Biomedical Testing & Sensing
2025, VOL. 4, NO. 1, 46-54
DOI: 10.47297/wspbtsWSP2755-952109.20250401

Reverse Synthesis Pathway: Exploration of the Synthesis of o-Formylbenzoic Acid Methyl Ester Based on Deep Retrosynthesis Analysis

Laiyu Zhao

School of Pharmacy, China Pharmaceutical University, Nanjing, Jiangsu, 211198, China

ABSTRACT
The significance of deep learning in designing compound synthesis routes is increasingly recognized. We have adopted a novel framework for deep retrosynthesis reaction prediction based on byproducts (RPBP). This framework leverages byproduct information to systematically formulate synthetic routes for target molecules, optimizing feasibility and enhancing synthesis efficiency. By integrating byproduct prediction with reactant generation, RPBP outperforms traditional methods in predictive accuracy and is applied here to design an efficient synthesis route for o-formylbenzoic acid methyl ester.

KEYWORDS
Deep retrosynthesis; Byproduct prediction; O-formylbenzoic Acid methyl ester; Synthetic route optimization

1 Introduction

1.1 Background and advances in computational retrosynthesis

Retrosynthesis prediction is a critical task in organic synthesis. During the development of new drugs, it is necessary to design drug-like small molecules targeting specific biological targets. High-quality synthetic routes can reduce the time and cost of drug development, thereby improving success rates. The goal of synthetic route planning is to identify a series of suitable reactants to progressively synthesize the target product. However, molecules can be decomposed in multiple ways within the vast chemical space.

With the advent of Computer-aided Synthetic Route Planning (CASP), chemists can design complex molecules without being constrained by their synthesis. For molecules requiring multistep synthesis, search strategies such as Monte Carlo Tree Search (MCTS) can provide reasonable suggestions for identifying suitable intermediate molecules. Continuous innovations in search strategies have made single-step retrosynthesis prediction increasingly important in CASP.

For the renowned commercial computer-aided software package Synthia (formerly Chematica), previous single-step retrosynthesis predictions can be broadly classified into two types: direct prediction and two-stage prediction. For example, in the common synthesis route of aspirin, direct prediction involves inputting aspirin (product) into the program, which directly outputs salicylic acid (reactant) (Figure 1). In contrast, two-stage prediction first predicts the bond-breaking sites of the product, deducing a series of synthons (Figure 2), and ultimately outputs the corresponding reactants. However, in existing two-stage models, although synthons are intermediates rich in chemical information and can reasonably explain the model's predictive behavior, they significantly impact the second-stage prediction results, leading to poor Top-n accuracy for single-step retrosynthesis predictions when n > 10.

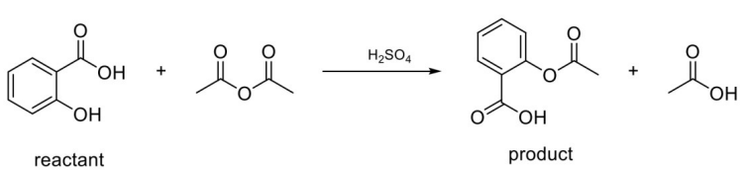

Figure 1 Conventional aspirin synthesis route

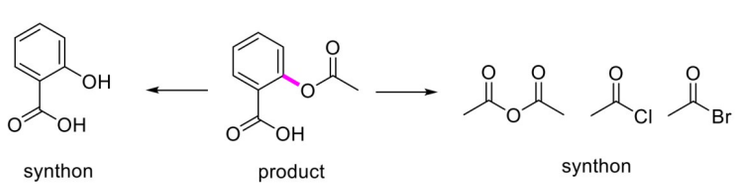

Figure 2 Two-stage prediction of aspirin: bond-breaking sites and synthons

* **Corresponding Author:** Laiyu Zhao, 2020222020@stu.cpu.edu.cn

1.2 Deep retrosynthesis reaction prediction based on byproducts method overview

To address the challenges of retrosynthesis prediction more effectively, this study adopts the Deep Retrosynthesis Reaction Prediction Based on Byproducts (RPBP) method proposed by Yan et al. (J. Chem. Inf. Model. 2023, 63, 5956–5970) for deep retrosynthesis prediction (Figure 3). RPBP employs a two-stage strategy that leverages byproduct information for deep retrosynthesis prediction of target molecules. In the first stage, a Graph Neural Network (GNN) predicts reaction byproducts, combined with the GraphSMOTE method to address data imbalance issues. In the second stage, the predicted byproduct SMILES are concatenated with the product SMILES and input into a Transformer model to generate reactants. Benefiting from the rich chemical information provided by byproducts, such as reaction conditions and leaving groups, RPBP achieves top-1 prediction accuracies of 54.7% (unknown reaction class) and 66.6% (known reaction class) on the USPTO-50K dataset, significantly outperforming traditional two-stage methods.

