-pairs) were extracted and represented as Morgan fingerprints with labels 1 for a synthesis reaction and 0 for a decomposition reaction separately. With a text-based convolutional neural network, the model predicts RTscore with the input of RP-pairs. (c) Hybrid synthesis search algorithm. The chemical synthesis route is generated for the target molecule at first and RTscore$_{(USPTO)}$ is employed to find the reaction that does not break down the target molecule effectively and searches the product molecule in that reaction with retrobiosynthesis tools; then, the biological precursor is searched again with the chemical method to form the chemoenzymatic synthesis routes.

synthesis routes have been reported. To the best of our knowledge, there are only three chemoenzymatic synthesis planning tools reported so far. In 2022, Coley and co-workers designed ASKCOS (hybrid version),[31] which extracts templates from the Reaxys (chemical) and BKMS (biological) databases,[32] and then uses two prioritizers to rank those templates in each step and expand the search tree. Since the prioritizers are based on reaction templates, this tool is limited to only rule-based methods. Moreover, because of their step by step searching algorithm, there could be multiple transitions between chemical and biological reactions in their predicted routes, increasing difficulties in experiments. Later, Jensen and co-workers[33] developed an alternative tool in which they first used ASKCOS to generate chemical synthesis routes and then identified enzymes to carry out the same transformation for every step using templates in RetroBioCat. This tool did not aim to predict new enzymatic transformations, and it is time-consuming to exhaustively search for alternative enzymatic reactions to replace every chemical step in the original routes. Recently, Wu and co-workers developed BioNavi[34] which only used ML-based methods, and their biological reaction predictor was specifically trained for natural product synthesis. Notably, in the first two studies, no experimental validation was conducted for the newly proposed reactions, and only partial synthesis routes were validated in BioNavi.

In this work, we have developed a reaction type score (RTscore)-guided chemoenzymatic synthesis planning (RTS-CESP) strategy (Figure 1). The existing retrosynthesis and retrobiosynthesis tools (e.g., Aizynthfinder,[26,29] RXN4Chem-istry[30,35]) were employed to identify chemical and biological reactions, which were then integrated by a custom-designed automatic searching algorithm to generate chemoenzymatic synthesis routes for the target molecules. To start with, an RTscore is designed by training a text-based convolutional neural network (TextCNN) to distinguish synthesis reactions from decomposition reactions, which could achieve an F1 score of 0.971 using the USPTO data set.[36] For each target molecule, the chemical synthesis route is generated first, and then RTscore is applied to every step to predict reaction type and determine which steps should be replaced by enzymatic reactions. As proof of concept, we used this tool to predict the synthesis routes for 10 molecules with known chemoenzymatic synthesis routes in literature and found all these reported chemoenzymatic synthesis routes were successfully predicted and six of them were ranked as the most preferred routes. Furthermore, a validation set with 1000 molecules was adopted to evaluate the performance of this tool on a large scale. Among the 1000 molecules from the boutique database,[37] our tool could predict chemoenzymatic synthesis routes for 554 molecules, versus 493 molecules by the state-of-art tool named ASKCOS, and we proposed shorter pathways for 30.2% molecules that were found by both tools. In addition to reproducing literature results, we predicted a novel shorter pathway for an FDA-approved drug Alclofenac[38] and experimentally validated the synthesis route with 69% yield.

■ RESULTS

RTscore for Ranking Synthesis Routes and Guiding Hybrid Synthesis. Retrosynthesis tools are used to break down target molecules into commercially available compounds through multiple steps. To improve synthesis efficiency, each step should ideally decompose the target molecules into simpler molecules rather than synthesizing more complex products (Figure 1a). The synthesis efficiency of each reaction can be evaluated based on several factors, including changes in molecular weight and atom numbers of the reactants and products, the introduction of chirality, and the number of references supporting the reaction. We processed the database to retain reactions where the product has a larger molecular weight or atom number than the reactant. Then we defined reactions as recorded in the database (the reaction from reactant to product) as synthesis reactions, and the reversed reaction (from product to reactant) as decomposition reactions. We designed an RTscore using a text-based convolutional neural network (TextCNN) to distinguish these two reaction types, while also learning chirality changes and synthesis preferences from the database (Supplementary Figure 1). RTscore was applied to each step of the synthesis routes. After summing the scores of all reactions, the entire route can be scored and ranked. Additionally, in each route, the reaction with the lowest score is replaced by an alternative biological step to generate hybrid synthesis routes. Finally, hybrid routes with fewer steps and faster access to commercially available materials are selected as the top options.

