SITE-SPECIFIC TEMPLATE GENERATIVE APPROACH FOR RETROSYNTHETIC PLANNING

Yu Shee[1], **Haote Li**[1], **Pengpeng Zhang**[1], **Andrea M. Nikolic**[1], **Sanil Sreekumar**[2], **Frédéric Buono**[2], **Jinhua J. Song**[2], **Timothy R. Newhouse**[1], and **Victor S. Batista**[1]

[1]Department of Chemistry, Yale University, New Haven, Connecticut 06511, USA
[2]Chemical Development, Boehringer Ingelheim Pharmaceuticals Inc, Ridgefield, Connecticut 06877, USA

ABSTRACT

Retrosynthesis, the strategy of devising laboratory pathways for small molecules by working backwards from the target compound, remains a rate limiting step in multi-step synthesis of complex molecules, particularly in drug discovery. Enhancing retrosynthetic efficacy requires overcoming the vast complexity of chemical space, the limited known interconversions between molecules, and the challenges posed by limited experimental datasets. In this study, we introduce generative machine learning methods for retrosynthetic planning that generate reaction templates. Our approach features three key innovations. First, the models generate complete reactions, known as templates, instead of reactants or synthons. Through this abstraction, novel chemical transforms resembling those in the training dataset can be generated. Second, the approach optionally allows users to select the specific bond or bonds to be changed in the proposed reaction, enabling human interaction to influence the synthetic approach. Third, one of our models, based on the conditional kernel-elastic autoencoder (CKAE) architecture, employs a latent space to measure the similarity between generated and known reactions, providing insights into their chemical viability. Together, these features establish a coherent framework for retrosynthetic planning, as validated by our experimental work. We demonstrate the application of our machine learning methodology to design a synthetic pathway for a simple yet challenging small molecule of pharmaceutical interest. The pathway was experimentally proven viable through a 3-step process, which compares favorably to previous 5-9 step approaches. This improvement demonstrates the utility and robustness of the generative machine learning approaches described herein and highlights their potential to address a broad spectrum of challenges in chemical synthesis.

***K*eywords** Retrosynthesis · Reaction Template · Generative Model

1 Introduction

Retrosynthesis is the design of breaking down complex molecules into simpler building blocks, a concept originally developed by Corey as a means to educate students to conduct multi-step synthesis [1]. This intellectual framework laid the foundation for the development of Computer-Aided Synthesis Planning (CASP), a field that emerged to assist chemists in navigating various paths of synthesis, playing a pivotal role in augmenting human capabilities for refining a synthetic approach [2, 3]. In the earlier stages, systems based on expert rules provided valuable insights for chemists [4, 5, 6, 7]. As organic chemistry advanced, encompassing broader chemical space and synthetic methodologies, recent advancements in CASP have shifted from rule-based to precedent-based approaches [8]. This shift was facilitated by the large-scale extraction of reaction rules [9]. The process progressed from manual creation to automated extraction from extensive chemical datasets. Several extraordinary software packages had emerged due to this transition which empowered CASP tools to tap into vast repositories of historical reaction data [8, 10, 11]. Grzybowski and others [12, 13] further introduced user-purpose-driven tools for route optimization, demonstrating remarkable success through experimental validations [14, 15, 16]. Furthermore, the integration of machine learning (ML) methods has marked the latest chapter in the ongoing evolution of CASP [17, 18]. ML

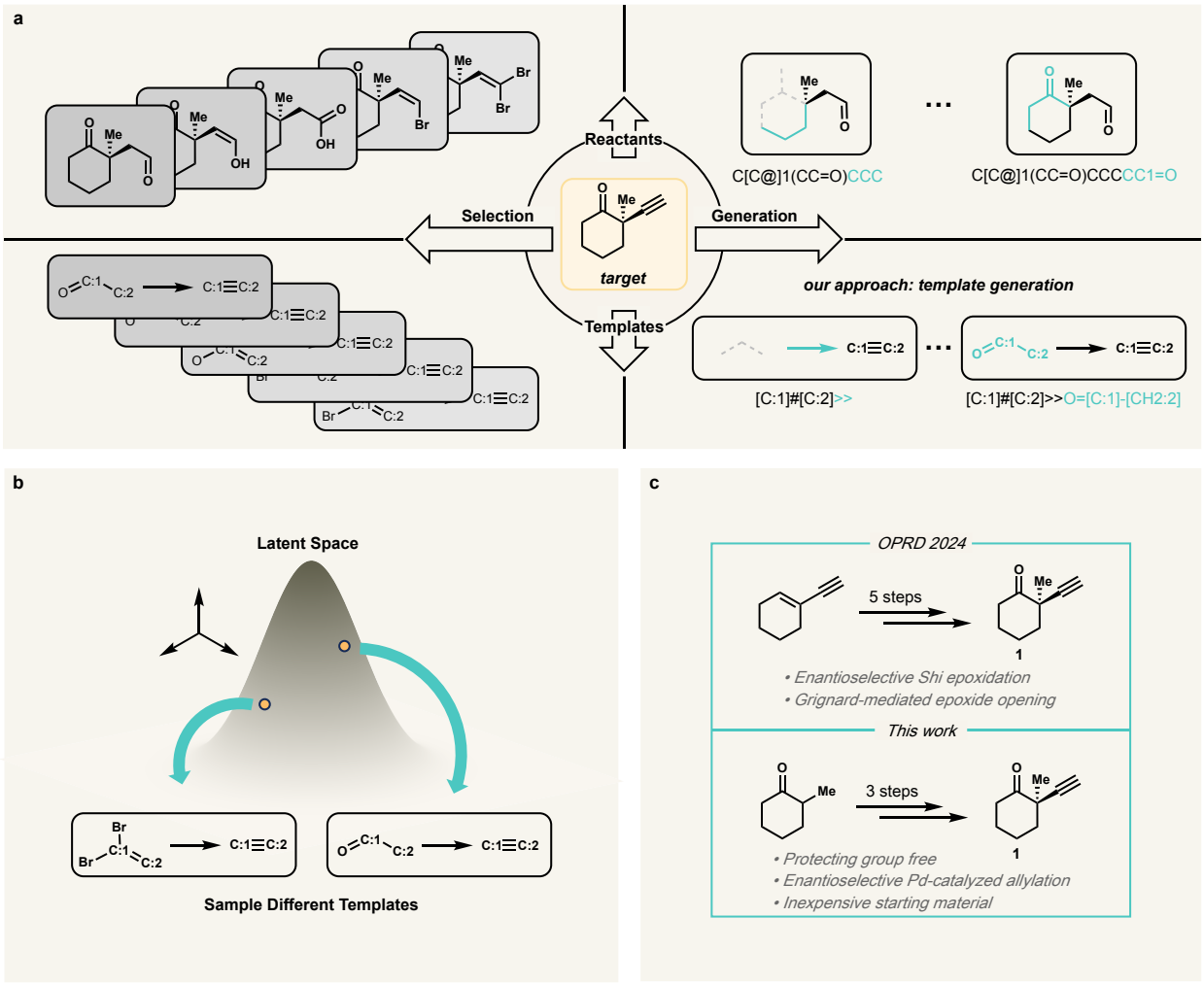


Figure 1: Common machine learning methods for retrosynthesis and our approach. (a) Reactants and templates can be selected or generated based on a target compound using different machine learning models. Template generation is used in this work. (b) A structured latent space is incorporated in one of the models in this work. Sampling in the latent space can give different reaction templates for given products. (c) Reduction of synthetic steps for a key intermediate for active pharmaceutical ingredients (API).

models offer promising alternatives and can be broadly categorized as selection-based, semi-template, or generation-based methods [19] (see Fig. 1a).

Selection-based methods, such as reactant selection and template selection methods, aim to choose appropriate molecules or reaction rules from the given sets. Reactant selection methods [20, 21] involve ranking molecules from a collection of candidates based on the target compounds. While reactant selection methods have the advantage of ensuring the chosen molecules are valid, their effectiveness relies on the availability of reactants in the candidate sets. Template selection methods [22, 23, 24, 25, 26, 27, 28] rank the reaction templates in terms of their applicability to the target molecules. These templates capture subgraph patterns representing the change in atoms and bonds during a reaction. Notably, the RDChiral repository by Coley et al. [29] offers template extraction methods and a collection of reaction templates in the form of SMARTS strings. Template selection methods simplify the reaction representation to a single template instead of multiple reactants. Additionally, the same template can be applied to different target compounds instead of having multiple sets of reactants for the target compounds, thereby providing a higher coverage of reaction space. However, like reactant selection methods, template selection methods are dependent on the coverage and diversity of available templates within predefined reaction rules.

Semi-template methods [30, 31, 32, 33, 34, 35] involve the identification of reaction centers, synthons, or leaving groups, followed by the prediction of corresponding reactants based on these rules. Some semi-template methods [32, 34, 35] are akin to selection-based methods, where reactants are obtained by predicting reaction centers and selecting from a collection of leaving groups, or by selecting necessary edits on molecular graphs. Other semi-template methods adopt generation components, in which reactants are generated from products and identified synthons or rules.

Generation-based methods are not bound by the sets of available reactants or templates and hold promise to mapping wider areas of chemical space. These include template-free methods [36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51] that treat reactant generation as a translation task, aiming to predict the reactants directly from the given products without having in-dataset reaction rules. They therefore bear the potential to explore a wider range of possible reactions.

In this study, we introduce *template generation* which represents a new distinct category of generation-based methods for retrosynthetic planning. Template generation models employ the Sequence-to-Sequence (S2S) architecture trained to translate product information into reaction templates, as opposed to generating reactants. The capability of template generation thus extends beyond the available templates or predefined reaction rules of template selection-based approaches, enabling the discovery of novel reaction templates that expand the scope of retrosynthetic planning. The combination of generated reaction templates and the "RunReactants" function from RDKit, offer an efficient means to swiftly identify templates that yield grammatically coherent reactants from given products. This facilitates the exploration of previously uncharted chemical reactions and pathways.

One of the major benefits of using template generation is the ease of checking the reaction validity. During the transformation of a reaction template, the product is guaranteed to be converted to the reactant with exact matching of atoms indices and relevant functional groups from the description of the template. In comparison to reactant generative models, this feature greatly reduces the uncertainty in the produced reactants which might not correspond to any known reactions or have key atom mismatches due to problems during decoding.

The second design is a sampling generative model (sampling model) for template generation conditioned by target compounds. S2S models, such as those employed in the template-free methods, predict pathways deterministically and do not have a sampling process or definition of latent space. In contrast, our sampling model has a latent space, enabling the generation, interpolation, extrapolation, and distance measurement of various templates (Fig. 1b). Deterministic models that take target compounds as inputs and generate templates are also developed in this work. Importantly, the encoder of the model can incorporate positional embedding for reaction centers, enabling users to specify specific reacting sites during prediction. Results are benchmarked on the USPTO-FULL dataset.

Our sampling model, based on the conditional kernel-elastic autoencoder (CKAE) [52], is the first of its kind in the field of retrosynthesis. This model conditions on corresponding products during training, allowing interpolation and extrapolation of reaction templates in latent space to generate new reaction templates during the sampling process. The latent space also provides a measure of distances between reaction templates, allowing us to identify the closest reaction reference within the dataset, or determine the similarity between two chemical reactions.

Our template generation method introduces a design where site-specific templates (SST) are generated along with target compounds with labeled reaction centers (i.e., center-labeled products, CLP) that specifies the reaction centers. This results in the generation of concise and informative sets of templates that are different from the templates available in the RDChiral repository [29]. Through benchmarking with a public dataset, the performance of our approach is demonstrated.

With SSTs and generation methods in place, our approach is validated through the practical application of synthesis. A library of potent anti-cancer agents was recently reported by Boehringer Ingelheim [53]. One of the key intermediates for the synthesis of these anti-cancer compounds is compound **1**, a cyclohexanone with a quaternary chiral center in the α-position containing an alkyne moi-

ety (Fig. 1c) [54, 55]. Our objective was to develop a more step-efficient route to synthesize compound **1**. The route proceeds over 3 steps, compared to prior approaches that required 5-9 steps, including a recent process involving Grignard-mediated epoxide opening as a key step in a 5-step route starting from commercial starting materials [54, 55]. Reducing the number of steps in a synthetic process is enabling to develop scalable and more sustainable approaches, while also reducing the amount of time necessary for each batch [56, 57]. Our experimental validation demonstrates the practicality and reliability of the retrosynthetic predictions, suggesting their underlying promise to address a wide spectrum of synthetic challenges.

