JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

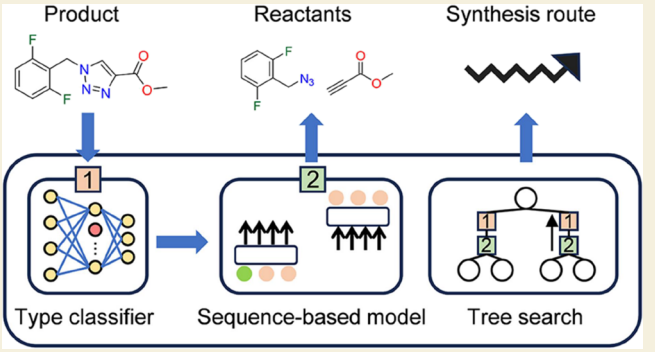
pubs.acs.org/jacsau
Article

SynCluster: Reaction Type Clustering and Recommendation Framework for Synthesis Planning

Tiantao Liu, Zheng Cao, Yuansheng Huang, Yue Wan, Jian Wu, Chang-Yu Hsieh,* Tingjun Hou,* and Yu Kang*

Cite This: JACS Au 2023, 3, 3446−3461
Read Online

ABSTRACT: AI-assisted synthesis planning has emerged as a valuable tool in accelerating synthetic chemistry for the discovery of new drugs and materials. The template-free approach, which showcases superior generalization capabilities, is seen as the mainstream direction in this field. However, it remains unclear whether such an end-to-end approach can achieve problem-solving performance on par with experienced chemists without fully revealing insights into the chemical mechanisms involved. Moreover, there is a lack of unified and chemically inspired frameworks for improving multitask reaction predictions in this area. In this study, we have addressed these challenges by investigating the impact of fine-grained reaction-type labels on multiple downstream tasks and propose a novel framework named SynCluster. This framework incorporates unsupervised clustering cues into the baseline models and identifies plausible chemical subspaces which is compatible with multitask extensions and can serve as model-independent indicators to effectively enhance the performance of multiple downstream tasks. In retrosynthesis prediction, SynCluster achieves significant improvements of 4.1 and 11.0% in top-1 and top-10 prediction accuracy, respectively, compared to the baseline Molecular Transformer, and achieves a notable enhancement of 13.9% in top-10 accuracy when combined with Retroformer. By incorporating simplified molecular-input line-entry system augmentation, our framework achieves higher top-10 accuracy compared to state-of-the-art sequence-based retrosynthesis models and improves over the baseline on the diversity and validity of reactants. SynCluster also achieves 94.9% top-10 accuracy in forward synthesis prediction and 51.5% top-10 Maxfrag accuracy in reagent prediction. Overall, SynCluster provides a fresh perspective with chemical interpretability and reinforcement of domain knowledge in the synthesis design. It offers a promising solution for improving the accuracy and efficiency of AI-assisted synthesis planning and bridges the gap between template-free approaches and the problem-solving abilities of experienced chemists.



KEYWORDS: *synthesis planning, fine-grained type, transformer, unsupervised clustering*

INTRODUCTION

Synthesis planning[1] is a critical process in drug discovery and chemical industry that involves fabricating reasonable pathways for synthesizing given compounds. It entails multiple reasoning tasks including forward synthesis prediction, reagent selection, and retrosynthesis. The first two tasks deduce the possible underlying product or reagents by providing building blocks or complete reactions; while retrosynthesis operates in reverse and requires logical disconnection of the desired starting molecules, searching a vast space of possible chemical transformations from a given state.[2] With the continuous expansion of accessible chemical space, the rapid growth in the number of synthetic molecules has made traditional manual treatment of this process inefficient.[3] As a result, computer-aided approaches are now in urgent demand to overcome the limitations of manual methods.

Computer-aided retrosynthetic analysis strategies can be divided into two main categories,[4] template-based and template-free models. The template-based model involves matching generalized reaction rules with target molecules to produce one or more candidate precursors based on defined subgraph patterns of the chemical reaction.[5] Most template-based models[6−9] leverage a classification model for selecting a suitable reaction template, which can be interpreted as a process of searching and matching within established domain knowledge, rather than generating new chemical knowledge

Received: October 7, 2023
Revised: November 7, 2023
Accepted: November 8, 2023
Published: November 17, 2023

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

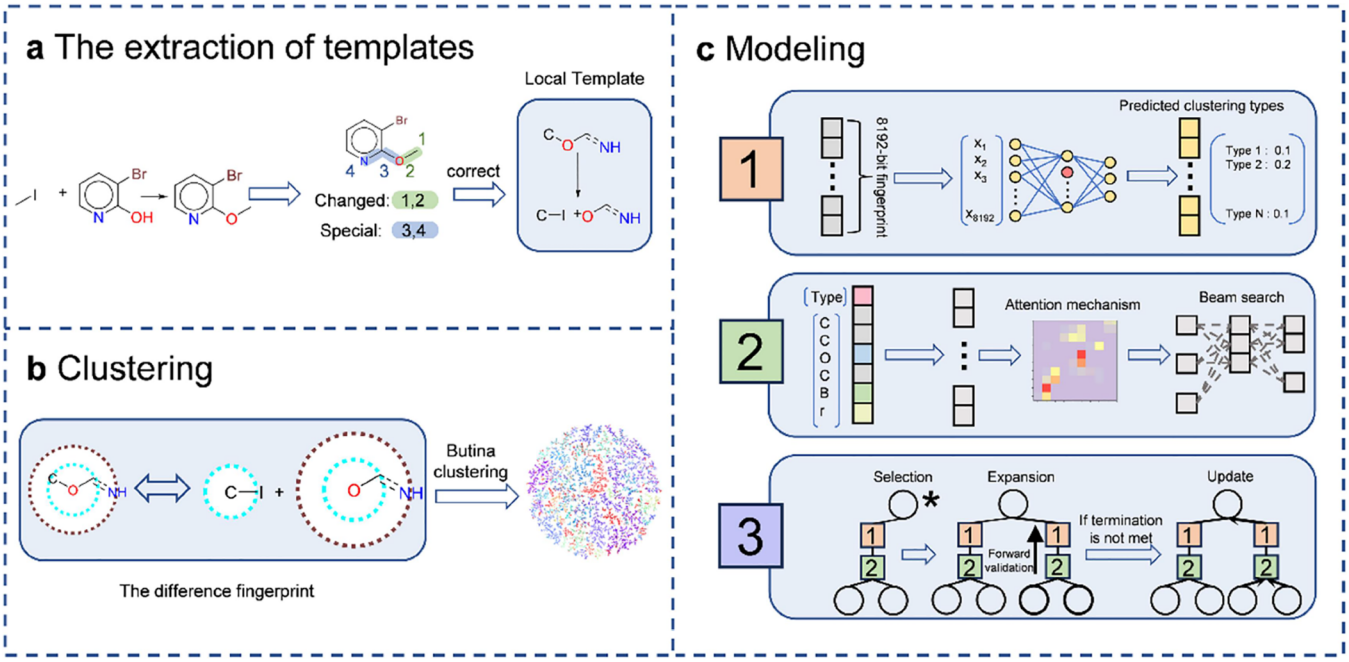


Figure 1. Workflow of our two-stage framework. Workflow consists of three parts. (a) Template extraction and correction process. In the standardization procedure, an automatic template extraction method[72] based on the radius and functional group is employed. The reaction center (Changed) and extended part (Special) of the compound are identified respectively, followed by round-trip verification and canonization.[39] (b) Reaction clustering is developed by substructure fingerprints in which the reaction is compressed as a local template. (c) Modeling procedure. In the phase of inference, types are predicted (not included in reagent prediction) by a feed-forward neural network (part 1 of Figure 1c) and then embedded and entered into a downstream model such as Transformer or Retroformer (part 2 of Figure 1c). In the phase of the multistep application, the SynCluster framework is combined with Retro* search[67] (part 3 of Figure 1c) to achieve the selection, expansion, and update.

thoroughly.[10] For example, Chen et al.[11] have employed the LocalRetro model, which uses a global attention mechanism upon message passing neural networks[12] (MPNN) for searching the suitable atom-based or bond-based reaction rules. Their approach represents the current state-of-the-art (SOTA) in the template-based field. The template-free model, in contrast, does not rely on predefined templates, but takes a single product as input and uses deep learning techniques, such as the Graph Neural Network[13−15] (GNN) or the Transformer,[16] to represent reactants as heterogeneous molecular graphs or simplified molecular-input line-entry system (SMILES) strings. While many template-free approaches rely on predefined leaving groups[13] or action space[15] to generate the precursors, sequence-based methods use SMILES strings to capture the molecule representation and introduce powerful language models like the Transformer.[17−19] The Molecular Transformer (MT),[20] with its multihead attention mechanism, has achieved impressive performance in the retrosynthesis task based on the top-*k* accuracy evaluation. MT is even generalizable to reactant-reagent mixed data, surpassing human chemists in a benchmark test.[17] Moreover, improved models such as Retroformer[21] have achieved remarkable accuracy for end-to-end template-free retrosynthesis based on the fundamental algorithm design of the Transformer.

Despite the efficiency and cost reduction offered by the present retrosynthesis prediction algorithms compared to the manual approach,[8] they continue to face several challenges. The rule-based approach in template-based models requires significant domain expertise and limits the generalizability of novel reactions.[22] Although the template-free method has superior generalization capability,[13] it remains unclear whether AI-assisted synthesis can achieve problem-solving performance on par with experienced chemists.[23] Moreover, most template-free models report high benchmark performance without fully revealing insights into the transition process of chemical groups and other organic mechanisms, limiting the interpretability of these models.[8] To address this, one approach is to split the retrosynthesis task into two steps: synthon prediction and precursor prediction,[13,24] where the model identifies the reaction centers and generates the minimum edit distance in SMILES accordingly. However, this approach may result in additional computational overhead due to the specified requirements for the reaction center.[23] Additionally, there is a lack of a unified framework for adapting multitask prediction in synthesis design, which limits the model's fine-tuning possibilities for other tasks like yield prediction and reagent prediction.[11,25]

In this work, we propose a novel two-stage framework, named SynCluster, as a more integrated and efficient solution for multiple synthesis design tasks, including retrosynthesis, forward synthesis, and reagent prediction. This framework is centered around the core concept of integrating data-driven reaction-type predictions for target molecules into a comprehensive AI-assisted synthesis planning workflow (Figure 1). Unlike conventional template-based methods, SynCluster for downstream task inference does not rely on pre-existing template libraries. Unlike purely template-free methods, our framework enforces output regulation via type clustering, which imparts enhanced interpretability to the obtained results. Compared with various baseline models, SynCluster exhibits significant improvements in top-*k* accuracy and generation diversity. It introduces several novel features: (1) Fine-grained categorization of reactions through the clustering of condensed reaction templates. This approach

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

enables a more precise and visualizable classification of reactions compared to the widely used NameRxn superclasses.[26] (2) Interactive application for retrosynthesis, forward synthesis, and reagent prediction, providing novel and robust multiscale expansion capabilities. (3) Enhanced reaction diversity and validity offered by SynCluster, which brings about more abundant chemical insight and addresses realistic chemical needs more effectively. (4) Excellent compatibility with various single-step models, which leads to significant improvements in prediction accuracy across a broad range of applications. (5) Great adaptability with tree-based searching in multistep retrosynthesis prediction, enabling the generation of shorter paths compared to SOTA models. (6) Minimal additional computational burden.

MATERIALS AND METHODS

Data Set and Baseline

In this study, the benchmark data set for retrosynthesis and forward synthesis prediction comprising the reaction SMILES was derived from Lowe.[27] It was further cleaned into the subset USPTO-50K[28] (about 50,000 reactions) for our experiments. The training, validation, and testing split for our models was 80, 10, and 10%, respectively. Through our experimental procedure, we further cleaned the data set by discarding the reactions featuring invalid template representations, e.g., the reactions with an error in chirality. In detail, we cleaned 1439, 189, and 176 reactions for training, validation, and testing, respectively (Table S1). Additionally, to push the limits of the reaction reasoning ability of large-scale verification, USPTO-MIT[29] (about 460k reaction) was investigated in the clustering phase. The benchmark data set for the reagents prediction was provided by Lu et al.[25] 15902 reactions were extracted from the USPTO_500_MT data set, and the number of training and testing reactions was set to 14903 and 999, respectively.