Since chemical and biological reactions occupy different reaction spaces, we trained two separate scores using the USPTO (chemical) and ECREACT (biological) databases,

forming RTscore$_{(USPTO)}$ and RTscore$_{(ECREACT)}$, respectively (Figure 1b). Our training input comprised the extracted reactant-product pairs (RP-pairs) from both databases, processed as follows. First, common cofactors were removed to avoid interference (Supplementary Figure 2), as our goal was to rank the routes based on the transformation between the major reactant and product rather than predicting a complete reaction. Next, since all the reactions in both databases contain only a single product, it was regarded as the major product. For reactions with multiple reactants, we calculated the fingerprint similarity of all reactants to the major product and the reactant with the highest similarity score to the major product was selected as the major reactant. In addition, we canonicalized all SMILES strings in the data set, including tautomer canonicalization, to ensure consistent representation of molecular structures and deduplicated the data set to remove redundant entries. Finally, we retained the reactions where the major product had a larger molecular weight or more atoms than the major reactant and discarded the others. The retained reactions were labeled as synthesis reactions, while the reversed reactions labeled as decomposition reactions. This systematic approach allowed us to construct a comprehensive data set containing 367,000 chemical and 47,000 enzymatic RP-pairs.

The differentiation of reaction type was defined as a classification problem and the data set was split into training, validation, and test sets (8:1:1). A performance comparison of four models—Random Forest (RF),[39] Support Vector Machine (SVM),[40] Feedforward Neural Network (FNN),[41] and Text-based Convolutional Neural Network (TextCNN)[42,43]—was conducted using the ECREACT data-base. The models were evaluated using both the F1 score and the Matthews Correlation Coefficient (MCC), with the TextCNN model achieving the highest values for both metrics (Supplementary Figure 3). Therefore, the TextCNN model was selected for further study. A two-layer TextCNN was employed, utilizing ReLU (Rectified Linear Unit) activation functions after each convolutional and fully connected layer, except for the final layer which uses a sigmoid activation function. In addition, we evaluated the model's performance using both binary and count-based fingerprints, with the count-based fingerprint achieving higher F1 and MCC scores (Supplementary Figure 4). To further improve the model, we performed hyperparameter tuning using a grid search approach with 5-fold cross-validation. The hyperparameter grid included output channels of 2, 3, and 4 for the first and second convolutional layers and hidden sizes of 256, 512, and 1024 for the fully connected layers that follow the convolutional layers. The heatmaps for F1 scores on the biological and chemical data sets illustrate the performance of different hyperparameter combinations. For the biological data set ECREACT, the best F1 score of 0.921 was achieved with a hidden size of 512 and output channels of 4 (Supplementary Figure 5). For the chemical data set USPTO, the highest F1 score of 0.971 was obtained with a hidden size of 256 and output channels of 3 and 4 for the first and second convolutional layer separately (Supplementary Figure 6). To further assess the model's robustness, we evaluated its performance using 10 different random seeds, which demonstrated stability across runs with a standard deviation of 0.0034 for the ECREACT database and 0.00089 for the USPTO database, confirming the model's reliability under different initializations (Supplementary Figure 7). After that, we evaluated RTscore$_{(ECREACT)}$ using the BKMS
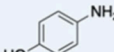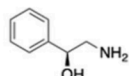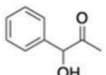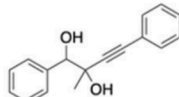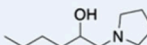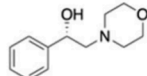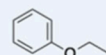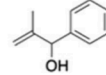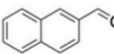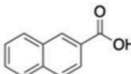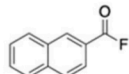
Entry	Hybrid routes					Rank

Figure 2. Validation with literature-reported hybrid synthesis routes. For 10 selected target molecules, we searched for literature reactions to build a hybrid synthesis route data set in which each route contains both chemical and enzymatic reactions. All the hybrid synthesis routes were successfully reproduced by our tool, and when ranked with RTscore, 6 out of 10 routes were the highest among all the predicted routes for a target molecule.

database,[32] an external enzymatic database. We first deduplicated the reactions with the ECREACT database and selected the major RP-pairs, then $RTscore_{(ECREACT)}$ was used to predict reaction type and reached 75% accuracy, while the SCScore[44] reached only 64% accuracy (Supplementary Figure 8).

