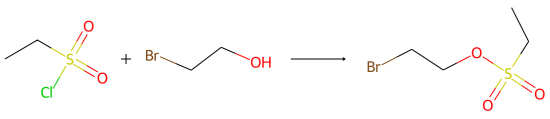
SUPPLEMENTARY INFORMATION FOR "SITE-SPECIFIC TEMPLATE GENERATIVE APPROACH FOR RETROSYNTHETIC PLANNING"

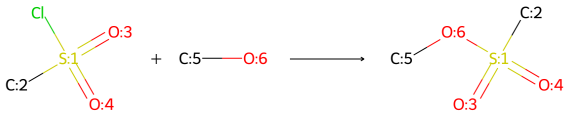
Yu Shee[1, 3], **Haote Li**[1, 3], **Pengpeng Zhang**[1], **Andrea M. Nikolic**[1], **Wenxin Lu**[1], **H. Ray Kelly**[2], **Vidhyadhar Manee**[2], **Sanil Sreekumar**[2], **Frederic G. Buono**[2], **Jinhua J. Song**[2], **Timothy R. Newhouse**[1, 4], and **Victor S. Batista**[1, 4]

[1]Department of Chemistry, Yale University, New Haven, Connecticut 06511, USA
[2]Chemical Development, Boehringer Ingelheim Pharmaceuticals Inc, Ridgefield, Connecticut 06877, USA
[3]These authors contributed equally to the method development.
[4]Corresponding authors: timothy.newhouse@yale.edu and victor.batista@yale.edu

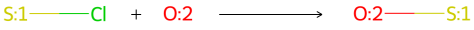
1 Reaction Template: RDChiral Template vs Site-Specific Template



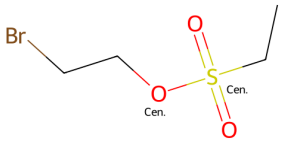
(a) A Reaction Example.



(b) RDChiral Template.



(c) Site-Specific Template.



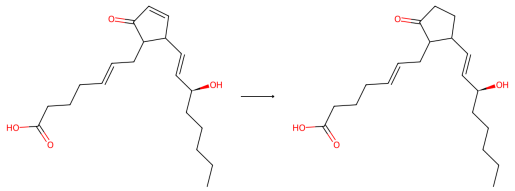
(d) Center-Labeled Product.

Supplementary Figure 1: Reaction template visualization. (a) Reaction SMILES 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. Supplementary Fig. 1 illustrates an example reaction template visualization for the reaction SMILES string in Supplementary Fig. 1a.

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 Supplementary Fig. 1b. However, in our work, we have modified the template extraction process to focus exclusively on the reaction centers as depicted in Supplementary Fig. 1c. 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. Supplementary Fig. 1d showcases an example of a reaction center-labeled target molecule.

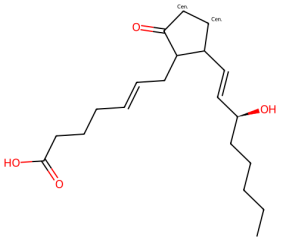
2 Specificity from Reaction Center-Labeled Products



(a) A Reaction Example.

(b) RDChiral Template.

(c) Site-Specific Template.



(d) Center-Labeled Product.

Supplementary Figure 2: An example when site-specific template requires target compound with reaction centers labeled to get the reaction SMILES string. (a) Reaction SMILES string: CCCCC[C@H](O)C=CC1C=CC(=O)C1CC=CCCCC(=O)O>>CCCCC[C@H](O)C=CC1CCC(=O)C1CC=CCCCC(=O)O. The target compound is CCCCC[C@H](O)C=CC1CCC(=O)C1CC=CCCCC(=O)O. (b) RDChiral template: [C:1]=[C:2]-[C:3]-[CH2;D2;+0:4]-[CH2;D2;+0:5]-[C:6]=[O;D1;H0:7]>>[C:1]=[C:2]-[C:3]-[CH;D2;+0:4]=[CH;D2;+0:5]-[C:6]=[O;D1;H0:7] (c) Site-specific template: [CH2:1]-[CH2:2]>>[CH:1]=[CH:2]. (d) Target compound with reaction centers labeled: CCCCC[C@H](O)C=CC1**C(=O)C1CC=CCCCC(=O)O.

