Chemical Engineering Journal 420 (2021) 129845

Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

RetroPrime: A Diverse, plausible and Transformer-based method for Single-Step retrosynthesis predictions

Xiaorui Wang [a,1], Yuquan Li [a,1], Jiezhong Qiu [b], Guangyong Chen [c], Huanxiang Liu [d], Benben Liao [e,*], Chang-Yu Hsieh [e,*], Xiaojun Yao [a,*]

[a] College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou, PR China
[b] Department of Computer Science and Technology, Tsinghua University, Beijing, PR China
[c] Guangdong Provincial Key Laboratory of Computer Vision and Virtual Reality Technology, Shenzhen, Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, PR China
[d] School of Pharmacy, Lanzhou University, Lanzhou, PR China
[e] Tencent Quantum Laboratory, Shenzhen, Tencent, PR China

ARTICLE INFO

Keywords:
Deep Learning
Natural Language Processing
Template-free Single-Step Retrosynthesis

ABSTRACT

Retrosynthesis prediction is a crucial task for organic synthesis. In this work, we propose a single-step template-free and Transformer-based method dubbed RetroPrime, integrating chemists' retrosynthetic strategy of (1) decomposing a molecule into synthons then (2) generating reactants by attaching leaving groups. These two stages are accomplished with versatile Transformer models, respectively. RetroPrime achieves the Top-1 accuracy of 64.8% and 51.4%, when the reaction type is known and unknown, respectively, in the USPTO-50 K dataset. And the Top-1 accuracy is close to the state-of-the-art transformer-based method in the large dataset USPTO-full. It is known that outputs of the Transformer-based retrosynthesis model tend to suffer from insufficient diversity and high chemical implausibility. These problems may limit the potential of Transformer-based methods in real practice, yet few works address both issues simultaneously. RetroPrime is designed to tackle these challenges.

1. Introduction

Organic synthesis is not only an essential part of organic chemistry but also a cornerstone for a wide array of modern scientific disciplines such as drug discovery, environmental science, and materials science, etc. Retrosynthetic analysis is the most common method to design synthetic routes by iteratively decomposing molecules into potentially simpler and easier-to-synthesize precursors via applying known reactions [1]. In recent years, with the development of artificial intelligence technology, computer-aided synthesis planning (CASP) has further empowered chemists to contemplate even more complex molecules and save tremendous amounts of time and energy from designing synthetic experiments [2–12].

At present, purely machine-learning retrosynthesis models are classified into two categories [13]: the template-based [14–16] and template-free [17–22] methods. A template-based algorithm extracts reaction templates from chemical reaction data [23,24], matches the subgraph in the product part of the template to a target molecule, decomposes the target molecules as prescribed by the matched template, and completes the leaving group through the atomic changes indicated by the template to obtain the reaction precursors. Despite being interpretable in terms of why certain templates are preferred, template-based methods can only predict reactions if corresponding templates have been curated in a database [14,15]. With the ever-growing list of reaction templates, it is certainly desirable to contemplate alternative approaches.

Template-free methods may predict chemical reactions not present in a training set. Chen et al. [25] have tried to study the generalization ability of the template-free method relative to the template-based method using a special data splitting approach. As a complement, in **Supplementary Information** Section S3, we present one experiment to support this intuition. In particular, we showed template-free methods perform much better in predicting reactions when the corresponding templates never appear in the training set. In this work, we focus on the

* Corresponding authors.
E-mail addresses: bliao@tencent.com (B. Liao), kimhsieh@tencent.com (C.-Y. Hsieh), xjyao@lzu.edu.cn (X. Yao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.cej.2021.129845
Received 23 December 2020; Received in revised form 6 April 2021; Accepted 9 April 2021
Available online 21 April 2021

Transformer-based template-free method.

Liu et al. [18] treated the one-step retrosynthesis as a translation task, using SMILES [26] to represent molecules and using an LSTM [27] model, a venerable tool in natural language processing (NLP), to convert SMILES of a product to SMILES of reactant(s). Later on, many researchers [17,25,28–30] adopted a more advanced NLP model, Transformer [31], for predicting retrosynthesis. Transformer-based methods easily outperform baselines established by prior arts. Furthermore, the same model architecture can be directly applied for "forward prediction" [32], i.e., predicting a product molecule given a set of reactants and reagents. In another study [30], Lee et al. showed the generalizability of the Transformer model across chemical spaces. Transformer not only performs well in single-step retrosynthesis but also in multi-step retrosynthesis. Lin et al. [17] combined Transformer and Monte-Carlo tree search, and re-discovers the reported retrosynthetic route of four molecules. Furthermore, Schwaller et al. [33] also used the Transformer-based model combined with hyper-graph exploration strategy to complete the prediction of the multi-step retrosynthesis pathway and required reagents. These works greatly promote the development of template-free methods in the application of multi-step retrosynthesis planning.

While Transformer-based models possess so many desiderata, they suffer from two severe shortcomings: (1) lack of diverse outputs [25], and (2) chemically implausible outputs. So far, these difficulties have not been intensively discussed in the chemistry literature and are partially diverted by the fact that Transformer-based models perform well under the metric of Top-N accuracy. However, this metrics is not entirely appropriate for retrosynthesis. Schwaller et al. [33] proposed a multifaceted evaluation scheme to replace the Top-N accuracy that could capture these two subtle issues to some extent. In this work, we still stick to Top-N accuracy to offer a consistent comparison with other methods reported in the literature, but we also propose strategies to deal with these two shortcomings.

There are only a few studies [25,28] set out to address either of these two shortcomings. For instance, to reduce the number of grammatically invalid SMILES outputted by a Transformer, Zheng et al. [28] proposed a self-correction learning scheme. While this method reduces the number of invalid SMILES, which can be easily detected, it does not guarantee corrected outputs are chemically implausible reactants. In a separate study, Chen et. al. [25] attempted to coax a Transformer into giving more diverse outputs covering a broader set of reactions. This successful demonstration by Chen et al. is encouraging, but this work could not compete with the most recent single-step retrosynthesis methods in Top-N accuracy. Further details on these two shortcomings are elaborated in **Section 3**.

Herein, we set out to alleviate both shortcomings while achieving an accuracy that is competitive with advanced models. We named our single-step method the RetroPrime. Following a recent trend [19,21] to imitate a chemist's approach to retrosynthesis in two stages: (1) disconnecting a molecule at a reaction center, and (2) converting synthons into reactants; RetroPrime relies on two Transformers to predict reaction center and synthons-to-reactants, respectively. This two-stage framework simplifies the complex pattern of chemical reactions for Transformer to learn in a divide-and-conquer manner. To enhance output diversity and chemical plausibility, we introduce the "mix and match" and "label and align" strategies in the RetroPrime workflow. To estimate chemical plausibility, we adopt forward reaction prediction model verification method, which is similar to the round-trip accuracy used by Schwaller et al. [33]. Details can be found in **Section 2.**

In this work, we have evaluated our methods on a standard dataset USPTO-50 K [34] and the large-scale USPTO-full [35]. By substantially improving Transformer's shortcoming while achieving great performances, RetroPrime is a reliable tool and points out a promising direction to further develop more advanced template-free methods that, hopefully, may enable fully automated and data-driven retrosynthetic planning of complex molecules in the future.

2. Methodology

2.1. Bird's-eye view

Following chemists' approach, we solve a one-step retrosynthesis in two stages. 1. Given a molecule, identify possible reaction centers and disconnect relevant bonds to produce synthons (P → S). 2. Transform synthons to reactants (S → R). Both tasks can be accomplished with advanced deep-learning techniques. In previous work, Shi et.al. [19] and Somnath et.al. [21] have used two graph neural networks to complete the above two stages. Different from the above works, we employ the powerful Transformer model, commonly used for natural language processing, and integrate domain knowledge through token tagging in both stages to complete the single-step retrosynthesis predictions.

In this work, we refer to the two Transformers as the product-to-synthons (P2S) model and the synthons-to-reactants (S2R) model, respectively. Fig. 1 provides a bird's-eye view of our proposed method pipeline. Firstly, the P2S model tags atoms in a molecule that may potentially participate in a reaction. Multiple possibilities are returned by the P2S model. For each case, a set of synthons are converted from the tagged SMILES according to the rules defined in **Section 2.2.1**. Subsequently, SMILES strings for these synthons are preprocessed (explained in **Section 2.2.2**) before feeding them as input to the S2R model to predict possible reactants containing these synthons as substructures. We used the same regularization token as Schwaller et al. [36] in the training data processing. The parameters and training details of the transformer model are shown in the **Supplementary Information** Section S1. All of our methods are coded in Python 3.6. We used Open NMT [37] to build the transformer model and RDKit [38] to process the molecular structure. We open-source the code at the end of the article so that you can handle your own dataset using our data processing methods.

2.2. Data preparation

To train the two transformers in Fig. 1, we generate two new datasets (Reaction-Center dataset and Synthons-to-Reactants dataset) by processing information derived from the publicly available reaction dataset USPTO-50 K, which contains ~ 50,000 records of atom-mapping reactions that have been classified into ten distinct reaction types [34]. Following other previous studies, we consider two settings for the predictive task depending on whether the reaction type for each data record is provided as part of the input to the model. Furthermore, we adopt the same training/validation/test split as reported in Coley et al. [15], which recommends a split of 80%/10%/10% of 50 K reactions. Table S2 succinctly summarizes the USPTO-50 K dataset. In these new datasets, each data entry is prepared in the format of < source>-<output > pair, following the standard data format for NLP tasks. Further details on these datasets are elaborated thoroughly in the following sections.

2.2.1. Reaction-Center dataset generation and augmentation

For each atom-mapping reaction record in the USPTO-50 K, we analyze and tag the essential atoms of the product molecule involved in a reaction. The P2S model is trained to identify these tagged atoms for each reaction. Hence, the source of the reaction-center dataset is the SMILES of the product, and the target is the SMILES representing the same order of atoms as the input and with tags added to the reaction center atoms.