Multistep Searching Algorithm. To achieve effective chemoenzymatic synthesis by integrating retrosynthesis and retrobiosynthesis tools, we designed an automatic hybrid synthesis search algorithm (Figure 1c). In this algorithm, a target molecule is first input into a retrosynthesis tool, and $RTscore_{(USPTO)}$ is used to rank the predicted routes. Since $RTscore_{(USPTO)}$ can be calculated for each reaction, the reaction with the worst score in each route is selected for improvement. The product molecule in that reaction is then input into retrobiosynthesis tools to find alternative enzymatic

Figure 3. Validation and benchmark study on a large-scale data set. (a) Benchmark with ASKCOS on the 1000 target molecules from the boutique database using same calculation time (three minutes) and stock molecules. (b) Analysis of the molecules whose chemoenzymatic synthesis routes were predicted by RTS-CESP and ASKCOS. (c) Comparison of the lengths of synthesis routes predi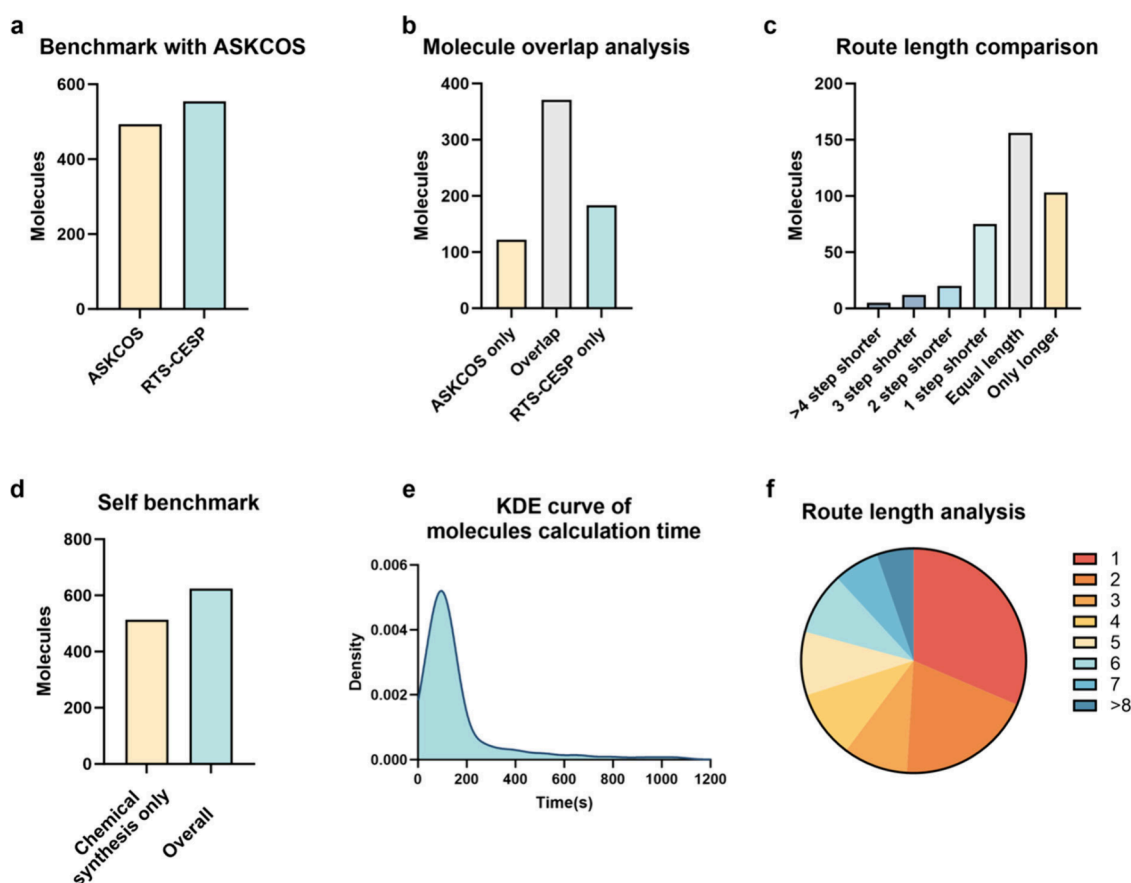cted by RTS-CESP and ASKCOS. RTS-CESP identified more shorter synthesis routes than ASKCOS. (d) Using the hybrid synthesis search algorithm of RTS-CESP identified synthesis routes for more molecules than using the chemical synthesis search algorithm only. (e) The kernel density estimation (KDE) curve for molecule calculation time. (f) Distribution of the lengths of predicted synthesis routes.

reactions. All available enzymatic templates are explored due to their limited number. $RTscore_{(ECREACT)}$ is then applied to rank the enzymatic reactions and identify intermediates suitable for chemical synthesis. Finally, the hybrid synthesis routes are collected, and the RTscore for each reaction in the entire route is added up to rank all the predicted routes. The interface between the different tools is fully programmatic. Once a target molecule is input, the algorithm automatically generates a list of chemoenzymatic synthesis routes with ranking.