2 Results and Discussion

2.1 Site-Specific Templates and Center-Labeled Products

Reaction templates that only apply to reaction centers within the target compounds are referred to as site-specific templates (SST). These are different from RDChiral templates which involve a broader structural context [29] since SSTs do not differentiate neighboring atoms or special functional groups when matching substructures within the target compounds. The presence of center-labeled products (CLP) is a pre-requisite for the effective use of SSTs. Such labeling is essential to avoid ambiguity when a SST can be applied to multiple sites within a target compound. Examples of SST and CLP are shown in Fig. 2a where the "*" symbol represents the reaction centers. To prepare SSTs, the radius parameter in RDChiral is set to 0 (while RDChiral normally sets radius to 1 which captures 1 bond away from the reaction centers) and special functional groups are removed. Therefore, neighboring atoms and distal functional groups are not included in SSTs. Also, explicit degrees and explicit numbers of hydrogens are not included in the SSTs. To prepare CLPs, RDChiral also has implementations to capture the changed atoms, so the centers can be labeled for target compounds. Further explanations and examples are provided in the SI (Sec. 5.1 and Sec. 5.2).

2.2 Deterministic Model Performance

Figure 2 shows a schematic representation of the deterministic models, Model A and Model B. Model A takes a target compound as an input and translates it into SSTs and CLPs. CLPs specify how the SSTs should be applied to the target compound. Model B takes the target and the specific reaction centers and generates templates corresponding to those specific sites (see SI Sec. 5.3 for more information and Fig. 8 for a comparison of the models).

Figure 3 shows the comparative analysis of the performance for Models A and B (highlighted in red), in terms of Top-K accuracy, as compared to state-of-the-art methods. Top-K accuracy measures the percentage of top K

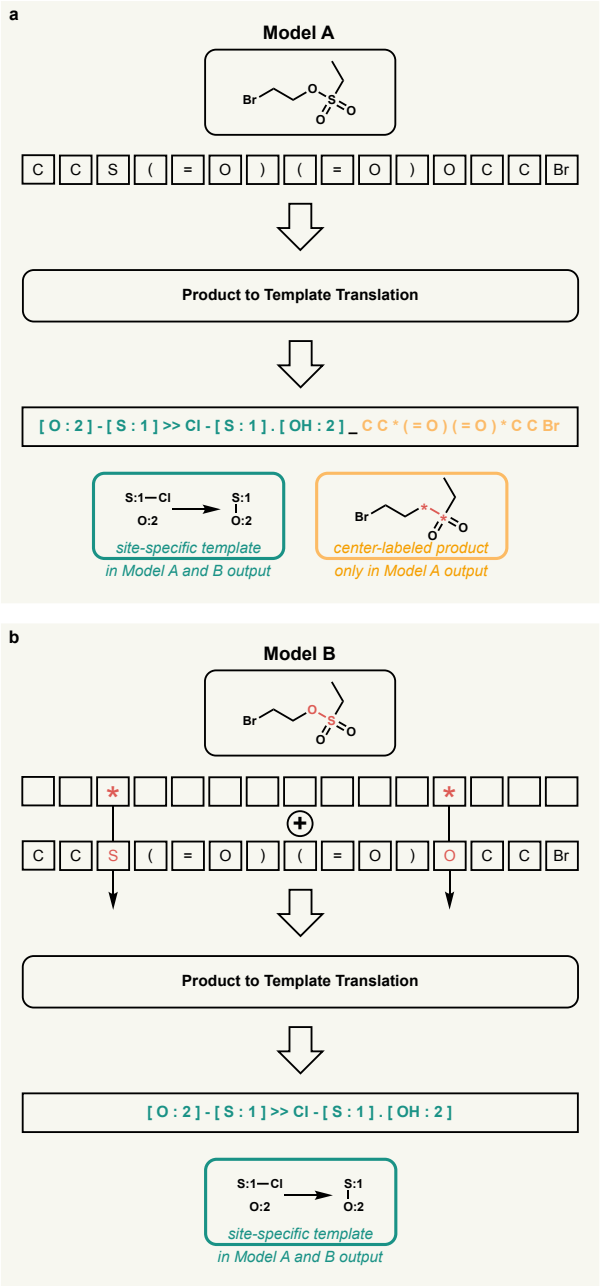


Figure 2: Schematic retrosynthetic workflow for Models A and B. (a) Workflow of Model A. (b) Workflow of Model B. Model B has reaction center embeddings and does not have center-labeled products in the output. Detailed descriptions of the models are provided in the SI (Sec. 5.3 and Sec. 5.4)

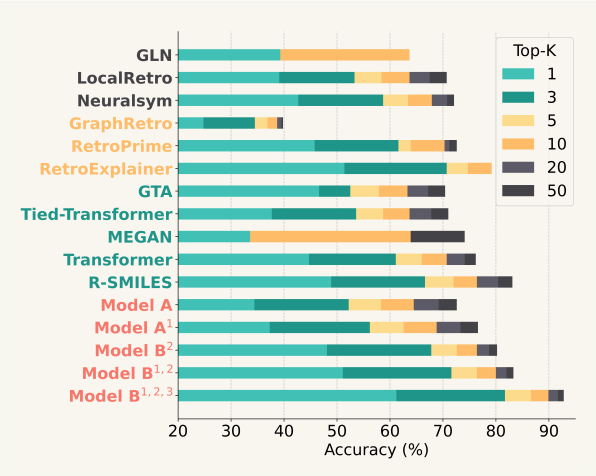


Figure 3: USPTO-Full Top-K accuracy of retrosynthesis models. GLN [26], LocalRetro [27], and Neuralsym [22] in black are template-based selection methods. GraphRetro [32], RetroPrime [33], and RetroExplainer [34] in yellow are semi-template methods. GTA [43], Tied-Transformer [48], MEGAN [45], Transformer [42], and R-SMILES [50] in green are template-free generation methods. This work (in red) uses a template-generation method. Reactant-based selection methods are not included due to out-of-memory for the USPTO-FULL dataset [19]. [1] indicates that if the correct reactants contain one of the 50 most commonly seen spectators in the USPTO-Full dataset, the reaction is removed from the test set. [2] indicates that reaction centers are provided. [3] indicates that the maximum number of reaction centers is two.

predictions containing reactants that precisely match the ground truth reactants in the testing set. Figure 3 includes results for the original USPTO-Full testing dataset as well as for a cleaned testing set to address errors related to atom-mappings such as solvent and reagent atoms erroneously considered as part of the reactions. The cleaned testing set is prepared by removing reactions containing reactants that are the 50 most frequently observed spectators in the USPTO-FULL dataset. The size of the cleaned testing set is 90.7% of the original set of 95k reactions. Top-K results are obtained by using the beam search method.

Model A, which does not use reaction centers, performs comparably well to other methods. The cleaned set allows for higher accuracy although it may inadvertently exclude some reactions where the common spectators actually participate as reactants. Model B leverages reaction center information. On the cleaned set, Model B reaches a performance milestone, achieving an accuracy rate as high as 80% for Top-10 predictions (see SI Sec. 5.7 for details).

RetroExplainer [34], with semi-template components, demonstrates remarkable prediction accuracy owing to its data modeling approach and the utilization of a set of leaving groups. However, this approach may experience variations in performance when handling uncommon scenarios or leaving groups not explicitly represented in the dataset. R-SMILES [50], a template-free generation-based method, introduced the root-aligned SMILES representation to ensure minimal edit distances between product and reactant SMILES. Through this customized string representation and data augmentation, they achieved the highest accuracy among template-free methods. Nonetheless, data augmentation is not utilized in this work, leaving room for potential improvements in accuracy for future endeavors.

Additionally, an analysis of the Top-K accuracy considering different numbers of reaction centers for Model B is shown. Over half of the test reactions possess one or two reaction centers, following the same distribution of reaction center counts of the training set. Consequently, for test reactions with a maximum of two reaction centers, Model B achieved the highest Top-K accuracy compared to other center counts, with the Top-10 accuracy reached 90% (see last row of Fig. 3 and SI Sec. 5.7), showcasing exceptional predictive capabilities in scenarios characterized by a limited number of reaction centers. The high Top-K accuracy achieved by Model B for reactions with few reaction centers is particularly significant, as it corresponds to real-world applications where a majority of reactions feature a low number of reaction centers. For instance, 90% of the dataset comprises reactions with no more than four reaction centers (see SI Sec. 5.7).

2.3 Sampling Model with Latent Space

A sampling generative model, which exploits a sampling process with a latent space, is different from the deterministic approach. To the best of our knowledge, the application of a sampling model for retrosynthetic planning has not been explored. Model C is built upon the architecture of Conditional Kernel-Elastic Autoencoder (CKAE) [52]. In Model C, both the input and output consist of combinations of SSTs and CLPs. The goal of Model C, akin to a variational autoencoder, is to reconstruct the input with latent space compression. Comparing to previous CKAE molecular generation models where conditions are represented by specific values or molecular properties, the CKAE model as applied to Model C utilizes the SMILES representation of target molecules as conditions. During the sampling process, a target compound is provided as the condition and latent vectors are sampled, different SSTs and CLPs, which correspond to the same target compound condition, can then be generated for different latent vectors.

In addition to generative sampling, the encoder of Model C offers a valuable referencing feature. It maps the input into a latent space with a distance regularized by a modified maximum mean discrepancy loss (m-MMD) [52]. This distance furnishes a quantifiable metric for assessing the similarity between reactions, aiding in evaluating and understanding the differences between chemical transformations. Such capability enables the identification of similar reactions within the dataset.

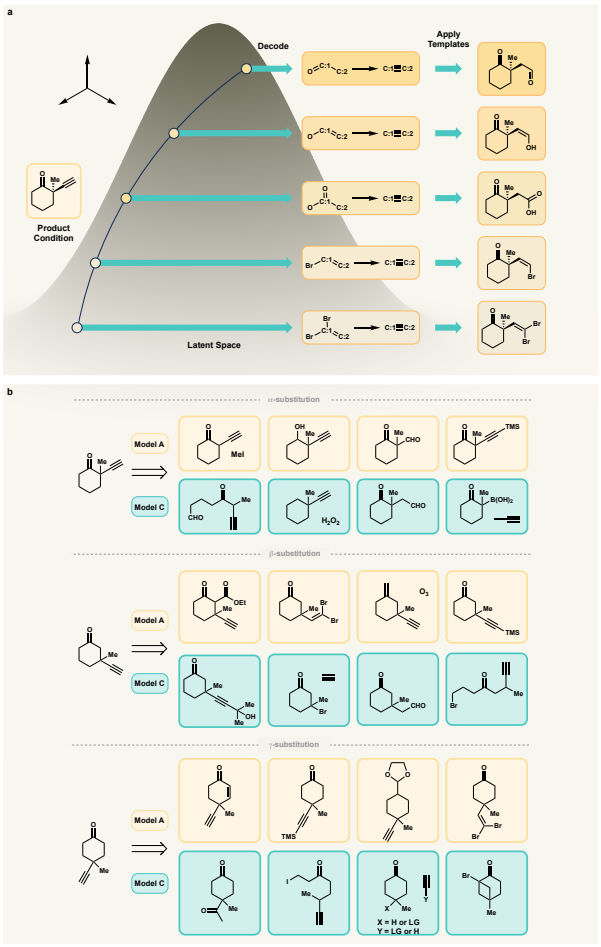


Figure 4: Interpolation of templates in the latent space of Model C and reactants from Modal A and Model C outputs. (a) The intermediates of the top and bottom latent representations are decoded. (b) Selected reactants for 2-, 3-, 4-substituted cyclohexanone derivatives as target compounds.

The distance between various chemical transformations in latent space can be used to interpolate between chemical reactions. This can be useful when searching for a reaction that could be an intermediate between two known chemical reactions. In Fig. 4a, an interpolation process is visualized. Initially, two reaction templates are selected, represented by the top and bottom templates and the latent vectors in the latent space. These templates serve as the starting points to explore the intermediates. This interpolation allowed the discovery of the templates corresponding to each of the latent vectors along the path between the two originals. It can be observed that the middle templates and reactants form a blending of the starting templates and reactants. This observation provides evidence that the latent space captures chemical information, showing the distance measure between various chemical transformations.