In this research, our model was evaluated for multiple downstream tasks. In the retrosynthesis prediction, LocalRetro,[11] SCROP,[30] Graph2SMILES,[31] Graphretro,[14] MEGAN,[15] and AT[32] were chosen and trained from scratch. Considering that our model adds fine-grained reaction types to MT,[17] it was set as our baseline in the retrosynthesis, forward synthesis, and reagent prediction. Besides, we have investigated the performance of our additional framework when combined with Retroformer,[21] a new SOTA model for the end-to-end template-free retrosynthesis. Thus, the baseline models of retrosynthesis also include Retroformer. More details are provided in Supporting Information, Part 1.

Overall Workflow of SynCluster

The workflow consists of three parts: template extraction/correction, reaction clustering, and modeling. First, we automatically extract and correct the templates (Figure 1a). Then, we clustered the accessible chemical space into groups of similar reactions based on the distance between their template difference fingerprints (Figure 1b). During modeling (Figure 1c), we train a downstream model (e.g., MT or Retroformer) through inputting the SMILES of products (retrosynthesis), reactants (forward synthesis), or reactions (reagent prediction) with corresponding clustering types. We then recommend the reaction type based on the clustering using a trained feed-forward neural network (FNN). Specifically, the inference pipeline involves the following two steps: (1) Recommending (forward and retrosynthesis) or retrieving (reagent prediction) the candidate reaction clusters (i.e., types) based on the input of the task; (2) Accomplishing the target task by given downstream models conditioned on the chosen types and the previous input. Finally, for multistep prediction of retrosynthesis, we have utilized a tree-based search to get the complete routes.

Reaction Representation and Clustering

Several techniques have been proposed to gather similar reactions together.[33−35] The commercial software NameRxn[26] has preset 11 superclasses, 69 classes/categories, and more than 300 named types to match reactions. However, these relative and predetermined reaction sets are invariably too specific to cover the accessible chemical space for the hand-crafted templates.[36] It has been reported[37] that NameRxn could only cover 54% of the extracted reactions in the NextMove's data set.[38] To tag this problem, Schneider et al.[37] have revealed a data-driven chemical fingerprint to cluster the reactions. Nevertheless, there remains space for trade-offs in which the reaction representation at different levels (i.e., using the complete reaction or specifying different sizes of reaction) contains multilevel information.[39] In our study, Schneider's method was fine-tuned in the scenario that the reaction was compressed as the local template.

In detail, the templatecorr toolkit[39] was utilized to cut down the number of extraction templates. With the advent of automated template extraction and canonization software, reactions featuring an invalid template representation were discarded. Besides, the template difference fingerprint was embraced to map the template into a vector, as expressed in eq 1.

$$\text{Template Fp} = \left(\sum_{\text{Pro}} \text{ProTem FP} - \sum_{\text{reac}} \text{ReacTem FP} \right) \tag{1}$$

On this basis, the diverse types of chemical fingerprints (Atompairs,[40] ECFP,[47] FCFP,[41] and TopologicalTorsions[42]) were investigated. The template fingerprints were clustered by Butina's algorithm[43] integrated with RDkit to evaluate the similarity of the fingerprints. In this phase, the first step is the generation of standard Daylight's fingerprints (ASCII), followed by the identification of potential cluster centroids and clustering based on the exclusion spheres by similarity. The similarity was calculated with the Tanimoto coefficient:[44]

$$\text{Tanimoto} = \frac{\sum x_i y_i}{\sum x_i^2 + \sum y_i^2 - \sum x_i y_i} \tag{2}$$

where x_i, y_i represent the subpattern of the object of comparison. Furthermore, the influence of the cutoff of Butina's clustering was included in the analysis, and the purities metric was calculated as follows:

$$\text{Purity}\ (\Omega, C) = \frac{1}{N} \sum_{\text{k}} \max|\omega_{\text{k}} \cap c_j| \tag{3}$$

where Ω, C represent the set of the NameRxn superclasses and the clustering results, respectively. Furthermore, ω_{k} and c_j refer to the set of detailed superclasses and clustering types, respectively.

For the sake of evaluating the overlap ratio of the 10 superclasses given by NameRxn and our clustering types to further interpret the clustering, TMAP tree-based[45] and t-distributed stochastic neighbor embedding (t-SNE) reduction[46] are investigated. The details for the reduction are provided in Supporting Information, Part 2.

SMILES Augmentation

Typical data enhancement methods in natural language processing include back translation, synonym substitution, random insertion, random deletion, and random exchange.[47] However, these methods are not compatible with our model because minor character changes may induce semantic errors regarding SMILES. An alternative choice is to introduce the SMILES Augmentation.[48] Since SMILES is not unique (i.e., multiple SMILES can represent the same molecule), we have augmented the data set by creating new SMILES strings with different starting points and traversal orders. In the setting of augmentation in MT, we have adopted the 20-fold augmentation to expand the data set ahead with randomly permuted SMILES. The standard procedure of preprocessing is illustrated in Figures S1 and S2. Otherwise, in the setting of Retroformer, we obeyed the original on-the-fly augmentation in which a probability of 50% to permute the product SMILES and the reactants ordering every epoch is adopted.

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

Model Construction

Our model construction consists of three components: (1) the feed-forward neural network and (2) the downstream model, including Molecular Transformer (MT) or Retroformer. In the inference of retrosynthesis or forward synthesis, the assignment of clustering type was achieved through the utilization of an FNN. Conversely, in the reagent prediction model, the type was directly retrieved for the information about the complete reactions of the test set was given. (3) A multistep application for retrosynthesis using a tree-based search.

In the first step, the dimension of an input vector was set as 8192. A single layer with 2048 nodes was taken as the hidden. The output layer nodes were set as the number of clustering types (77 dimensions). Furthermore, an Adam optimizer and dropout technique were utilized in the training phase. The cross-entropy loss and log-softmax activation were applied to the output layer. Furthermore, we have checked the chemical interpretability of our type of predictor with the integrated gradient. The standard calculation procedure is as follows:

(1) Masking out one nonzero dimension of the fingerprint and retrieving the predicted probability $P(s)$ for the predicting clustering type.
(2) Using the original fingerprint and retrieving the predicted probability $\hat{P}(s)$ for the predicting clustering type.
(3) The integrated gradient is computed as $\hat{P}(s) - P(s)$.

Based on this, the top-k positive/negative bits that own the top-k biggest/smallest $\hat{P}(s) - P(s)$ were chosen. For the sake of making a quantitative inspection of those bits, we have computed the coverage for active/negative fingerprint bits by the reaction center. In detail, the reaction center is defined by template's radius of 3 (merely on the side of the product).

To compare the results of the first step where the imbalanced test set is attached, we have calculated the confusion entropy of the CEN[49] and then computed the Kappa coefficient and overall MCC score[50] as follows:

$$\text{kappa} = \frac{p_0 - p_e}{1 - p_e} \tag{4}$$

$$\text{Overall\_MCC} = \frac{\text{cov}(X, Y)}{\sqrt{\text{cov}(X, X) \times \text{cov}(Y, Y)}} \tag{5}$$

where p_0 represents the overall accuracy, p_e is the weighted accuracy by imbalanced classes, cov is the covariance function, and X, Y represent the two matrices where $X_{sn} = 1$ if the sample s is predicted to be of class n otherwise 0 and $Y_{sn} = 1$ if the sample s belongs class n otherwise 0.

The first choice of downstream model in this work was performed with the versatile Transformer,[16] which shares a consistent structure with MT model.[17] The details of the model are exhibited in Supporting Information Part 3 and keep consistency in retrosynthesis, forward synthesis, and reagents' prediction.

The second choice of downstream model in this work was performed with the Retroformer.[21] It has proposed a special form of local heads in the multihead mechanism to support efficient information exchange between the local region of reactive importance and the global reaction context. The details of the model are provided in Supporting Information, Part 3.3.

The final part is a multistep application for retrosynthesis. We endowed the A*-based algorithm model Retro* to our single-step model. The ideal setting and details are presented in Supporting Information, Part 4.

Output Ranking and Evaluation

During the process of generating the predictions, we adopted a beam search and a hybrid scoring function when SMILES is augmented to rank the output. The details are stated in Supporting Information, Part 5 and Figure S1.

Consequently, a crucial phase is to make a reasoning evaluation of the output. Under the hypothesis that the ground truth is always the only choice,[28,51] most research on evaluation index selection has examined the degree of matching between the predicted results and the ground truth. This metric, which is called the top-k accuracy, fails to consider the diversity of chemical pathways. Although there are some efforts to assess the diversification of results, the ease of these new metrics is hampered by the request of the pretrained model.[52] Hence, the Similarity distance metric was introduced to evaluate retrosynthesis models thoroughly:

$$\text{dis}_\alpha = \text{Mean}\left(\frac{\sum_{o,p\in c_i} \text{dist}(o, p)}{|c_i| - 1}\right) \tag{6}$$

$$\text{dist}(o, p) = \begin{cases} 1 - \text{sim}(o, p), & \text{if } o, p \text{ exist} \\ 0, & \text{else} \end{cases} \tag{7}$$

where o, p represents the different reaction fingerprints for top-k prediction, and c_i denotes the top-10 outputs. Concerning diversity evaluation, the prediction inputting is set as a pair to calculate the reaction fingerprint, and the similarity is obtained using eq 2.

For the sake of a comprehensive assessment, the traditional top-k accuracy was reported in both retrosynthesis and forward synthesis. The Maxfrag coverage[32] (defined as the top-k accuracy between the maximum fragment of the output reactants and the ground truth) and the Roundtrip accuracy[52] (quantifies what percentage of the retrosynthetic suggestions is valid through forward model) were induced in retrosynthesis prediction. Finally, the Maxfrag coverage was also considered in reagent prediction to encourage the exploration of major solvents and catalyzers.

RESULTS AND DISCUSSION

Reaction Clustering

The number of the clustering types and their purities are analyzed with different radii of reaction templates and cutoffs in Butina's clustering algorithm (Tables S2−S4) to evaluate the quality of reaction clusters on the USPTO-50K data set. A higher purity indicates better interpretability and a stronger connection to human knowledge. The positive correlation between the number of clustering types and the purity suggests that fine-grained clustering leads to highly dispersed data. When selecting clustering parameters, it is important to consider the tradeoff between the cutoff in Butina's clustering and the number of clusters. A moderate cutoff of 0.6 is used for further study, which provides an appropriate number of clusters and relatively high purity. In addition, it is worth noting the difference between ECFP and FCFP. With a template radius of 1, which represents a moderate central atom expansion capturing the essential information from the entire reaction, FCFP showed a generally higher or equivalent purity compared to ECFP. This can be attributed to FCFP's ability to acquire more abstract role-based substructural features, such as the consistent identification of halogens.[41] Therefore, we utilized FCFP and a cutoff of 0.6 in Butina's clustering to achieve powerful interpretability and predictability. The average purity of all the clusters is 0.784 (Table S3), which suggests a significant correspondence between the 10 NameRxn[26] superclasses (Table 1) and our clustering types, which refer to 77 fine-grained types clustered by SynCluster.