Different from the forward synthesis prediction, retrosynthetic prediction requires not only the correct decomposition of the product molecule, but also the complementation of the leaving group in many cases. We defined four tags based on the number of reactants (one or more) and the presence or absence of leaving groups (ignore the hydrogen atoms) by investigating the training reaction dataset. The definitions for these four tags are summarized below, and further details can be found in Fig. 2. And Table S3 shows the number of reactions

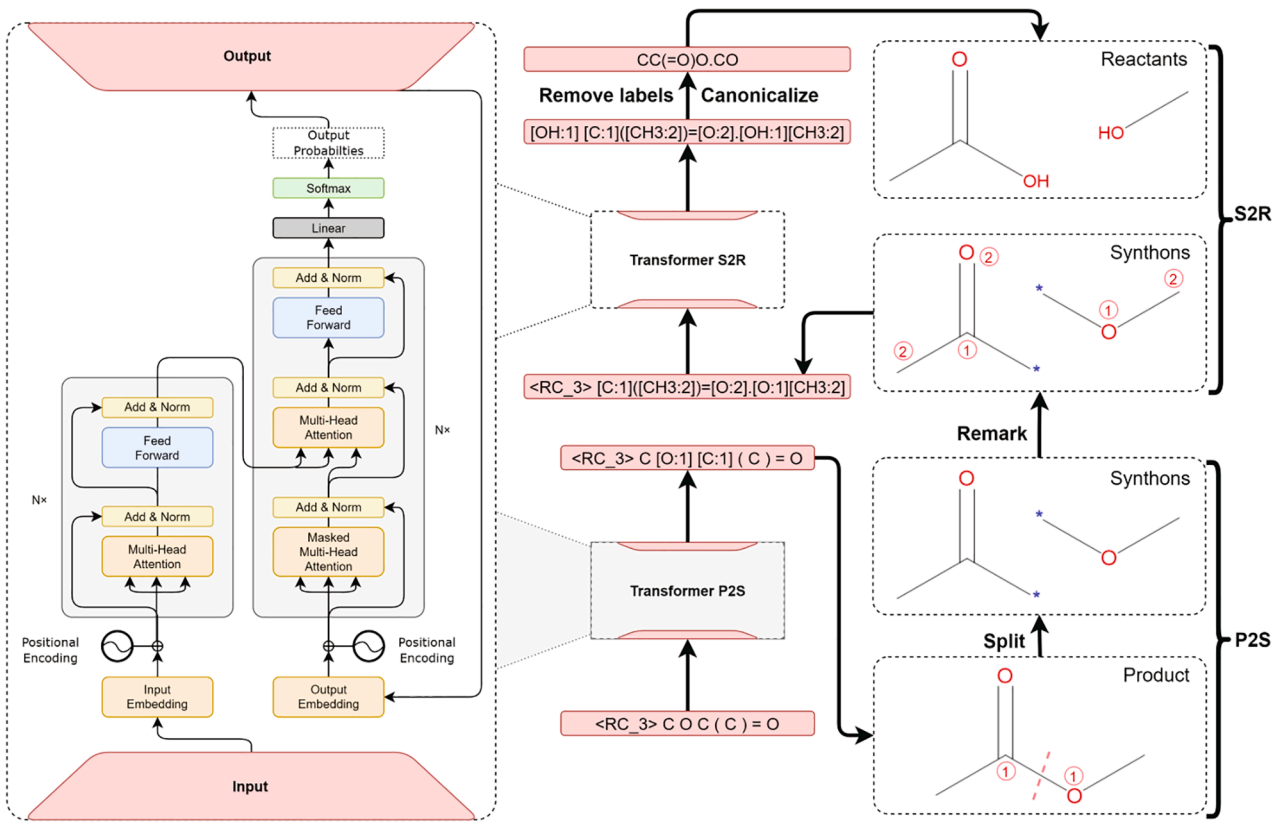

Fig. 1. Method pipeline. First, the canonical SMILES of the product is input into the Transformer P2S to obtain the product SMILES with the reaction center tag. The second step, if the predicted sequence is tagged with a disconnected tag then the bonds between the tagged atoms are broken to form synthons, otherwise the product itself is treated as a synthon. The third step, re-label the synthon(s) sequence and place the reaction center at the front of the sequence. The fourth step, input the synthon(s) into the Transformer S2R to predict the corresponding reactant(s). Finally, remove labels and convert reactant(s) into canonical SMILES. (<RC_i > is the reaction type if applicable.)

represented by the four types of tag.

- Tag 1, Tag two atoms. Disconnect bond between these atoms to form two synthons. This type of tag represents 68.7% of reaction data of the whole USPTO-50 K training set. Reactions represented by this tag have two reactants, at least one of the reactants contains the leaving group, including most of the substitution reactions, simple coupling reactions, inter-molecular esterification reactions, inter-molecular amination reactions, etc. (There is no ring formation in these reactions). The general formula and example of these reactions are shown in Fig. 3.
- Tag 2, Tag at least two atoms but generate synthons without breaking any bonds. The product itself is a synthon. This type of tag represents 5.8% of reaction data of the whole USPTO-50 K training set. Reactions represented by this tag have one reactant and multiple reaction-center atoms, including most rearrangement reactions, multi-site deprotection reactions (such as the deprotection reaction of acetals and ketals.), and intramolecular cyclization reactions, etc. The general formulas and examples of these reactions are shown in Fig. 4.
- Tag 3, Tag one atom. The product itself is a synthon. Reactions represented by this tag have only one reactant, and there must be a leaving group. This type of tag represents 23.3% of reaction data of the whole USPTO-50 K training set. These reactions are mostly single-site deprotect reactions. The general formula and example of these reactions are shown in Fig. 5.
- Tag 4, Tag multiple atoms. Disconnect bonds between these atoms to form at least two synthons. This type of tag represents 2.2% of reaction data of the whole USPTO-50 K training set. Reactions represented by this tag have at least two reactants and multiple reaction sites, including multi-molecular cyclization and multi-component reactions. The general formulas and examples of these reactions are shown in Fig. 6.

There always exist multiple valid SMILES to represent one molecule. It has been reported that NLP models, such as various RNN architectures, tend to perform better for applications in the molecular science when the dataset is augmented with the same molecules represented in multiple SMILES. In this case, we augment the Reaction-Center dataset by using the SMILES enumerator [39] to randomly generate nine additional SMILES for each canonical one. An illustration is given in **Supplementary Information Fig. S2**. Note that the source and the target of each data entry only differ by the tags attached to the reaction-center atoms on the target side. Table S4 provides the amount of data in the augmented Reaction-Center dataset.

2.2.2. *Synthons-to-Reactants dataset generation and augmentation*

According to the pipeline depicted in Fig. 1, synthons are converted from the product molecules by following the instructions implied by the tags introduced in **Section 2.2.1**. These synthons need to be further processed with labels before feeding to the S2R model. The labeling principle is that the reaction-center atoms (the atoms tagged in **Section 2.2.1**) are marked as 1, the adjacent atoms (connected via chemical bonds) are marked as 2, and the remaining atoms are marked as 3. This is how we prepare the source (input) part of this dataset. As for the corresponding target (output) part, we take the reactants from the original USPTO-50 K dataset and furnish the SMILES with labels according to the above principle. Additionally, the atoms of leaving groups are also marked as 1 for the reactants. Finally, for each synthon-reactant pair, we calculate the edit distance and attempt to minimize it by

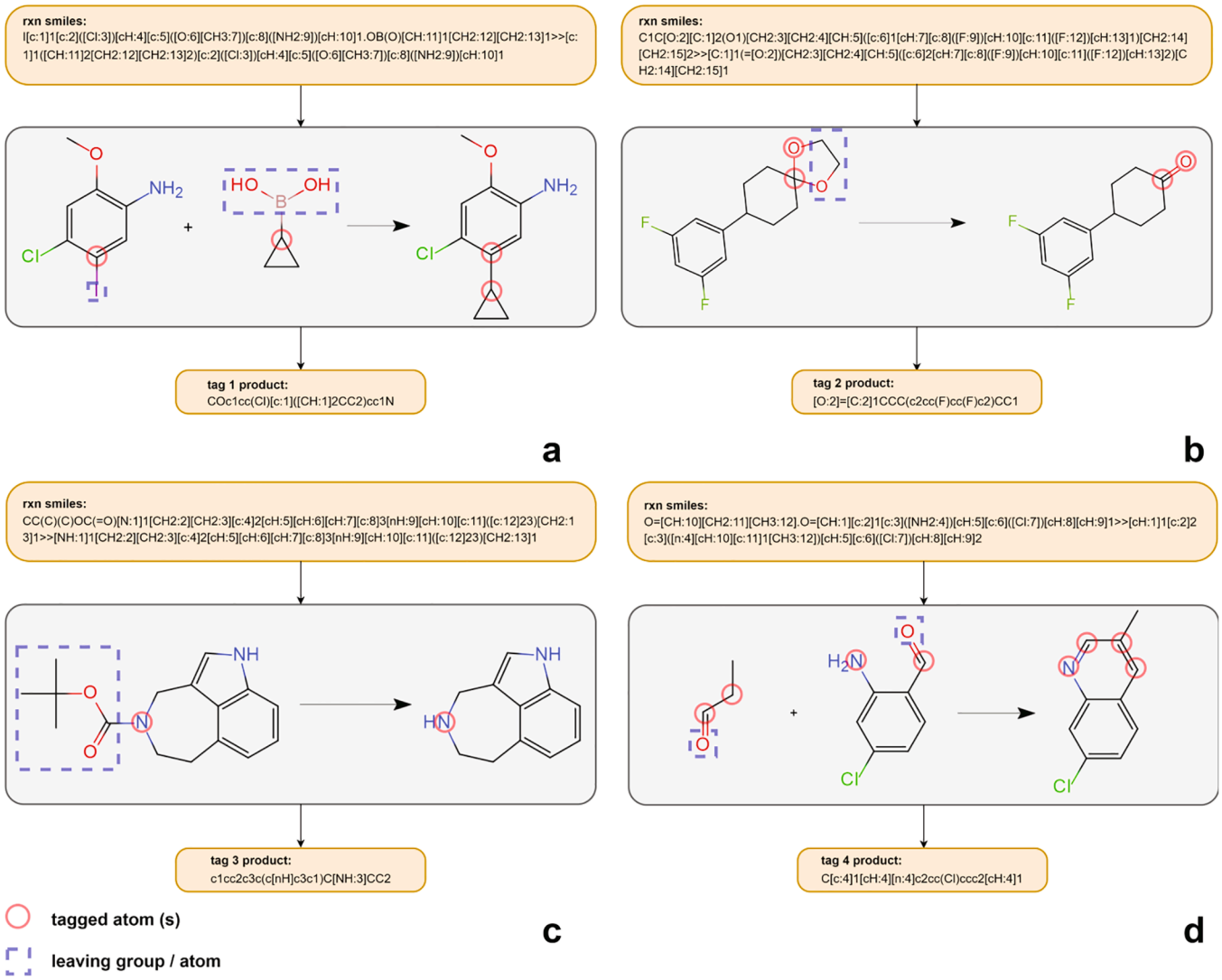

Fig. 2. The interpretation of the reaction center tag. (a) Tag 1, Tag two atoms. Disconnect bonds between these atoms to form two synthons, (b) Tag 2, Tag at least two atoms but do not disconnect any bonds. The product itself is a synthon. (c) Tag 3, Tag one atom. The product itself is a synthon, there must be a leaving group, this tag is a non-disconnected mark and (d) Tag 4, Tag multiple atoms. Disconnect bonds between these atoms to form at least two synthons. Multi-component reactions also fall under this tag.