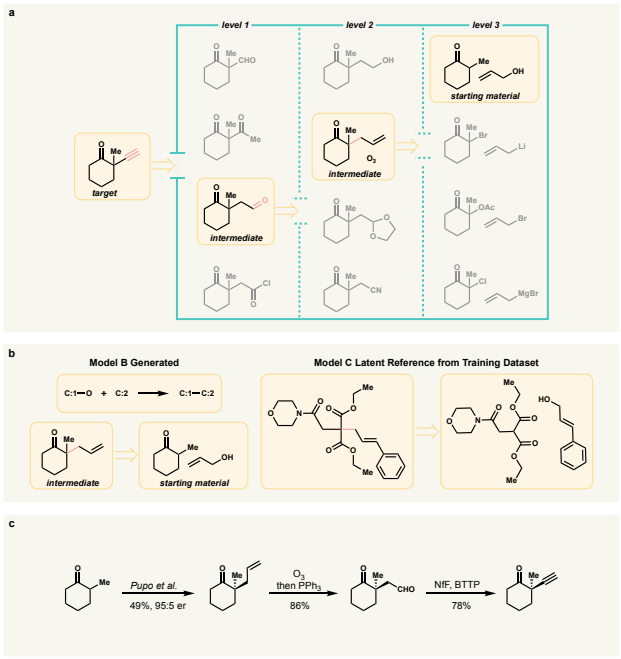


Figure 5: Retrosynthesis tree for compound **1** and its synthesis. (a) A synthetic route is selected from the retrosynthesis tree generated by Model B. (b) Reference found with Model C for the allylation step. (c) Experimental procedure for the selected route: i) 2-methylcyclohexanone (1 equiv.), allyl methyl carbonate (3 equiv.), $Pd_2(dba)_3$ (5 mol % Pd), *t*-BuXPhos (11 mol %), *R*-TRIP (10 mol %), 3Å MS, CyH, 45 ºC, 5 days, 49%; ii) O_3, CH_2Cl_2, - 78 ºC, then PPh_3 (2 equiv.), -78 ºC → rt, 16 h, 86%; iii) NfF (1.05 equiv.), BTTP (6 equiv.), DMF, - 30 ºC → rt, 19 h, 78%.

To illustrate the differences between Model A (deterministic) and Model C (generative sampling), both without reaction center information, the single-step predictions of 2-, 3- and 4-substituted cyclohexanone derivatives are examined. Based on the acquired results, representative precursors are selected for all three target molecules. As shown in Fig. 4b, Model A suggestions are primarily based on functional group interconversions and protection reactions. While Model C also proposes these transformations, diverse precursors and reactions are also proposed. These examples complement the intuitive bias of many synthetic chemists and point to areas of opportunity for creative development of novel chemical transforms.

2.4 Experimental Validation

Developing inexpensive, rapid and robust methods for the synthesis of bioactive molecules is one of the key goals in the pharmaceutical chemistry [58]. Herein, we utilized our Model B, chosen because of its high accuracy and reaction center embedding, for establishing the shortest route for

the synthesis of a target compound. Fig. 5a highlights how Model B can be used to navigate multiple options for retrosynthesis. Each level corresponds to a new prediction by the single-step model to reach an intermediate. Importantly this tool highlights the interactive nature of this model with a human expert who selects intermediates for further analysis.

Fig. 5b serves as a reference point derived from Model C. The left-hand side illustrates the allylation step that we employed in our synthesis. On the right-hand side, the reference is obtained by encoding the allylation template and the product labeled with the reaction center into Model C's latent space. This process allows us to identify the closest latent vectors from the training dataset, and that closest reference corresponds to the reaction shown on the right-hand side of Fig. 5b. Interestingly, the exact chemical transformation that was suggested had previously been conducted, but is not in the USPTO-FULL dataset. This highlights how our approach compliments other synthetic planning tools, such as Reaxys and SciFinder.

In order to synthesize the enantiomerically enriched target molecule, we applied the enantioselective Pd-catalyzed Tsuji-Trost allylation of a ketone and applied conditions recently reported by Pupo et al. [59]. The prior literature protocol for this substrate reported an enantiomeric ratio of 95.5:4.5. The allylated intermediate was treated with ozone in order to obtain the ketoaldehyde derivative in good yield (Fig. 5c). For the final step, a modified procedure by Boltukhina et al. [60] was applied to form the alkyne in 78% yield. The overall yield of our 3-step route is 33%, despite our route not having undergone process optimization. It should be stated that further process optimization is expected to improve the efficiency of this approach, although this proof-of-concept demonstrates the ability to develop step-efficient routes. This experimental procedure serves as evidence that the newly developed ML models can facilitate the development of synthetic routes for pharmaceutically significant molecules and enhance existing routes.

An alternative to the route presented in Fig 5c, an even shorter route to compound **1**, could be one entailing direct α-alkynylation of 2-methylcyclohexanone. Methods for direct introduction of an alkyne moiety next to a ketone are scarce and rely on substitution with electrophilic alkyne species (selected examples: [61, 62, 63, 64, 65, 66]). Most commonly used in modern organic chemistry are hypervalent iodine reagents such as Waser's or Ochiai's reagent [67]. While this method would furnish the target molecule in fewer synthetic steps, it would have to be followed by separation of two enantiomers since enantioselective α-alkynylation of ketones has not yet been reported.

3 Conclusions

In this work, a string-based approach for retrosynthesis planning is introduced, utilizing generative models to address the challenges posed by the vast chemical space and synthesis complexity. Specifically, this work introduces ***template generation*** as a new category in machine learning methods for computer-aided synthesis planning. Two types of generative models are developed, including deterministic generative models (Model A and Model B) and a sampling generative model that utilizes CKAE (Model C).

Model A and Model B are benchmarked on the USPTO-FULL dataset. Particularly, Model B can incorporate reaction center information, enabling the generation of templates that apply to the specified reacting sites. On the other hand, Model C represents a pioneering application of sampling method from latent space, capable of generating diverse reactions. The design of Model C defines distances between reactions, which allows Model C to identify the closest reference from the dataset for newly generated templates, making it a suitable tool for generating and validating a wide range of potential reactions.

This work presents two approaches for single-step synthetic planning, high-accuracy deterministic models and high-diversity sampling models. The capability of specifying reacting sites, the availability of relevant reaction references, and the successful results of experimental validations on an important pharmaceutically-relevant intermediate make the models valuable tools in guiding retrosynthetic analysis.

4 Methods

4.1 Training Details

10% dropout was applied to all attention matrices and embedding vectors. ADAM optimizer [68] was used with a learning rate of 5×10^{-5}. Gradient normalization [69] was set to 1.0. During training, each token in the input to the encoders has a 15% chance of being replaced by a mask token for Model A and Model B.

4.2 Model Architecture

Models A, B, and C each has 6 layers of transformer encoders and decoders as implemented in [70]. For Models A and B, 8 attention heads and an embedding size of 256 are used. For Model C, 16 attention heads and an embedding size of 512 are used.

The reaction center embeddings for Model B are achieved by adding the embedding of the reaction center token "*" at the specific position of the atoms similar to the concept of positional embeddings.

Model C is constructed based on the conditional kernel-elastic autoencoder model [52], with a 5120-dimensional latent space. The conditions are embeddings of target compounds and are also achieved by 6 layers of transformer encoders and 16 heads with an embedding size of 512 [70]. These embeddings are then compressed into 10 embedding vectors by a linear layer and concatenated with the input

embedding and the latent space. See SI Sec. 5.4 and Fig. 10 for more details and visualization of the architecture.

4.3 Beam Search

To derive multiple possible predictions, beam search [42] is used across all models. During decoding, the transformer decoder attends to the encoder output and the sequence that had been generated. The decoder outputs probabilities of all possible tokens for the next position in the sequence. Beam search maintains a fixed-size set of candidate sequences, the number that the method keeps is called the beam size B. The top B most probable sequences at each decoding step are selected to proceed to the next step of decoding until the stopping criteria of maximum allowed length is reached or an End Of Sequence (<EOS>) token is output.

For the Top-K accuracy test, beam search with a beam size of 50 is used during all decoding processes. At each decoding step, the model outputs the 50 most probable candidate tokens and continues the sequence until the stopping criteria is met.

The diversity of deterministic models is solely derived from the beam search process, as this type of model lacks a latent space for sampling. Consequently, generating novel reactions using a deterministic model through beam search can be challenging. In contrast, the sampling model, equipped with a latent space, can generate diverse and novel reactions more effectively.

Acknowledgments

VSB acknowledges a generous allocation of high-performance computing time from NERSC. The development of the methodology was supported by the NSF CCI grant (VSB, Award Number 2124511). The applications and experiments were supported by Boehringer Ingelheim.

Data Availability

The 50 most commonly seen spectators are obtained from the USPTO-Full reaction file on RDChiral GitHub Repository [29]. While the train-validation-test split of the USPTO-Full dataset is obtained from the GitHub repository of [42].

Author Contributions

The machine learning methods are developed by YS and HL, with equal contributions, under the guidance from VSB. The experimental validations are conducted by AMN and PZ, with equal contribution, under the guidance from TRN. The experimental design and execution were advised and supervised by SS, FB, JJS, and TRN. The initial draft of the manuscript was primarily written by YS, with contributions from all authors during the final draft preparation.

References

[1] E. J. Corey and W. Todd Wipke. Computer-Assisted Design of Complex Organic Syntheses: Pathways for molecular synthesis can be devised with a computer and equipment for graphical communication. *Science*, 166(3902):178–192, 1969.

[2] Ajit J. Thakkar. The coming of the computer age to organic chemistry: Recent approaches to systematic synthesis analysis. 39/1:3–18, 1973.

[3] Alan R. Katritzky, Minati Kuanar, Svetoslav Slavov, C. Dennis Hall, Mati Karelson, Iiris Kahn, and Dimitar A. Dobchev. Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *Chem. Rev.*, 110(10):5714–5789, 2010.

[4] W. Todd Wipke and W. Jeffrey Howe, editors. *Computer-Assisted Organic Synthesis*, volume 61. AMERICAN CHEMICAL SOCIETY, 1977.

[5] H. L. Gelernter, A. F. Sanders, D. L. Larsen, K. K. Agarwal, R. H. Boivie, G. A. Spritzer, and J. E. Searleman. Empirical Explorations of SYNCHEM: The methods of artificial intelligence are applied to the problem of organic synthesis route discovery. *Science*, 197(4308):1041–1049, 1977.

[6] Johannes Bauer, Eric Fontain, Dietmar Forstmeyer, and Ivar Ugi. Interactive generation of organic reactions by IGOR 2 and the PC-assisted discovery of a new reaction. *Tetrahedron Computer Methodology*, 1(2):129–132, 1988.

[7] S. Hanessian, Jonathan Franco, and Benoit Larouche. The psychobiological basis of heuristic synthesis planning - man, machine and the chiron approach. *Pure and Applied Chemistry*, 62(10):1887–1910, 1990.

[8] Orr Ravitz. Data-driven computer aided synthesis design. *Drug Discovery Today: Technologies*, 10(3):e443–e449, 2013.

[9] Anthony Cook, A. Peter Johnson, James Law, Mahdi Mirzazadeh, Orr Ravitz, and Aniko Simon. Computer-aided synthesis design: 40 years on. *WIREs Comput Mol Sci*, 2(1):79–107, 2012.

[10] Anders Bøgevig, Hans-Jürgen Federsel, Fernando Huerta, Michael G. Hutchings, Hans Kraut, Thomas Langer, Peter Löw, Christoph Oppawsky, Tobias Rein, and Heinz Saller. Route Design in the 21st Century: The IC *SYNTH* Software Tool as an Idea Generator for Synthesis Prediction. *Org. Process Res. Dev.*, 19(2):357–368, 2015.

[11] Ian W. Davies. The digitization of organic synthesis. *Nature*, 570(7760):175–181, 2019.

[12] Chris M. Gothard, Siowling Soh, Nosheen A. Gothard, Bartlomiej Kowalczyk, Yanhu Wei, Bilge Baytekin, and Bartosz A. Grzybowski. Rewiring Chemistry: Algorithmic Discovery and Experimental Validation of One-Pot Reactions in the Network of Organic Chemistry. *Angew Chem Int Ed*, 51(32):7922–7927, 2012.

[13] Mikołaj Kowalik, Chris M. Gothard, Aaron M. Drews, Nosheen A. Gothard, Alex Weckiewicz, Patrick E. Fuller, Bartosz A. Grzybowski, and Kyle J. M. Bishop. Parallel Optimization of Synthetic Pathways within the Network of Organic Chemistry. *Angew Chem Int Ed*, 51(32):7928–7932, 2012.