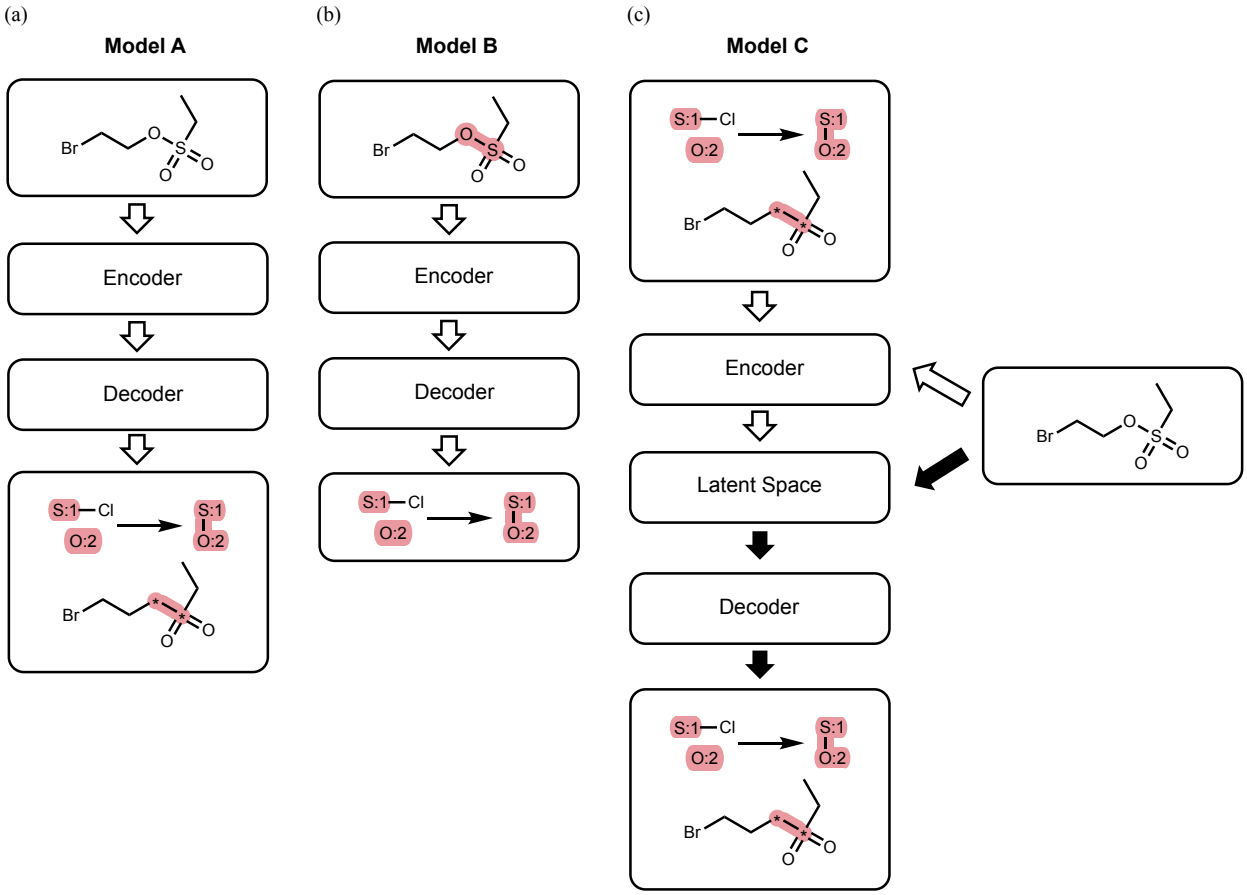
An important aspect of our site-specific template approach is that the specificity is given by reaction center-labeled products. While the site-specific templates focus exclusively on the reaction centers, they lack the necessary information to determine the precise locations/atoms within the target compound where the template should be applied. Supplementary Fig. 2 provides an illustrative example of how the reaction center-labeled target compound plays a crucial role in achieving specificity.

In Supplementary Fig. 2a, we present a specific chemical reaction involving a carbon-carbon double bond reduction. The RDChiral template (Supplementary Fig. 2b) 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 (Supplementary Fig. 2c), 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 (Supplementary Fig. 2d). 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 SMILES 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.