Validation of RTS-CESP Using a Data Set Containing Known Hybrid Synthesis Routes. To determine the prediction accuracy of RTS-CESP, we sought to use the hybrid synthesis routes reported in literature as a test case. We chose 10 target molecules and discovered the corresponding synthesis reactions in literature from a chemoenzymatic synthesis routes data set.[45−62] The retrosynthesis tools used in this task were RXN4Chemistry (chemical) and RXN4Chemistry (enzymatic), while RTscore was used to guide the search and rank all the predicted chemoenzymatic synthesis routes. As shown in Figure 2, all the reported chemoenzymatic synthesis routes for these 10 molecules were identified and six of them were ranked as the top by RTscore, indicating that RTscore can be used to prioritize the chemoenzymatic synthesis routes and help researchers choose the most efficient routes.

Large-Scale Validation and Benchmark of RTS-CESP with ASKCOS. To evaluate the performance of RTS-CESP on a large scale, we selected 1000 molecules from the boutique database as target molecules and used Aizynthfinder and RetroBioCat for synthesis planning. Aizynthfinder utilized rules extracted from the Reaxys database and performed well in breaking down complex molecules, so it was used to generate the chemical synthesis routes for a target molecule at first. We then applied the biological reaction templates from the RetroBioCat database to intermediates in these chemical synthesis routes to predict enzymatic reactions. For the target molecules that Aizynthfinder failed to generate a complete chemical synthesis route (a route from target to stock molecules), we selected the intermediates in the generated partial route as targets for biological reactions. Finally, the biological precursors were searched with Aizynthfinder again to finish the whole routes and we ranked all the synthesis routes using RTscore.

Using the same calculation time (three minutes) and stock molecules[31] (i.e., molecules less than \$100/g from eMolecules and Sigma-Aldrich) (Supplementary Table 1), RTS-CESP predicted the chemoenzymatic synthesis routes for 554 molecules, while the state-of-the-art tool ASKCOS hybrid predicted the chemoenzymatic synthesis routes for 493 molecules (Figure 3a) (See Supplementary Figure 9 for examples of chemoenzymatic synthesis routes to target
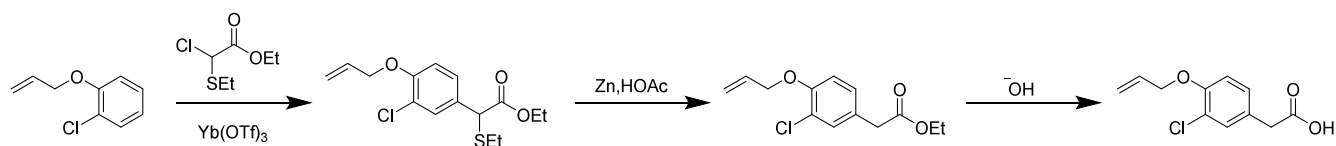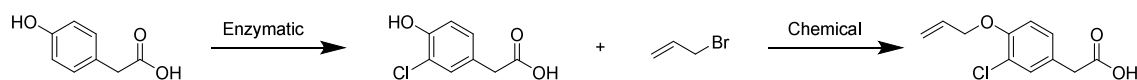
a Literature–reported synthesis route:

b Predicted route:

Figure 4. Experimental validation of the predicted synthesis route for an FDA-approved drug. (a) Literature-reported synthesis routes for Alclofenac. (b) Our predicted route. Our predicted route is one step shorter than the literature-reported route and starts with a cheaper precursor.

molecules identified by our tool but not by ASKCOS). There were 371 molecules that both tools have identified synthesis routes for, among which RTS-CESP predicted shorter synthesis routes for 112 of them (30.2%) (Figure 3b, c). Besides, as a self-benchmark, employing the chemoenzymatic search algorithm could predict synthesis routes for more molecules than using the chemical synthesis algorithm only (Figure 3d). The calculation time and route length varied for different target molecules (Figure 3e, f). Most target molecules were solved within three minutes, and more than half of the routes consisted of one or two steps, showing RTS-CESP's effectiveness in planning synthesis routes. Moreover, RTS-CESP could find routes with more than eight steps, indicating its competence in breaking down complex molecules.