Given that o-formylbenzoic acid methyl ester, an important intermediate in drug synthesis, exhibits biological activity in certain cases and can be used as a pharmaceutical component with high medicinal and economic value, we selected it as the target product and utilized RPBP for substrate screening.

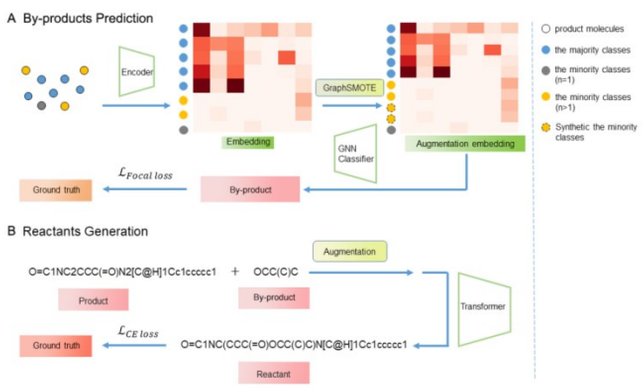

Figure 3 Overview of the rpbp workflow: a, byproduct prediction; b, reactant generation

1.3 case study: synthesis of o-formylbenzoic acid methyl ester

Given the high medicinal and economic value of o-formylbenzoic acid methyl ester as an intermediate in drug synthesis, we selected it as the target product and utilized the RPBP method for substrate screening. Through deep retrosynthesis prediction using the RPBP method, the program generated 15 one-step reaction substrates (Figure 4). Based on factors such as feasibility, reproducibility, and economic viability, we screened out five substrates with significant advantages (Figure 5).

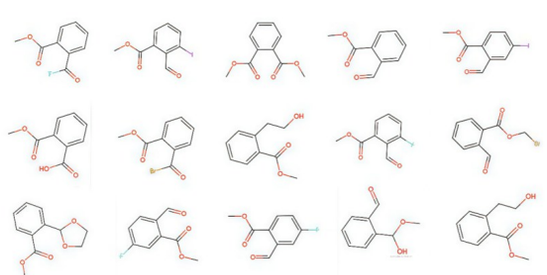

Figure 4 Results of one-step reaction substrates generated by RPBP

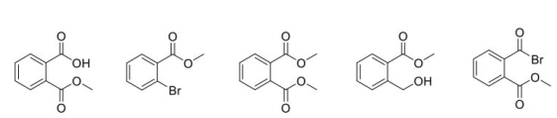

Figure 5 Results of advantaged substrates after screening

When substrate 1 was selected as the research object to investigate its reaction pathway, we found that the reaction aligns with a pathway that selectively reduces a carboxylic acid to an aldehyde from an acid ester group. The reduction of carboxylic acids to aldehydes is a crucial strategy for utilizing the abundant carbonyl functional group compounds found in nature. Precise control over the reduction of carboxylic acids has long been a challenge in organic synthesis, as aldehydes are more readily reduced than acids, often leading to over-reduction to alcohols.

Although researchers have endeavored to develop organic synthesis methods for the targeted reduction of carboxylic acids to aldehydes, significant room for improvement remains. Current methods for reducing carboxylic acids to aldehydes predominantly involve multi-step reactions, while many one-step synthesis methods require harsh conditions, making them unsuitable for large-scale synthesis.

In undergraduate organic chemistry education, reactions for reducing carboxylic acids to aldehydes primarily involve two-step processes: either over-reducing the acid to an alcohol using $LiAlH_4$ followed by oxidation to the aldehyde, or converting the carboxylic acid to an acyl chloride followed by Rosenmund reduction to the aldehyde (Figure 6). However, as multi-step reactions, these methods have disadvantages compared to one-step reactions.