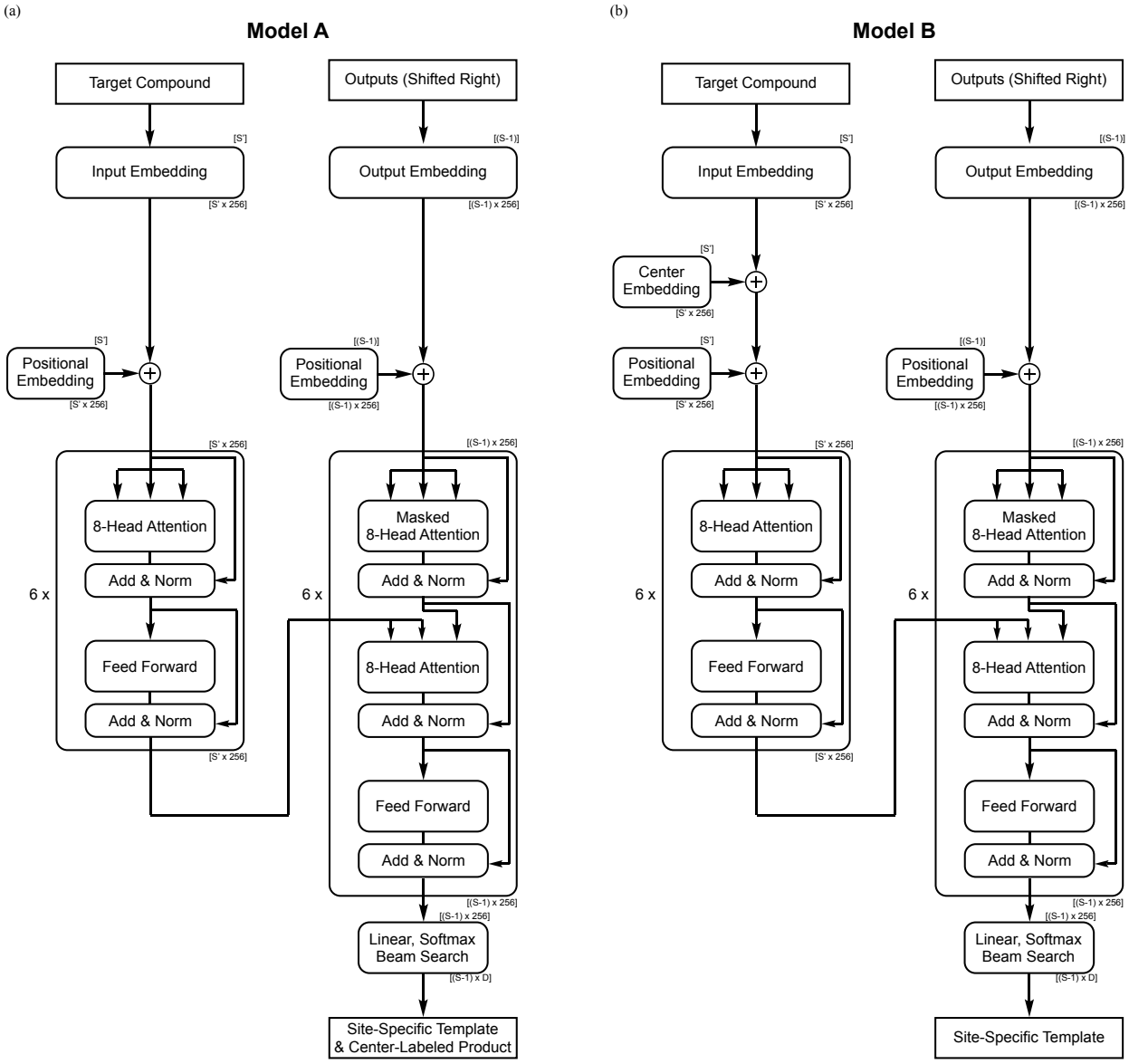
3 Template Generation Deterministic Model Architecture



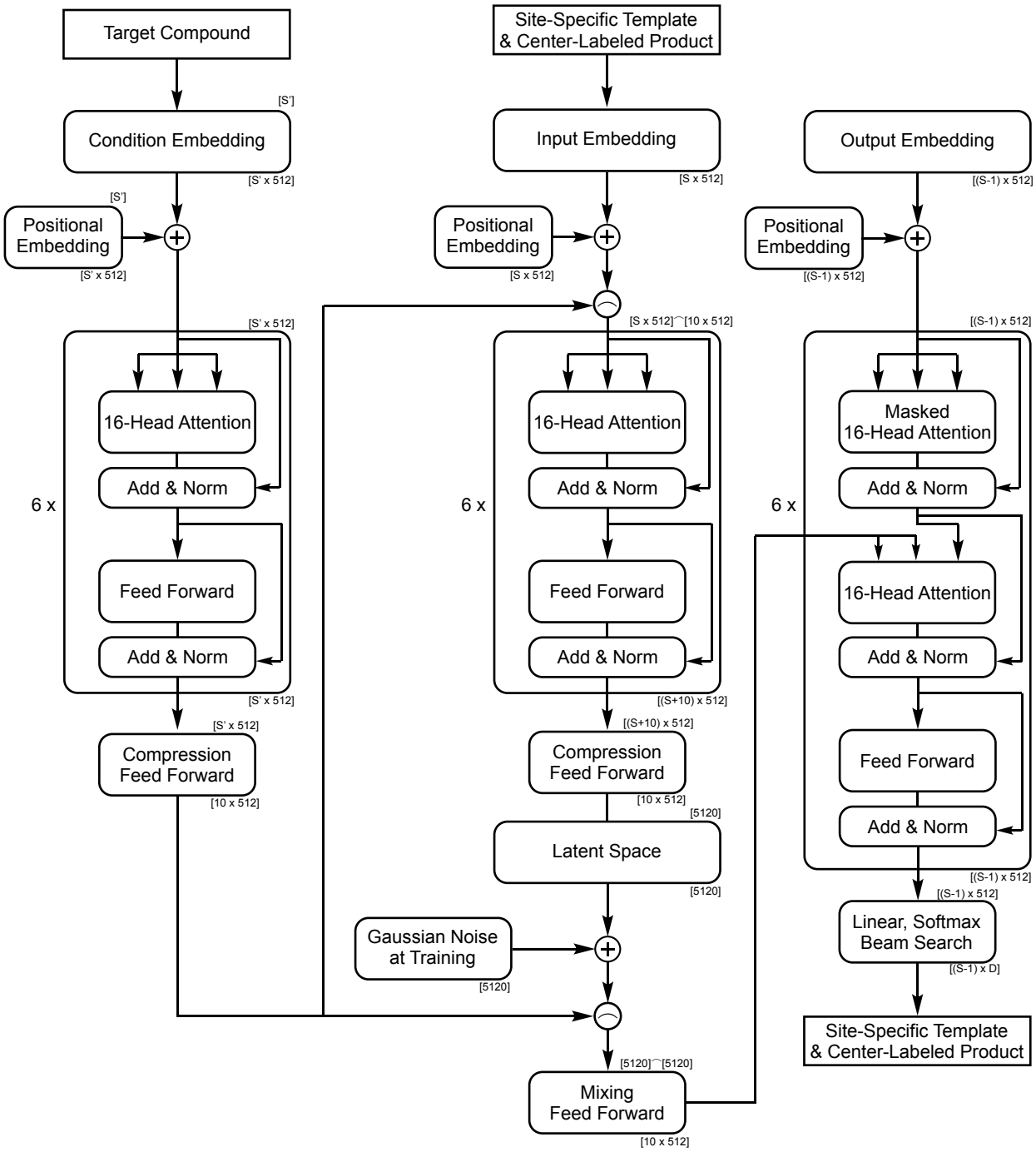
Supplementary Figure 3: Model architectures of the generative models for retrosynthesis planning. (a) Model A is a deterministic generative model that takes in target products and output site-specific templates and labeled products. (b) Model B, a variant of Model A, incorporating positional embeddings for conditioning on specific reacting sites. (c) Model C is a sampling generative model based on the conditional kernel-elastic autoencoder (CKAE) approach.