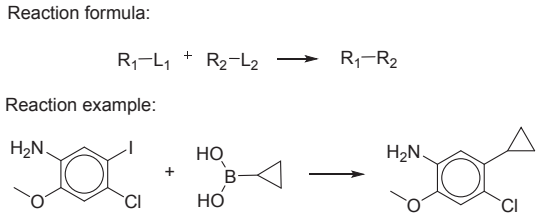

Fig. 3. The general formula and example of the reaction represented by reaction tag 1.

manipulating the target sequence in order to align the two SMILES strings as closely as possible. As shown in Fig. 7, after alignment, a typical input–output pair in the S2R dataset share a relatively large and identical subsequence. We called this strategy "Label and Align".

As shown in **Supplementary Information Fig. S3**, when a SMILES contains multiple entities, we permute the SMILES to generate additional data. For each augmented data entry, we still align the source and target sequences to minimize the edit distance. The amount of data in the augmented Synthons-to-Reactants dataset is given in Table S4.

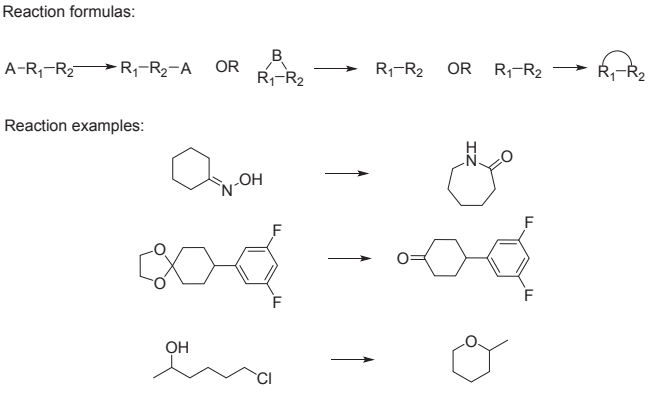

Fig. 4. The general formulas and examples of the reaction represented by reaction tag 2.

2.2.3. *Large-scale experiments on USPTO-full*

To more comprehensively test our method, we produce the USPTO-full dataset from USPTO (1976-Sep2016) [35] following the cleaning

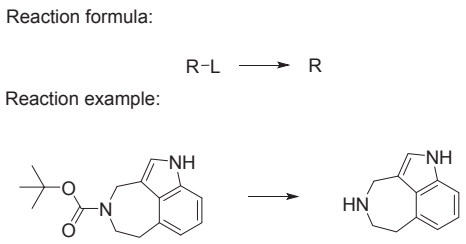

Fig. 5. The general formula and example of the reaction represented by reaction tag 3.

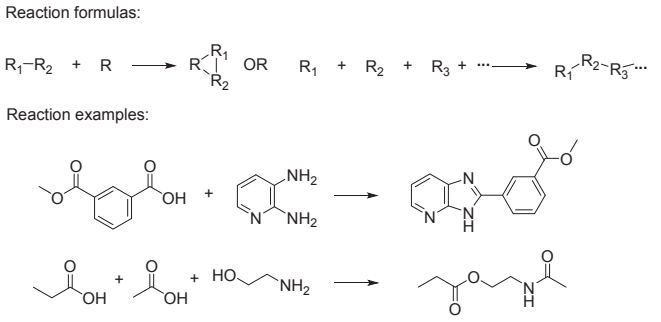

Fig. 6. The general formulas and examples of the reaction represented by reaction tag 4.

method of previous researchers [14]. There are 1,808,937 raw records in USPTO (1976-Sep2016). For reactions involving multiple products, we duplicate the same entry as many times as the number of products. In each copy, we remove all products but one to create additional data with a unique product molecule for the same reaction. We use exactly the same training/validation/test splits as Dai et al. [14], which contain 80%/10%/10% of the total 1 M unique reactions. See https://github.com/Hanjun-Dai/GLN/blob/master/gln/data_process/clean_uspto.py for data cleansing and split script.

Repeating the procedures given in USPTO-50 K dataset processing, we further produce the Reaction-Center dataset and the Synthons-to-Reactants dataset from the USPTO-full. Note, this dataset generation procedure includes the data augmentations described in the previous section. Further details of these large-scale datasets are summarized in Table S2-S4.

2.3. Evaluation metrics

The evaluation metrics we used are slightly different for the two stages. The essence of the P2S stage is to obtain reaction tag information. We sequentially extracted the tags in the predicted sequence with the same number of atoms of the input product, and performed preliminary screening according to the tag rules defined in **Section 2.2.1**. Finally, we re-labeled the tags that meet the rules into the input SMILES to get the results of the P2S stage, and used these results to evaluate the P2S stage. The evaluation method and valid prediction standard of the P2S stage are shown in Fig. 8.

For the Transformer S2R, it is expected to translate synthons to reactants. To boost accuracy, we propose to mark atoms in order to facilitate the alignment of the source and target sequences for this

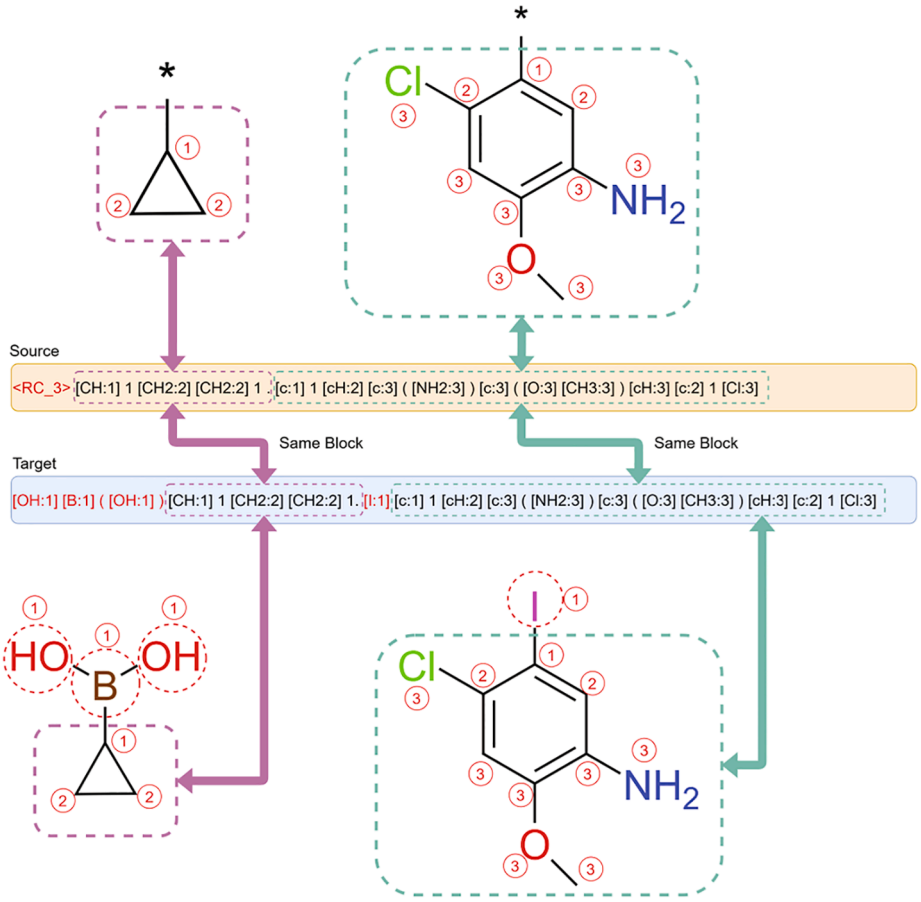

Fig. 7. Label and Align. We use labeled SMILES that minimize the editing distance in the S2R stage so that the source and target SMILES have many blocks that are exactly the same. (<RC_i > is the reaction type if applicable.).

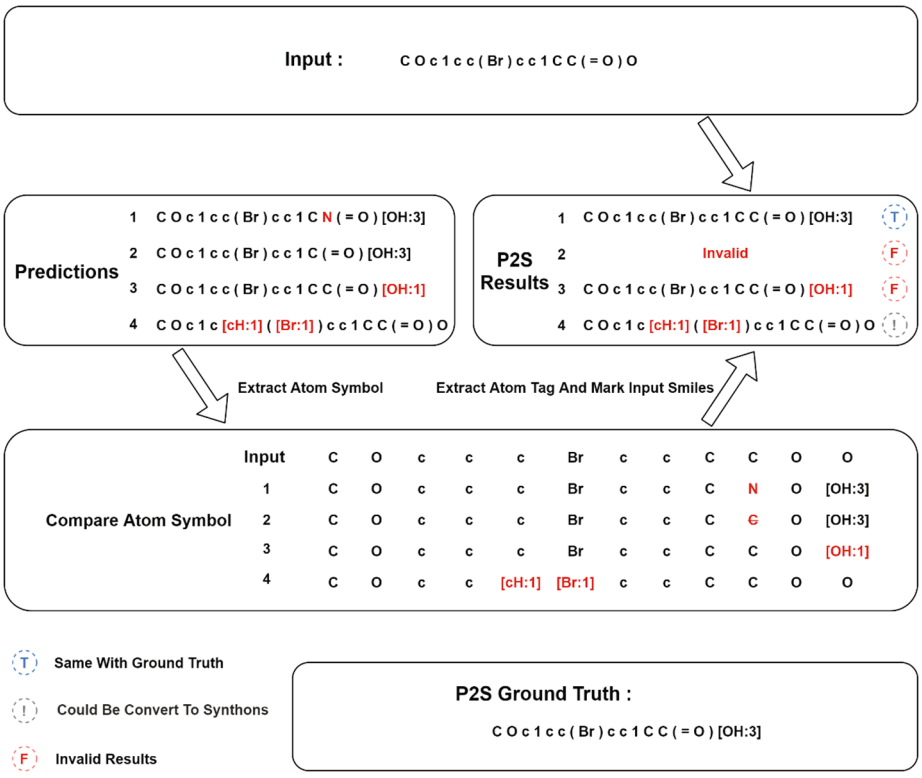

Fig. 8. The evaluation method and valid prediction standard of the P2S stage. After generating the tagged SMILES, we first determined whether the number of atoms represented by the predicted SMILES is the same as the input. If the number of atoms is not the same (e.g., No.2), the prediction is invalid. If the extracted tag does not conform to the rules defined in Section 2.2.1 (e.g., No. 3), then this prediction is also invalid. If the tags conform to the rules (e.g., No.1 and No.4), then these tags are extracted and relabeled into the input SMILES sequence to obtain the final prediction result.

translation task. Hence, one should remove these labels and convert the target sequence (given by the S2R model) back to canonical SMILES before comparing to the ground truth for a given reaction in the USPTO-50 K/full dataset.