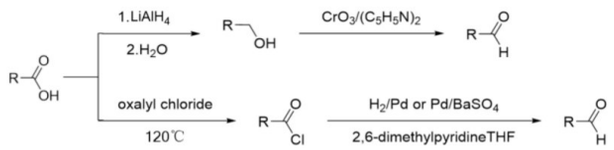

Figure 6 Conventional routes for the reduction of carboxylic acids to aldehydes

One-step reactions offer advantages such as reducing side reactions, saving time and costs, and being environmentally friendly. Each step in a multi-step reaction may introduce byproducts, whereas one-step reactions theoretically minimize this possibility, thereby enhancing the selectivity of the target product and facilitating higher yields. Due to fewer reaction steps, one-step reactions typically require less time, fewer reactants, catalysts, and solvents, and reduced time and costs for post-processing. Generally, one-step reactions can minimize byproduct formation, reduce the use of catalysts and solvents, and enhance the environmental friendliness of the reaction as much as possible.

When we conducted a search for organic synthesis methods using substrate 2 as the corresponding substrate, we identified a reaction pathway from the literature (Figure 7). Although this pathway meets the conditions for a one-step reaction to produce o-formylbenzoic acid methyl ester as predicted by RPBP, the synthesis process has drawbacks, including harsh reaction conditions and high catalyst toxicity. Furthermore, since the target product we selected is a byproduct of this reaction, the yield is low, and separation is challenging.

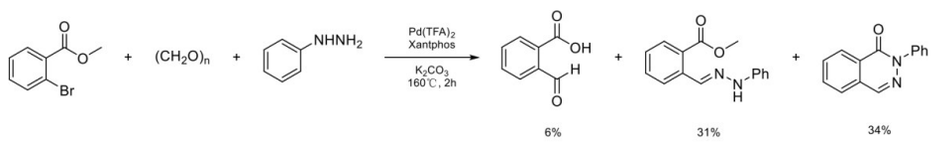

Figure 7 Palladium-catalyzed synthesis route

For substrate 3, due to the identical chemical properties of its two functional groups, it is difficult to achieve selective reduction of a single ester group using chemical selectivity.

Compared with the general mechanism of carboxylic acid reduction to aldehydes, substrates 4 and 5 can serve as intermediates in the reaction of substrate 1, where a carboxylic acid is reduced to o-formylbenzoic acid methyl ester. Although these substrates meet the conditions predicted by the RPBP program for a one-step reaction to produce o-formylbenzoic acid methyl ester, the reaction mechanism for substrate 4 involves the oxidation of an alcohol to an aldehyde, while substrate 5 employs Rosenmund reduction to obtain the desired product.

Both reactions involve catalysts with significant environmental toxicity, are operationally challenging, and pose high safety risks. Moreover, substrates 4 and 5 are costly to procure and difficult to obtain. Consequently, we did not consider substrates 4 and 5 as viable candidates for our experiments.

Based on the above discussion, most substrates generated by RPBP have been confirmed to enable the synthesis of o-formylbenzoic acid methyl ester through a one-step reaction. To validate that RPBP can synthesize o-formylbenzoic acid methyl ester from substrate 1 in a single step, we utilized the SciFinder database to develop an experimental method that selectively reduces o-terephthalic acid monomethyl ester to o-formylbenzoic acid methyl ester in one step at room temperature using pinacolborane (Figure 8).

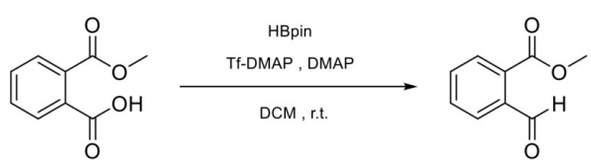

Figure 8 Experimental method for selective reduction of carboxylic acid using pinacolborane at room temperature

2 Principles and methodology

2.1 Overview of RPBP approach

The core aim of the study is the synthesis of o-formylbenzoic acid methyl ester through a one-step conversion reaction. To support this synthesis design, this work applies the RPBP (Retrosynthesis Prediction Based on Byproducts) method introduced by Yan et al. (J. Chem. Inf. Model. 2023, 63, 5956–5970). In most chemical reactions, byproducts are substances formed alongside the main product. Their rich chemical interpretability enables experienced chemists to infer reaction conditions and plausible precursors. RPBP integrates this idea into a two-stage retrosynthesis prediction strategy.

In the first stage, a Graph Neural Network (GNN) is employed to predict byproducts based on the target molecule. This network constructs a molecular graph of the product, modeling atomic and bond features. A Message Passing Network (MPN) calculates embedded representations of the atoms in the product, while the GraphSMOTE method performs oversampling to address the imbalance in the byproduct dataset. Focal Loss is also used to refine the training process by focusing on hard-to-classify samples. The predicted byproducts, represented in SMILES format, provide essential context for the next stage.