Supplementary Fig. 3a 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 [1] with an embedding size of 256.

In our example, referring back to Supplementary Fig. 1, we consider the example reaction in Supplementary Fig. 1a, the site-specific template in Supplementary Fig. 1c, and the product with reaction centers labeled in Supplementary Fig. 1d. 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



Supplementary Figure 4: 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 [1] 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.



Supplementary Figure 5: 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 [1] with an embedding dimension of 512.

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 Supplementary Fig. 1a.

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 Supplementary Fig. 3b, 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 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 [1] with an embedding size of 256. Supplementary Fig. 4 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 [1].

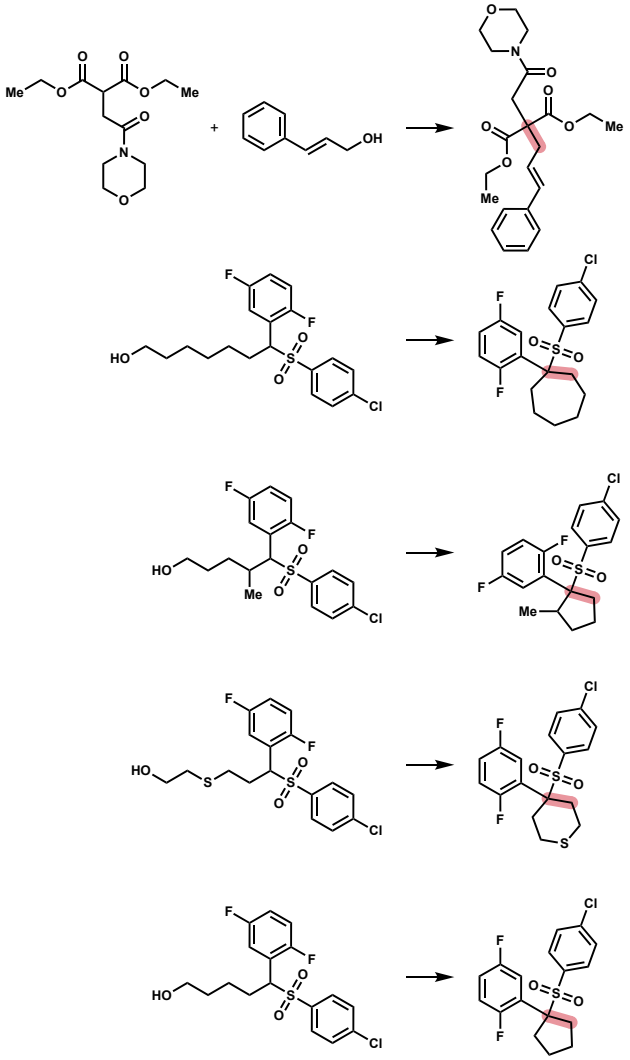
4 Template Generation Sampling Model Architecture

Supplementary Fig. 3c illustrates the model architecture of our sampling approach (Model C). Using the example in Supplementary Fig. 1 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 regioselectivity. During the sampling phase (filled arrows in Supplementary Fig. 3c), 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 [1] and 16 heads with an embedding size of 512. The latent space and regularization loss is implemented in the CKAE paper [2]. Supplementary Fig. 5 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 Supplementary Fig. 5). 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 [1].

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 [2].

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 [2] showcases the superior interpolation and extrapolation capabilities of the sampling model, highlighting its capacity to sample a wider range of diverse reactions.



Supplementary Figure 6: Top 5 references for the allylation step in the synthesis of 2-substituted cyclohexanone provided by Model C.