2.4. Reaction diversity

For predictions with unknown reaction types, we check whether RetroPrime can offer diverse reaction outcomes. To estimate diversity, we used a reaction type counting method similar to that used by Schwaller et al. [33]. We use a reaction type predictor [25] based on a typical message-passing neural network to predict the reaction type of a predicted reaction. Taken RetroPrime's Top-n predictions for each test case, we use reaction type predictor to estimate the number of unique reaction types. Then we calculated the average number (D_n) of reaction types corresponding to the predicted reactants for each product:

$$D_n = \frac{1}{I}\sum_i^I T_{i,n} \quad (1)$$

Where $T_{i,n}$ is all the reaction types included in the **valid prediction results** for one input when the Top-n prediction results are taken, and I is the number of products for the test. In **Section 3.4**, we compared the diversity by taking n as 10.

In addition, through the investigation of the USPTO-50 K dataset, we used all the reaction data, which are different reactions but lead to the same product in the validation dataset and the test dataset, as the multi-ground-truth test set. We collected 22 groups of products with multiple ground truth reactants as another indicator to test and compare the models' diversity. To increase the difficulty, we also tested the diversity of retrosynthesis models using all the multi-ground-truth data in USPTO-full, which has a total of 34,003 groups of products, and we call this dataset as multi-ground-truth USTO-full. The collection method for this dataset is described in **Supplementary Information** Section S1. Note that all diversity tests used model parameters obtained by training in USPTO-50 K under the condition of unknown reaction type.

2.5. Mix and match

The P2S model predicts how a molecule can be decomposed into simpler constituents. Various decompositions imply different chemical reactions. In other similar studies, one would simply take synthons for the Top-1 decomposition to make further predictions of reactants. However, we reckon that processing multiple decompositions down the pipeline of Fig. 1 is a simple yet highly effective method to enormously enhance the overall output diversity. We call this strategy "mix and match". Here we present a schematic Fig. 9 to illustrate the "mix and match" strategy. See **Supplementary Information Fig. S4** for further details on the "mix and match".

2.6. Label and align

While preparing the S2R dataset, we meticulously minimized the edit distance for the input–output sequences and inserted extra labels as detailed in **Section 2.2.2**. These efforts aim to expose as much similarity between the source and target sequences as possible and facilitate the translational model's learning to capture the chemistry behind the data. Indeed, the "Label and Align" strategy not only improves the Transformer's overall accuracy but also decreases the number of chemically implausible outputs. We design experiments to provide more proof in **Section 3.5**.

3. Results and discussion

3.1. Challenges for translation-based retrosynthetic model

In recent years, the sequence-to-sequence (seq2seq) based generative models have been widely used in the prediction of single-step retrosynthesis because of its low requirements on data processing (for example, not requiring curation of atom-mapped reactions templates) and strong generalization ability. However, it is not an entirely error-free approach. Liu et al. have summarized three types of prediction errors of the translation-based retrosynthesis model: [18]

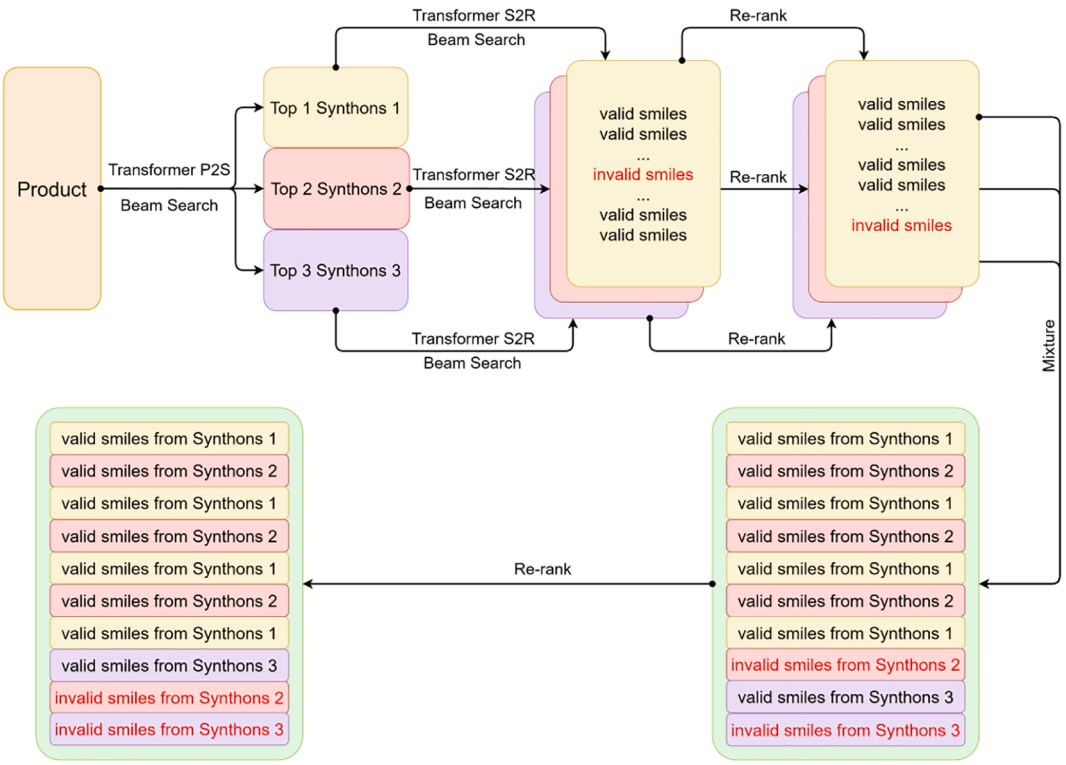

Fig. 9. Mix and Match. We select the rank 1–3 synthons predicted by Transformer P2S and send them to Transformer S2R to predict the reactants, and the obtained results are alternately combined. Use the re-rank approach to rank invalid SMILES at the end.

(1) The SMILES for predicted reactants are grammatically invalid. This problem can be resolved with a simple filter. A small amount of invalid SMILES outputs does not pose a severe challenge.
(2) The SMILES for predicted reactants are grammatically valid and chemically plausible, yet the predicted reactants are not identical to the ground-truth reactants specified in the dataset. This is due to the fact that each molecule may contain multiple reaction centers corresponding to multiple sets of possible reactants, but the dataset does not record all possible cases.
(3) The SMILES for predicted reactants are valid, but the product-reactants pair does not constitute a chemically plausible reaction. Nowadays, this type of error can be identified with the forward models [33] or judged with an in-scope filter [40,41]. For training the in-scope filter, the critical part is the data. We can easily obtain a large amount of positive reaction data, but the negative data are usually not recorded in the reaction database. In any case, these methods require additional models to judge whether the predicted reactions are plausible or not. It is essential that to explore how we can reduce this type of error generated by retrosynthesis models.

The second type of error should not be viewed as a real mistake in the context of synthesis planning. Rather, this notion of "error" does expose the fact that there are always multiple valid approaches to synthesize an organic compound. Hence, one expects a single-step retrosynthetic model to enumerate as many valid options as possible. Unfortunately, upon close inspection, the seq2seq translation-based models do not exhibit much diversity in their outputs. Furthermore, they also tend to produce the third type of errors as shown in Fig. 10. Fig. 10**a** presents a single-stage Transformer (Referred to as S-Transformer, it will be introduced in the following section.)'s Top-6 recommendations of possible reactants for a selected molecule, while an alternative view of these six recommendations (in terms of SMILES which directly output by S-Transformer) is presented in Fig. 10**b**. As clearly shown in the figure, S-Transformer's predictions are lack of diversity. Furthermore, many predicted reactants are completely unreasonable in the chemical reaction sense. In order to tackle these challenges, we propose the "Mix and Match" and "Label and Align" strategies in RetroPrime to alleviate the problems of poor diversity and high chemical implausibility, respectively. "Mix and Match" explicitly takes into account the different ways of decomposition of the products and the diverse options of the synthons. "Label and Align" uses labeled tokens to distinguish and align reaction center and conservative groups between synthons and reactants. For the same example as Fig. 10, our models' direct outputs process are shown in Fig. 11. After adopting the above two strategies, the diversity and chemical plausibility of model's predictions can be significantly increased.

3.2. Baseline

We benchmark our method against seven baselines, which do not use correction methods like used by RetroXpert [22] and SCROP [28], including five template-free and two template-based methods. Specifically, Seq2Seq [18] is a template-free approach that trains an LSTM model to translate the SMILES of target molecules to SMILES of reactants. RetroSim is a template-based method that recommends templates for target molecules based on the molecular similarity between the present molecule and the ones in the dataset. S-Transformer is similar to the Seq2Seq translation model but using a single-stage transformer instead of LSTM architecture at the core. G2Gs [19] and GraphRetro [21] are template-free approaches using graph neural networks to predict retrosynthesis. Under the premise of the model without correction methods, GraphRetro achieved state-of-the-art Top-n accuracy in the USPTO-50 K dataset. GLN [14] is a template-based method, which samples templates and reactants jointly form a distribution learned by a conditional graphical model. AT [42] is a state-of-the-art transformer-based single-step retrosynthesis model, which adopts a very effective data augmentation strategy.