[14] Tomasz Klucznik, Barbara Mikulak-Klucznik, Michael P. McCormack, Heather Lima, Sara Szymkuć, Manishabrata Bhowmick, Karol Molga, Yubai Zhou, Lindsey Rickershauser, Ewa P. Gajewska, Alexei Toutchkine, Piotr Dittwald, Michał P. Startek, Gregory J. Kirkovits, Rafał Roszak, Ariel Adamski, Bianka Sieredzińska, Milan Mrksich, Sarah L.J. Trice, and Bartosz A. Grzybowski. Efficient Syntheses of Diverse, Medicinally Relevant Targets Planned by Computer and Executed in the Laboratory. *Chem*, 4(3):522–532, 2018.

[15] Barbara Mikulak-Klucznik, Patrycja Gołębiowska, Alison A. Bayly, Oskar Popik, Tomasz Klucznik, Sara Szymkuć, Ewa P. Gajewska, Piotr Dittwald, Olga Staszewska-Krajewska, Wiktor Beker, Tomasz Badowski, Karl A. Scheidt, Karol Molga, Jacek Mlynarski, Milan Mrksich, and Bartosz A. Grzybowski. Computational planning of the synthesis of complex natural products. *Nature*, 588(7836):83–88, 2020.

[16] Melissa A. Hardy, Bozhao Nan, Olaf Wiest, and Richmond Sarpong. Strategic elements in computer-assisted retrosynthesis: A case study of the pupukeanane natural products. *Tetrahedron*, 104:132584, 2022.

[17] A. Filipa De Almeida, Rui Moreira, and Tiago Rodrigues. Synthetic organic chemistry driven by artificial intelligence. *Nat Rev Chem*, 3(10):589–604, 2019.

[18] Thomas J. Struble, Juan C. Alvarez, Scott P. Brown, Milan Chytil, Justin Cisar, Renee L. DesJarlais, Ola Engkvist, Scott A. Frank, Daniel R. Greve, Daniel J. Griffin, Xinjun Hou, Jeffrey W. Johannes, Constantine Kreatsoulas, Brian Lahue, Miriam Mathea, Georg Mogk, Christos A. Nicolaou, Andrew D. Palmer, Daniel J. Price, Richard I. Robinson, Sebastian Salentin, Li Xing, Tommi Jaakkola, William. H. Green, Regina Barzilay, Connor W. Coley, and Klavs F. Jensen. Current and Future Roles of Artificial Intelligence in Medicinal Chemistry Synthesis. *J. Med. Chem.*, 63(16):8667–8682, 2020.

[19] Zipeng Zhong, Jie Song, Zunlei Feng, Tiantao Liu, Lingxiang Jia, Shaolun Yao, Tingjun Hou, and Mingli Song. Recent advances in artificial intelligence for retrosynthesis. 2023. arXiv:2301.05864 [physics, q-bio].

[20] Zhongliang Guo, Stephen Wu, Mitsuru Ohno, and Ryo Yoshida. Bayesian Algorithm for Retrosynthesis. *J. Chem. Inf. Model.*, 60(10):4474–4486, 2020.

[21] Hankook Lee, Sungsoo Ahn, Seung-Woo Seo, You Young Song, Eunho Yang, Sung-Ju Hwang, and Jinwoo Shin. RetCL: A Selection-based Approach for Retrosynthesis via Contrastive Learning. 2021. arXiv:2105.00795 [cs].

[22] Marwin H. S. Segler and Mark P. Waller. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chem. Eur. J.*, 23(25):5966–5971, 2017.

[23] Connor W. Coley, Luke Rogers, William H. Green, and Klavs F. Jensen. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Cent. Sci.*, 3(12):1237–1245, 2017.

[24] Shoichi Ishida, Kei Terayama, Ryosuke Kojima, Kiyosei Takasu, and Yasushi Okuno. Prediction and Interpretable Visualization of Retrosynthetic Reactions Using Graph Convolutional Networks. *J. Chem. Inf. Model.*, 59(12):5026–5033, 2019.

[25] Michael E. Fortunato, Connor W. Coley, Brian C. Barnes, and Klavs F. Jensen. Data Augmentation and Pretraining for Template-Based Retrosynthetic Prediction in Computer-Aided Synthesis Planning. *J. Chem. Inf. Model.*, 60(7):3398–3407, 2020.

[26] Hanjun Dai, Chengtao Li, Connor W. Coley, Bo Dai, and Le Song. Retrosynthesis Prediction with Conditional Graph Logic Network. 2020. arXiv:2001.01408 [cs, stat].

[27] Shuan Chen and Yousung Jung. Deep Retrosynthetic Reaction Prediction using Local Reactivity and Global Attention. *JACS Au*, 1(10):1612–1620, 2021.

[28] Philipp Seidl, Philipp Renz, Natalia Dyubankova, Paulo Neves, Jonas Verhoeven, Jörg K. Wegner, Marwin Segler, Sepp Hochreiter, and Günter Klambauer. Improving Few- and Zero-Shot Reaction Template Prediction Using Modern Hopfield Networks. *J. Chem. Inf. Model.*, 62(9):2111–2120, 2022.

[29] Connor W. Coley, William H. Green, and Klavs F. Jensen. RDChiral: An RDKit Wrapper for Handling Stereochemistry in Retrosynthetic Template Extraction and Application. *J. Chem. Inf. Model.*, 59(6):2529–2537, 2019.

[30] Chaochao Yan, Qianggang Ding, Peilin Zhao, Shuangjia Zheng, Jinyu Yang, Yang Yu, and Junzhou Huang. RetroXpert: Decompose Retrosynthesis Prediction like a Chemist. 2020. arXiv:2011.02893 [cs, q-bio].

[31] Chence Shi, Minkai Xu, Hongyu Guo, Ming Zhang, and Jian Tang. A Graph to Graphs Framework for Retrosynthesis Prediction. 2021. arXiv:2003.12725 [cs, stat].

[32] Vignesh Ram Somnath, Charlotte Bunne, Connor W. Coley, Andreas Krause, and Regina Barzilay. Learning Graph Models for Retrosynthesis Prediction. 2021. arXiv:2006.07038 [cs, stat].

[33] Xiaorui Wang, Yuquan Li, Jiezhong Qiu, Guangyong Chen, Huanxiang Liu, Benben Liao, Chang-Yu Hsieh, and Xiaojun Yao. RetroPrime: A Diverse, plausible and Transformer-based method for Single-Step retrosynthesis predictions. *Chemical Engineering Journal*, 420:129845, 2021.

[34] Yu Wang, Chao Pang, Yuzhe Wang, Junru Jin, Jingjie Zhang, Xiangxiang Zeng, Ran Su, Quan Zou, and Leyi Wei. Retrosynthesis prediction with an interpretable deep-learning framework based on molecular assembly tasks. *Nat Commun*, 14(1):6155, 2023.

[35] Weihe Zhong, Ziduo Yang, and Calvin Yu-Chian Chen. Retrosynthesis prediction using an end-to-end graph generative architecture for molecular graph editing. *Nat Commun*, 14(1):3009, 2023.

[36] Bowen Liu, Bharath Ramsundar, Prasad Kawthekar, Jade Shi, Joseph Gomes, Quang Luu Nguyen, Stephen Ho, Jack Sloane, Paul Wender, and Vijay Pande. Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models. *ACS Cent. Sci.*, 3(10):1103–1113, 2017.

[37] Pavel Karpov, Guillaume Godin, and Igor Tetko. A Transformer Model for Retrosynthesis. preprint, Chemistry, 2019.

[38] Benson Chen, Tianxiao Shen, Tommi S. Jaakkola, and Regina Barzilay. Learning to Make Generalizable and Diverse Predictions for Retrosynthesis. 2019. arXiv:1910.09688 [cs, stat].

[39] Alpha A. Lee, Qingyi Yang, Vishnu Sresht, Peter Bolgar, Xinjun Hou, Jacquelyn L. Klug-McLeod, and Christopher R. Butler. Molecular Transformer unifies reaction prediction and retrosynthesis across pharma chemical space. *Chem. Commun.*, 55(81):12152–12155, 2019.

[40] Kangjie Lin, Youjun Xu, Jianfeng Pei, and Luhua Lai. Automatic retrosynthetic route planning using template-free models. *Chem. Sci.*, 11(12):3355–3364, 2020.

[41] Shuangjia Zheng, Jiahua Rao, Zhongyue Zhang, Jun Xu, and Yuedong Yang. Predicting Retrosynthetic Reactions Using Self-Corrected Transformer Neural Networks. *J. Chem. Inf. Model.*, 60(1):47–55, 2020.

[42] Igor V. Tetko, Pavel Karpov, Ruud Van Deursen, and Guillaume Godin. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat Commun*, 11(1):5575, 2020.

[43] Seung-Woo Seo, You Young Song, June Yong Yang, Seohui Bae, Hankook Lee, Jinwoo Shin, Sung Ju Hwang, and Eunho Yang. GTA: Graph Truncated Attention for Retrosynthesis. *AAAI*, 35(1):531–539, 2021.

[44] Kelong Mao, Xi Xiao, Tingyang Xu, Yu Rong, Junzhou Huang, and Peilin Zhao. Molecular graph enhanced transformer for retrosynthesis prediction. *Neurocomputing*, 457:193–202, 2021.

[45] Mikołaj Sacha, Mikołaj Błaż, Piotr Byrski, Paweł Dąbrowski-Tumański, Mikołaj Chromiński, Rafał Loska, Paweł Włodarczyk-Pruszyński, and Stanisław Jastrzębski. Molecule Edit Graph Attention Network: Modeling Chemical Reactions as Sequences of Graph Edits. *J. Chem. Inf. Model.*, 61(7):3273–3284, 2021.

[46] Vipul Mann and Venkat Venkatasubramanian. Retrosynthesis prediction using grammar-based neural machine translation: An information-theoretic approach. *Computers & Chemical Engineering*, 155:107533, 2021.

[47] Umit V. Ucak, Taek Kang, Junsu Ko, and Juyong Lee. Substructure-based neural machine translation for retrosynthetic prediction. *J Cheminform*, 13(1):4, 2021.

[48] Eunji Kim, Dongseon Lee, Youngchun Kwon, Min Sik Park, and Youn-Suk Choi. Valid, Plausible, and Diverse Retrosynthesis Using Tied Two-Way Transformers with Latent Variables. *J. Chem. Inf. Model.*, 61(1):123–133, 2021.

[49] Ross Irwin, Spyridon Dimitriadis, Jiazhen He, and Esben Jannik Bjerrum. Chemformer: a pre-trained transformer for computational chemistry. *Mach. Learn.: Sci. Technol.*, 3(1):015022, 2022.

[50] Zipeng Zhong, Jie Song, Zunlei Feng, Tiantao Liu, Lingxiang Jia, Shaolun Yao, Min Wu, Tingjun Hou, and Mingli Song. Root-aligned SMILES: a tight representation for chemical reaction prediction. *Chem. Sci.*, 13(31):9023–9034, 2022.

[51] Umit V. Ucak, Islambek Ashyrmamatov, Junsu Ko, and Juyong Lee. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments. *Nat Commun*, 13(1):1186, 2022.

[52] Haote Li, Yu Shee, Brandon Allen, Federica Maschietto, and Victor Batista. Kernel-Elastic Autoencoder for Molecular Design. 2023.

[53] Jason ABBOTT, Joachim BROEKER, Jianwen CUI, Stephen W. Fesik, Andreas Gollner, Tim HODGES, Jale KAROLYI-OEZGUER, Andrew Little, Andreas Mantoulidis, Jason Phan, Dhruba Sarkar, Christian Alan Paul Smethurst, Qi Sun, Matthias Treu, and Alex Waterson. Annulated 2-amino-3-cyano thiophenes and derivatives for the treatment of cancer, U.S. Patent 2023/0212164 A1, 2023.

[54] Zhulin Tan, Xuan Ju, Hao Wu, Weitong Dong, Joyce C. Leung, Xiaowen Hou, Heewon Lee, Alice Granger, Joshua M. Paolillo, Susan V. DiMeo, Chengsheng Chen, Linglin Wu, Jon C. Lorenz, Max Sarvestani, Frederic Buono, Rogelio Frutos, Thomas G. Tampone, Xiaojun Huang, Gulin Zhang, Yuwen Wang, Earl Spinelli, Zhen Lei, and Jinhua J. Song. Development of a Scalable Synthesis toward

a KRAS G12C Inhibitor Building Block Bearing an All-Carbon Quaternary Stereocenter, Part 2: Asymmetric Synthesis via Shi Epoxidation. *Org. Process Res. Dev.*, 2024.