Experimental Validation of the Predicted Synthesis Route for an FDA-Approved Drug. Alclofenac is an FDA (Food and Drug Administration) approved anti-inflammatory drug. To the best of our knowledge, only chemical methods were used to synthesize this compound and the shortest synthesis route in the literature consisted of three steps (Figure 4).[37] Here, we used RTS-CESP to predict synthesis routes for this target compound. Aizynthfinder was used to generate the chemical synthesis route first, and templates from RetroBioCat and BKMS databases were used to generate the enzymatic step. A synthesis route with only two steps was found, including one chemical and one enzymatic step and both have not been reported in the literature. Therefore, we sought to experimentally validate this predicted synthesis route.

For the first step (enzymatic halogenation), we tested a previously reported chloroperoxidase[63] and were able to isolate the target product with 76% yield. For the second step, we performed the reaction in ethanol with base added, generating the final product with 91% yield. Besides that, when we compared the prices for the starting material in our route and the literature-reported route from the same supplier, the compound in literature-reported route is 10 times more expensive than ours.

■ DISCUSSION

In this work, we have developed RTS-CESP, a versatile, robust, and reliable chemoenzymatic synthesis planning tool. It starts with the predicted chemical synthesis routes for a target molecule and identifies the steps that do not break down the target molecule efficiently and replaces them with enzymatic reactions, which saves searching time and minimizes transitions

between chemical reactions and enzymatic reactions. RTS-CESP has been validated using a small-scale database with 10 known chemoenzymatic synthesis routes, a large-scale database with 1000 molecules, and an FDA-approved drug.

To develop this chemoenzymatic synthesis planning tool, we used a deep learning model to design a score function named RTscore that can distinguish synthesis reactions from decomposition reactions and evaluate reaction effectiveness. In principle, RTscore could also be used separately to rank predicted synthesis routes, and RTscore$_{(ECREACT)}$ was a specifically trained score for enzymatic reactions performing better than SCScore on an external enzymatic database. While our method focuses on selecting efficient reactions based on molecular transformations, the lack of reaction condition data in the USPTO and ECREACT databases may limit its practical applicability in some cases. Additional metrics, such as price, reaction conditions, and toxicity of chemicals, could be incorporated in the future for more comprehensive evaluations. In our hybrid synthesis search algorithm, RTscore was used to guide transitions between chemical and enzymatic reactions. This search method enabled the discovery of new enzymatic reactions while deepening the search tree to reach stock molecules efficiently. Since only the selected intermediate was searched for each route, RTscore also helped to save searching time.

The versatility of RTS-CESP was demonstrated in the combination with both ML-based and rule-based synthesis planning tools. In this work, five different tools were used to generate synthesis routes. The RXN4Chemistry (chemical) and RXN4Chemistry (enzymatic) employed a Molecular Transformer architecture and were continuously updated by the IBM researchers. The Aizynthfinder was trained with rules extracted from the Reaxys database and used MCTS as a multistep search algorithm, making it a robust retrosynthesis tool. RetroBioCat used curated enzymatic reaction rules and has been verified by enzymatic cascades reported in literature, while BKMS extracted abundant enzymatic templates from four biological databases.

The robustness and reliability of RTS-CESP were validated by different tasks. For robustness, predicting synthesis routes for molecules in a large data set could examine the competence of a synthesis planning tool in generating abundant transformations and reaching the commercially available molecules efficiently. Guided by RTscore, RTS-CESP predicted chemo-enzymatic synthesis routes for more target molecules and

suggested more shorter synthetic routes than ASKCOS. For reliability, it could be evaluated either using literature data or experimental validation. Other than reproducing and prioritizing the hybrid synthesis routes in our data set with literature support, we carried out experiments to validate the synthesis route for the FDA-approved drug Alclofenac. The validated route is shorter than the literature-reported route, providing a promising synthesis option.

■ ASSOCIATED CONTENT

Data Availability Statement

The source code is available at https://github.com/Zhao-Group/RTS-CESP.

ⓈⒾ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.4c01525.