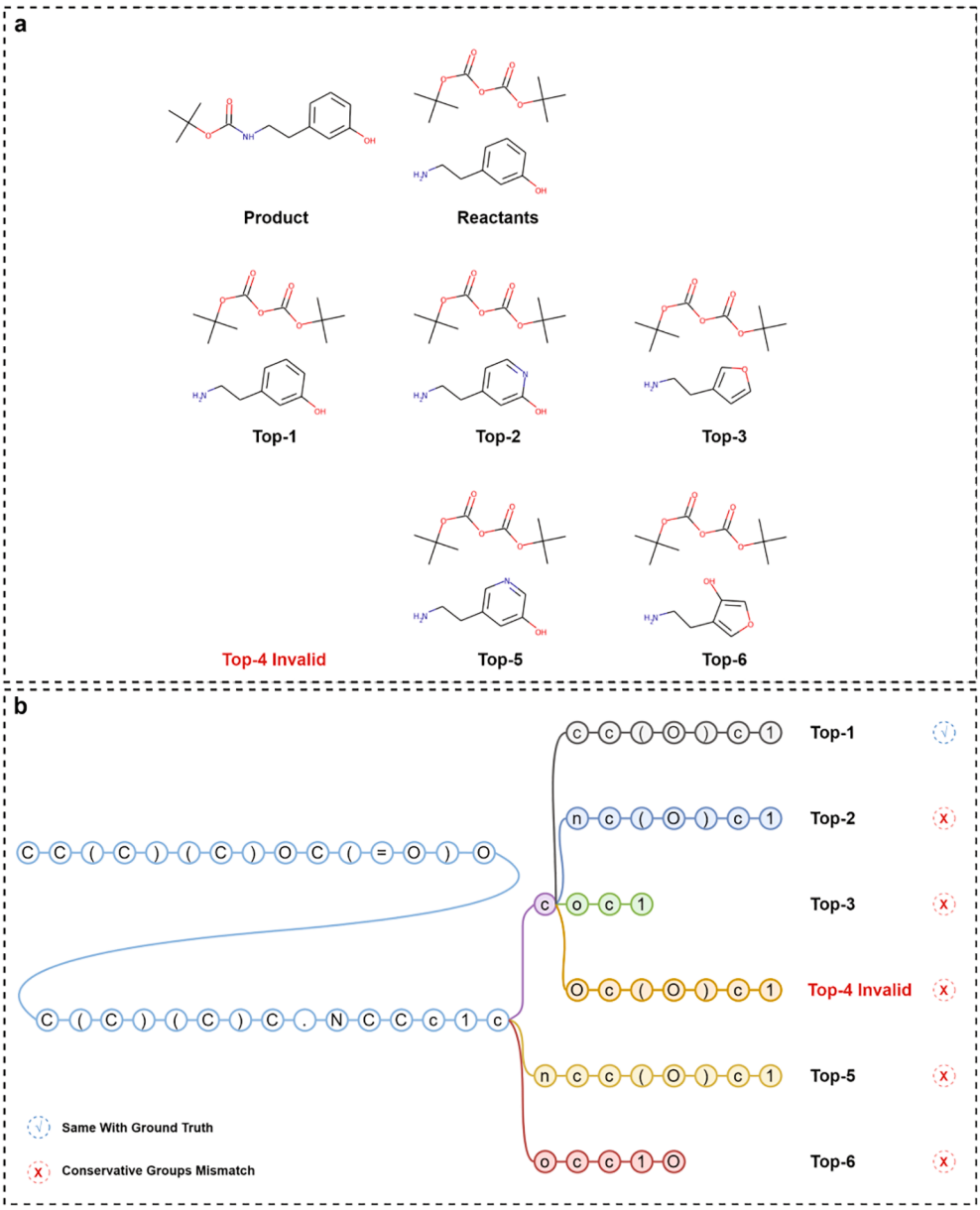

Fig. 10. (a) Visualization of a set of predicted results selected from S-Transformer test dataset. The first row contains the input molecule and the ground truth reactants in the dataset, the second row and the third row are the Top-6 predicted results. In this example, we can observe that S-Transformer predicts that one of the reactants is exactly the same as the ground truth reactant, and the other reactant is very similar to the ground truth reactant, but the atomic changes on the conjugated membered ring makes the results completely unreasonable. (Top-1 hits the ground truth reactants.) **(b)** Visualization of the sequence directly output by the S-Transformer. The results show that most part of the sequences predicted by the S-Transformer are the same, and the model did not capture the reaction center and conservative groups of the molecule in the chemical reaction. In most cases, different parts of the predicted sequence often lead to chemically implausible results, rather than diversification.

3.3. Top-N accuracy

We evaluated the method in two datasets, USPTO-50 K and USPTO-full, which contain ~ 50000 and ~ 1 M reaction data, respectively. For the USPTO-50 K dataset, our results are presented in Table 1. Our method achieves the Top-1 accuracy of 64.8% and 51.4% when the reaction type is known and unknown, respectively. It can be seen from the table that RetroPrime is completely superior to these two models, Seq2Seq and S-Transformer, which only use canonical SMILES to express molecules. Compared to the template-based methods, except for the Top-10 accuracy when the reaction type is known, RetroPrime is far more accurate than RetroSim and competes with GLN. It is worth noting that template-based methods perform well at Top-10 accuracy because template-based methods are not limited to predicting similar reaction precursors, and the depth search approach (Top-n, $n \geq 10$) is conducive to finding the reaction precursors recorded in the dataset. Compared with the graph-based template-free methods, RetroPrime is entirely superior to G2Gs when the reaction type is unknown. And its accuracy in Top-1 and Top-3 is also better than G2Gs when the reaction type is known. GraphRetro is very well designed at capturing and learning representations on small dataset, achieving the state-of-the-art performance on USPTO-50 K. AT applies the multiple effective data augmentation strategies to achieve great performance on the transformer-based models, which indicates that how to effectively and completely express the molecular structure information is very important for the accuracy of prediction.

The performance of each method in the noisier USPTO-full dataset is shown in Table 2. The performance of RetroPrime is close to AT, and the Top-1 accuracy is 44.1%. It is worth noting that S-Transformer trained using only canonical SMILES also significantly outperforms the template-based method GLN in this dataset, showing the transformer-based approach's advantages in the noisy dataset. In addition, we also list these two separate stages Top-N accuracy of RetroPrime in Table S5 and S6.

3.4. More diverse predictions

We also investigate whether our method provides outputs covering a broad range of chemical reactions. This is crucial if these single-step predictors were to be integrated into a multi-step retrosynthetic route

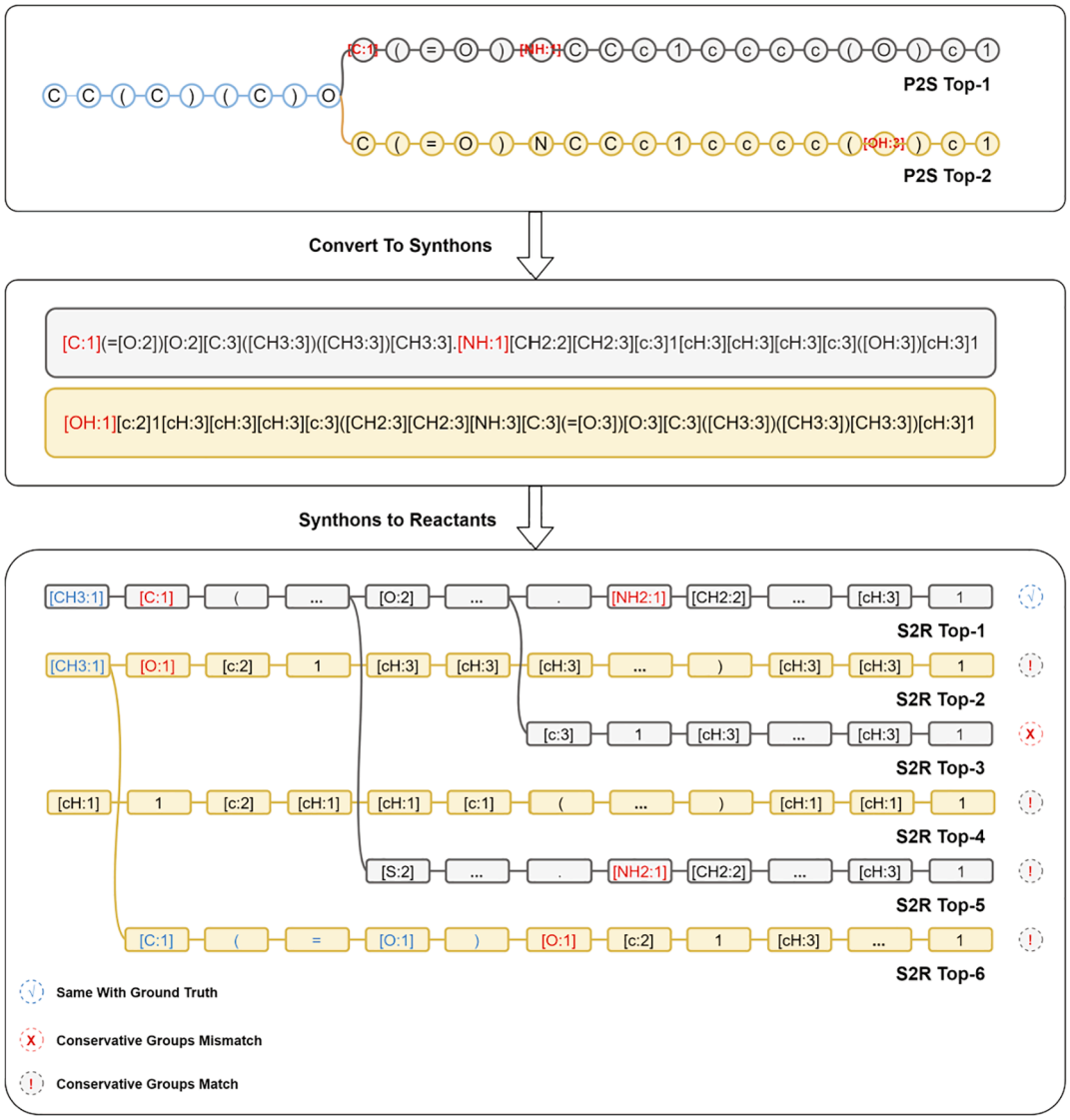

Fig. 11. The two stages of RetroPrime directly output sequence display. Top black box shows the tagged results of the two types of reaction centers directly predicted by the P2S stage. Middle black box shows the sequences representing the two sets of synthons. (These two sequences are displayed in different colored boxes.) The tokens marked with red are reaction center atoms, which are fixed at the forefront of the sequence representing a structure. Bottom black box shows the sequences of the direct output of the model in the S2R stage, which are the combination of two synthons corresponding results (Top-6). The tokens marked with blue indicate the predicted leaving groups, while the red tokens are still the reaction center. "Mix and match" can increase diversity. "Label and align" can enhance the conservative groups relationship between synthons and reactants, thereby improving the chemical credibility of predictions. Corresponding molecular pictures are shown in Supplementary Information Fig. S5.

Table 1
USPTO-50 K dataset Top-N exact match accuracy.

Methods	Top-N accuracy %							
	Reaction type known				Reaction type unknown			
	1	3	5	10	1	3	5	10
Seq2Seq [18]	37.4	52.4	57.0	61.7	–	–	–	–
RetroSim [15]	52.9	73.8	81.2	88.1	37.3	54.7	63.3	74.1
S-Transformer [30]	57.3	71.6	75.2	78.0	43.5	59.2	63.9	68.2
G2Gs [19]	61.0	81.3	86.0	88.7	48.9	67.6	72.5	75.5
GLN [14]	64.2	79.1	85.2	90.0	52.5	69.0	75.6	83.7
RetroPrime	64.8	81.6	85.0	86.9	51.4	70.8	74.0	76.1
AT [42]	–	–	–	–	53.2	–	80.5	85.2
GraphRetro [21]	67.8	82.7	85.3	87.0	63.8	80.5	84.1	85.9

Table 2
USPTO-full dataset Top-N exact match accuracy when the reaction type is unknown.

Methods	Top-N accuracy %			
N	1	3	5	10
RetroSim [15]	32.8	–	–	56.1
GLN [14]	39.3	–	–	63.7
S-Transformer [30]	42.9	58.1	61.0	66.8
RetroPrime	44.1	59.1	62.8	68.5
AT [42]	46.2	–	–	73.3

Table 3
Average number of Reaction types between S-Transformer and RetroPrime on USPTO-50 K dataset when the reaction type is unknown. (Top-10).