[55] Joyce C. Leung, Yibo Xu, Suttipol Radomkit, Jaehee Lee, Wan Shin Kim, Jonathan T. Reeves, Weitong Dong, Hwanjong Jang, Xiaowen Hou, Jon C. Lorenz, Xiaole Shao, Denis Byrne, Joe Johnson, Anthony Brundage, Clement Valentin, Phouvieng Beyer, Susan V. DiMeo, Bo Qu, Ruoshi Li, Max Sarvestani, and Jinhua J. Song. Development of a Scalable Synthesis toward a KRAS G12C Inhibitor Building Block Bearing an All-Carbon Quaternary Stereocenter, Part 1: From Discovery Route to Kilogram-Scale Production. *Org. Process Res. Dev.*, 2024.

[56] Timothy Newhouse, Phil S. Baran, and Reinhard W. Hoffmann. The economies of synthesis. *Chem. Soc. Rev.*, 38(11):3010, 2009.

[57] Juan Colberg, King Kuok (Mimi) Hii, and Stefan G. Koenig. Importance of Green and Sustainable Chemistry in the Chemical Industry: A joint virtual issue between *ACS Sustainable Chemistry & Engineering* and *Organic Process Research & Development*. *ACS Sustainable Chem. Eng.*, 10(26):8239–8241, 2022.

[58] M. D. Eastgate, M. A. Schmidt, and K. R. Fandrick. On the design of complex drug candidate syntheses in the pharmaceutical industry (vol 1, 0016, 2017). *Nature Reviews Chemistry*, 1(3), 2017.

[59] G. Pupo, R. Properzi, and B. List. Asymmetric catalysis with co2 : The direct -allylation of ketones. *Angewandte Chemie-International Edition*, 55(20):6099–6102, 2016.

[60] E. V. Boltukhina, A. E. Sheshenev, and I. M. Lyapkalo. Convenient synthesis of non-conjugated alkynyl ketones from keto aldehydes by a chemoselective one-pot nonaflation-base catalyzed elimination sequence. *Tetrahedron*, 67(30):5382–5388, 2011.

[61] A. S. Kende and P. Fludzinski. Chloroacetylenes as michael acceptors .2. direct ethynylation and vinylation of tertiary enolates. *Tetrahedron Letters*, 23(23):2373–2376, 1982.

[62] Y. Nishimura, R. Amemiya, and M. Yamaguchi. α-ethynylation reaction of ketones using catalytic amounts of trialkylgallium base. *Tetrahedron Letters*, 47(11):1839–1843, 2006.

[63] A. Utaka, L. N. Cavalcanti, and L. F. Silva. Electrophilic alkynylation of ketones using hypervalent iodine. *Chemical Communications*, 50(29):3810–3813, 2014.

[64] M. Wegener and S. F. Kirsch. The reactivity of 4-hydroxy- and 4-silyloxy-1,5-allenynes with homogeneous gold(i) catalysts. *Organic Letters*, 17(6):1465–1468, 2015.

[65] J. Wang, B. K. Hong, D. C. Hu, Y. Kadonaga, R. Y. Tang, and X. G. Lei. Protecting-group-free syntheses of -kaurane diterpenoids: [3+2+1] cycloaddition/cycloalkenylation approach. *Journal of the American Chemical Society*, 142(5):2238–2243, 2020.

[66] D. Jang, M. Choi, J. L. Chen, and C. Lee. Enantioselective total synthesis of (+)-garsubellin a. *Angewandte Chemie-International Edition*, 60(42):22735–22739, 2021.

[67] D. P. Hari, P. Caramenti, and J. Waser. Cyclic hypervalent iodine reagents: Enabling tools for bond disconnection via reactivity umpolung. *Accounts of Chemical Research*, 51(12):3212–3225, 2018.

[68] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2014.

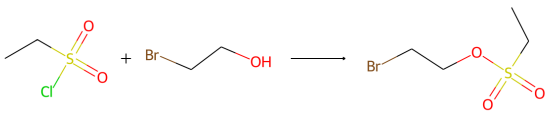
[69] Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In *International conference on machine learning*, pages 794–803. PMLR, 2018.

[70] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

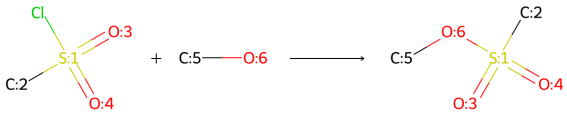
[71] D. C. Behenna and B. M. Stoltz. The enantioselective tsuji allylation. *Journal of the American Chemical Society*, 126(46):15044–15045, 2004.

5 Supplementary Information

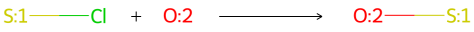
5.1 Reaction Template: RDChiral Template vs Site-Specific Template



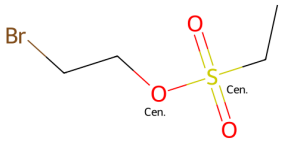
(a) A Reaction Example.



(b) RDChiral Template.



(c) Site-Specific Template.



(d) Center-Labeled Product.

Figure 6: Reaction template visualization. (a) Reaction SMARTS string: CCS(=O)(=O)Cl.OCCBr>>CCS(=O)(=O)OCCBr. the target compound is CCS(=O)(=O)OCCBr (b) RDChiral template: [C:5]-[O;H0;D2;+0:6]-[S;H0;D4;+0:1](-[C:2])(=[O;D1;H0:3])=[O;D1;H0:4]>>Cl-[S;H0;D4;+0:1](-[C:2])(=[O;D1;H0:3])=[O;D1;H0:4].[C:5]-[OH;D1;+0:6] (c)Site-specific template: [O:2]-[S:1]>>Cl-[S:1].[OH:2] (d) Target compound with reaction centers labeled: CC*(=O)(=O)*CCBr.

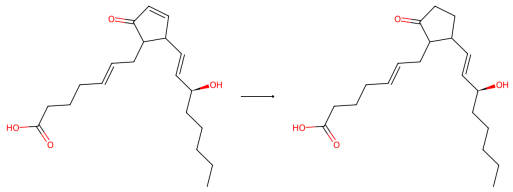
A reaction template is a concise representation of a chemical reaction, capturing the essential information about the substructure changes occurring during the reaction. In the context of retrosynthesis, reaction templates provide a valuable tool for generating potential pathways to synthesize target molecules. The format of reaction templates is typically represented as PRODUCT>>REACTANT in the retro-direction, indicating the transformation from the product back to the reactant. However, for the purpose of visualization in this work, we adopt the forward-direction format since it is more intuitive for understanding the reaction process. Fig. 6 illustrates an example reaction template visualization for the reaction SMARTS string in Fig. 6a.

Previous template-based methods have commonly utilized template extraction codes from the RDChiral repository to extract reaction templates. These templates include not only the reaction centers but also neighboring atoms and special functional groups, providing a comprehensive representation of the chemical transformations as demonstrated in Fig.6b. However, in our work, we have modified the template extraction process to focus exclusively on the reaction centers as depicted in Fig. 6c. We refer to these modified templates as site-specific templates since they specifically apply to the reacting sites (reaction centers) of the target compounds. To incorporate this specificity, we introduce additional input in the form of reaction center labels by replacing the atoms by "*" symbol. These labels indicate the specific sites within the target compound where the template should be applied. Fig. 6d showcases an example of a reaction center-labeled target molecule.

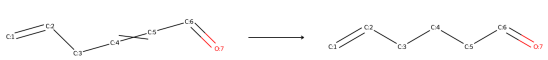
5.2 Specificity from Reaction Center-Labeled Products

An important aspect of our site-specific template approach is that the specificity is given by reaction center-labeled products. While the site-specific templates focus exclusively on the reaction centers, they lack the necessary information to determine the precise locations/atoms within the target compound where the template should be applied. Fig. 7 provides an illustrative example of how the reaction center-labeled target compound plays a crucial role in achieving specificity.

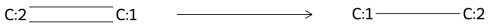
In Fig. 7a, we present a specific chemical reaction involving a carbon-carbon double bond reduction. The RDChiral template (Fig. 7b) offers a comprehensive representation of the transformation, including the reaction centers, neighboring atoms, and special functional groups. It is evident from the RDChiral template that the carbon-carbon double bond reduction occurs at a specific location within the molecule. However, when we consider the site-specific template



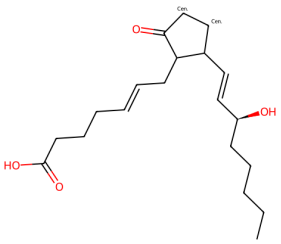
(a) A Reaction Example.



(b) RDChiral Template.



(c) Site-Specific Template.



(d) Center-Labeled Product.

Figure 7: An example when site-specific template requires target compound with reaction centers labeled to get the reaction SMARTS string: CCCCC[C@H](O)C=CC1C=CC(=O)C1CC=CCCCC(=O)O>>CCCCC[C@H](O)C=CC1CCC(=O)C1CC=CCCCC(=O).

(Fig. 7c), which solely captures the reaction centers, we observe a lack of specificity. Multiple carbon pairs in the product can potentially undergo the same transformation, resulting in ambiguity.

To resolve this ambiguity and introduce specificity, we utilize the reaction center-labeled target compound (Fig. 7d). By labeling the specific reaction centers within the product molecule, we indicate the precise locations where the site-specific template should be applied. In this example, the labeled reaction centers specify the carbon-carbon double bond that needs to be reduced. By combining the site-specific template and the labeled product molecule, we can obtain the accurate reaction SMARTS string that represents the desired chemical transformation. Note that the "RunReactants" function in RDKit still has to take in a site-specific template and a target compound without the "*". The center labels are used provide the atom numbers within a target compound SMILES that have to be changed. This ensures that correct reactants are selected from the output of the "RunReactants" function.

5.3 Template Generation Deterministic Model Architecture

Fig. 8a illustrates the model architecture of our deterministic approach (Model A). The model employs a transformer encoder to capture the relevant features and representations of the target molecule. Subsequently, these encoded features are fed into a transformer decoder, which generates the site-specific template and the product with reaction centers labeled. Models A and B both have 6 layers of Transformer encoders and decoders and 8 heads as implemented in [70] with an embedding size of 256.

In our example, referring back to Fig. 6, we consider the example reaction in Fig. 6a, the site-specific template in Fig. 6c, and the product with reaction centers labeled in Fig. 6d. The input of the deterministic model consists of the target compound of the reaction, CCS(=O)(=O)OCCBr in our example. The output of the deterministic model is structured in the following format: [O:2]-[S:1]>>Cl-[S:1].[OH:2]_CC*(=O)(=O)*CCBr where the underscore symbol "_" is a separator (also in the output). Here, the site-specific template is represented by [O:2]-[S:1]>>Cl-[S:1], indicating the breaking of the S-Cl bond and the formation of an S-O bond. The product with reaction centers labeled, CC*(=O)(=O)*CCBr, highlights the third and sixth atoms as the reaction centers using asterisks. With the generated site-specific template and the labeled product, we can reconstruct the original reaction depicted in Fig. 6a.