We have then evaluated the overlap ratio between the 10 NameRxn superclasses and our clustering types in order to interpret the clustering in the USPTO-50K data set. Figure 2a,b shows the trees generated using the TMAP package,[45] a knock-down toolkit for compressing high-dimensional data sets into two-dimensional connection trees. Trees shown in Figure 2a,b are colored by the superclasses and our clustering

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

Table 1. Distribution of USPTO-50K

NameRxn's superclass	reaction type	percent of USPTO-50K (%)
1	heteroatom alkylation and arylation	30.3
2	acylation and related process	23.8
3	C−C bond formation	11.3
4	heterocycle formation	1.8
5	protections	1.3
6	deprotections	16.5
7	reductions	9.2
8	oxidations	1.6
9	functional group interconversion (FGI)	3.7
10	functional group addition (FGA)	0.5

types, respectively, and adjacent tree nodes can be considered as the same type of reaction. This indicates that the main characteristics of our clustering types are consistent with the chemical domain knowledge, and the assignment of our clustering types is more fine-grained compared with the broader superclasses given by NameRxn (Table 1). Taking the superclass 6 as an example, clustering types 1, 2, and 10 exclusively spans the subtree of this superclass.

We then sorted the clustering types by their size (i.e., the amount of reaction they cover). In this sorting, larger clustering type numbers correspond to smaller cluster sizes. This relationship is illustrated in Figure 2c−e, where the purity is negatively correlated with the cluster size, as expected. It is necessary to evaluate the clusters with low purity and explore the contribution of different superclasses to these clusters. For the purpose of illustration, we have selected clustering type 67, which consists of 47% superclass 1, 40% superclass 3, and 3% superclass 2. Figure 2f displays a typical reaction from this clustering type, where S1, S2, and S3 represent the superclasses 1−3 given by NameRxn, namely, heteroatom alkylation and arylation, acylation and related process, and C−C bond formation, respectively. Interestingly, the nucleophilic reaction can be captured even if the reactions come from different superclasses. In superclasses 1 and 2, the nucleophilic substitution by secondary amine is reported, while the Grignard compound is reported as the nucleophile in superclass 3. More generally, the carbonyl group is identified as a vital center in the aforementioned examples. This indicates that our clustering types have the potential to detect and rectify certain inconsistencies in classifications that arise from the limitations of the expert-defined NameRxn superclasses.

In addition, we provide experimental evidence for the reasonable chemical interpretation of our clustering through t-SNE reduction.[46] We have performed the t-SNE analysis on the 2048-bit FCFP fingerprints of the reaction templates in the USPTO-50K cleaned data set. Figure 3a−f display the clusters of large, medium, and small sizes, respectively, in line with the cluster classification shown in Figure 2. The reactions are color-coded to represent our clustering types (Figure 3a,c,e) and the NameRxn superclasses (Figure 3b,d,f) for comparison. Figure 3a−d demonstrates that our clustering types provide a more distinct low-dimensional distribution for clusters of large or medium size compared with the NameRxn superclasses. Figure 3g−l presents several noteworthy case studies to provide evidence for the benefits of this distinct distribution: (1) the reactions shown in Figure 3g−j are all grouped into superclass 3 (i.e., "C−C bond formation") according to the purpose of the reaction. In contrast, our clustering system has divided these reactions into separate clusters, which is in line with the organic reaction mechanism as well as the transformation of reaction sites. (2) The reactions shown in Figure 3k,l are both about the acylation reaction between amine and acid, but they are confusingly grouped into different NameRxn superclasses (superclass 9 and superclass 2). It is corrected by our clustering, as they are both grouped in clustering type 4, which is similar to the refactoring of superclass 1 in Figure 2f. Remarkably, despite the fact that the SynCluster classification system is entirely derived in a data-driven manner, there is a significant global-level consistency between it and NameRxn, owing to the high purity of our clustering types compared to the NameRxn superclasses (as shown in Tables S2−S4). It is evident that SynCluster is capable of rectifying some inconsistencies in the classification made by NameRxn.

We have extended our analysis to the larger USPTO-MIT data set (Figure S3) by applying TMAP-tree and t-SNE reduction. This allows us to visualize whether the model can still distinguish the reactions within a highly diversified and large-scale database of chemical reactions. The same parameter settings are used as those on the USPTO-50K data set, and 433 fine-grained clustering types are obtained. As shown in Figure S3a−c, t-SNE visualization indicates that our clustering effectively distinguishes the reactions in general, especially for the cluster with the larger size. This distinction is further supported by the TMAP tree (Figure S3d,e). It is also worth noting that several subtrees in the TMAP (Figure S3f,g) closely group unique clustering types, suggesting the potential for our clustering strategy to be applied to large-scale applications.

Clustering Type Classifier

The performance of the clustering type classification should be taken into account prior to the exploration of the effect of the fine-grained reaction types on the downstream synthesis tasks. We have benchmarked several clustering type classifiers and compared their performance, including the FNN with ECFP, XGBoost model with the same fingerprint,[53] the basic Message Passing Neural Network[12] (MPNN) model including or excluding global attention,[11] MPNN model with Attentive FP passing meachanism[54] or Weave predictor,[55] and path-augmented graph Transformer predictor.[56] For all graph-based models, we have utilized the default hyperparameters recommended by DGL-Life.[57] For FNN and XGBoost, we have performed a grid search to meticulously ascertain the hyperparameters, which can be found in the Supporting Information Part 6.1. As shown in Table 2, for our reaction type prediction, the FNN achieves the best statistical performance with the average top-1 accuracy of 0.654 and average macro F1 of 0.518. Besides, the best statistical performance with an overall MCC of 0.605 and kappa score of 0.606 states the best ability of the FNN model to deal with imbalanced data among investigated models. XGBoost performs slightly worse than the FNN, reaching the top-1 accuracy of 0.640 and an overall MCC of 0.602. On the contrary, MPNN and Graph Transformer models fail to show satisfying results, where low accuracy and low metrics of imbalance are perceived. In conclusion, we have found traditional fingerprints to yield superior performance compared to several graph-based models for the specific task of reaction

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

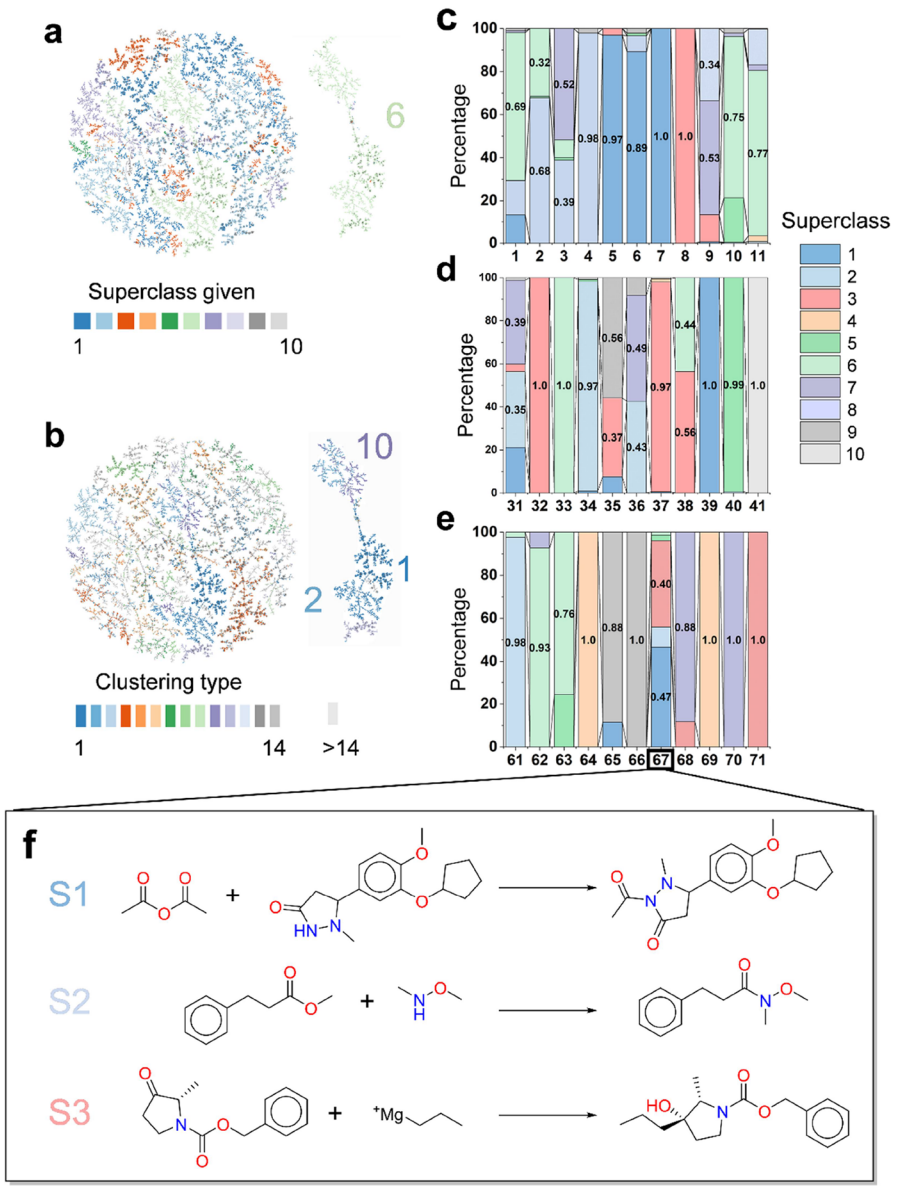


Figure 2. Analysis of clustering results. Distribution of the ground truth superclasses given by NameRxn (a) and the clustering types (b) are displayed in the TMAP tree. The cooperative region of superclass 6, corresponding to the clustering types 1, 2 and 10, is displayed individually for comparison. The superclass composition in various clustering types numbered by their size are presented in (c−e), i.e., the larger the clustering type number, the smaller the amount of reactions that the cluster covers. Clusters are classified into large (c, 1−11), medium (d, 31−41), and small (e, 61−71) sizes. (f) Typical reaction obtained from the clustering type 67. S1, S2, and S3 represent the superclass 1−3 given by NameRxn, i.e., heteroatom alkylation and arylation, acylation and related process, and C−C bond formation as listed in Table 1, respectively. All experiments are carried out using FCFP fingerprint and the cutoff of 0.6 in Butina's clustering.

type prediction in our work. The high expressivity of graph-based models is accompanied by the requirement for a large amount of data, which may hinder their performance in our task. The macro F1-score is unsatisfying in general, which may be caused by the poor prediction of small clusters (as shown in Figure S4a,b). Considering the highly unbalanced distribution of the original USPTO-50K data set (both in superclass and clustering types, Figure S4c), an unsatisfying macro F1-score is a comprehensible phenomenon[58] in multiclassification prediction. Nevertheless, we have chosen the FNN model and moved on to the next stage due to its highest top-*k* accuracy and metrics of imbalance.