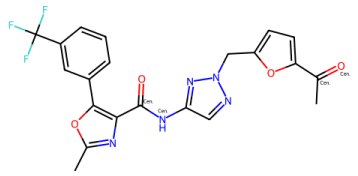
5 Other References for the Allylation Step

Supplementary Fig. 6 presents a compilation of the top 5 references for the allylation step depicted in main text Fig. 6a. 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.

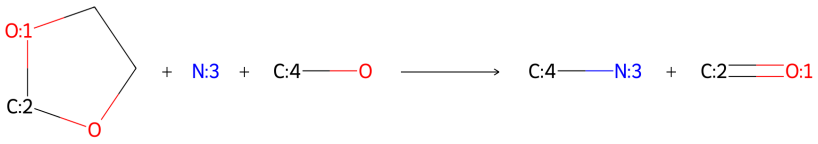
6 Encoder-Decoder Attention for Site-Specific Templates and Center-Labeled Product

In this section, we present a demonstration of our model's attention mechanism in Supplementary Fig. 7, 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 Supplementary Fig. 7a (without the labels of reaction centers). The output of the model is the template shown in Supplementary Fig. 7b along with the labeled product where the reaction centers are labeled in Supplementary Fig. 7a. The corresponding reaction is shown in Supplementary Fig. 7c where it is an amide bond formation and a removal of protection group for the ketone.

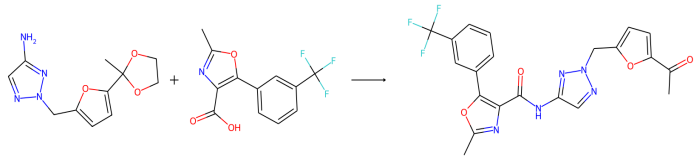
In Supplementary Fig. 7d, we provide a encoder-decoder attention matrix from one of the attention heads, where the column labels on top represent the encoder input target compound SMILES, and the row labels on the right represent the decoder output site-specific template and center-labeled product. The 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.



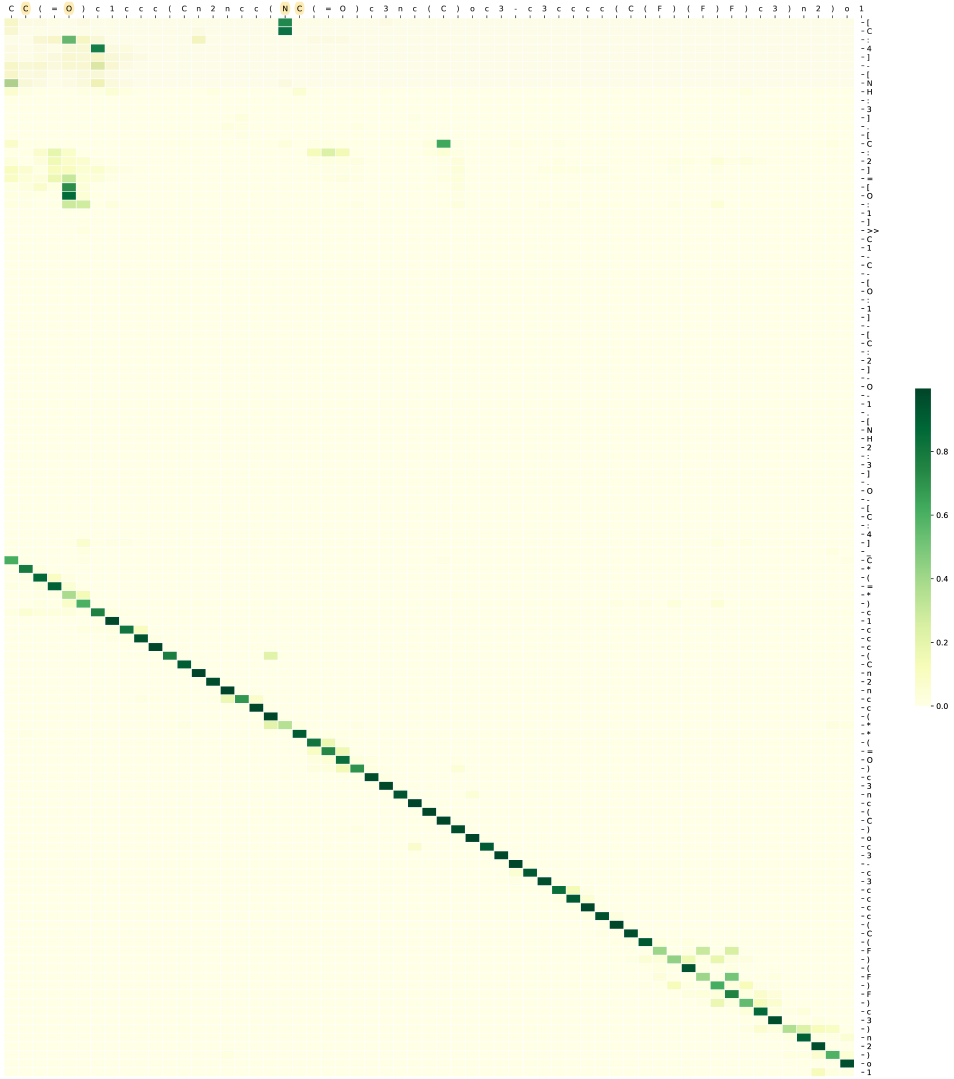
(a) Encoder Input Product (centers are from decoder output).