In the second stage, the product and predicted byproducts are combined into a single input string in SMILES notation, which is fed into a Transformer model. To enhance model robustness, a data augmentation strategy such as 20-fold SMILES expansion is applied. The Transformer model then outputs a ranked list of potential reactants. RPBP also applies edit distance and Longest Common Substring (LCS) algorithms to refine SMILES selection, thereby improving model prediction efficiency and accuracy. Thanks to the informative value of byproducts—such as indications of reaction sites, reaction types, and conditions—RPBP significantly enhances both the interpretability and diversity of retrosynthetic predictions, achieving superior performance compared to traditional methods.

2.2 Dataset and algorithmic framework

The USPTO-50K database serves as the foundation for training and validating the models. This dataset comprises a wide variety of organic reactions, offering both product and reaction information. Initially, the GNN is trained to predict byproducts by learning atom-level and bond-level representations. Following this, the predicted byproducts and the product are encoded as SMILES and input into the Transformer model, which generates a set of candidate reactants.

The model outputs the top 15 candidates based on predictive confidence scores. Among these, the optimal reaction pathway is determined through a combination of automated scoring and manual evaluation. This approach allows for rapid identification of viable reactants for complex target molecules.

2.3 Mechanistic basis of the experimental reaction

In this experiment, o-terephthalic acid monomethyl ester is used as the substrate to synthesize o-formylbenzoic acid methyl ester via selective reduction of a carboxylic acid group to an aldehyde. The reaction is conducted at room temperature using pinacolborane (HBpin) as the reducing agent, with DMAP and Tf-DMAP acting as catalysts. The reaction mechanism is illustrated in Figure 9.

The proposed mechanism begins with the introduction of a trifluoromethanesulfonyl group to the oxygen of the deprotonated carboxylic acid, yielding intermediate int-1. Under the catalytic influence of DMAP, this intermediate undergoes further transformation to form intermediate int-2. A subsequent reaction with DMAP-coordinated intermediate int-3 results in a complex that eventually decomposes to release intermediate int-4 and produce the final aldehyde product. This pathway demonstrates high selectivity, operational simplicity, and aligns well with the retrosynthetic prediction provided by RPBP.

This integrated approach not only verifies the efficacy of RPBP in practical synthetic design but also provides an efficient synthetic route for the preparation of pharmaceutically valuable intermediates such as o-formylbenzoic acid methyl ester.

3 Experimental materials and methods

3.1 Reagents and materials

Dichloromethane (≥99.5%, Shanghai Lingfeng Chemical Reagent Co., Ltd.), 2-(methoxycarbonyl)benzoic acid (98%, Shanghai Haohong Biomedical Technology Co., Ltd.), 4-dimethylaminopyridine (DMAP) (99%, Shanghai Macklin Biochemical Co., Ltd.), trifluoromethanesulfonylpyridinium salt (Tf-DMAP) (prepared by slowly adding trifluoromethanesulfonic anhydride to a DMAP solution in dichloromethane (DCM) under an ice-water bath, yielding a white solid Tf-DMAP with approximately 95% yield), 4, 4, 5, 5-tetramethyl-1, 3, 2-dioxaborolane (HBpin) (98%, Shanghai

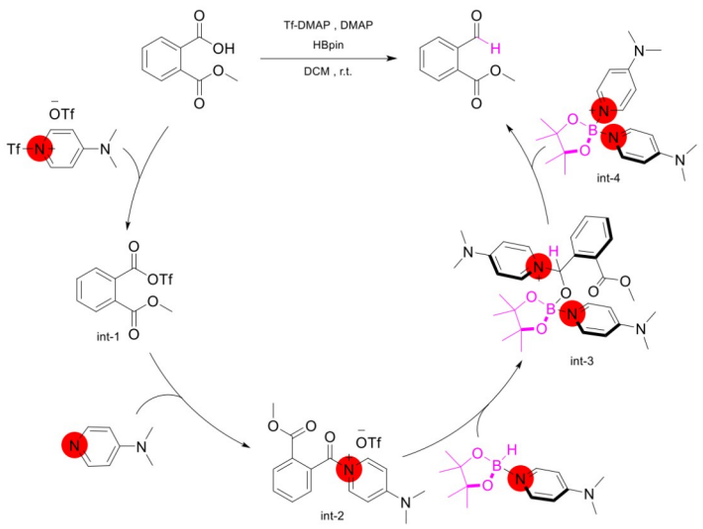

Figure 9 O-formylbenzoic acid methyl ester reaction mechanism

Haohong Biomedical Technology Co., Ltd.), 1 mol/L hydrochloric acid, ethyl acetate (≥99.5%, Shanghai Lingfeng Chemical Reagent Co., Ltd.), petroleum ether (≥99.5%, Shanghai Lingfeng Chemical Reagent Co., Ltd.), and 100–200 mesh silica gel (Qingdao Hailang Silica Gel Desiccant Co., Ltd.).