> Materials and methods, data cleaning, model development and evaluation, calculation parameters, and experimental procedures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Huimin Zhao − *NSF Molecule Maker Lab Institute, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States; Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, Department of Chemistry, and DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States;* ◎ orcid.org/0000-0002-9069-6739; Email: zhao5@illinois.edu

Authors

Hongxiang Li − *NSF Molecule Maker Lab Institute, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States; Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, and Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States*

Xuan Liu − *NSF Molecule Maker Lab Institute, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States; Department of Chemical and Biomolecular Engineering and Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States;* ◎ orcid.org/0000-0003-1590-8276

Guangde Jiang − *Department of Chemical and Biomolecular Engineering, Carl R. Woese Institute for Genomic Biology, and DOE Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.4c01525

Author Contributions

H.Z. coordinated the project. H.L. and H.Z. conceived the presented idea. H.L. conducted the computational and experimental studies. X.L. contributed to the benchmark study on the boutique database, while G.J. assisted in analyzing products in the experiment. H.L. and H.Z. wrote the manuscript with input from all authors.

■ REFERENCES

(1) Schreiber, S. L. Target-oriented and diversity-oriented organic synthesis in drug discovery. *Science* **2000**, *287*, 1964−1969.

(2) Schreiber, S. L. Organic synthesis toward small-molecule probes and drugs. *Proc. Natl. Acad. Sci.* **2011**, *108*, 6699−6702.

(3) Bull, O. S.; Bull, I.; Amadi, G. K.; Odu, C. O.; Okpa, E. A review on metal-organic frameworks (MOFs): synthesis, activation, characterisation, and application. *Orient. J. Chem.* **2022**, *38*, 490−516.

(4) Li, L.; Chen, Z.; Zhang, X.; Jia, Y. Divergent strategy in natural product total synthesis. *Chem. Rev.* **2018**, *118*, 3752−3832.

(5) Arnold, F. H. Directed evolution: Bringing new chemistry to life. *Angew. Chem., Int. Ed.* **2018**, *57*, 4143−4148.

(6) Fu, H.; Cao, J.; Qiao, T.; Qi, Y.; Charnock, S. J.; Garfinkle, S.; Hyster, T. K. An asymmetric sp^3-sp^3 cross-electrophile coupling using 'ene'-reductases. *Nature* **2022**, *610*, 302−307.

(7) Huang, X.; Wang, B.; Wang, Y.; Jiang, G.; Feng, J.; Zhao, H. Photoenzymatic enantioselective intermolecular radical hydroalkylation. *Nature* **2020**, *584*, 69−74.

(8) Loskot, S. A.; Romney, D. K.; Arnold, F. H.; Stoltz, B. M. Enantioselective total synthesis of Nigelladine A via late-stage C-H oxidation enabled by an engineered P450 enzyme. *J. Am. Chem. Soc.* **2017**, *139*, 10196−10199.

(9) Sheldon, R. A.; Pereira, P. C. Biocatalysis engineering: the big picture. *Chem. Soc. Rev.* **2017**, *46*, 2678−2691.

(10) Yi, D.; Bayer, T.; Badenhorst, C. P. S.; Wu, S.; Doerr, M.; Hohne, M.; Bornscheuer, U. T. Recent trends in biocatalysis. *Chem. Soc. Rev.* **2021**, *50*, 8003−8049.

(11) Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to Sitagliptin manufacture. *Science* **2010**, *329*, 305−309.

(12) Sheldon, R. A.; Woodley, J. M. Role of biocatalysis in sustainable chemistry. *Chem. Rev.* **2018**, *118*, 801−838.

(13) Shoda, S.; Uyama, H.; Kadokawa, J.; Kimura, S.; Kobayashi, S. Enzymes as green catalysts for precision macromolecular synthesis. *Chem. Rev.* **2016**, *116*, 2307−2413.

(14) Losada-Garcia, N.; Cabrera, Z.; Urrutia, P.; Garcia-Sanz, C.; Andreu, A.; Palomo, J. M. Recent advances in enzymatic and chemoenzymatic cascade processes. *Catalysts* **2020**, *10*, 1258.

(15) Ricca, E.; Brucher, B.; Schrittwieser, J. H. Multi-enzymatic cascade reactions: overview and perspectives. *Adv. Synth. Catal.* **2011**, *353*, 2239−2262.

(16) Siedentop, R.; Claassen, C.; Rother, D.; Luetz, S.; Rosenthal, K. Getting the most out of enzyme cascades: strategies to optimize *in vitro* multi-enzymatic reactions. *Catalysts* **2021**, *11*, 1183.

(17) Walsh, C. T.; Moore, B. S. Enzymatic cascade reactions in biosynthesis. *Angew. Chem., Int. Ed.* **2019**, *58*, 6846−6879.