(b) Decoder Output Template.



(c) Corresponding Reaction.



(d) Encoder-Decoder Attention Matrix.

Supplementary Figure 7: Visualization of Model A's encoder-decoder-attention obtained from the product: CC(=O)c1ccc(Cn2ncc(NC(=O)c3nc(C)oc3-c3cccc(C(F)(F)F)c3)n2)o1. (a) Input product, where the centers are shown but not included in the input. (b) Output template. The center-labeled product is also in the output, but the centers are labeled in (a). (c) The reaction obtained by applying the template to the input product. (d) Encoder-decoder attention matrix, with column labels representing input tokens and row labels representing output tokens.

7 Template Generation Extends Beyond the Available Templates

Supplementary Table 1: Number of unique precursors and novel reaction rate for different models, beam sizes, and numbers of latent vectors, calculated using a random sample of 1000 molecules from the test set. Standard deviations are indicated after the ± symbol.

Model	Beam Size	# of Latent Vector	# of Unique Precursors	Novelty
Model A	10	NA	8.25 ± 2.21	35.2% ± 29.5%
Model A	50	NA	36.51 ± 9.72	57.2% ± 22.3%
Model B	10	NA	8.76 ± 3.40	61.7% ± 35.5%
Model B	50	NA	41.51 ± 16.08	83.5% ± 20.2%
Model C	1	10	3.31 ± 1.85	49.8% ± 36.0%
Model C	1	50	11.38 ± 4.48	56.2% ± 23.4%
Model C	5	10	10.25 ± 5.56	57.7% ± 25.9%

The number of unique precursors and the novel reaction rate for different models and settings are reported in Supplementary Table 1. The number of unique precursors refers to how many unique sets of reactants are obtained from the reaction templates, as a reaction can have multiple reactants. SSTs do not carry information about neighboring atoms, but these templates can still reconstruct the atom mapping of the entire reaction strings. Therefore, whole reaction strings with atom mappings can be obtained from SSTs and CLPs, and these strings can then be transformed into default RDChiral templates (radius equals one, including special functional groups, explicit degrees, and explicit numbers of hydrogens) [3]. To calculate the novel reaction rate, SSTs are transformed back to default RDChiral templates and checked if they are in the training set in the default RDChiral template format.

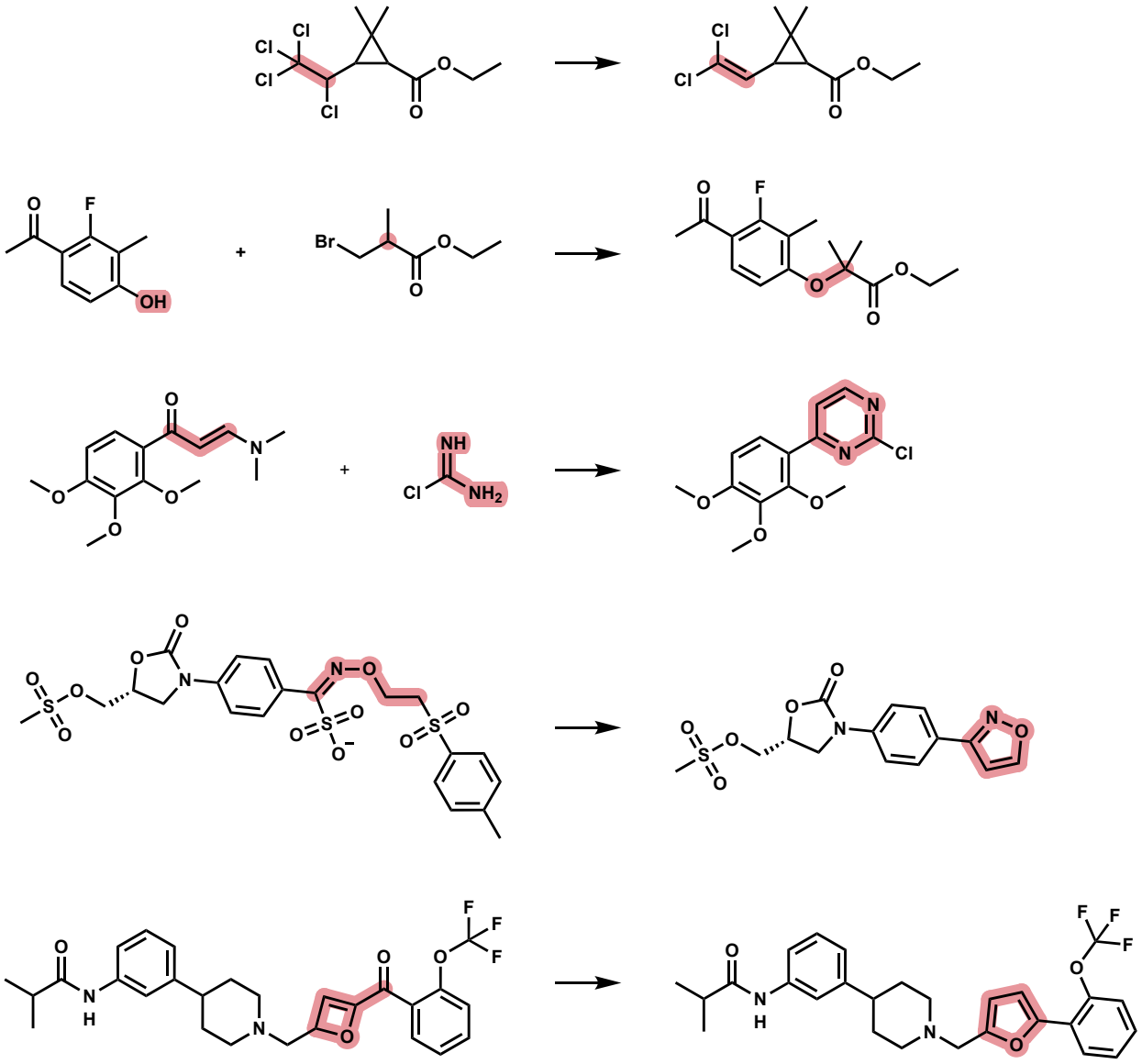
The table shows that new reaction transformations are proposed about half of the time. When beam sizes or the number of latent vectors increase, the novelty also increases. More unintuitive reaction transformations may be generated by Model C if more latent vectors are sampled, but in this experiment, the total number of generated templates is set to be 10 or 50. For example, a beam size of 5 and 10 latent vectors would give 50 pairs of SSTs and CLPs. Also, the CLPs must match the atoms and bond connections provided in the SSTs, so even though the generated SST SMARTS strings are mostly valid, some generated SSTs could be discarded. It can also be seen that Model B has more unique reactants and higher novelty. This could be due to the fact that reaction centers are already provided to Model B, and Model B only has to generate unique and novel templates based on the correct reaction centers from the test set (Model B does not need to identify possible reaction centers like Model A or Model C). This table demonstrates that template generation methods have the capability to extrapolate outside of the training set and extend beyond the available templates or predefined reaction rules.