Methods	Avg. Reaction Type (D_{10})
S-Transformer [30]	1.74
RetroPrime	2.40

planning. As the setting of unknown-reaction-type is more natural for this purpose, we choose this setting and compare our method against the S-Transformer (as both approaches mainly use Transformer to make predictions). This diversity estimation, based on the metrics (top-10 average reaction type D_{10}) introduced in **Section 2.4**, is shown in Table 3.

In addition, we used a multi-ground-truth test set to compare RetroPrime and S-Transformer's ability to predict multiple correct answers at the same time. Retroprime can completely match 77% (17 groups) of results, and S-Transformer can match 32% (7 groups). All 22 products can be found at least one correct reaction pathway by two methods. (The standard of "completely match" is that all correct reactants appear in the Top-10 prediction results of the model.) An example of the complete match is shown in Fig. 12**.** In the more difficult multi-ground-truth USPTO-full, RetroPrime can completely match 6.74%, which is also higher than S-Transformer's 4.28%. In addition, RetroPrime can find at least one correct reaction pathway for 46.16% of the test products, while S-Transformer only found 40.86%. The results show that, compared with S-Transformer, our method can still find multiple correct reaction pathways for products more effectively in this dataset, even when there are few data of multi-ground-truth in the training set. (See **Supplementary Information** Section S1 for detailed test results on multi-ground-truth reaction data.)

Our method can also generate more diversified prediction results for target molecules that do not belong to the multi-ground-truth reaction data. As shown in Fig. 13, We visualized the prediction results of RetroPrime and S-Transformer for some products of multiple reaction centers to demonstrate that RetroPrime can predict more diverse results. See **Supplementary Information** Section S2 for more similar examples. The more diverse prediction results of Retroprime are due to the mix and match of Top-3 synthons results.

3.5. *The effects of the "Label and Align" strategy*

Recall that we did two things while building the S2R dataset. We

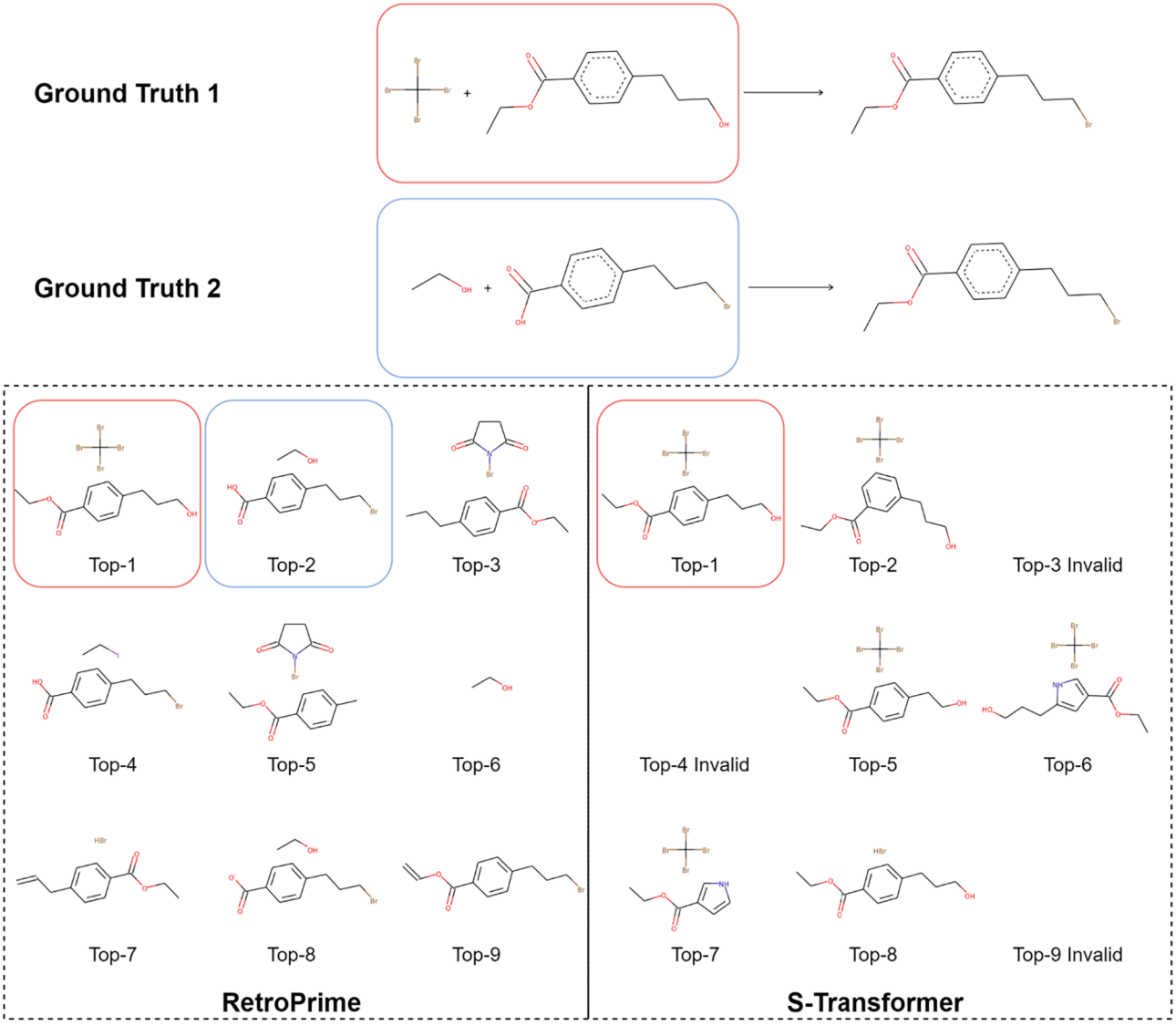

Fig. 12. An example of completely match in multi-ground-truth reaction data. In this example, two different reactions lead to the same product, Retroprime matched two ground truth results in the first and second place of the predicted results, and S-Transformer only matched one.

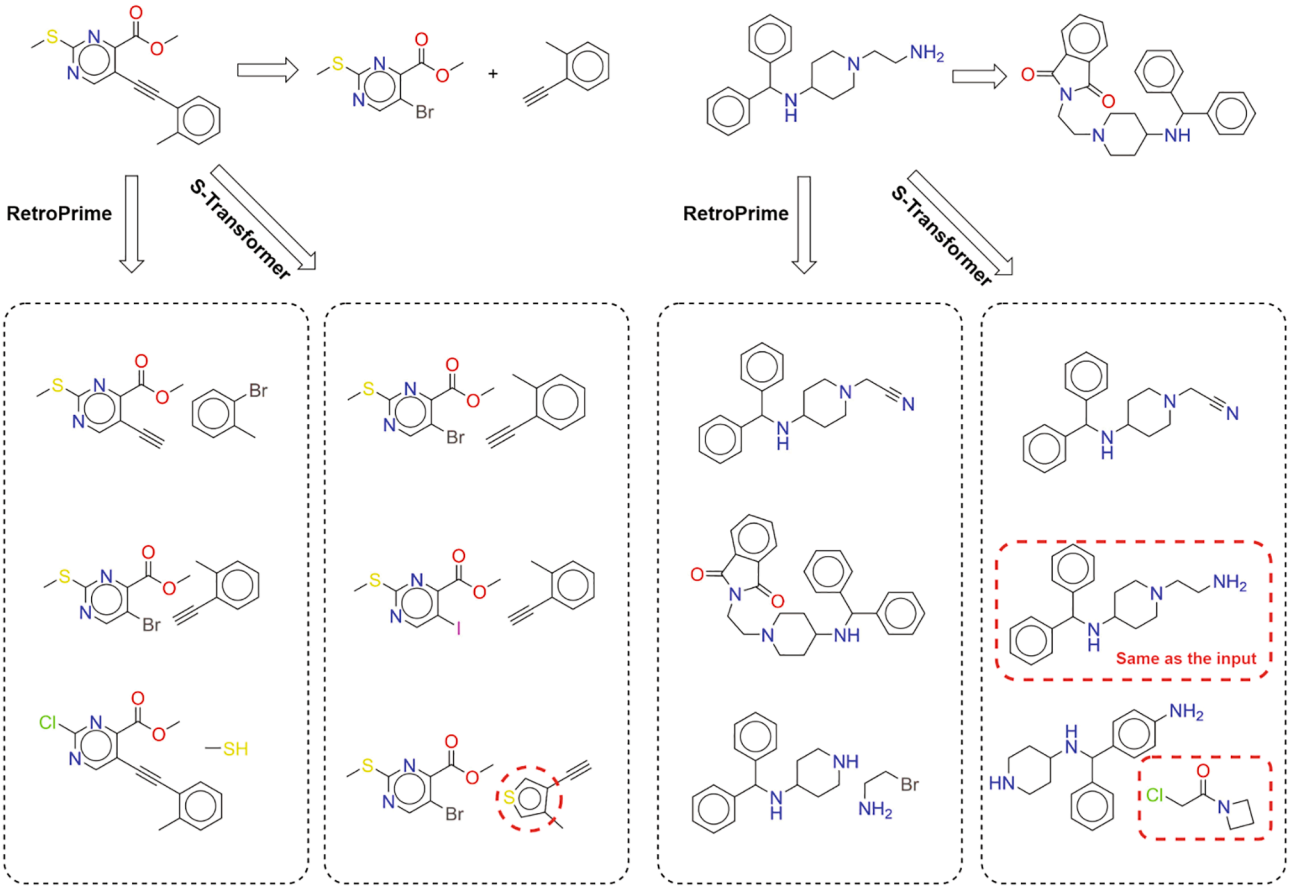

Fig. 13. Comparing the predictions of RetroPrime and S-Transformer when the reaction type is unknown. **(a)** The ground truth results in this example can be predicted by both RetroPrime and S-Transformer. The difference is that RetroPrime gives three different disassembly methods. The first two S-Transformer dismantling methods are the same except for halogen atoms, and the last prediction of the S-Transformer incorrectly predicted a thiophene ring instead of a benzene. **(b)** Retroprime not only correctly matched the ground truth result, but also recommended two additional reactions, reduction and nucleophilic substitution. S-Transformer did not match the ground truth. It only predicted the reduction reaction, and the other two predictions are chemically implausible.

Table 4
Compare the Top-N accuracy of the two methods in the S2R stage. The same reaction type input configuration uses the same P2S model.

S2R Methods	Pipeline Top-N accuracy %							
	Reaction type known				Reaction type unknown			
N	1	3	5	10	1	3	5	10
Label and align	**64.8**	**81.6**	**85.0**	**86.9**	**51.4**	**70.8**	**74.0**	**76.1**
Canonical smiles	60.2	75.2	78.8	81.2	48.4	66.2	70.0	72.5

align input–output sequences and mark atoms with extra labels. In this section, we attempt to elucidate the benefits these efforts provide.