To check the chemical interpretability of our type predictor, the integrated gradient based on the captum toolkits[59] has been leveraged to analyze the developed FNN model. A Grignard reaction (Figure 4a) is randomly chosen from clustering type 67 as an example for further analysis. Herein, the fingerprints dimension's largest five (positive impact on ground-truth type prediction) and smallest five attribution values (negative impact on ground-truth type prediction) are calculated and shown in Figure 4b, where the bit dimensions 2461, 4903, 3506, 3759, and 5250 are calculated as the five highest contributing bits for the ground truth clustering type, which corresponds to the Grignard reaction center. The FNN successfully captures this knowledge, which is highly consistent with human experts' cognition (as analyzed in Figure 2f). The inspection of the reaction fingerprint bits with a positive contribution explicates the complete reaction framework (Bit

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

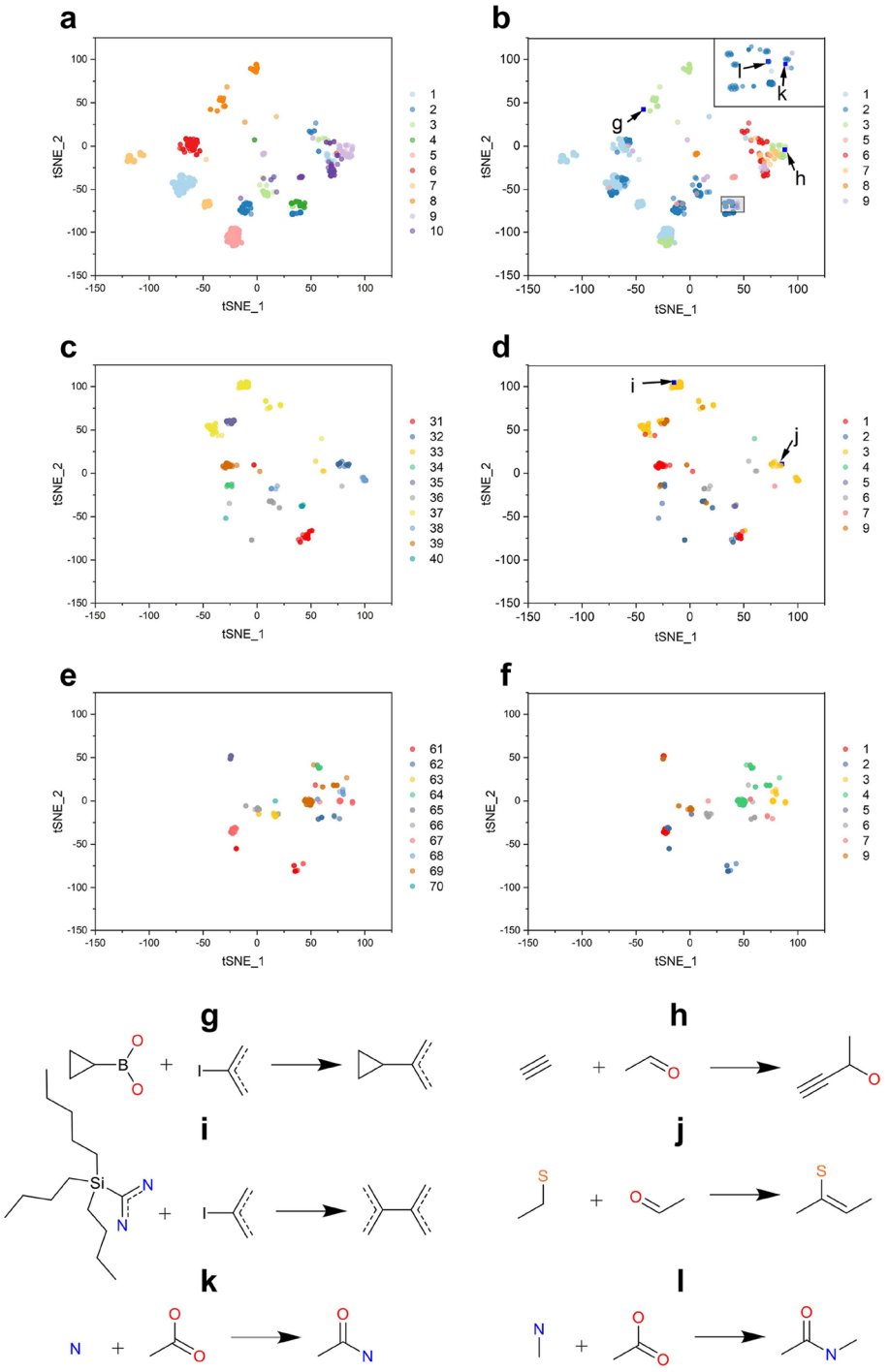


Figure 3. t-SNE visualization of 2048-bit FCFP fingerprints of the reaction templates on USPTO-50K cleaned data set, in the clusters with (a, b) large, (c, d) medium, and (e, f) small sizes. Reactions are colored by (a, c, e) our clustering types, or by (b, d, f) the NameRxn superclasses. Visualized reactions are listed as (g–l) for comparison: (g) Suzuki reaction, classified in superclass 3 (i.e., C−C bond formation) and clustering type 8; (h) reductive coupling reaction of Alkynes and Aldehydes,[73] classified in superclass 3 and clustering type 9; (i) Stille coupling reaction in superclass 3 and clustering type 37; (j) a Knoevenagel condensation reaction in superclass 3 and clustering type 32; (k) amide bond formation reaction in superclass 9 (i.e., FGI) and clustering type 4; (l) another amide bond formation reaction in superclass 2 (i.e., acylation and related process) and clustering type 4.

2461, Bit 3506, and Bit 3759 in Figure 4c) and the hydroxyl in the tertiary alcohol (Bit 4903 and Bit 5250). When it comes to the bits with a negative contribution (Figure 4d), structures similar to the carbamate are found (Bit 1033, Bit 7209, and Bit 5371). In a sense, these negative bits correspond to yet another active reaction center in which the combination of isocyanate and primary alcohol could be applied. To obtain an in-depth appreciation, we have randomly collected more instances in which the positive bits have a significant correlation to the reaction center (Figure S5).

The statistical measures related to the integrated gradients are axiomatic methods that are not affected by any bias. Thus,

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

Table 2. Performance Comparison in Type Prediction of Seven Models (Average in 3 Times Independent Runs)

method	top-1 acc	top-3 acc	top-5 acc	top-10 acc	Marco-F1	Kappa	overall_Mcc
FNN	0.654	0.871	0.928	0.970	0.518	0.605	0.606
Xgboost	0.640	0.863	0.924	0.967	0.523	0.602	0.603
MPNN	0.577	0.745	0.888	0.944	0.360	0.553	0.554
MPNN+Attention	0.605	0.765	0.900	0.952	0.419	0.583	0.584
weave	0.522	0.691	0.865	0.939	0.298	0.493	0.494
AttentionFP	0.488	0.644	0.820	0.914	0.215	0.458	0.460
GraphTransformer	0.553	0.730	0.898	0.961	0.390	0.530	0.530

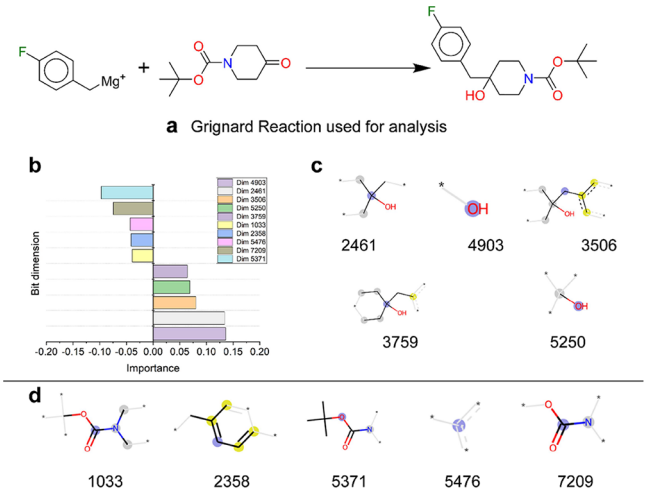


Figure 4. Visualization of FNN prediction results. (a) Grignard reaction randomly selected from the clustering type 67 for further analysis. (b) Largest 5 and smallest 5 attribution values of the fingerprints dimension. "Importance" refers to integrated gradient value. (c) Reaction fingerprint bits with a positive contribution to prediction probability. The center atom, aromatic atom, and aliphatic atom are colored blue, yellow, and gray, respectively. (d) Reaction fingerprint bits with a negative contribution to prediction probability.

the coverage for top-1 or top-3 positive/negative fingerprint bits by the reaction center is calculated (Tables S5 and S6). When the reaction center highly overlaps top positive bits and rarely overlaps top negative bits, the capability of our classifier to capture the local information on the veritable reactive center and overlook the interfering substructure could be verified. Thus, we also computed the difference between the coverage for active and negative fingerprint bits. The comparative experimental results in Tables S5 and S6 reveal significant coverage differences among most clustering models. Particularly, the selected parameter (FCFP fingerprint, Butina's cutoff of 0.6, and template's radius of 1) achieves the utmost difference of 0.427 in the top-1 assessment (Table S5). Furthermore, the coverage difference between top-k (k = 1, 3, 5 or 10) positive and negative bits is displayed in Figure S6 according to individual fine-grained clustering types to measure the bias of unbalanced data. It is apparent from this figure that irregular profiles are obtained, which may indicate that clear features are captured by our classifier even for highly generalized types due to the existence of a large number of different reactions (such as clustering types 1−10).

Taken together, these results suggest a high relevance between the learning of the model and the real chemical meaning. These findings provide strong support for the conceptual premise that chemical concepts can be understood by deep learning models.

Retrosynthesis Prediction

Built on top of the clustering type prediction, we applied our SynCluster framework to the task of retrosynthesis prediction.

Table 3. Top-k Retrosynthesis Accuracy and Maxfrag Accuracy on the USPTO-50K Cleaned Dataset

	top-1 acc		top-3 acc		top-5 acc		top-10 acc	
	acc	MF[a]	acc	MF	acc	MF	Acc	MF
Sequence-Based Model								
MT	0.444	0.491	0.591	0.643	0.630	0.680	0.675	0.728
Retroformer$_{base}$[b]	0.467		0.616		0.661		0.696	
Retroformer+SynCluster	0.477	0.525	0.679	0.718	0.751	0.788	0.835	0.867
Retroformer+SynCluster$_{aug}$[c]	0.507	0.553	0.704	0.746	0.776	0.813	**0.858**[d]	0.890
MT+SynCluster	0.485	0.531	0.657	0.708	0.722	0.772	0.785	0.834
MT+SynCluster$_{aug}$	0.536	0.576	0.708	0.753	0.782	**0.823**	0.853	**0.891**
MT+SynCluster$_{superclass}$[e]	0.476	0.529	0.641	0.698	0.698	0.746	0.726	0.785
AT[f]	**0.537**	**0.585**	**0.739**	**0.776**	**0.786**	0.820	0.825	0.860
Graph2SMILES[g]	0.522	0.575	0.668	0.717	0.702	0.751	0.728	0.782
SCROP	0.445	0.490	0.606	0.649	0.651	0.695	0.680	0.728
Graph-Based Model								
GraphRetro	**0.553**	**0.608**	0.703	0.737	0.743	0.770	0.775	0.798
LocalRetro	0.535	0.575	**0.765**	**0.797**	**0.851**	**0.878**	**0.921**	**0.989**
MEGAN	0.492	0.549	0.724	0.768	0.801	0.839	0.871	0.899

[a]"MF" refers to "Maxfrag accuracy". [b]Retroformer$_{base}$ refers to the nonaugmented edition. [c]The augmentation applied in Retroformer is dynamic permutation. [d]Best results of top-k and Maxfrag accuracy are bolded. [e]"Superclass" represents using NameRxn superclasses to replace clustering types in whole process. [f]In the scheme of reproduction, the parameter settings are consistent with the original text (×5 M training augmentation and 100× testing augmentation). [g]In Graph2SMILES, Molecular graph guides translation model generation.

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

We take advantage of top-*k* accuracy, Maxfrag accuracy,[32] the Tanimoto Similarity distance among the predictions, and the Roundtrip accuracy,[52] in order to thoroughly evaluate the reactants generation. To give an ablation study on the fine-grained reaction type, MT,[17] and Retroformer[21] are set as our baseline. Here, the SynCluster framework is added to the retrosynthesis task and denoted as "MT+SynCluster" or "Retroformer+SynCluster" for comparison. Besides, the following models are also taken into consideration: LocalRetro,[11] the SOTA template-based model; SCROP,[30] a SMILES correction vision of Molecular Transformer; Graphretro,[14] a graph to graph translation model; MEGAN,[15] a sequential graph editing model; Graph2SMILES,[31] a graph to SMILES translation model without the need of data augmentation; and AT,[32] a SMILES to SMILES translation model with data augmentation. The hyperparameters of "MT +SynCluster" are given in the Supporting Information, Part 6.