(18) Chakrabarty, S.; Romero, E. O.; Pyser, J. B.; Yazarians, J. A.; Narayan, A. R. H. Chemoenzymatic total synthesis of natural products. *Acc. Chem. Res.* **2021**, *54*, 1374−1384.

(19) Li, J.; Amatuni, A.; Renata, H. Recent advances in the chemoenzymatic synthesis of bioactive natural products. *Curr. Opin. Chem. Biol.* **2020**, *55*, 111−118.

(20) Sheldon, R. A.; Brady, D.; Bode, M. L. The hitchhiker's guide to biocatalysis: recent advances in the use of enzymes in organic synthesis. *Chem. Sci.* **2020**, *11*, 2587−2605.

(21) Zhang, X.; King-Smith, E.; Dong, L.-B.; Yang, L.-C.; Rudolf, J. D.; Shen, B.; Renata, H. Divergent synthesis of complex diterpenes through a hybrid oxidative approach. *Science* **2020**, *369*, 799−806.

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

(23) Shen, Y. N.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Primers* **2021**, *1*, 23.

(24) Yu, T.; Boob, A. G.; Volk, M. J.; Liu, X.; Cui, H.; Zhao, H. Machine learning-enabled retrobiosynthesis of molecules. *Nat. Catal.* **2023**, *6*, 137−151.

(25) Szymkuc, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-assisted synthetic planning: the end of the beginning. *Angew. Chem., Int. Ed.* **2016**, *55*, 5904−5937.

(26) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555*, 604−610.

(27) Segler, M. H. S.; Waller, M. P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chem. Eur. J.* **2017**, *23*, 5966−5971.

(28) Kumar, A.; Wang, L.; Ng, C. Y.; Maranas, C. D. Pathway design using de novo steps through uncharted biochemical spaces. *Nat. Commun.* **2018**, *9*, 184.

(29) Finnigan, W.; Hepworth, L. J.; Flitsch, S. L.; Turner, N. J. RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. *Nat. Catal.* **2021**, *4*, 98−104.

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

(31) Levin, I.; Liu, M.; Voigt, C. A.; Coley, C. W. Merging enzymatic and synthetic chemistry with computational synthesis planning. *Nat. Commun.* **2022**, *13*, 7747.

(32) Lang, M.; Stelzer, M.; Schomburg, D. BKM-react, an integrated biochemical reaction database. *BMC Biochem.* **2011**, *12*, 42.

(33) Sankaranarayanan, K.; Jensen, K. F. Computer-assisted multistep chemoenzymatic retrosynthesis using a chemical synthesis planner. *Chem. Sci.* **2023**, *14*, 6467−6475.

(34) Zeng, T.; Jin, Z.; Zheng, S.; Yu, T.; Wu, R. Developing BioNavi for Hybrid Retrosynthesis Planning. *JACS Au* **2024**, *4*, 2492−2502.

(35) 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.* **2020**, *11*, 3316−3325.

(36) Coley, C. W.; Jin, W.; Rogers, L.; Jamison, T. F.; Jaakkola, T. S.; Green, W. H.; Barzilay, R.; Jensen, K. F. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **2019**, *10*, 370−377.

(37) Sterling, T.; Irwin, J. J. ZINC 15 - Ligand discovery for everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324−2337.

(38) Sinha, S.; Mandal, B.; Chandrasekaran, S. Synthesis of ethyl arylacetates by means of Friedel-Crafts reactions of aromatic compounds with ethyl α-chloro-α-(ethylthio)acetate catalysed by ytterbium triflate. *Tetrahedron Lett.* **2000**, *41*, 9109−9112.

(39) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.

(40) Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273−297.

(41) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533−536.

(42) Bai, S.; Kolter, J. Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, 1803.01271.

(43) Tetko, I. V.; Karpov, P.; Bruno, E.; Kimber, T. B.; Godin, G. Augmentation is what you need! *Artificial Neural Networks and Machine Learning - ICANN 2019: Workshop and Special Sessions* **2019**, *11731*, 831−835.

(44) Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F. SCScore: synthetic complexity learned from a reaction corpus. *J. Chem. Inf. Model.* **2018**, *58*, 252−261.

(45) Akasaka, R.; Mashino, T.; Hirobe, M. Hydroxylation of benzene by horseradish peroxidase and immobilized horseradish peroxidase in an organic solvent. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 1861−1864.

(46) Babu, S.; Kumar, A.; Parella, R. Magnetic nano Fe_3O_4 catalyzed solvent-free stereo- and regioselective aminolysis of epoxides by amines; a green method for the synthesis of β-amino alcohols. *Synlett* **2014**, *25*, 835−842.