We designed experiments to clarify the benefits of these efforts. In this experiment, we train a modified Transformer that is asked to translate synthons to targets in canonical SMILES, i.e., without sequence alignments and labels. Table 4 compares the original experiment's (Label and Align S2R) outcome (as depicted in Fig. 1) and the new experiment with the S2R model replaced with this newly trained one (Canonical Smiles S2R). The results of Top-1 of the RetroPrime (Label and Align S2R) are 4.6% more accurate than the RetroPrime (Canonical Smiles S2R). This accuracy gain for the Top-1 result is 3.0% when the reaction type is unknown. Moreover, the accuracy gap widens between the two experiments when the comparison is expanded to consider Top-10 results, which is 5.7% and 3.6%, respectively, when the reaction is known and unknown.

In addition to increasing Top-N accuracy, we further elaborate on more subtle effects brought upon by the labels. It is easy to corroborate that not all outputs of grammatically valid SMILES by a Transformer model are chemically plausible, i.e. the input–output pair does not constitute a valid chemical reaction.

To estimate how many chemically implausible but grammatically valid SMILES are outputted by RetroPrime, we propose to use a forward reaction predictor to diagonalize potential errors. This verification method is similar to the round-trip accuracy used by Schwaller et al. [33]. In short, we feed the predictions results (i.e. reactants) of retrosynthetic method to a forward reaction prediction model, the Molecular Transformer [32]. If the forward model predicts correct product molecule at Top-1 results, the retrosynthesis is deemed successful. We refer to this metric as the forward check plausibility of the retrosynthesis model. Without taking into account of chirality, the USPTO-MIT mixed version of the Molecular Transformer reaches 88.6% [32] for the Top-1 accuracy. Although it is not absolutely accurate to use Molecular Transformer to check the chemical plausibility of the predicted reactants, the inspection results are highly correlated to the chemical plausibility. This

Table 5
Compare the forward check forward check plausibility in USPTO-50 K test dataset prediction Top-10 results.

Methods	All predictions	Reaction Type	Grammatically valid predictions	Forward Check Plausibility %
RetroPrime (Label and align S2R)	50,060	Known	48,053	45.2
		Unknown	49,786	46.8
RetroPrime (Canonical smiles S2R)		Known	48,637	33.7
		Unknown	49,790	42.0
S-Transformer [30]		Known	47,121	32.9
		Unknown	48,004	36.4

approach is designed to provide a lower bound guarantee.

Results of this scrutiny on the chemical plausibility of our method are summarized in Table 5. Recall that our test set consists of 5,006 cases. For each case, we took the first ten results (Top-10) predicted by the retrosynthesis model to perform the forward check, that is, we entered 50,060 results into the forward model to test the chemical plausibility.

Based on these results, RetroPrime (Label and align S2R) yields slightly more grammatically invalid SMILES in comparison to RetroPrime (Canonical Smiles S2R) in which the S2R model is trained with input–output pair given in canonical SMILES without extra labels. However, the potential number of chemically implausible cases is significantly reduced for our proposed method regardless of whether the reaction type is given as part of the input. In addition, as shown in Fig. 14, we provide a visual example to show what the "label and align" strategy works. The "label and align" strategy allows the model to learn how to maintain the relationship between the synthons and the predicted reactants' subgraph (conservative group) to reduce chemically implausible predictions. The corresponding sequences for this example are shown in **Supplementary Information Fig. S6**.

Clearly, our two-stage method has significantly ameliorated this deficiency of the rudimentary workflow using a single Transformer in an end-to-end fashion that directly translates a product molecule into a batch of reactants.

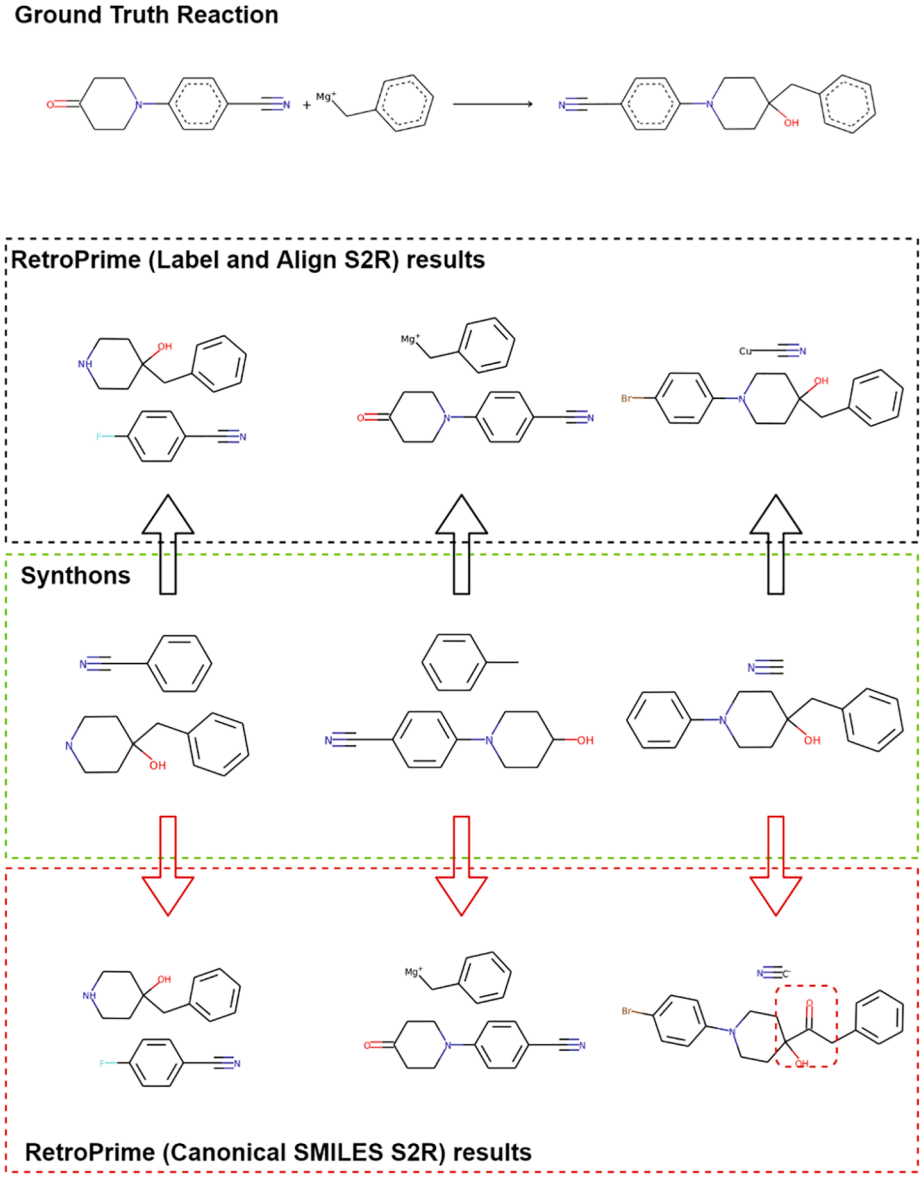

Fig. 14. The effects of "Label and align" strategy. The green box shows the three synthons predicted by the P2S model. The black box shows the reactants corresponding to the three synthons predicted by the S2R model using the "label and align" strategy. The red box shows the reactants corresponding to the three synthons predicted by the S2R model without the "label and align" strategy. In both the left and the middle examples, the model with the two strategies can predict the same reactants, which can correspond to the synthons. But in the right example, the prediction from the model without the "label and align" strategy does not correspond to the synthons and is a chemically implausible result, while the prediction from the model with the "label and align" strategy does not show this error. The corresponding sequences for this example are shown in Supplementary Information Fig. S6.

3.6. Limitation

Similar to several recent works based on two-stage single-step retrosynthesis methods [19,21], this work relies on high-quality atom-mapping reaction dataset to extract reaction centers. Earlier, the reaction atom-mapping labeling method combines graph theory methods, heuristic methods, and rule-based methods, but many quality issues still exist. In the open-source reaction dataset that relies on the above atom-mapping methods, especially in USPTO-full, the atom-mappings are not reliable in many cases, which leads to problems in the downstream retrosynthesis model trained on these atom-mapping datasets. Using more reliable atom-mapping algorithms is a way to alleviate the quality problem of dataset. With the development of deep learning technology, automatic atom-mapping algorithms are becoming more and more accurate and reliable in recent years. For example, Schwaller et al. proposed a method to extract hidden atom mapping information from an unsupervised model trained on unlabeled data [43], which can provide more high-quality atom-mapping reaction data, thereby improving the prediction quality of the retrosynthesis model.

In addition, although this work merges the predictions corresponding to multiple synthons, which effectively improves the diversity of model predictions, the fusion approach could be achieved in a learnable way. This is also one of the directions for our future work to be improved.

4. Conclusion

In summary, we propose a new Transformer-based method, RetroPrime, to tackle retrosynthesis. In the standard USPTO-50 K dataset, when the reaction type is known and unknown, RetroPrime's Top-1 accuracy reached 64.8% and 51.4%, respectively. In the large dataset USPTO-full, RetroPrime achieved the Top-1 accuracy of 44.1%, which is significantly higher than the template-based method, and is close to the state-of-the-art Transformer-based method AT. These encouraging results seems to concur with an earlier observation [30] that Transformer-based predictions possess excellent generalizability and robustness.

However, it is easy to show that Transformer suffers from two severe deficiencies: (1) lack of reaction diversity and (2) high percentage of chemically implausible solutions. Without further improvements on these two issues, one cannot trust Transformer's outputs beyond the first few ones. In this work, we make conscious efforts to deal with these challenges by proposing the "mix and match" and the "label and align" strategies as part of RetroPrime's two-stage workflow, inspired by a chemist's approach to retrosynthesis. The results show that our "mix and match" strategy can significantly improve the diversity of the model, and the "label and align" strategy could also reduce the proportion of chemically implausible prediction results. While improvements are substantial as reported, further innovations are urgently desired.

Given vast amount of chemical reaction data and new knowledge are generated on a daily basis, the benefits of building a reliable template-free method are obvious. Hopefully, without having to be explicitly trained on all reaction templates, these modern machine-learning methods can generalize more easily and guide us toward better synthetic routes.

CRediT authorship contribution statement

Xiaorui Wang: Writing - original draft, Software, Methodology, Conceptualization, Data Curation, Validation, Visualization, Investigation. **Yuquan Li:** Formal analysis, Writing - review & editing, Methodology, Visualization. **Jiezhong Qiu:** Validation. **Guangyong Chen:** Investigation. **Huanxiang Liu:** Conceptualization, Supervision. **Benben Liao:** Formal analysis, Conceptualization, Supervision, Project administration. **Chang-Yu Hsieh:** Project administration, Supervision, Conceptualization. **Xiaojun Yao:** Project administration, Supervision, Conceptualization, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This material is based upon work supported by the National Natural Science Foundation of China (Grant No. 21775060).