As shown in Table 3, the SynCluster framework greatly improves the performance of the baseline models, as is evident from the improvements in both top-*k* and Maxfrag evaluation metrics. Especially for the top-10 accuracy, our framework outperforms the MT and Retroformer models by 11.0 and 13.9%, respectively. Under the same nonaugmented setting for SMILES, Retroformer combined with SynCluster demonstrates superior top-5 and top-10 accuracy compared to Graph2SMILES, and convincingly surpasses SCROP. We further incorporate the SMILES augmentation techniques into the two baseline models and achieve outstanding performance. When the MT model is combined with SynCluster and trained with SMILES augmentation, namely, "MT+SynCluster$_{aug}$", the top-1 and top-5 accuracy reaches 53.6 and 78.2%, respectively, which would be as competitive as the state-of-the-art sequence-based model AT. Furthermore, the top 10 accuracy achieves 85.3%, surpassing the performance of the AT model. These results also highlight the competitiveness of our model compared to the graph-based model Graphretro, with better top-*k* accuracy in top-3, top-5, and top-10 evaluations. Another significant finding to emerge from the analysis is the superiority of the clustering type over NameRxn's superclass for the higher top-*k* accuracy when combined with MT. When using superclass as a substitute and completing the entire retrosynthetic prediction pipeline, the top-10 accuracy suffers a dramatic drop of 5.9%. Although LocalRetro is still leading on the board, it is noteworthy that the reactant generation of our model is constructed on the basis of translation and template-free formulation, which are frequently regarded as more efficient ways to generate unseen compounds without additional expert knowledge.[13,14,24] Furthermore, our strategies are independent of any deployment model and present a promising technique for all single-step models, especially considering consistent improvement with either MT or Retroformer as the backbone model.

We further adopted the Roundtrip accuracy to evaluate the validity of the predicted reactants. Roundtrip accuracy is calculated by checking if it is the predicted product, which is predicted by the forward model using MT, and can match the original product. Here, we utilized a forward Molecular Transformer model trained on the USPTO-50K data set.[17] Consistent with the above results, our framework has elevated MT and Retroformer noticeably in top-*k* Roundtrip accuracy. After incorporating the SMILES augmentation technique for training and the SynCluster framework, MT outperforms all competitive methods listed in Table 4 on top-1 Roundtrip accuracy (79.6%), top-3 Roundtrip accuracy (69.4%), and top-5 Roundtrip accuracy (63.9%), which indicate the promising improvement in chemical validity of sequence-based method using our procedure. The single striking observation to emerge from the comparison is that MEGAN's top-1, top-3, and top-10 Roundtrip accuracies lag behind our best model. This is an impressive result considering that MEGAN is highly dependent on the predefined action space and our SynCluster framework is completely data-driven and unsupervised.

Table 4. Top-*k* Retrosynthesis Roundtrip Accuracy on the USPTO-50K Cleaned Dataset

	top-*k* acc			
	k = 1	*k* = 3	*k* = 5	*k* = 10
MT	0.756	0.554	0.450	0.336
MT+SynCluster	0.776	0.607	0.542	0.449
MT+SynCluster$_{aug}$	**0.796**	**0.694**	**0.639**	0.534
Retroformer$_{base}$	0.729	0.655	0.592	0.486
Retroformer$_{base}$+SynCluster	0.774	0.691	0.638	**0.560**
Retroformer$_{base}$+SynCluster$_{aug}$	0.775	0.682	0.628	0.553
Graph2SMILES	0.755	0.551	0.447	0.323
MEGAN	0.751	0.686	0.639	0.553

Best results of Roundtrip accuracy are in bold.

In terms of the diversity measure, we calculated the Tanimoto similarity distance distribution for four types of fingerprints among the top 10 predictions (Figure 5). According to the higher average distance, our strategy has improved upon the chemical diversity of the predicted reactants for both MT and Retroformer. A surprising aspect of the result is the different effects on the diversity by SMILES augmentation. The average similarity distance in the top-10 reaction pairs is obviously smaller, while MT combines SynCluster and augmentation. In contrast, such a phenomenon is less obvious using Retroformer. The fundamental aspects of augmentation in Retroformer are significantly different from those in MT (as described in Method). This combination of findings provides more information about optimizing the augmentation strategy, which is worthy of further study.

To investigate the chemical validity and interpretability of "MT+SynCluster," we have inspected the self-attention weights of SMILES tokens between input and output. When comparing the "MT+SynCluster" model with the MT model, the comprehensive impact of clustering type is observed in Figures 6, S7, and S8. While adding the SynCluster framework, MT has made a decision consistent with ground truth due to the most significant concern between clustering type and output token. For example, our "MT+SynCluster" model has predicted the accurate Hantzsch thiazole synthesis in which the clustering is associated with the bulk of outputting tokens (Figure 6a). The MT model has reported an erroneous reaction and simply copies the input SMILES in some areas (Figure 6b).

Turning now to the experimental evidence on how augmentation influences the focus of the clustering type. Figure 7 measures the distribution of top attention weight between the clustering type and outputting tokens. Under the circumstance that MT merely attaches SynCluster framework, the average top attention is much higher, which illustrates an explicit positive correlation between the type-token and outputting tokens. On the contrary, the additional augmentation results in a lower average top weight, which indicates that it is not conducive to capturing the information of the

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

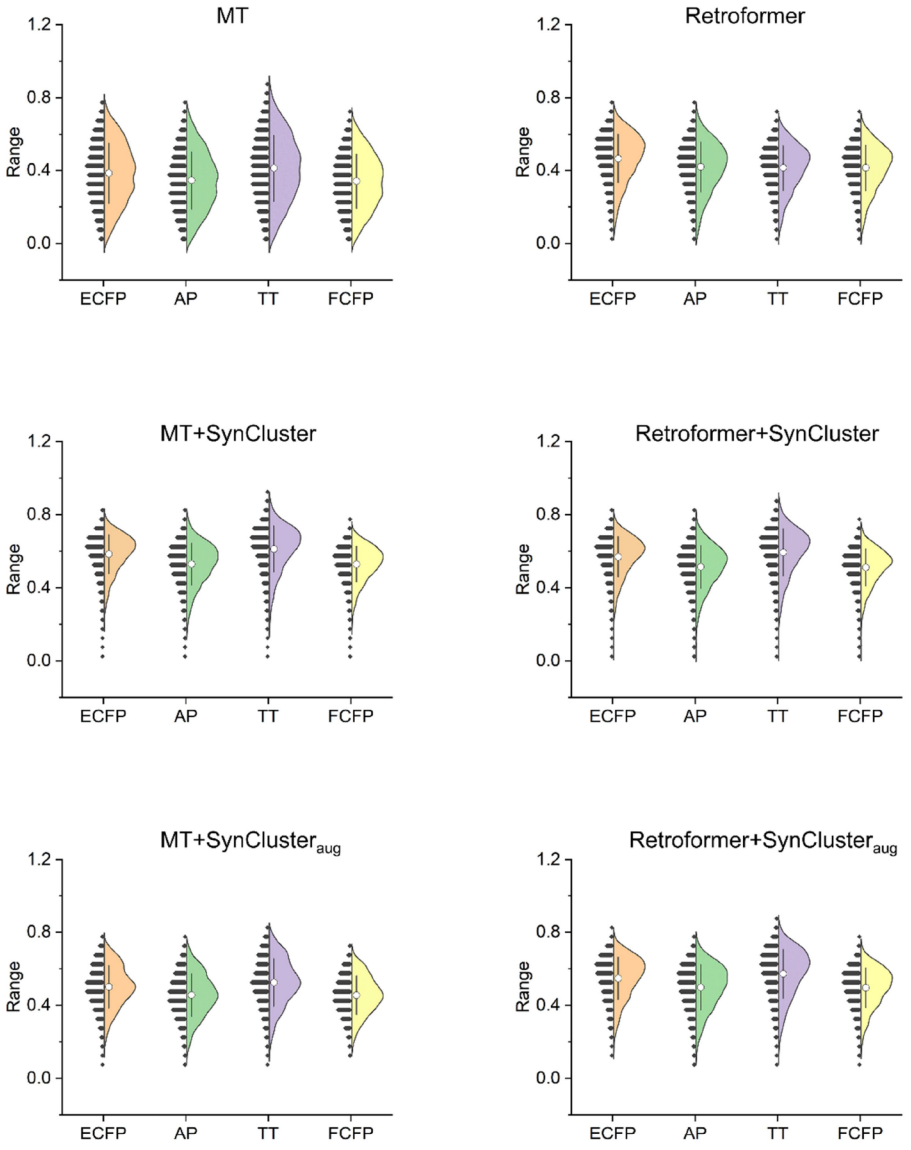


Figure 5. Distribution of Tanimoto similarity distance for six models using ECFP, AP, TT, and FCFP fingerprints.

clustering type when commonly used permutation is leveraged. Besides, this extra augmentation would cause more repetitions while directly embedding top-10 clustering types and assembling top-1 results, as displayed in Figure S9. More examples are shown in Figure S10 for a multifaceted understanding of how models benefit from the embedding of clustering types. For instance, the top-3 high probability clustering types of targeting molecule in Figure S10b are embedded and related to the generations of Wittig reaction, Substitution Nucleophilic reaction, and reduction reaction of amide respectively in the "MT+Syncluster" model, which shows that the information extracted from the clustering type may represent a crucial structure of precursors. In contrast, the model would not capture the information of the third type and generate duplicate reactants while adding augmentation. These findings raise intriguing questions regarding the reasoning application of augmentation for the implication that it may not make a positive contribution to all indicators (consistent with the results in diversity, Figure 5).

Finally, we also demonstrate predictions for difficult cases. Reactions involving multiple reactants (more than two reactants) are challenging to predict due to the multiple sites of disconnection and rarity in the data set. However, our "MT +SynCluster" model has predicted a cohesive chemical space for instances involving multiple reactants, thereby producing the highest top-1 accuracy (Figure S11). The above results elucidate again that our model has acquired the capability to perceive obvious characteristics of reactions and achieve intricate disconnection.

Reference-Pathway Reproductivity and Multistep Application

The synthetic examples in real-world situations are further investigated to highlight the reliability of our retrosynthesis procedure. To be more specific, we selected one testing molecule and three FDA-approved drug molecules for evaluation (Figure 8). All involved molecules are not included in the USPTO50K training data set. In detail, we have checked whether the reference route's intermediates and building blocks would appear in our single-step model's top-10 output. To give a thorough analysis, the ordering of every compound in the LocalRetro model, our "MT+SynCluster", and AT model are compared. It is worth noting that "MT+SynCluster"

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

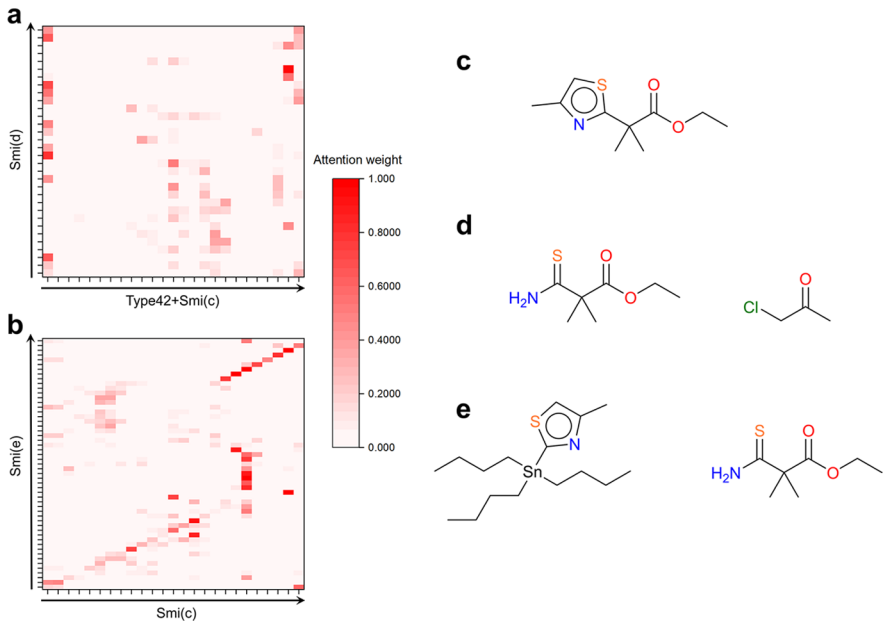


Figure 6. Distribution of attention weight of the input-output pairs, a case study of retrosynthesis prediction using the MT model with or without SynCluster. (a) Attention weight of the input-output pair (c, d) using the "MT+SynCluster" model. "Type" indicates the predicted corresponding cluster label. "Smi" refers to the tokens of canonical SMILES. (b) Attention weight of the input-output pair (c, e) using the MT model.