Note that these novel templates may represent known transformations with slightly different neighboring atoms for the reaction centers, or they could be completely new reaction transformations. To clarify, **known methods** refer to previously reported and documented procedures used to achieve specific chemical transformations, involving well-established reaction conditions, reagents, and mechanisms widely recognized in the scientific literature. **Expansion of the scope of a known method** involves slight changes to reaction conditions, reagents, or substrates—such as altering neighboring atoms or functional groups—while maintaining the core transformation. This broadens the method's applicability to a wider range of substrates or conditions without fundamentally changing the chemical conversion type. A **novel transformation** refers to a completely new chemical process or type of chemical conversion not previously reported, involving the development of new procedures, techniques, or reaction conditions (**novel methods**) that result in unique chemical outcomes or new chemical products.

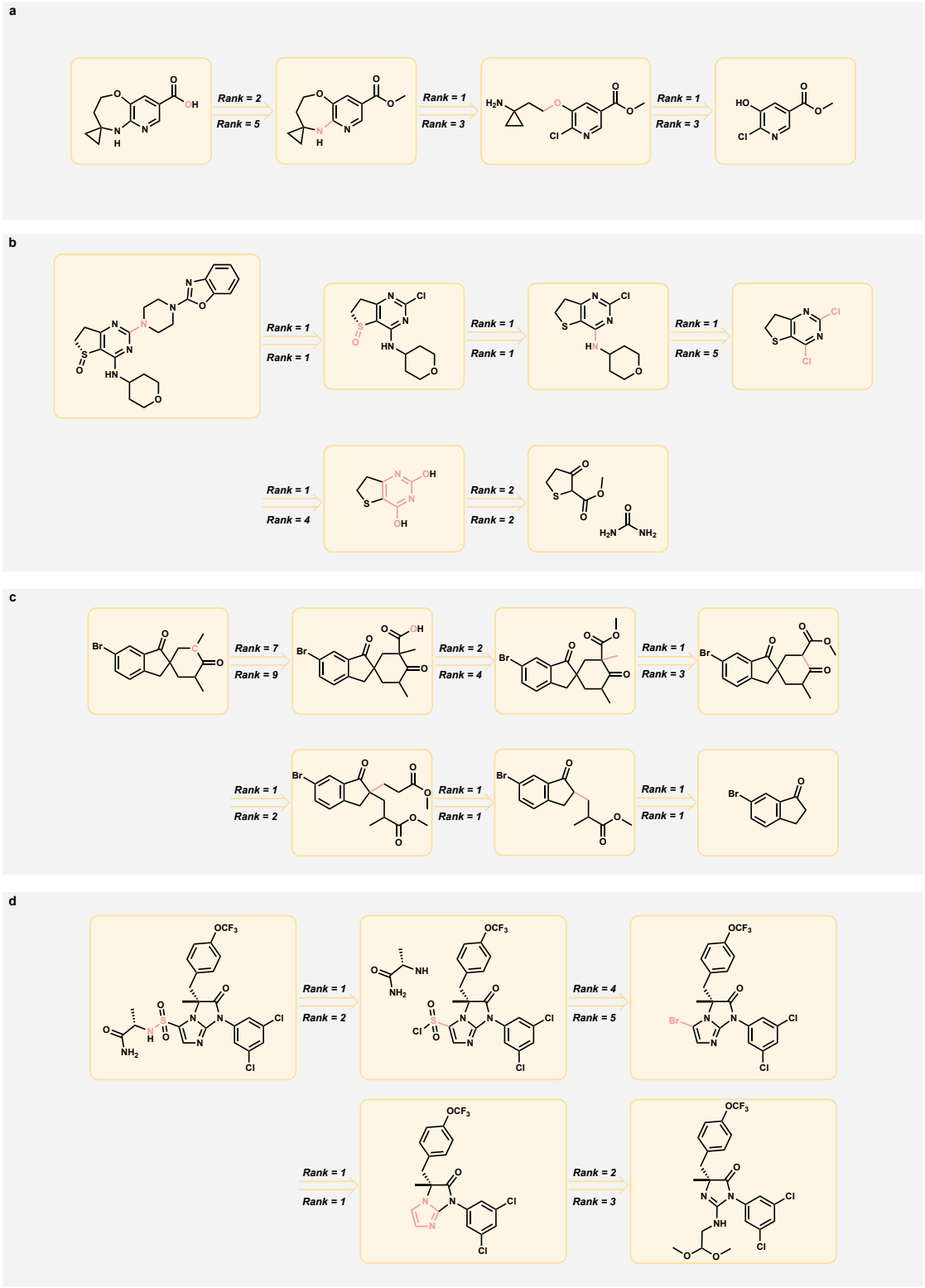
The novel templates proposed by the models would require new synthetic methods and expansions of the scope of known methods to achieve the proposed transformations. Supplementary Fig. 8 presents several examples of reactions from novel reaction templates obtained from the experiment in Supplementary Table 1 for Model A. The first three reactions represent examples of expansions of the scope of known methods, while the last two reactions are novel transformations requiring the development of new synthetic methods.