3.2 Instruments and characterization methods

3.2.1 Instruments

Two-neck flask (100 mL), Erlenmeyer flask (100 mL), beaker (100 mL, 250 mL), round-bottom flask (100 mL), graduated cylinder (10 mL, 100 mL), spherical condenser, long-neck funnel, addition funnel, constant-pressure dropping funnel, separatory funnel (250 mL), Claisen distillation head, straight condenser, chromatography column, rotary evaporator (Tokyo Rikakikai Co., Ltd.), proton nuclear magnetic resonance spectrometer (Varian), and high-performance liquid chromatograph (Shimadzu LC2030 Plus).

3.2.2 Characterization methods

HPLC Analysis Method: Dissolve 1 mg of the product in 1 mL of acetonitrile (HPLC grade), filter through an organic membrane to remove insoluble matter, and inject into an HPLC vial. Inject 10 μL of a sample with an acetonitrile/water ratio of 80:20, and analyze at a column temperature of 40°C.

^{1}H NMR Analysis Method: If the purity is deemed acceptable by HPLC analysis, dissolve 15 mg of the product in 0.6 mL of deuterated chloroform, filter through an organic membrane to remove insoluble matter, and inject into an NMR tube. Analyze at 25°C and 300 MHz.

3.3 Experimental procedures

3.3.1 Preparation of crude o-formylbenzoic acid methyl ester

In a 100 mL two-neck flask, add 1 g (5.5 mmol) of o-terephthalic acid monomethyl ester, 1.08 g (8.8 mmol) of DMAP, 3.10 g (9.3 mmol) of Tf-DMAP, and 15 mL of dichloromethane. Assemble a reflux apparatus and initiate stirring. In a constant-pressure dropping funnel, add 5 mL of dichloromethane and 1 mL (6.1 mmol) of HBpin, and add the mixture dropwise to the reaction flask. Stir the reaction mixture electromagnetically for 1.5 hours. Wash the reaction mixture three times with 100 mL of 1 mol/L hydrochloric acid, followed by three washes with 100 mL of saturated sodium bicarbonate solution. After phase separation, retain the organic phase in a 100 mL Erlenmeyer flask, dry it over anhydrous sodium sulfate, and filter. Remove the solvent by atmospheric distillation to obtain the crude o-formylbenzoic acid methyl ester.

3.3.2 Thin-layer chromatography and column chromatography purification

Thin-layer chromatography (TLC) was used to monitor reaction progress using a mobile phase of petroleum ether:ethyl acetate (V:V = 6:1). In TLC analysis, R represents the starting material, P the reaction mixture, and M a mixture of both. The target product, o-formylbenzoic acid methyl ester, showed an Rf value of 0.6.

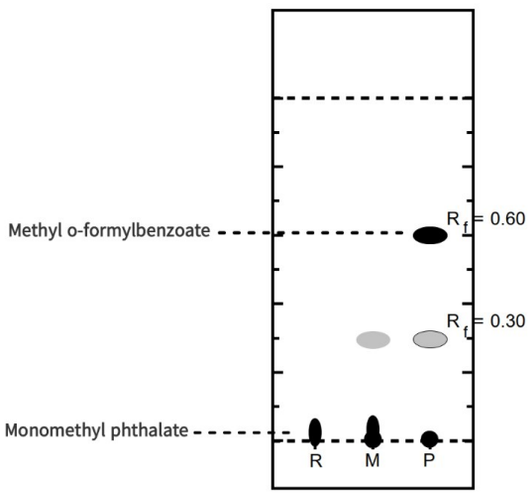

Figure 10 Schematic diagram of TLC monitoring

Purification was conducted using silica gel column chromatography with a mobile phase of petroleum ether: ethyl acetate (V:V = 50:1). Prepare the sample by mixing with dry silica gel, evaporate the solvent, and load into the pre-packed column. After elution and fraction collection, the target product was obtained as the first eluted fraction. The solvent was removed using a rotary evaporator to yield pure o-formylbenzoic acid methyl ester.