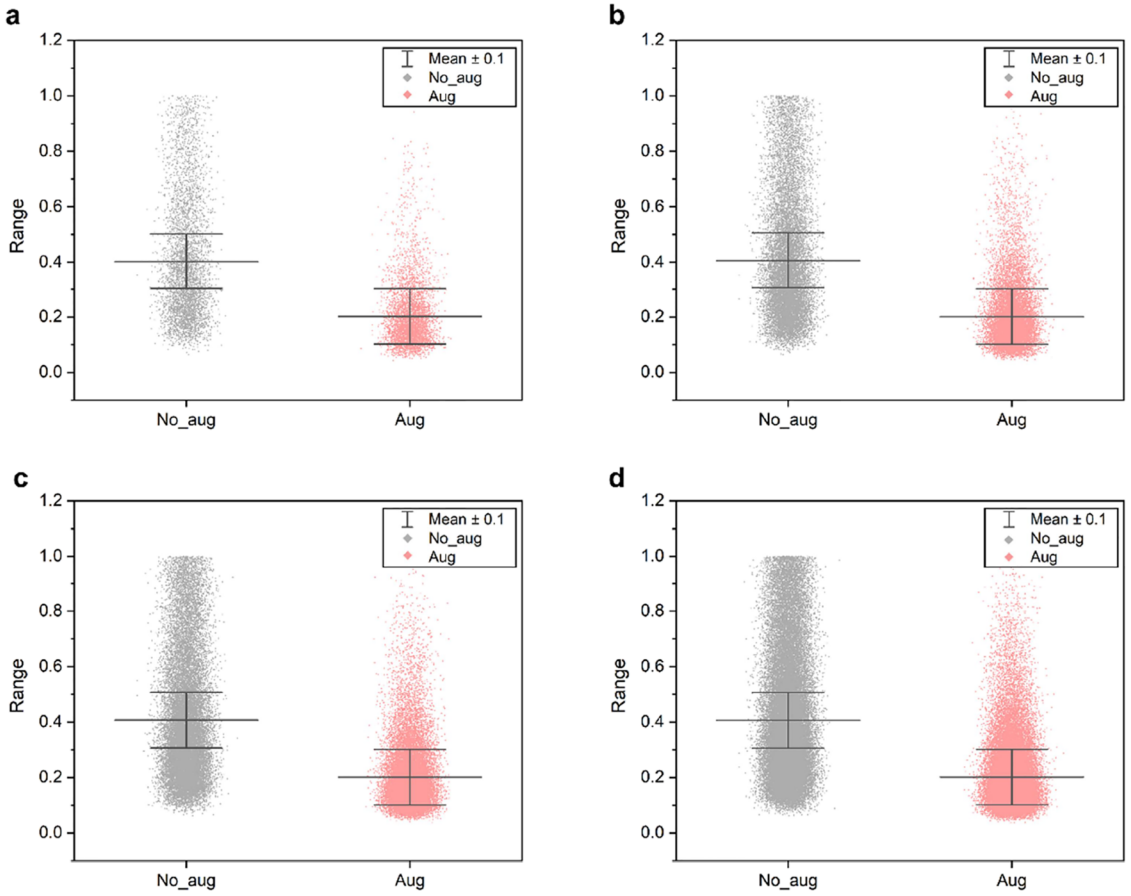


Figure 7. Distribution of top attention weight between the clustering type and outputting SMILES. (a–d), respectively, represent the distribution of top-1, 3, 5, and 10 biggest weights. "No_aug" refers to the "MT+SynCluster" model, and "Aug" refers to "MT+SynCluster$_{aug}$" model.

dominates less time consumption for inference than "MT +SynCluster_aug" and AT. As the experiments were unified as the batch size of 64 and inputting SMILES numbers of 100, the total time consumption for the "MT+SynCluster", "MT +SynCluster_aug", and AT models would be 32 s, 316 s, and 656 s, respectively (Figure S12). Thus, we only take "MT

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446–3461

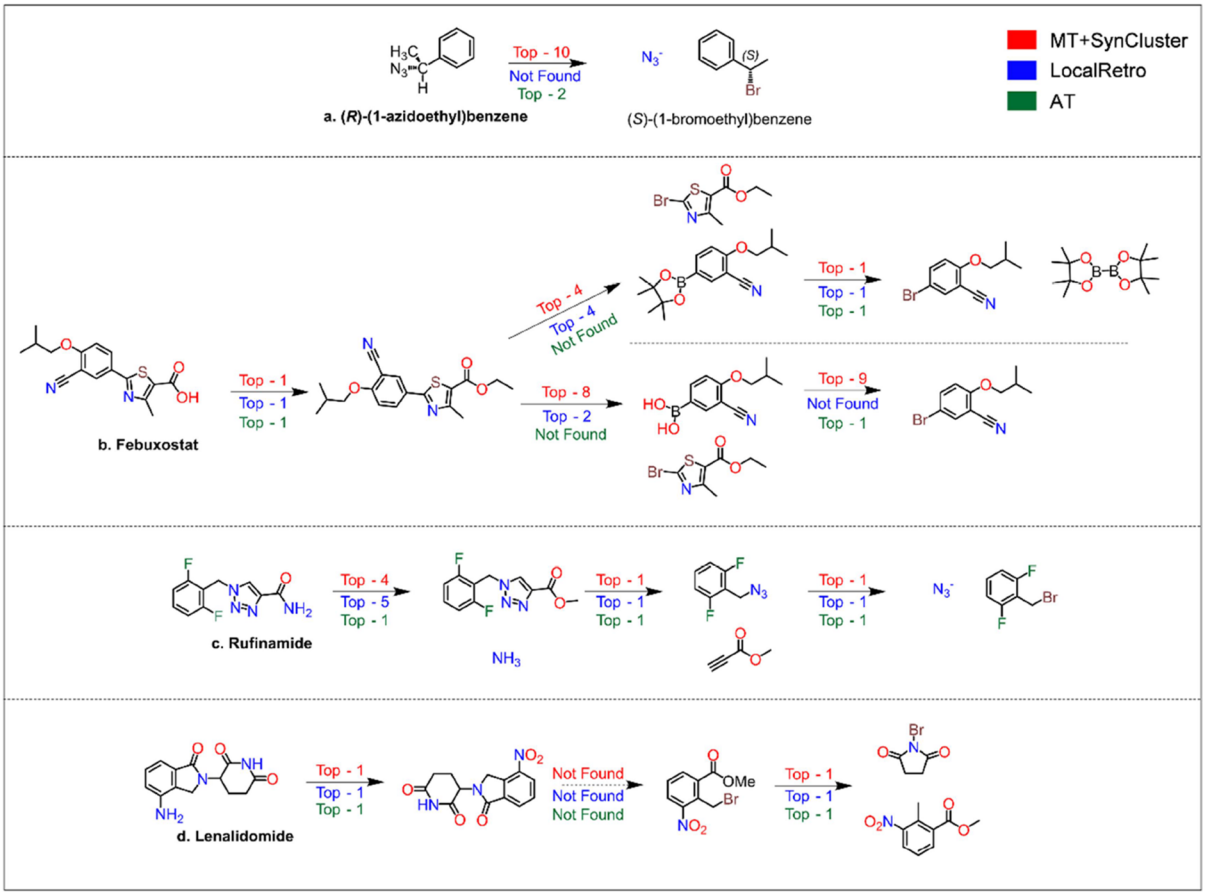


Figure 8. Multistep retrosynthesis pathway reproduction for (a) (R)-(1-azidoethyl)benzene, (b) Febuxostat, used clinically as an antigout drug for its orally effective inhibition of xanthine oxidase. (c) Rufinamide, used in adjuvant therapy for epilepsy Lennox-Gastaut syndrome (LGS). (d) Lenalidomide, an immunomodulator that can be used as anticancer drug. The dotted line and arrow represent the unsuccessfully reproduced step in the prediction. The red, blue, and green fonts in the figure represent the order of each step in the "MT+SynCluster," LocalRetro, and AT models, respectively.

+SynCluster" into account in the repetition of the reference's route and hierarchical search but forsake any form of augmentation.

These four examples provided in Figure 8 showcase the synthesis of R-(1-azidoethyl)benzene (Figure 8a), Febuxostat[60] (Figure 8b), Rufinamide[61] (Figure 8c), Lenalidomide[62] (Figure 8d). The results clearly demonstrate the versatility and effectiveness of our SynCluster in recovering real-world retrosynthetic pathways. Specifically, our method demonstrates its capability in handling reactions involving stereochemistry, as evidenced by the paradigmatic process of nucleophilic substitution depicted in Figure 8a. In addition, SynCluster shows the ability to form C−C bonds through Suzuki coupling, as exhibited in Figure 8b. The successful prediction of the disconnection of Triazole derivatives through a Huisgen reaction is presented in Figure 8c. Furthermore, Figure 8d demonstrates a robust preparation method for the preparation of halogenated hydrocarbons.

The example that deserves deep inspection is the synthesis of Febuxostat, an antigout drug to inhibit xanthine oxidase[60] (Figure 8b). Our framework has replicated the core strategy of the literature[63−65] and makes appropriate improvements in the selection of the Suzuki reaction. The first reported step is to complete the hydrolysis of ester, which is retrieved by our top-1 prediction. Then, the following cross-coupling of the thiazole and benzene rings is the core of the entire route. Here, our framework precisely predicts two candidate Suzuki Cross-coupling processes. In detail, a BPin ester and boronic acid are reported in the top-4 and top-8 output, respectively. Significantly, the preference for aryl boronic esters is proven to contribute to the improvement of yield. It has been reported that reductive dehalogenation is the major side reaction of this step, and the introduction of aryl boronic esters reduces byproduct generation.[66] Previous work has demonstrated that the structure of the boronic ester companies with higher yield (75%) during the synthesis of Febuxostat,[64] while the yield decreases to 65.4%[63] and 63%[65] in the application of boronic acid. To state the performance on the selection of boronic esters or boronic acid is not caused by popularity trends, the distribution of Boronic Esters (or Boronic acid) in the training set, testing set, and top output of prediction is searched further for additional illustration in Figure S13. Boronic acid appears more frequently in training sets than BPin ester, which is discrepant from our selection. Another interesting observation that emerged from the figure was the discriminative trend between our top output and training set. Actually, our output demonstrates a distribution of boronic esters that closely aligns with the testing set, indicating that the predictions are not biased toward what the model has observed in the training set.

If we now turn our attention to the comparison with the AT and LocalRetro models, we find that "MT+SynCluster" has a higher coverage of intermediate candidates, indicating that our

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

Table 5. Top-*k* Forward Synthesis Accuracy on the USPTO-50K Cleaned Dataset through 3 Times Independent Runs

	top-*k* acc			
	k = 1	*k* = 3	*k* = 5	*k* = 10
MT	0.817 ± 0.001	0.888 ± 0.002	0.902 ± 0.002	0.911 ± 0.002
MT+SynCluster	**0.825**[a] ± 0.001	**0.917** ± 0.002	**0.940** ± 0.000	**0.949** ± 0.000

[a]Best results of top-*k* accuracy are in bold.

two-stage strategy bears the potential to meet the real route node compared with the SOTA model. Although the AT model has higher or commensurate orders of candidates in some cases (cf. Figure 8a,c), it would be denoted that retrieving is considerably more significant because the value function mainly determines the actual orders in the multistep search. In the synthesis of Febuxostat (as shown in Figure 8b), the AT model fails to replicate the core strategy of Suzuki cross-coupling. Likewise, LocalRetro exhibits a similar performance in most steps but fails to show a satisfying result where no coincidental prediction is perceived in the first case.