(47) Bernier, D.; Blake, A. J.; Woodward, S. Improved procedure for the synthesis of enamine N-oxides. *J. Org. Chem.* **2008**, *73*, 4229−4232.

(48) Bian, H.; Feng, J.; Xu, W. Synthesis and biological evaluation of novel AM80 derivatives as antileukemic agents. *Med. Chem. Res.* **2013**, *22*, 175−185.

(49) Birrell, J. A.; Desrosiers, J. N.; Jacobsen, E. N. Enantioselective acylation of silyl ketene acetals through fluoride anion-binding catalysis. *J. Am. Chem. Soc.* **2011**, *133*, 13872−13875.

(50) Chen, F. F.; Liu, Y. Y.; Zheng, G. W.; Xu, J. H. Asymmetric amination of secondary alcohols by using a redox-neutral two-enzyme cascade. *ChemCatChem* **2015**, *7*, 3838−3841.

(51) Engel, S.; Vyazmensky, M.; Berkovich, D.; Barak, Z.; Chipman, D. M. Substrate range of acetohydroxy acid synthase I from *Escherichia coli* in the stereoselective synthesis of α-hydroxy ketones. *Biotechnol. Bioeng.* **2004**, *88*, 825−831.

(52) Juršić, B. Synthetic application of micellar catalysis. williamson's synthesis of ethers. *Tetrahedron* **1988**, *44*, 6677−6680.

(53) Kakoti, A.; Kumar, A. K.; Goswami, P. Microsome-bound alcohol oxidase catalyzed production of carbonyl compounds from alcohol substrates. *J. Mol. Catal. B: Enzym.* **2012**, *78*, 98−104.

(54) McKenna, S. M.; Leimkühler, S.; Herter, S.; Turner, N. J.; Carnell, A. J. Enzyme cascade reactions: synthesis of furandicarboxylic acid (FDCA) and carboxylic acids using oxidases in tandem. *Green Chem.* **2015**, *17*, 3271−3275.

(55) Mendieta-Wejebe, J. Catalytic activity of cytochrome P-450 using $NADP^+$ reduced by an anionic hydride organosiloxane. *Sci. Pharm.* **2008**, *76*, 241−257.

(56) Pickl, M.; Fuchs, M.; Glueck, S. M.; Faber, K. Amination of ω-functionalized aliphatic primary alcohols by a biocatalytic oxidation-transamination cascade. *ChemCatChem* **2015**, *7*, 3121−3124.

(57) Sello, G.; Orsini, F.; Bernasconi, S.; Gennaro, P. D. Synthesis of enantiopure 2-amino-1-phenyl and 2-amino-2-phenyl ethanols using enantioselective enzymatic epoxidation and regio- and diastereoselective chemical aminolysis. *Tetrahedron: Asymmetry* **2006**, *17*, 372−376.

(58) Slagbrand, T.; Lundberg, H.; Adolfsson, H. Ruthenium-catalyzed tandem-isomerization/asymmetric transfer hydrogenation of allylic alcohols. *Chem. Eur. J.* **2014**, *20*, 16102−16106.

(59) Song, S.; Wang, Y.; Yan, N. A remarkable solvent effect on reductive amination of ketones. *Mol. Catal.* **2018**, *454*, 87−93.

(60) Toda, H.; Imae, R.; Itoh, N. Bioproduction of chiral epoxyalkanes using styrene monooxygenase from Rhodococcus sp. ST-10 (RhSMO). *Adv. Synth. Catal.* **2014**, *356*, 3443−3450.

(61) Xu, Y.; Jia, X.; Panke, S.; Li, Z. Asymmetric dihydroxylation of arylolefins by sequential enantioselective epoxidation and regioselective hydrolysis with tandem biocatalysts. *Chem. Commun.* **2009**, 1481−1483.

(62) Zhang, H.; Peng, X.; Dai, Y.; Shao, J.; Ji, Y.; Sun, Y.; Liu, B.; Cheng, X.; Ai, J.; Duan, W. Discovery of a pyrimidinedione derivative as a potent and orally bioavailable Axl inhibitor. *J. Med. Chem.* **2021**, *64*, 3956−3975.

(63) Getrey, L.; Krieg, T.; Hollmann, F.; Schrader, J.; Holtmann, D. Enzymatic halogenation of the phenolic monoterpenes thymol and carvacrol with chloroperoxidase. *Green Chem.* **2014**, *16*, 1104−1108.