Data Availability.

The source code and datasets are available in GitHub: https://github.com/wangxr0526/RetroPrime.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cej.2021.129845.

References

[1] E.J. Corey, The logic of chemical synthesis: Multistep synthesis of complex carbogenic molecules (nobel lecture), Angew. Chemie Int. Ed. English. 30 (5) (1991) 455–465.
[2] M.H. Todd, Computer-aided organic synthesis, Chem. Soc. Rev. 34 (2005) 247–266, https://doi.org/10.1039/b104620a.
[3] A. Cook, A.P. Johnson, J. Law, M. Mirzazadeh, O. Ravitz, A. Simon, Computer-aided synthesis design: 40 years on, Wiley Interdiscip. Rev. Comput. Mol. Sci. 2 (1) (2012) 79–107.
[4] C.W. Coley, W.H. Green, K.F. Jensen, Machine learning in computer-aided synthesis planning, Acc. Chem. Res. 51 (5) (2018) 1281–1289, https://doi.org/10.1021/acs.accounts.8b00087.
[5] W.A. Warr, A short review of chemical reaction database systems, computer-aided synthesis design, reaction prediction and synthetic feasibility, Mol. Inform. 33 (6-7) (2014) 469–476, https://doi.org/10.1002/minf.201400052.
[6] T.J. Struble, J.C. Alvarez, S.P. Brown, M. Chytil, J. Cisar, R.L. DesJarlais, O. Engkvist, S.A. Frank, D.R. Greve, D.J. Griffin, X. Hou, J.W. Johannes, C. Kreatsoulas, B. Lahue, M. Mathea, G. Mogk, C.A. Nicolaou, A.D. Palmer, D. J. Price, R.I. Robinson, S. Salentin, L.i. Xing, T. Jaakkola, W.H. Green, R. Barzilay, C.W. Coley, K.F. Jensen, Current and future roles of artificial intelligence in medicinal chemistry synthesis, J. Med. Chem. 63 (16) (2020) 8667–8682.
[7] W.-D. Ihlenfeldt, J. Gasteiger, Computer-assisted planning of organic syntheses: The second generation of programs, Angew. Chemie (International Ed. English) 34 (2324) (1996) 2613–2633.
[8] O. Engkvist, P.-O. Norrby, N. Selmi, Y.-H. Lam, Z. Peng, E.C. Sherer, W. Amberg, T. Erhard, L.A. Smyth, Computational prediction of chemical reactions: Current status and outlook, Drug Discov. Today. 23 (6) (2018) 1203–1218, https://doi.org/10.1016/j.drudis.2018.02.014.
[9] F. Feng, L. Lai, J. Pei, Computational chemical synthesis analysis and pathway design, Front. Chem. 6 (2018) 199, https://doi.org/10.3389/fchem.2018.00199.
[10] S.V. Ley, D.E. Fitzpatrick, R.J. Ingham, R.M. Myers, Organic synthesis: March of the machines, Angew. Chemie - Int. Ed. 54 (11) (2015) 3449–3464, https://doi.org/10.1002/anie.201410744.
[11] D. Caramelli, J. Granda, D. Cambié, H. Mehr, A. Henson, L. Cronin, An Artificial Intelligence that Discovers Unpredictable Chemical Reactions, (2020). https://doi.org/10.26434/chemrxiv.12924968.v1.
[12] Florian Häse, Loïc M. Roch, Alán Aspuru-Guzik, Next-generation experimentation with self-driving laboratories, Trends Chem. 1 (3) (2019) 282–291, https://doi.org/10.1016/j.trechm.2019.02.007.
[13] Vishnu H Nair, Philippe Schwaller, Teodoro Laino, Data-driven chemical reaction prediction and retrosynthesis, Chimia (Aarau). 73 (12) (2019) 997–1000, https://doi.org/10.2533/chimia.2019.997.
[14] H. Dai, C. Li, C.W. Coley, B. Dai, L. Song, Retrosynthesis prediction with conditional graph logic network, ArXiv. (2020) 1–15. http://arxiv.org/abs/2001.01408.
[15] Connor W. Coley, Luke Rogers, William H. Green, Klavs F. Jensen, Computer-assisted retrosynthesis based on molecular similarity, ACS Cent. Sci. 3 (12) (2017) 1237–1245, https://doi.org/10.1021/acscentsci.7b00355.
[16] Marwin H.S. Segler, Mark P. Waller, Neural-symbolic machine learning for retrosynthesis and reaction prediction, Chem. - A Eur. J. 23 (25) (2017) 5966–5971, https://doi.org/10.1002/chem.201605499.
[17] Kangjie Lin, Youjun Xu, Jianfeng Pei, Luhua Lai, Automatic retrosynthetic route planning using template-free models, Chem. Sci. 11 (12) (2020) 3355–3364, https://doi.org/10.1039/C9SC03666K.
[18] Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender, Vijay Pande, Retrosynthetic reaction prediction using neural sequence-to-sequence models, ACS Cent. Sci. 3 (10) (2017) 1103–1113, https://doi.org/10.1021/acscentsci.7b00303.

[19] C. Shi, M. Xu, H. Guo, M. Zhang, J. Tang, A graph to graphs framework for retrosynthesis prediction, ArXiv Prepr. ArXiv2003.12725 (2020). http://arxiv.org/abs/2003.12725.
[20] J. Nam, J. Kim, Linking the Neural Machine Translation and the Prediction of Organic Chemistry Reactions, ArXiv Prepr. ArXiv1612.09529. (2016). http://arxiv.org/abs/1612.09529.
[21] V.R. Somnath, C. Bunne, C.W. Coley, A. Krause, R. Barzilay, Learning Graph Models for Template-Free Retrosynthesis, ArXiv Prepr. ArXiv2006.07038 (2020). http://arxiv.org/abs/2006.07038.
[22] C. Yan, Q. Ding, P. Zhao, S. Zheng, J. Yang, Y. Yu, J. Huang, RetroXpert: Decompose retrosynthesis prediction like a chemist, ArXiv. (2020). https://doi.org/10.26434/chemrxiv.11869692.
[23] Connor W. Coley, William H. Green, Klavs F. Jensen, RDChiral: An RDKit wrapper for handling stereochemistry in retrosynthetic template extraction and application, J. Chem. Inf. Model. 59 (6) (2019) 2529–2537.
[24] James Law, Zsolt Zsoldos, Aniko Simon, Darryl Reid, Yang Liu, Sing Yoong Khew, A. Peter Johnson, Sarah Major, Robert A. Wade, Howard Y. Ando, Route designer: A retrosynthetic analysis tool utilizing automated retrosynthetic rule generation, J. Chem. Inf. Model. 49 (3) (2009) 593–602, https://doi.org/10.1021/ci800228y.
[25] B. Chen, T. Shen, T.S. Jaakkola, R. Barzilay, Learning to make generalizable and diverse predictions for retrosynthesis, ArXiv Prepr. ArXiv1910.09688 (2019). http://arxiv.org/abs/1910.09688.
[26] David Weininger, SMILES, a chemical language and information system: 1: Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1) (1988) 31–36, https://doi.org/10.1021/ci00057a005.
[27] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[28] Shuangjia Zheng, Jiahua Rao, Zhongyue Zhang, Jun Xu, Yuedong Yang, Predicting retrosynthetic reactions using self-corrected transformer neural networks, J. Chem. Inf. Model. 60 (1) (2020) 47–55.
[29] P. Karpov, G. Godin, I. V. Tetko, A Transformer Model for Retrosynthesis, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer, 2019: pp. 817–830. https://doi.org/10.1007/978-3-030-30493-5_78.
[30] Alpha A. Lee, Qingyi Yang, Vishnu Sresht, Peter Bolgar, Xinjun Hou, Jacquelyn L. Klug-McLeod, Christopher R. Butler, Molecular transformer unifies reaction prediction and retrosynthesis across pharma chemical space, Chem. Commun. 55 (81) (2019) 12152–12155, https://doi.org/10.1039/C9CC05122H.
[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Adv. Neural Inf. Process. Syst., 2017: pp. 5999–6009.
[32] Philippe Schwaller, Teodoro Laino, Théophile Gaudin, Peter Bolgar, Christopher A. Hunter, Costas Bekas, Alpha A. Lee, Molecular transformer: A model for uncertainty-calibrated chemical reaction prediction, ACS Cent. Sci. 5 (9) (2019) 1572–1583, https://doi.org/10.1021/acscentsci.9b00576.
[33] Philippe Schwaller, Riccardo Petraglia, Valerio Zullo, Vishnu H. Nair, Rico Andreas Haeuselmann, Riccardo Pisoni, Costas Bekas, Anna Iuliano, Teodoro Laino, Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy, Chem. Sci. 11 (12) (2020) 3316–3325.
[34] Nadine Schneider, Nikolaus Stiefl, Gregory A. Landrum, What's what: The (nearly) definitive guide to reaction role assignment, J. Chem. Inf. Model. 56 (12) (2016) 2336–2346.
[35] D. Lowe, Chemical reactions from US patents (1976-Sep2016), URL Https//Figshare. Com/Articles/Chemical_ React. (2017). https://doi.org/10.6084/m9.figshare.5104873.v1.
[36] P. Schwaller, T. Gaudin, D. Lanyi, C. Bekas, T. Laino, "Found in translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models, Chem. Sci. 9 (2018) 6091–6098.
[37] G. Klein, Y. Kim, Y. Deng, J. Senellart, A.M. Rush, OpenNMT: Open-source toolkit for neural machine translation, ACL 2017–55th Annu, Meet. Assoc. Comput. Linguist. Proc. Syst. Demonstr. (2017) 67–72. https://doi.org/10.18653/v1/P17-4012.
[38] G. Landrum, RDKit: Open-source cheminformatics, (2006).
[39] E.J. Bjerrum, SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules, ArXiv Prepr. ArXiv1703.07076. (2017). http://arxiv.org/abs/1703.07076.
[40] Marwin H.S. Segler, Mike Preuss, Mark P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI, Nature. 555 (7698) (2018) 604–610, https://doi.org/10.1038/nature25978.
[41] S. Genheden, A. Thakkar, V. Chadimová, J.L. Reymond, O. Engkvist, E. Bjerrum, AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning, J. Cheminform. 12 (2020) 70, https://doi.org/10.1186/s13321-020-00472-1.
[42] I.V. Tetko, P. Karpov, R. Van Deursen, G. Godin, State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis, Nat. Commun. 11 (2020) 1–11.
[43] P. Schwaller, B. Hoover, J.L. Reymond, H. Strobelt, T. Laino, Unsupervised attention-guided atom-mapping, ChemRxiv. (2020). https://doi.org/10.26434/chemrxiv.12298559.v1.