In addition, our deterministic generative model offers the flexibility to control the exact atoms participating in reactions by incorporating the relevant information within the encoder. This variation, denoted as Model B in Fig. 8b, introduces an embedding for the "*" token, representing the positions of the reacting atoms. Such positional information and the product SMILES input are passed in as model input. The output of Model B consists solely of site-specific templates, as the reaction centers are explicitly provided. This variant model allows researchers to customize the

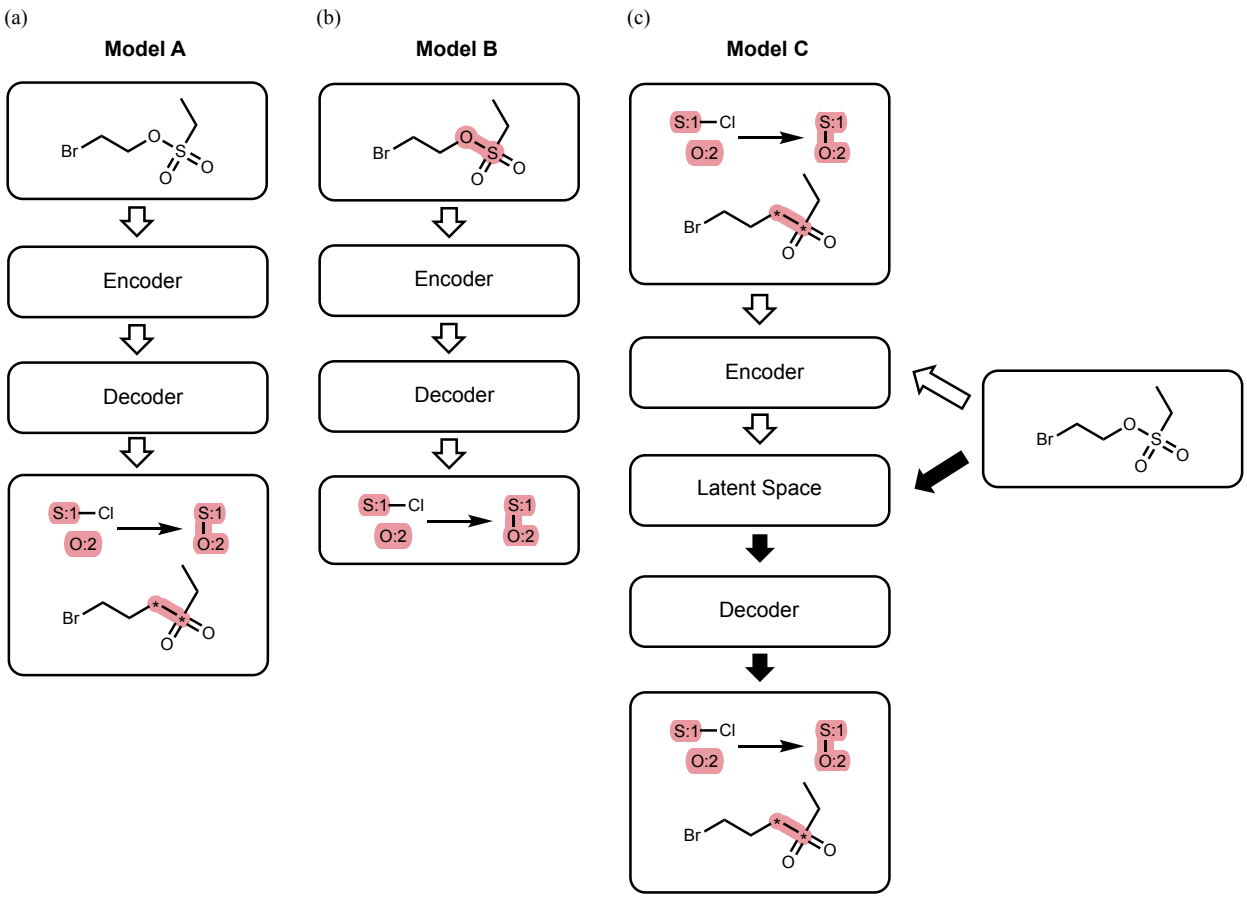


Figure 8: Model architectures of the generative models for retrosynthesis planning. (a) Model A is a deterministic generative model that takes in target products and output site-specific templates and labeled products. (b) Model B, a variant of Model A, incorporating positional embeddings for conditioning on specific reacting sites. (c) Model C is a sampling generative model based on the conditional kernel-elastic autoencoder (CKAE) approach.

reaction centers by specifying the atoms involved. Such unique feature allows for precise control over retrosynthetic disconnections/transformations.

Model A and Model B have 6 layers of Transformer encoders and decoders and 8 heads as implemented in [70] with an embedding size of 256. Fig. 9 has more detailed architectures with the dimension transformation at each layer. Note that the batch size dimension is ignored in the figure, but batch training and testing are used. The target compounds are first passed into the encoder where the embedding for the "*" token is added to specific positions of reaction centers for Model B. This is similar to the concept of positional embedding. Linear and Softmax layers after the decoder can then be used for predicting output probabilities as in [70].

5.4 Template Generation Sampling Model Architecture

Fig. 8c illustrates the model architecture of our sampling approach (Model C). Using the example in Fig. 6 again, by incorporating the product CCS(=O)(=O)OCCBr as the condition, the transformer encoder processes the site-specific template and the labeled product ([O:2]-[S:1]>>Cl-[S:1].[OH:2]_CC*(=O)(=O)*CCBr) at the same time and passes it through the latent space. The decoder is then tasked with reproducing the same input ([O:2]-[S:1]>>Cl-[S:1].[OH:2]_CC*(=O)(=O)*CCBr) as the output. This comprehensive encoding and decoding process where site-specific templates and center-labeled products are processed at the same time enables our attention model to capture essential information for single-step prediction, including the influence of functional groups on reactivity and regiose-

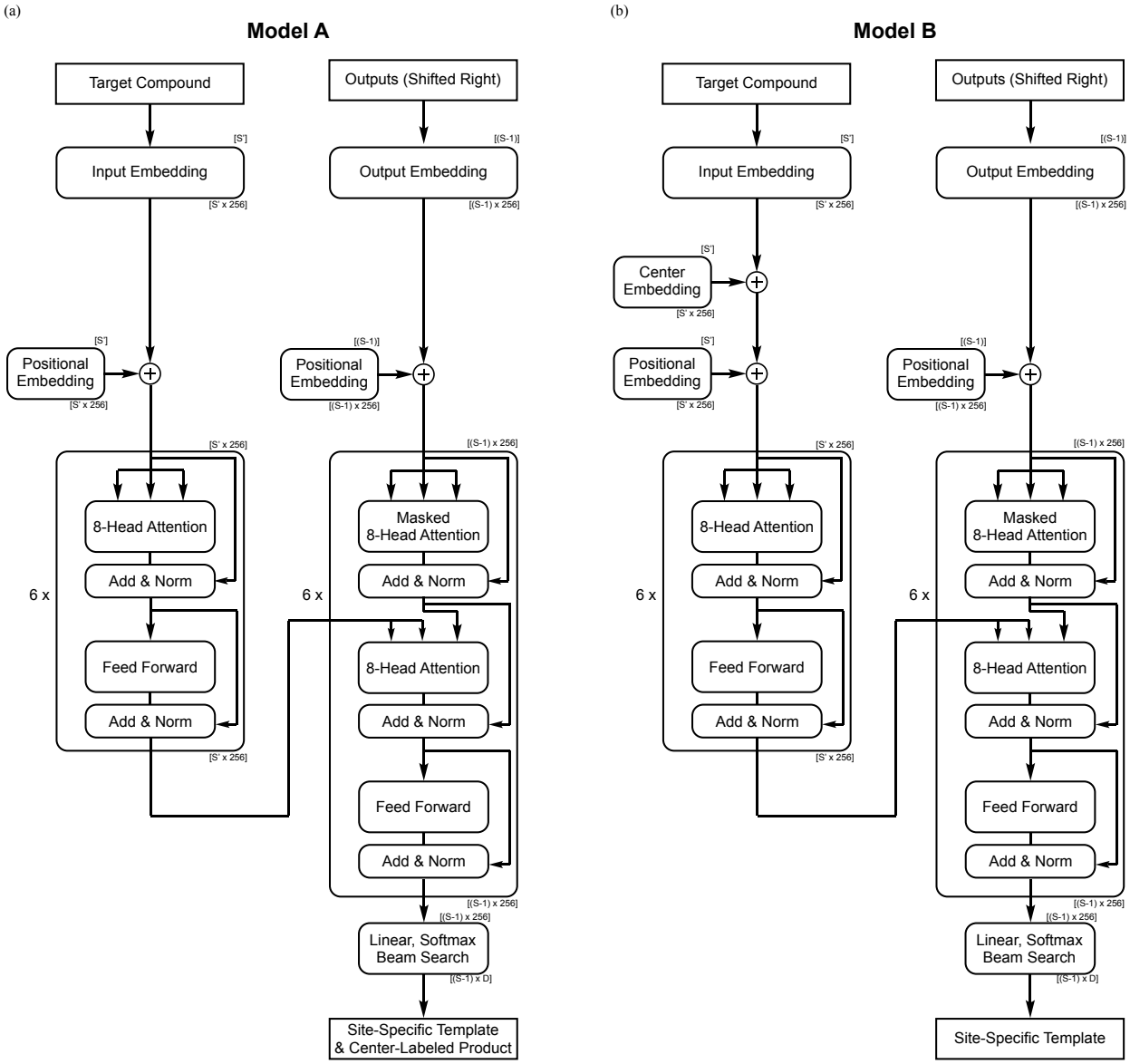


Figure 9: Model A and Model B detailed architectures and dimension transformations. (a) Model A detailed architecture. (b) Model B detailed architecture. The numbers in the brackets on the top and bottom of each box represent the input and output dimensions of the tensors for the boxed layers/components. S is the sequence length of outputs (site-specific templates and/or center-labeled products), S' is the sequence length of inputs (target compounds), and D is the size of the token dictionary. The $\frown$ symbol represents the concatenation of two tensors. The batch size dimension is ignored in this figure. The implementation of the transformer encoder and decoder layers are from [70] with an embedding dimension of 256. The reaction center embedding is a token embedding of the reaction character "*" which is added to the input embeddings at the specified positions.

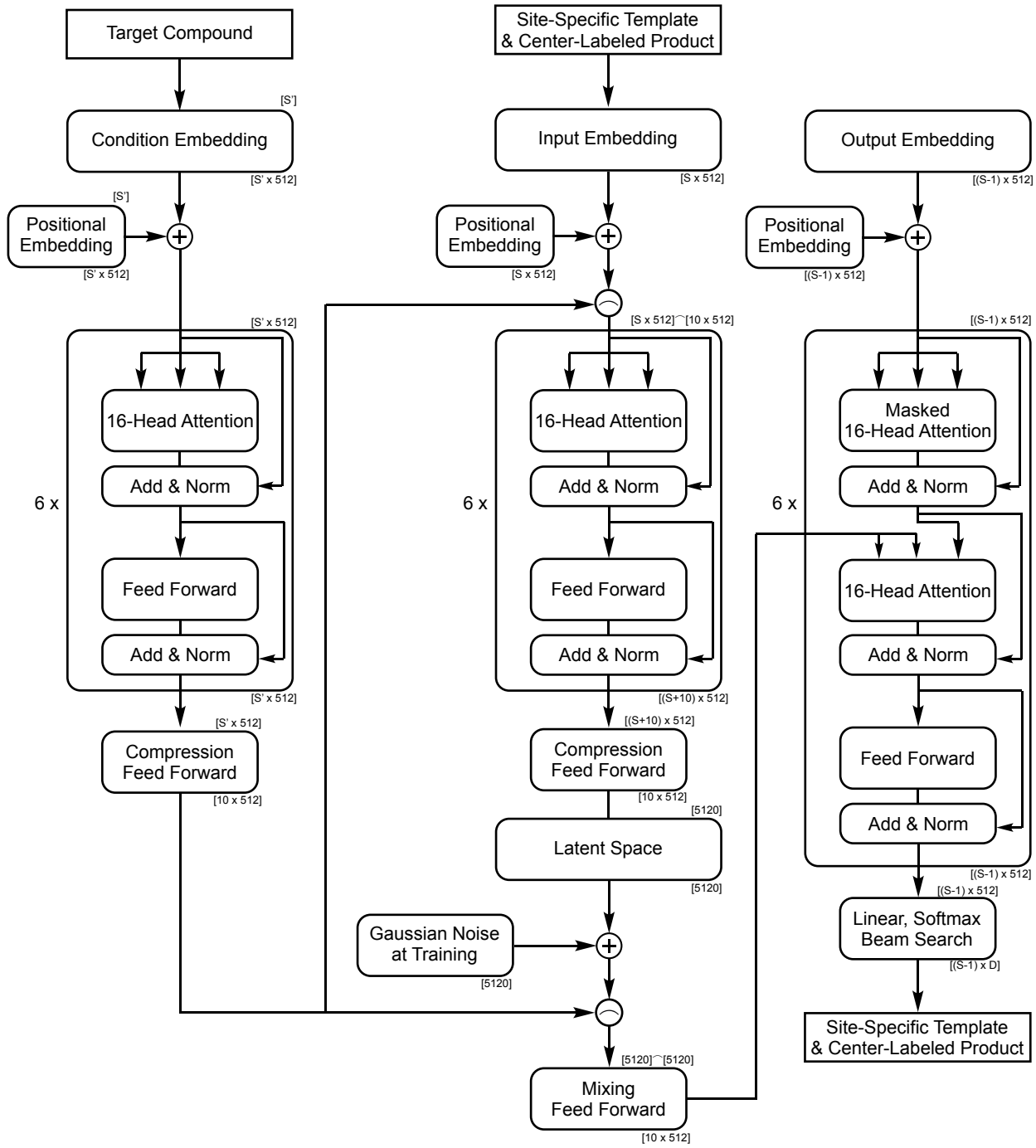


Figure 10: Model C detailed architecture and dimension transformations. The numbers in the brackets on the top and bottom of each box represent the input and output dimensions of the boxed layers/components. S is the sequence length of inputs (site-specific templates and center-labeled products), S' is the sequence length of conditions (target compounds), and D is the size of the token dictionary. The ⌒ symbol represents the concatenation of two tensors. The compression and mixing layers are fully-connected feed-forward networks. The batch size dimension is ignored in this figure. The implementation of the transformer encoder and decoder layers are from [70] with an embedding dimension of 512.

lectivity. During the sampling phase (filled arrows in Fig. 8c), given target products as conditions and random latent vectors, the model can generate a variety of templates and center-labeled products, leveraging the flexibility of the latent space and the conditioning on target molecules.