As the potential to reconstruct the reference-based pathway is particularly high, it is reasonable to apply robust search algorithms to recursively generate candidate pathways. Thus, we have examined the potential of the SynCluster framework in multistep retrosynthesis combined with the powerful search algorithm − Retro*.[67] In this experiment, we have utilized the benchmark testing data set from Chen et al.,[67] composed of 190 routes along with their corresponding target molecules. Our evaluation criteria include both the successful rate and planning quality, enabling a comprehensive comparison among AT, LocalRetro, and "MT+SynCluster" in combination with Retro*. As shown in Table S7, our model is able to achieve an impressive 27.9% success rate and an average route length of 3.2, surpassing AT in all aspects. Moreover, even without relying on the template library for generating candidate reactants, our SynCluster framework is able to achieve success rates and route lengths comparable to those of LocalRetro, providing evidence of its substantial diversity and rationality.

Finally, as part of the case study, we have taken three FDA-approved or withdrawal drugs as our target molecules and conducted the multistep application. The generated synthetic routes of LocalRetro, AT, and the "MT+SynCluster" model combined with Retro* are described and analyzed in Supporting Information Part 7. All results above further substantiate the remarkable potential of our framework in the realm of generalization, especially within the context of multistep synthesis.

Multitask Expansion

We have further evaluated the generalizability of the "MT +SynCluster" model to multitasks besides retrosynthesis. Here, we have applied our model to a forward synthesis problem. MT was set as the baseline. As shown in Table 5, our model achieves higher top-*k* accuracy, and the most significant difference is 3.8% on top-10. It indicates that our two-stage strategy bears the potential to deal with the single-step prediction problem of both retrosynthesis and forward synthesis.

In addition to the forward synthesis, another practical concern is the reagent prediction, where we have tailored our "MT+SynCluster" model to approach this problem. The learning scheme of the reagent prediction task is the same as the forward synthesis task. Being subject to the flexibility of reagent applications, the mapping between the reaction and the reagents becomes somehow indefinite for the model to capture. Such difficulty is confirmed by the Maxfrag accuracy, in which both the MT's (baseline) and "MT+SynCluster" model's top-1 accuracy is below 0.3 (Table 6), while the latter still achieves a much higher top-*k* accuracy. Detailed case studies are presented in Figure S14. The "MT+SynCluster" model matches the ground truth better than the baseline, revealing its potential for application in integrated synthesis planning involved conditions. It should be pointed out that the findings would be more persuasive if real-world validation had been performed because reactions are known to have a plethora of possible conditions. The metric (top-*k* accuracy) of these results is subject to encounter greater challenges than a retrosynthesis prediction while mismatching of MT does not mean the completely wrong prediction. As an investigation of methodology, we concurrently look forward to the broad application of SynCluster in reality.

Table 6. MaxFrag Accuracy of Transformer-Based Reagent Prediction Model

	top-1	top-3	top-5	top-10
MT	0.150	0.253	0.306	0.401
MT+SynCluster	**0.276**[a]	**0.400**	**0.457**	**0.515**

[a]Best results of top-*k* MaxFrag accuracy are in bold.

CONCLUSIONS

This study introduces an innovative two-stage synthesis prediction model that combines unsupervised clustering clues with two distinct baseline models, resulting in significant improvements in retrosynthesis, forward synthesis prediction, and reagent prediction. We validated that the clustering type aligns with chemical domain knowledge, while the classifier can effectively capture the relevance of substructures to each clustering type. Moreover, we successfully combined robust search algorithms with SynCluster to recursively generate short and applicable pathways, providing further evidence of the remarkable potential of our model for multistep retrosynthesis prediction.

Our research makes a significant contribution by dividing the end-to-end model, which has limited insights into core organic mechanisms, into two more accessible models. In combination with earlier studies[9,68] proposing a novel type generation model, this approach demonstrates the generation of data-driven, unsupervised, and fine-grained clustering types, which contribute to the comprehension of how AI models extract chemical meaning from data. SynCluster, which serves as a generally compatible framework, can be integrated with downstream models for various tasks and consistently delivers accurate, diverse, and interpretable results. Further investigation into the potential impact of the unsupervised types, which play a crucial role in interpretation, would be highly beneficial.

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

This research also highlights several issues, primarily related to the lack of clean data. For example, since the model is unaware of the asymmetric reaction conditions, the representation of regio- and stereoselectivity needs to be inquired further.[69] Another uncontrolled factor is the incompleteness of many of the collected reactions. Predicting the leaving group (such as the type of halogen) is complicated for the model to do without byproducts.[70] Additionally, mixing high-quality and low-quality data (with incomplete annotations for all reactions in the data set) introduces confusion, making these findings less applicable to real-world settings. Therefore, it is crucial to make additional efforts to standardize and enhance the accessibility of chemical reaction data records.

In terms of the model level, this study aligns with earlier research that suggests that the template-free generation model is slightly inferior to the template-based model in terms of top-k accuracy.[4] However, as a critical technology for reproducing human domain knowledge and driving reactions, the template-free model is gaining more attention.[22] Our standard procedure offers a new viewpoint on the template-free model, but further research is necessary to fully comprehend the implications of fine-grained reaction types in AI models. Furthermore, this study has not evaluated the use of different reaction representations for clustering, and the issue of clustering and reaction representation is fascinating and would benefit from further exploration within the context of various proposed methods.[68,71] Finally, due to the weak representation of SMILES, the occurrence of disordering (error SMILES) remains a vital issue. Nevertheless, we aim to provide a general technique for deep learning in chemical reactions, and we believe that it can be even more promising if the aforementioned limitations are addressed. Our future work will explore other applications of this framework, such as reaction yield prediction, to further demonstrate its generalizability. In conclusion, the SynCluster framework proposed in this study offers a fresh perspective for reasoning about synthesis prediction and has significant implications for our understanding of how models draw inspiration from chemical knowledge.

■ ASSOCIATED CONTENT

Data Availability Statement

USPTO-50K data set is obtained from the previous study by Dai et al.[6] (https://github.com/Hanjun-Dai/GLN). USPTO-500-MT data set is released in t5chem package[25] (https://github.com/HelloJocelynLu/t5chem). USPTO-MIT data set is proposed by Jin et al.[29] (https://github.com/wengong-jin/nips17-rexgen). The testing data set used for multistep synthesis comes from Chen et al.[67] (https://github.com/binghong-ml/retro_star).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.3c00607.

Baseline and models used in this study; details about TMAP tree-based and t-SNE reduction; details about the Transformer and Retroformer model; details about the search model; details for output ranking; hyperparameters of the model used in this study; case studies for multistep retrosynthesis; and supplementary Tables and Figures (PDF)

Accession Codes

Our code is through https://github.com/Yoko1030/SynCluster.

■ AUTHOR INFORMATION

Corresponding Authors

Chang-Yu Hsieh − *Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences and Cancer Center, Zhejiang University, Hangzhou 310058 Zhejiang, China;* orcid.org/0000-0002-6242-4218; Email: kimhsieh2@gmail.com

Tingjun Hou − *Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences and Cancer Center, Zhejiang University, Hangzhou 310058 Zhejiang, China;* orcid.org/0000-0001-7227-2580; Email: tingjunhou@zju.edu.cn

Yu Kang − *Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences and Cancer Center, Zhejiang University, Hangzhou 310058 Zhejiang, China;* orcid.org/0000-0002-0999-8802; Email: yukang@zju.edu.cn

Authors

Tiantao Liu − *Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences and Cancer Center, Zhejiang University, Hangzhou 310058 Zhejiang, China*

Zheng Cao − *College of Computer Science and Technology, Zhejiang University, Hangzhou 310027 Zhejiang, China*

Yuansheng Huang − *Innovation Institute for Artificial Intelligence in Medicine of Zhejiang University, College of Pharmaceutical Sciences and Cancer Center, Zhejiang University, Hangzhou 310058 Zhejiang, China*

Yue Wan − *Tencent Quantum Laboratory, Shenzhen 518057 Guangdong, China*

Jian Wu − *Second Affiliated Hospital School of Medicine, and School of Public Health, Zhejiang University, Hangzhou 310058 Zhejiang, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.3c00607

Author Contributions

T.L. contributed to the main idea, manuscript writing and code; Z.C. contributed to the code reorganization. Y.H. contributed to the manuscript writing and inspection. Y.W. contributed to the collection of the data set and the corresponding experiment. J.W. contributed to the manuscript revision and experimental design. C.H., Y.K. and T.H. contributed to the essential financial support, and conception was responsible for the overall quality. T.L. and Z.C. are equivalent authors. CRediT: **Yuansheng Huang** investigation.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was financially supported by National Key Research and Development Program of China (2022YFF1203003), National Natural Science Foundation of China (22220102001, 82373791, 92370130), and Fundamental Research Funds for the Central Universities (226-2023-00059).

REFERENCES

(1) Thakkar, A.; Kogej, T.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. J. Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. *Chemical science* **2020**, *11* (1), 154−168.

(2) Corey, E. J. The logic of chemical synthesis: multistep synthesis of complex carbogenic molecules (nobel lecture). *Angewandte Chemie International Edition in English* **1991**, *30* (5), 455−465.

(3) Blum, L. C.; Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.* **2009**, *131* (25), 8732−8733.

(4) Dong, J.; Zhao, M.; Liu, Y.; Su, Y.; Zeng, X. Deep learning in retrosynthesis planning: datasets, models and tools. *Briefings Bioinf.* **2022**, *23* (1), No. bbab391.

(5) Coley, C. W.; Green, W. H.; Jensen, K. F. Machine learning in computer-aided synthesis planning. *Accounts of chemical research* **2018**, *51* (5), 1281−1289.

(6) Dai, H.; Li, C.; Coley, C. W.; Dai, B.; Song, L. Retrosynthesis prediction with conditional graph logic network. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 1−11.

(7) Fortunato, M. E.; Coley, C. W.; Barnes, B. C.; Jensen, K. F. Data augmentation and pretraining for template-based retrosynthetic prediction in computer-aided synthesis planning. *Journal of chemical information and modeling* **2020**, *60* (7), 3398−3407.

(8) Segler, M. H. S.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555* (7698), 604−610.

(9) Baylon, J. L.; Cilfone, N. A.; Gulcher, J. R.; Chittenden, T. W. Enhancing retrosynthetic reaction prediction with deep learning using multiscale reaction classification. *Journal of chemical information and modeling* **2019**, *59* (2), 673−688.

(10) Coley, C. W.; Rogers, L.; Green, W. H.; Jensen, K. F. Computer-assisted retrosynthesis based on molecular similarity. *ACS central science* **2017**, *3* (12), 1237−1245.

(11) Chen, S.; Jung, Y. Deep retrosynthetic reaction prediction using local reactivity and global attention. *JACS Au* **2021**, *1* (10), 1612−1620.

(12) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E., Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning, PMLR: Proceedings of Machine Learning Research*, 2017; Vol. *70*, pp 1263−1272.

(13) Shi, C.; Xu, M.; Guo, H.; Zhang, M.; Tang, J., A Graph to Graphs Framework for Retrosynthesis Prediction. In *Proceedings of the 37th International Conference on Machine Learning*, Hal, D., III; Aarti, S., Eds.; PMLR: Proceedings of Machine Learning Research; 2020; Vol. *119*, pp 8818−8827.

(14) Somnath, V. R.; Bunne, C.; Coley, C. W.; Krause, A.; Barzilay, R., Learning graph models for template-free retrosynthesis. *arXiv preprint arXiv:2006.07038* 2020.