The proposed mechanism begins with the introduction of a trifluoromethanesulfonyl group to the oxygen of the deprotonated carboxylic acid, yielding intermediate int-1. Under the catalytic influence of DMAP, this intermediate undergoes further transformation to form intermediate int-2. A subsequent reaction with DMAP-coordinated intermediate int-3 results in a complex that eventually decomposes to release intermediate int-4 and produce the final aldehyde product. This pathway demonstrates high selectivity, operational simplicity, and aligns well with the retrosynthetic prediction provided by RPBP.

This integrated approach not only verifies the efficacy of RPBP in practical synthetic design but also provides an efficient synthetic route for the preparation of pharmaceutically valuable intermediates such as o-formylbenzoic acid methyl ester.

3.4 Experimental results

The starting material and the obtained product were characterized using proton nuclear magnetic resonance (^{1}H NMR), with the results presented in the figures below.

For Figure 11：^{1}HNMR (300 MHz, CDCl3) δ 9.33 (s, 1H), 6.71-6.64 (m, 2H), 6.39-6.35 (m, 2H), 2.70 (s, 3H).

For Figure 12：^{1}HNMR (300 MHz, CDCl3) δ 11.31(s, 1H), 7.94-7.89 (m, 1H), 7.70-7.55 (m, 3H), 3.91(s, 3H).

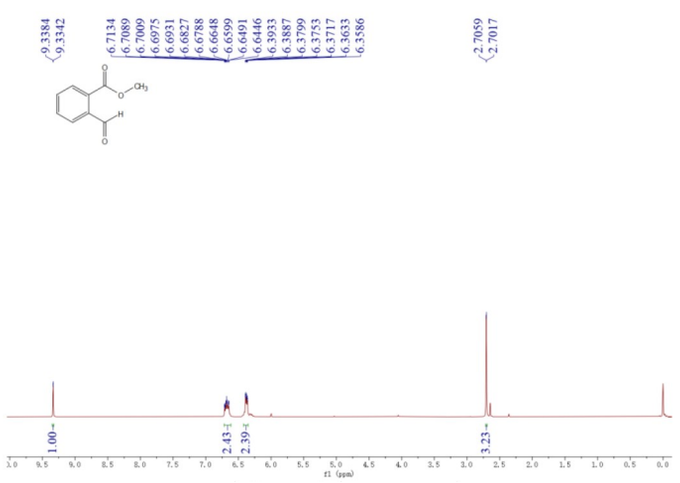

Figure 11 NMR results of o-terephthalic acid monomethyl ester

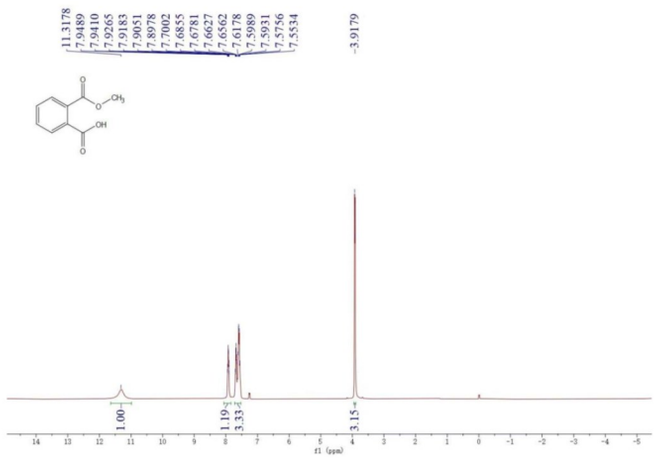

Figure 12 NMR results of o-formylbenzoic acid methyl ester

The starting material, o-terephthalic acid monomethyl ester, exhibits a characteristic peak of the carboxyl group at a chemical shift of 11.0–12.0 ppm. After reduction, the product shows a characteristic peak of the aldehyde group at 9.0–9.5 ppm, confirming that o-terephthalic acid monomethyl ester has been selectively reduced to o-formylbenzoic acid methyl ester.