Model C has 6 layers of Transformer encoders and decoders as implemented in [70] and 16 heads with an embedding size of 512. The latent space and regularization loss is implemented in the CKAE paper [52]. Fig. 10 is a more detailed architecture with the dimension transformation at each layer. Note that the batch size dimension is ignored in the figure, but batch training and sampling are used. The target compounds are first passed into the condition encoder, which also has 6 layers of Transformer encoders and 16 heads with an embedding size of 512 (same as the input encoder). These embeddings of the conditions are then compressed into 10 embedding vectors by a linear layer. This compressed tensor is concatenated with the input embedding and the latent space (see the $\frown$ symbol in Fig. 10). In order to represent inputs in the latent space, another linear compression layer is used and the tensors are flattened to be 5120-dimensional latent vectors. The mixing layer takes in latent vectors concatenated with compressed conditions and transforms the dimensions back to 10 embedding vectors using a feed-forward layer. These tensors are then used for the cross-attention for the decoder. Linear and Softmax layers after the decoder can then be used for predicting output probabilities as in [70].

CKAE incorporates a specially designed loss function known as modified Maximum Mean Discrepancy (m-MMD), which enhances the generative power of the model. CKAE also utilizes a weighted cross-entropy loss, with the weights controlled by the δ and λ parameters, to improve the reconstruction capability. Additionally, CKAE presents exceptional correlations between outputs and given conditions. Further details on these loss functions and correlation results can be found in the CKAE paper [52].

While both deterministic and sampling models aim to accurately predict templates and center-labeled products, the sampling model offers additional capabilities. By incorporating a latent space and conditioning on target molecules, the sampling model has the ability to generate diverse and novel reactions. Leveraging the latent space, the model can sample reactions beyond the provided templates, resulting in a broader range of potential transformations. In contrast, deterministic models lack a latent space, limiting its ability to extrapolate and generate innovative reactions. The CKAE paper [52] showcases the superior interpolation and extrapolation capabilities of the sampling model, highlighting its capacity to sample a wider range of diverse reactions.

5.5 Other References for the Allylation Step

Fig. 11 presents a compilation of the top 5 references for the allylation step depicted in Fig. 5b. The site-specific templates are the same for these 5 references. Therefore, the products of these reactions are the primary determinant for the ranking (latent distance) in this particular case.

5.6 Encoder-Decoder Attention for Site-Specific Templates and Center-Labeled Product

In this section, we present a demonstration of our model's attention mechanism in Fig. 12, highlighting its ability to capture essential chemical information like functional groups and regioselectivity during the generation of reaction templates. We illustrate this through an example using our deterministic model, without the inclusion of reaction center information from positional embedding (Model A). The input of the model is the product in Fig. 12a (without the labels of reaction centers). The output of the model is the template shown in Fig. 12b along with the labeled product where the reaction centers are labeled in Fig. 12a. The corresponding reaction is shown in Fig. 12c where it is an amide bond formation and a removal of protection group for the ketone.

In Fig. 12d, we provide a encoder-decoder attention matrix from one of the attention heads, where the column labels on top represent the encoder input target compound SMILES, and the row labels on the right represent the decoder output site-specific template and center-labeled product. The reaction centers from the row labels are highlighted in yellow for encoder input for better visualization (the column labels). The presence of the ketone oxygen, originating from the protection group removal, significantly affects the output. Also, the matrix reveals that the influence on the template output extends beyond the reaction centers. Furthermore, the product input affects the labeled product portion of the output, resulting in a distinct diagonal pattern in the bottom of the matrix. These findings demonstrate the model's integration of critical chemical features that enhance its ability to generate accurate and relevant reaction templates.

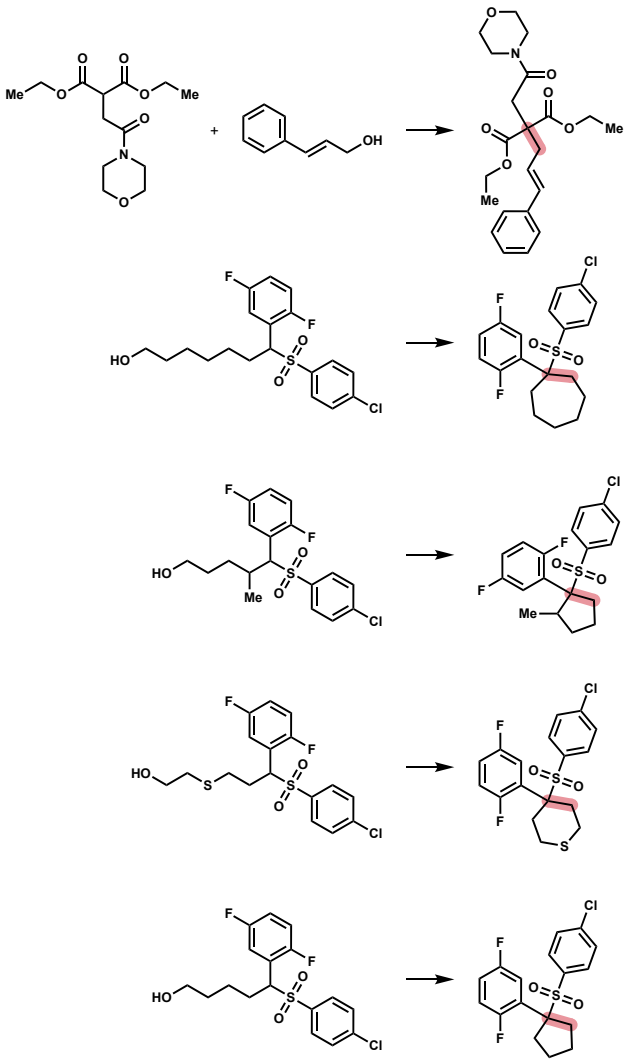
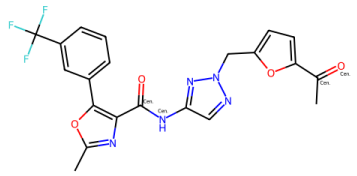
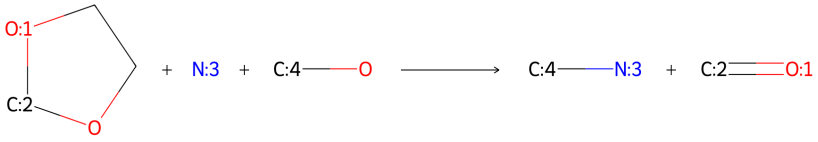


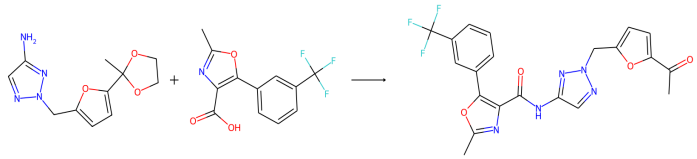
Figure 11: Top 5 references for the allylation step in the synthesis of 2-substituted cyclohexanone provided by Model C.



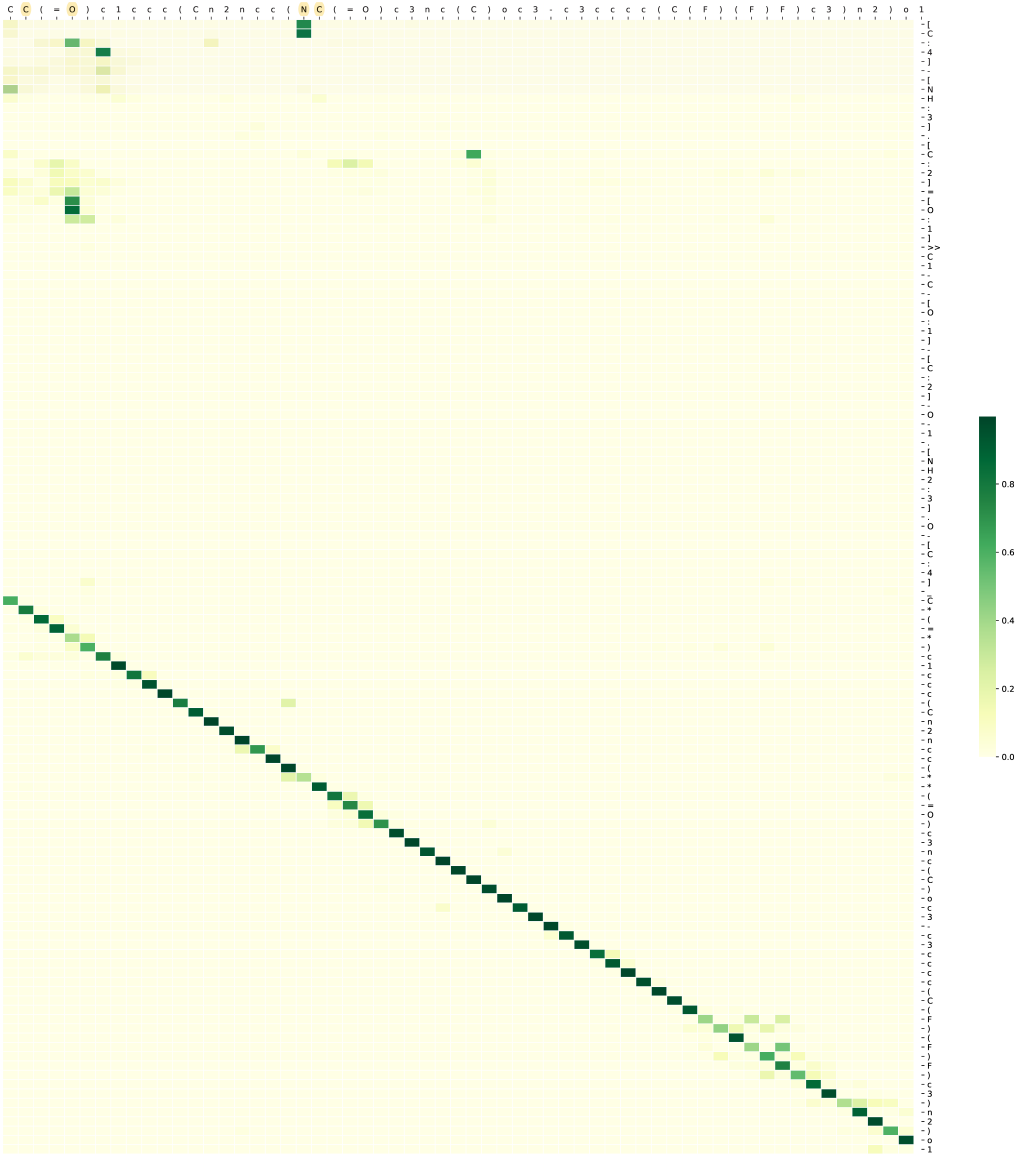
(a) Encoder Input Product (centers are from decoder output).



(b) Decoder Output Template.



(c) Corresponding Reaction.



(d) Encoder-Decoder Attention Matrix.

Figure 12: Visualization of the encoder-decoder-attention obtained from the product: CC(=O)c1ccc(Cn2ncc(NC(=O)c3nc(C)oc3-c3cccc(C(F)(F)F)c3)n2)o1.