(15) Sacha, M.; Błaz, M.; Byrski, P.; Dabrowski-Tumanski, P.; Chrominski, M.; Loska, R.; Włodarczyk-Pruszynski, P.; Jastrzebski, S. Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits. *Journal of Chemical Information and Modeling* **2021**, *61* (7), 3273−3284.

(16) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5999−6009.

(17) Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science* **2019**, *5* (9), 1572−1583.

(18) Kreutter, D.; Schwaller, P.; Reymond, J.-L. Predicting enzymatic reactions with a molecular transformer. *Chemical science* **2021**, *12* (25), 8648−8659.

(19) Mao, K.; Xiao, X.; Xu, T.; Rong, Y.; Huang, J.; Zhao, P. Molecular graph enhanced transformer for retrosynthesis prediction. *Neurocomputing* **2021**, *457*, 193−202.

(20) Karpov, P.; Godin, G.; Tetko, I. V., A Transformer Model for Retrosynthesis. In *Artificial Neural Networks and Machine Learning − ICANN 2019: Workshop and Special Sessions*; Tetko, I. V.; Kůrková, V.; Karpov, P.; Theis, F., Eds.; Springer International Publishing: Cham, 2019; Vol. *11731*, pp 817−830.

(21) Wan, Y.; Hsieh, C.-Y.; Liao, B.; Zhang, S. Retroformer: Pushing the Limits of End-to-end Retrosynthesis Transformer. In *International Conference on Machine Learning*; PMLR, 2022; pp 22475−22490.

(22) Lin, K.; Xu, Y.; Pei, J.; Lai, L. Automatic retrosynthetic route planning using template-free models. *Chemical science* **2020**, *11* (12), 3355−3364.

(23) Wang, X.; Li, Y.; Qiu, J.; Chen, G.; Liu, H.; Liao, B.; Hsieh, C.-Y.; Yao, X. RetroPrime: A Diverse, plausible and Transformer-based method for Single-Step retrosynthesis predictions. *Chemical Engineering Journal* **2021**, *420*, No. 129845.

(24) Yan, C.; Ding, Q.; Zhao, P.; Zheng, S.; Yang, J.; Yu, Y.; Huang, J., Retroxpert: Decompose retrosynthesis prediction like a chemist. *arXiv preprint arXiv:2011.02893* 2020.

(25) Lu, J.; Zhang, Y. Unified deep learning model for multitask reaction predictions with explanation. *Journal of Chemical Information and Modeling* **2022**, *62* (6), 1376−1387.

(26) NameRxn, Version 3.2.0 [May 2021]; https://www.nextmovesoftware.com/namerxn.html: 2022.

(27) Lowe, D. M. *Extraction of chemical structures and reactions from the literature*; University of Cambridge, 2012.

(28) Liu, B.; Ramsundar, B.; Kawthekar, P.; Shi, J.; Gomes, J.; Luu Nguyen, Q.; Ho, S.; Sloane, J.; Wender, P.; Pande, V. Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS central science* **2017**, *3* (10), 1103−1113.

(29) Jin, W.; Coley, C.; Barzilay, R.; Jaakkola, T. Predicting organic reaction outcomes with weisfeiler-lehman network. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 2607−2616.

(30) Zheng, S.; Rao, J.; Zhang, Z.; Xu, J.; Yang, Y. Predicting retrosynthetic reactions using self-corrected transformer neural networks. *Journal of chemical information and modeling* **2020**, *60* (1), 47−55.

(31) Tu, Z.; Coley, C. W. Permutation invariant graph-to-sequence model for template-free retrosynthesis and reaction prediction. *arXiv preprint arXiv:2110.09681* 2021, DOI: DOI: 10.1186/s13619-023-00179-2.

(32) Tetko, I. V.; Karpov, P.; Van Deursen, R.; Godin, G. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat. Commun.* **2020**, *11*, 15575.

(33) Kraut, H.; Eiblmaier, J.; Grethe, G.; Löw, P.; Matuszczyk, H.; Saller, H. Algorithm for reaction classification. *Journal of chemical information and modeling* **2013**, *53* (11), 2884−2895.

(34) Hendrickson, J. B.; Sander, T. COGNOS: A Beilstein-type system for organizing organic reactions. *Journal of chemical information and computer sciences* **1995**, *35* (2), 251−260.

(35) Law, J.; Zsoldos, Z.; Simon, A.; Reid, D.; Liu, Y.; Khew, S. Y.; Johnson, A. P.; Major, S.; Wade, R. A.; Ando, H. Y. Route designer: a retrosynthetic analysis tool utilizing automated retrosynthetic rule generation. *Journal of chemical information and modeling* **2009**, *49* (3), 593−602.

(36) Boström, J. Transformers for future medicinal chemists. *Nature Machine Intelligence* **2021**, *3* (2), 102−103.

(37) Schneider, N.; Lowe, D. M.; Sayle, R. A.; Landrum, G. A. Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. *Journal of chemical information and modeling* **2015**, *55* (1), 39−53.

(38) Nextmove, *Patent Data*. http://nextmovesoftware.com/blog/2014/02/27/unleashing-over-a-million-reactions-into-the-wild/, Ed. 2014.

(39) Heid, E.; Liu, J.; Aude, A.; Green, W. H. Influence of Template Size, Canonicalization, and Exclusivity for Retrosynthesis and Reaction Prediction Applications. *Journal of Chemical Information and Modeling* **2022**, *62* (1), 16−26.

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461

(40) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure-activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25* (2), 64−73.
(41) Rogers, D.; Hahn, M. Extended-connectivity fingerprints. *Journal of chemical information and modeling* **2010**, *50* (5), 742−754.
(42) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, *27* (2), 82−85.
(43) Butina, D. Unsupervised data base clustering based on daylight's fingerprint and Tanimoto similarity: A fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (4), 747−750.
(44) Lipkus, A. H. A proof of the triangle inequality for the Tanimoto distance. *J. Math. Chem.* **1999**, *26* (1−3), 263−265.
(45) Probst, D.; Reymond, J.-L. Visualization of very large high-dimensional data sets as minimum spanning trees. *J. Cheminf.* **2020**, *12* (1), 12.
(46) Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9* (11), 2579−2605.
(47) Wei, J.; Zou, K. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196* 2019.
(48) Bjerrum, E. J. SMILES enumeration as data augmentation for neural network modeling of molecules. *arXiv preprint arXiv:1703.07076* 2017, DOI: DOI: 10.1007/s00464-023-10221-8.
(49) Wei, J.-M.; Yuan, X.-J.; Hu, Q.-H.; Wang, S.-Q. A novel measure for evaluating classifiers. *Expert Systems with Applications* **2010**, *37* (5), 3799−3809.
(50) Jurman, G.; Riccadonna, S.; Furlanello, C. A comparison of MCC and CEN error measures in multi-class prediction. *PLoS ONE* **2012**, *7* (8), No. e41882.
(51) Segler, M. H.; Waller, M. P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chem. − Eur. J.* **2017**, *23* (25), 5966−5971.
(52) Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V. H.; Haeuselmann, R. A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chemical science* **2020**, *11* (12), 3316−3325.
(53) Chen, T.; Guestrin, C., Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016; Vol. *10*, pp 785−794.
(54) Xiong, Z.; Wang, D.; Liu, X.; Zhong, F.; Wan, X.; Li, X.; Li, Z.; Luo, X.; Chen, K.; Jiang, H. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *J. Med. Chem.* **2019**, *63* (16), 8749−8760.
(55) Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular graph convolutions: moving beyond fingerprints. *Journal of computer-aided molecular design* **2016**, *30* (8), 595−608.
(56) Chen, B.; Barzilay, R.; Jaakkola, T. *Path-augmented graph transformer network. arXiv preprint arXiv:1905.12712* 2019, DOI: DOI: 10.1186/s12864-023-09782-8.
(57) Li, M.; Zhou, J.; Hu, J.; Fan, W.; Zhang, Y.; Gu, Y.; Karypis, G. Dgl-lifesci: An open-source toolkit for deep learning on graphs in life science. *ACS omega* **2021**, *6* (41), 27233−27238.
(58) Dal Pozzolo, A.; Caelen, O.; Johnson, R. A.; Bontempi, G., Calibrating probability with undersampling for unbalanced classification. In *2015 IEEE symposium series on computational intelligence*; IEEE, 2015; pp 159−166.
(59) Kokhlikyan, N.; Miglani, V.; Martin, M.; Wang, E.; Alsallakh, B.; Reynolds, J.; Melnikov, A.; Kliushkina, N.; Araya, C.; Yan, S., Captum: A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv:2009.07896* 2020.
(60) Hair, P. I.; McCormack, P. L.; Keating, G. M. Febuxostat. *Drugs* **2008**, *68* (13), 1865−1874.
(61) Zhang, P.; Russell, M. G.; Jamison, T. F. Continuous flow total synthesis of rufinamide. *Organic Process Research & Development* **2014**, *18* (11), 1567−1570.
(62) Ponomaryov, Y.; Krasikova, V.; Lebedev, A.; Chernyak, D.; Varacheva, L.; Chernobroviy, A. Scalable and green process for the synthesis of anticancer drug lenalidomide. *Chem. Heterocycl. Compd.* **2015**, *51* (2), 133−138.
(63) Cao, Q.; Ma, X.; Xiong, J.; Guo, P. The preparation of febuxostat by Suzuki reaction. *Chin. J. New Drugs* **2016**, *25* (09), 1057−1060.
(64) Jian, H.; Li, J. The Synthesis of febuxostat. CN102285937A, 2011−12−21, 2011.
(65) Birari, D. R.; Rao, D. R. Rajendra Narayanrao Kankan Processes for the preparation of febuxostat and salts thereof. WO2011073617A1, 2011.
(66) Urawa, Y.; Naka, H.; Miyazawa, M.; Souda, S.; Ogura, K. Investigations into the Suzuki−Miyaura coupling aiming at multikilogram synthesis of E2040 using (o-cyanophenyl) boronic esters. *Journal of organometallic chemistry* **2002**, *653* (1−2), 269−278.
(67) Chen, B.; Li, C.; Dai, H.; Song, L., Retro*: learning retrosynthetic planning with neural guided A* search. In *Proceedings of the 37th International Conference on Machine Learning, PMLR: Proceedings of Machine Learning Research*, 2020; Vol. *119*, pp 1608−1616.
(68) Probst, D.; Schwaller, P.; Reymond, J.-L. Reaction Classification and Yield Prediction using the Differential Reaction Fingerprint DRFP. *Digital Discovery* **2022**, *1* (2), 91−97.
(69) Pesciullesi, G.; Schwaller, P.; Laino, T.; Reymond, J.-L. Transfer learning enables the molecular transformer to predict regio-and stereoselective reactions on carbohydrates. *Nature communications* **2020**, *11* (1), 4874.
(70) Guo, Z.; Wu, S.; Ohno, M.; Yoshida, R. Bayesian Algorithm for Retrosynthesis. *Journal of Chemical Information and Modeling* **2020**, *60* (10), 4474−4486.
(71) Nugmanov, R. I.; Mukhametgaleev, R. N.; Akhmetshin, T.; Gimadiev, T. R.; Afonina, V. A.; Madzhidov, T. I.; Varnek, A. CGRtools: Python library for molecule, reaction, and condensed graph of reaction processing. *Journal of chemical information and modeling* **2019**, *59* (6), 2516−2521.
(72) Coley, C. W.; Green, W. H.; Jensen, K. F. RDChiral: An RDKit wrapper for handling stereochemistry in retrosynthetic template extraction and application. *Journal of chemical information and modeling* **2019**, *59* (6), 2529−2537.
(73) Finnigan, W.; Hepworth, L. J.; Flitsch, S. L.; Turner, N. J. RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. *Nature catalysis* **2021**, *4* (2), 98−104.

https://doi.org/10.1021/jacsau.3c00607
JACS Au 2023, 3, 3446−3461