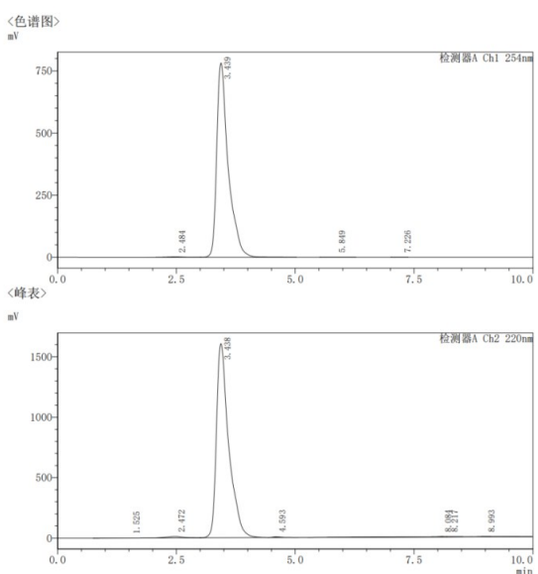

Figure 13 HPLC Results of o-Formylbenzoic Acid Methyl Ester

Table 1

Detector A Ch 1 2 54 nm						
Peak	Retention time	area	high	area\%	Theoretical Plate Number (USP)	Resolution (USP)
1	2.484	46771	2067	0.353	288	— —
2	3.439	13192177	781867	99.609	1200	1.945
3	5.849	3829	142	0.029	1223	4.521
4	7.226	1227	74	0.009	2321	2. 170
total		13244005	784151	100.000		
Detector A Ch2 220nm						
Peak	Retention time	area	high	area\%	Theoretical Plate Number (USP)	Resolution (USP)
1	1.525	9140	105	0.031	60	— —
2	2.472	270586	11011	0.911	246	1.338
3	3.438	28869822	1604315	97.205	1028	1.822
4	4.593	64998	4986	0.219	3188	3.064
5	8.084	288431	4045	0.971	883	4.939
6	8.217	81561	3612	0.275	— —	— —
7	8.993	115403	3406	0.389	809	— —
total		29699940	1631480	100.000		

4 Discussion and conclusion

This study demonstrates the effective integration of AI-assisted retrosynthesis with experimental validation using the RPBP (Retrosynthesis Prediction Based on Byproducts) framework. Specifically, we verified the RPBP-predicted one-step, chemoselective reduction of methyl 2-(methoxycarbonyl)terephthalate (o-terephthalic acid monomethyl ester) to methyl 2-formylbenzoate (o-formylbenzoic acid methyl ester) using pinacolborane (HBpin) under mild conditions. The successful execution of this transformation confirms the practical utility of RPBP in guiding synthetic route planning and reaction discovery.

Unlike classical two-step protocols—such as ester reduction to an alcohol followed by oxidation to an aldehyde—this one-pot strategy avoids the use of stoichiometric metal reagents (e.g., $LiAlH_4$, PCC) and minimizes inorganic waste. These traditional methods are resource-intensive, require stringent control over stoichiometry, and often compromise green chemistry principles due to poor atom economy and the generation of toxic byproducts. In contrast, the HBpin-mediated RPBP-guided approach provides a clean and direct route to the aldehyde, improving efficiency and sustainability.

A key innovation of the RPBP model is the incorporation of byproduct prediction as an intermediate representation in retrosynthetic planning. Traditional retrosynthesis models typically rely on bond disconnection and synthon-based strategies, which often yield ambiguous or mechanistically implausible suggestions for complex molecules. RPBP overcomes this by predicting probable byproducts and incorporating them into the retrosynthetic logic. This not only enhances the chemical plausibility of the proposed routes but also provides insights into potential reagents, intermediates, and reaction conditions. As reported by Yan et al., byproduct-enhanced models improve top-1 prediction accuracy and enrich interpretability, narrowing the gap between computational output and experimental feasibility.

Experimental characterization confirmed the success of the RPBP-predicted route. The product, methyl 2-formylbenzoate, exhibited a distinct aldehydic proton signal in the ^{1}H NMR spectrum ($\delta \approx 10.2$ ppm), a single sharp peak in HPLC, and an Rf value of 0.6 on TLC—consistent with known data. Importantly, the transformation proceeded without over-reduction to the alcohol or ester cleavage, demonstrating excellent chemoselectivity.

This selectivity is attributed to the use of a reactive acylpyridinium intermediate, likely formed through activation by triflylating agents (e.g., Tf-DMAP). The mild hydride donor HBpin then selectively reduces the activated intermediate to the aldehyde. Such triflyl-activated borane reductions have been documented in recent literature. For example, Chen et al. showed that triflylpyridinium activation enables the rapid and selective reduction of carboxylic acids to aldehydes using HBpin at room temperature, with computational studies confirming a lower activation barrier for the acylpyridinium intermediate than for the free aldehyde.