5.7 Supplementary Tables

Table 1: USPTO Full Top-K accuracy (in %) comparison.

Method[a]	Model	Top-1	Top-3	Top-5	Top-10	Top-20	Top-50
Template-Based	GLN[26]	39.3			63.7		
	LocalRetro[27][b]	39.1	53.3	58.4	63.7	67.5	70.7
	Neuralsym[22][b]	42.7	58.7	63.4	67.9	70.8	72.1
Semi-Template	GraphRetro[32][b]	24.8	34.5	36.9	38.7	39.5	39.8
	RetroPrime[33][b]	45.8	61.6	63.9	70.3	71.2	72.6
	RetroExplainer[34]	51.4	70.7	74.7	79.2		
Template-Free	GTA[43][b]	46.6	52.5	57.9	63.3	67.2	70.4
	Tied-Transformer[48][b]	37.7	53.6	58.7	63.7	67.8	71.0
	MEGAN[45]	33.6			63.9		74.1
	Transformer[42][b]	44.7	61.1	66.0	70.7	74.1	76.2
	R-SMILES[50]	48.9	66.6	72.0	76.4	80.4	83.1
Template-Generation (This Work)	**Model A**	34.4	52.2	58.3	64.5	69.2	72.6
	Model A[c]	37.3	56.2	62.6	68.8	73.3	76.6
	Model B[d]	**48.1**	**67.8**	**72.6**	**76.4**	**78.7**	**80.2**
	Model B[c,d]	**51.1**	**71.6**	**76.4**	**80.0**	**82.0**	**83.3**

[a] Reactant-based methods are not included due to out-of-memory for USPTO-Full dataset. [b] Results obtained from [19]. [c] If the correct reactants contain one of the 50 most commonly seen spectators in the USPTO Full dataset, the reaction is removed from the test set. [d] Positional embedding of the reaction centers are included.

Table 2: Top-K accuracy (in %) for different number of reaction centers using Model B.

Maximum Reaction Centers	Top-1	Top-3	Top-5	Top-10	Top-20	Top-50	% of Test Data[a]
No Limit[b]	51.1	71.6	76.4	80.0	82.0	83.3	90.7%
5	53.0	74.1	79.0	82.6	84.6	86.0	85.2%
4	54.7	76.0	80.9	84.5	86.5	87.8	81.5%
3	57.8	78.9	83.8	87.3	89.2	90.4	75.1%
2	**61.2**	**81.7**	**86.6**	89.9	91.7	92.8	58.3%
1	60.1	80.8	86.3	**90.3**	**92.7**	**93.5**	11.6%

[a] Reactions containing 50 most common spectators as reactants are removed for all these cases, so no limit does not mean 100% of the test data. [b] The maximum reaction center count in test set is 18, while the maximum for training set is 19.

5.8 Experimental Section

5.8.1 General

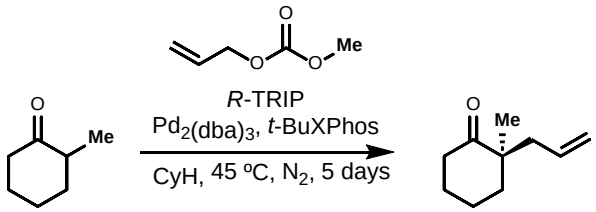
All reactions were carried out under an inert nitrogen atmosphere with dry solvents under anhydrous conditions unless otherwise stated. Stainless steel cannula or syringe was used to transfer solvent, and air- and moisture sensitive liquid reagents. Reactions were monitored by thin-layer chromatography (TLC) carried out on 0.25 mm Merck silica gel plates (60F254) using UV light as the visualizing agent and potassium permanganate and an acidic solution of *p*-anisaldehyde, on SiO_2 as developing agents. Flash column chromatography employed SiliaFlash® P60 (40-60 μm, 230-400 mesh) silica gel purchased from SiliCycle, Inc.

Materials: $Pd_2(dba)_3$ was purchased from Strem. *t*-BuXPhos was purchased from Sigma Aldrich. *R*-TRIP ((*R*)-3,3'-bis(2,4,6-triisopropylphenyl)-1,1'-binaphthyl-2,2'-diyl hydrogenphosphate) was purchased from AmBeed. NfF (nonafluorobutanesulfonyl fluoride) was purchased from Oakwood Products, Inc. BTTP (*tert*-butylimino-tri(pyrrolidino)phosphorane) was purchased from Sigma Aldrich. Dry cyclohexane and DMF were purchased from Sigma Aldrich. All other reagents were used as received without further purification, unless otherwise stated.

Instrumentation: All new compounds were characterized by means of ^{1}H NMR, ^{13}C NMR, FT-IR, and HR-MS. Optical rotations were measured on Polarimeter Rudolph Autopol IV at 589 nm, 22°C. Data are reported as: $[\alpha]_D^t$, concentration (c in g/100 mL) and solvent. The absolute configurations were determined by comparison between the

measured optical rotations and the reported values in literature. Copies of the ^{1}H- and ^{13}C-NMR spectra can be found after experimental procedures. NMR spectra were recorded using a Varian 400 MHz NMR spectrometer. All ^{1}H NMR data are reported in δ units, parts per million (ppm), and were calibrated relative to the signals for residual chloroform (7.26 ppm) in deuterochloroform ($CDCl_3$). All ^{13}C NMR data are reported in ppm relative to $CDCl_3$ (77.2 ppm) and were obtained with ^{1}H decoupling unless otherwise stated. The following abbreviations or combinations thereof were used to explain the multiplicities: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. All IR spectra were taken on an FT-IR Shimadzu IRTracer-100 (thin film. High resolution mass spectra (HRMS) were recorded on a LC-MS Shimadzu 9030 Quadrupole Time-of-Flight high resolution mass spectrometer.

5.8.2 Synthesis



(*R*)-2-Allyl-2-methylcyclohexan-1-one [71]. For the synthesis of (*R*)-2-allyl-2-methylcyclohexan-1-one, procedure reported by Pupo et. al was applied [59]. To a flame-dried microwave vial equipped with a magnetic stir bar were added $Pd_2(dba)_3$ (75.4 mg, 0.0824 mmol, 5 mol % Pd), *t*-BuXPhos (154 mg, 0.362 mmol, 11 mol %), *R*-TRIP (248 mg, 0.329 mmol, 10 mol %), 3Å molecular sieves (3.3 g), cyclohexane (33 mL), 2-methylcyclohexanone (400 μL, 3.29 mmol, 1 equiv) after which allyl methyl carbonate (1.12 mL, 9.88 mmol, 3 equiv) was added dropwise. The reaction vial was capped and placed into a pre-heated 45 ºC oil bath and stirred for 5 days. The reaction mixture was removed from the oil bath and cooled to ambient temperature before filtering through a short pad of celite. The celite was washed with Et_2O (30 mL) and the solution was concentrated under reduced pressure by rotary evaporation. Purification by flash column chromatography on silica gel (Et_2O/pentane = 1:99 to 5:95) afforded the product (245 mg, 49%) as a colorless oil.

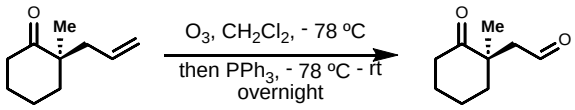
R_f: 0.50 (EtOAc/Hex= 1:9)

$[\alpha]_D^{22}$: 40.96 (c = 0.166, CH_2Cl_2, lit. 49.60 for c = 2.9, ee 95%)

^{1}H NMR (400 MHz, $CDCl_3$): δ 5.75 – 5.63 (m, 1H), 5.10 – 4.98 (m, 2H), 2.43 – 2.31 (m, 3H), 2.27 – 2.18 (m, 1H), 1.91 – 1.65 (m, 5H), 1.63 – 1.54 (m, 1H), 1.07 (s, 3H)

^{13}C NMR (100 MHz, $CDCl_3$): δ 215.5, 133.9, 118.0, 48.6, 42.1, 38.9, 38.7, 27.5, 22.8, 21.2

IR (cm^{-1}): 3076, 2932, 2864, 1704, 1640, 1451, 1437, 1428, 1314, 1124, 993, 912, 613



(*R*)-2-(1-Methyl-2-oxocyclohexyl)acetaldehyde. To a round-bottom flask equipped with a magnetic stir were added 2-allyl-2-methylcyclohexan-1-one (80 mg, 0.526 mmol, 1.0 equiv) and CH_2Cl_2 (5.5 mL). The solution was cooled to - 78 ºC in an acetone/dry ice bath and ozone was bubbled through until the solution turned blue. The excess ozone was removed by bubbling oxygen thorough the solution until it turned clear. To the solution was added PPh_3 (275 mg, 1.05 mmol, 2 equiv) at - 78 ºC and the reaction mixture was allowed to warm to room temperature and the stirring was continued for 16 h. The solution was concentrated under the reduced pressure by rotary evaporation. Purification by flash column chromatography on silica gel (Et_2O/pentane = 1:9 to 3:7) afforded the product (70 mg, 86%) as a colorless oil.

R_f: 0.67 (EtOAc/Hex= 2:8)

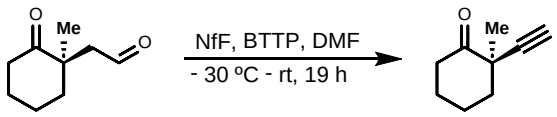
$[\alpha]_D^{22}$: -64.52 (c = 0.155, CH_2Cl_2)

^{1}H NMR (400 MHz, $CDCl_3$): δ 9.77 (t, J = 2.1 Hz, 1H), 2.64 – 2.32 (m, 4H), 2.07 – 1.96 (m, 1H), 1.88 – 1.68 (m, 5H), 1.28 (s, 3H)

^{13}C NMR (100 MHz, $CDCl_3$): δ 214.3, 201.5, 51.7, 47.8, 39.1, 38.4, 27.1, 23.7, 21.1

IR (cm^{-1}): 2934, 2862, 1703, 1462, 1448, 1431, 1178, 1159, 1128, 1080, 1042, 1014, 978, 934, 901, 868, 797, 735, 573

HRMS (*m/z*): calc'd. for $C_9H_{15}O_2^+$: 155.1067; detected: 155.1069



(*R*)-2-ethynyl-2-methylcyclohexan-1-one. For the synthesis of (*R*)-2-ethynyl-2-methylcyclohexan-1-one, procedure reported by Boltukhina et. al was applied [60]. To a flame-dried round-bottom flask equipped with a magnetic stir bar were added 2-(1-methyl-2-oxocyclohexyl)acetaldehyde (309 mg, 2.00 mmol, 1 equiv), NfF (380 μL, 2.10 mmol, 1.05 equiv) and dry DMF (2 mL). The solution was cooled to - 30 ºC in an acetonitrile/dry ice bath and the BTTP base (3.68 mL, 12.02 mmol, 6 equiv) was added dropwise. The reaction mixture was allowed to warm to room temperature and the stirring was continued for 19 h. The reaction was quenched with saturated solution of NH_4Cl (15 mL) and extracted with Et_2O (3 × 15 mL). The organic solution was washed with water (4 × 15 mL) and brine (15 mL) and dried over anhydrous Na_2SO_4. The solution was concentrated under the reduced pressure by rotary evaporation. Purification by flash column chromatography on silica gel (Et_2O/pentane = 1:99 to 3:97) afforded the product (212 mg, 78%) as a colorless oil.

R_f: 0.48 (EtOAc/Hex= 1/9)

$[\alpha]_D^{22}$: 274.14 (c = 0.116, CH_2Cl_2)

^{1}H NMR (400 MHz, $CDCl_3$): δ 3.01 – 2.90 (m, 1H), 2.38 – 2.25 (m, 2H), 2.16 – 2.03 (m, 3H), 1.78 – 1.49 (m, 3H), 1.31 (s, 3H)

^{13}C NMR (100 MHz, $CDCl_3$): δ 208.8, 86.4, 72.7, 45.8, 42.1, 38.6, 28.2, 23.3, 22.4

IR (cm^{-1}): 3290, 3271, 2982, 2936, 2864, 2112, 1717, 1462, 1448, 1427, 1375, 1333, 1312, 1277, 1258, 1232, 1155, 1121, 1111, 1090, 1063, 982, 905, 851, 829, 737, 688, 636, 569, 536, 519, 511, 498

HRMS(*m/z*): calc'd. for $C_9H_{13}O^+$: 137.0961; detected: 137.0964.