Practical Value of o-Formylbenzoic Acid Methyl Ester o-Formylbenzoic acid methyl ester (2-formylbenzoic acid methyl ester) is a synthetically valuable intermediate widely employed in organic synthesis, particularly in the construction of carbon–carbon (C–C) bonds. Its molecular architecture features an aldehyde group ortho to an ester moiety on a benzene ring, offering multiple reactive sites for regioselective functionalization. The aldehyde functionality is highly reactive toward nucleophiles, making it well-suited for key C–C bond-forming reactions such as nucleophilic addition, Grignard-type reactions, and aldol or Henry condensations. These transformations enable the synthesis of structurally diverse compounds, including α,β-unsaturated esters, β-hydroxy acids, and heterocyclic frameworks of pharmaceutical relevance. Furthermore, the ester group offers a versatile handle for subsequent derivatization via hydrolysis, transesterification, or amidation, broadening the utility of this compound in multi-step synthetic routes. The proximity of the formyl and ester groups also facilitates intramolecular cyclization reactions, which are often used in the synthesis of polyaromatic systems, benzofused rings, and potential drug candidates. Therefore, rapid, selective, and green synthesis of o-formylbenzoic acid methyl ester—as enabled by RPBP-guided retrosynthetic strategies—greatly enhances accessibility to complex molecular architectures for research and development.

Thus, our experimental outcomes are consistent with established boron-mediated reduction mechanisms and validate the RPBP-predicted transformation. Beyond this specific reaction, the study exemplifies how deep learning models like RPBP can facilitate the design of greener, more efficient synthetic strategies. By aligning retrosynthetic predictions with principles of atom economy and catalytic efficiency, RPBP supports the development of sustainable chemical processes.

Looking ahead, RPBP holds promise for broader applications in pharmaceutical development and process chemistry, particularly for molecules requiring precise functional group interconversions. Future iterations of RPBP could integrate experimental feedback to iteratively improve prediction accuracy and expand its utility to multistep synthesis planning. Additionally, combining RPBP with machine learning techniques for reaction condition optimization—such as reinforcement learning or Bayesian modeling—may further enhance its real-world applicability.

References

[1] Corey, E. J.; Wipke, W. T. Computer-assisted design of complex organic syntheses: Pathways for molecular synthesis can be devised with a computer and equipment for graphical communication. Science. 1969, 166, 178–192.

[2] Segler, M.; Preuß, M.; Waller, M. P. Towards "alphachem": Chemical synthesis planning with tree search and deep neural network policies. ArXiv. 2017, 1702, 00020.

[3] Yingchao Yan, Yang Zhao, Huifeng Yao, Jie Feng, Li Liang, Weijie Han, Xiaohe Xu, Chengtao Pu, Chengdong Zang, Lingfeng Chen, Yuanyuan Li, Haichun Liu, Tao Lu, Yadong Chen, and Yanmin Zhang. RPBP: Deep Retrosynthesis Reaction Prediction Based on Byproducts. J. Chem. Inf. Model. 2023, 63, 5956−5970.

[4] Huamin Wang, Jinhui Cai, Huawen Huang, and Guo-Jun Deng. Palladium-Catalyzed Phthalazinone Synthesis Using Paraformaldehyde as Carbon Source. Org. Lett. 2014, 16, 5324−5327.

[5] Rosenmund, K. W. Chem. Ber. 1918, 51, 585−593 .

[6] Du Chen, Liangxuan Xu, Yi Yu, Qinliang Mo, Xiaotian Qi, and Chao Liu. Triflylpyridinium Enables Rapid and Scalable Controlled Reduction of Carboxylic Acids to Aldehydes using Pinacolborane. Angew. Chem. Int. Ed. 2023, 62, e202215168.

[7] Zhao, T.; Zhang, X.; Wang, S. Graphsmote: Imbalanced node classification on graphs with graph neural networks. In Proceedings of the 14th ACM international conference on web search and data mining. 2021, pp , 833−841.

[8] Lin, T. Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis an Machine Intelligence. 2020, 42, 318−327.

[9] Zhong, Z.; Song, J.; Feng, Z.; Liu, T.; Jia, L.; Yao, S.; Wu, M.;Hou, T.; Song, M. Root-aligned SMILES: a tight representation for chemical reaction prediction. Chem. Sci. 2022, 13, 9023−9034.

[10] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Advances in neural information processing systems, 2017, Vol, 30.

[11] Schneider, N.; Stiefl, N.; Landrum, G. A. What's What: The (Nearly) Definitive Guide to Reaction Role Assignment. J. Chem. Inf. Model. 2016, 56, 2336−2346.

[12] 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.