8 More Examples of Retrosynthetic Planning

In the main text, the retrosynthetic planning of a cyclohexanone, with a quaternary chiral center in the α-position containing an alkyne moiety, is shown. In this section, we demonstrate that various literature-reported routes for pharmaceutical compounds can be reconstructed using Model B. In Supplementary Fig. 9, each pharmaceutical compound route is annotated with rankings derived from beam scores (above the arrows) and synthetic accessibility



Supplementary Figure 8: Reactions from templates outside of training set in default RDChiral template format provided by Model A. Reaction centers are highlighted in red.



Supplementary Figure 9: Synthetic routes for several pharmaceutical compounds reconstructed by Model B. The ranks above the arrows are from beam scores while the ranks below the arrows are from SA scores. Reaction centers are highlighted in red. (a) synthetic route of a oxazepane pyridine intermediate [4]. (b) synthetic route of a PDE4 inhibitor [5]. (c) synthetic route of an intermediate of a BACE inhibitor [6]. (d) synthetic route of a LFA-1 inhibitor [7].

(SA) scores (below the arrows) and the reaction centers are highlighted in red. The examples include an oxazepane pyridine intermediate [4], a PDE4 inhibitor [5], an intermediate of a BACE inhibitor [6], and an LFA-1 inhibitor [7]. These routes are all identified by Model B with high ranks.

9 Supplementary Tables

Supplementary Table 2: 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[8]	39.3			63.7		
	LocalRetro[9][b]	39.1	53.3	58.4	63.7	67.5	70.7
	Neuralsym[10][b]	42.7	58.7	63.4	67.9	70.8	72.1
Semi-Template	GraphRetro[11][b]	24.8	34.5	36.9	38.7	39.5	39.8
	RetroPrime[12][b]	45.8	61.6	63.9	70.3	71.2	72.6
	RetroExplainer[13]	51.4	70.7	74.7	79.2		
Template-Free	GTA[14][b]	46.6	52.5	57.9	63.3	67.2	70.4
	Tied-Transformer[15][b]	37.7	53.6	58.7	63.7	67.8	71.0
	MEGAN[16]	33.6			63.9		74.1
	Transformer[17][b]	44.7	61.1	66.0	70.7	74.1	76.2
	R-SMILES[18]	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.

Supplementary Table 3: 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.

10 Experimental Section

10.1 General

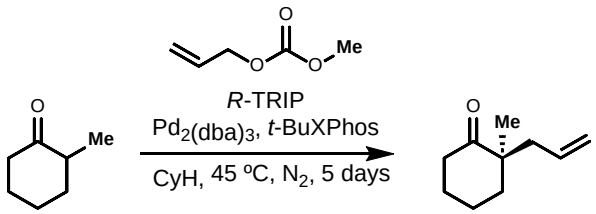
All reactions were carried out under an inert nitrogen atmosphere with dry solvents under anhydrous conditions unless otherwise stated. Stainless steel cannula or syringe was used to transfer solvent, and air- and moisture sensitive liquid reagents. Reactions were monitored by thin-layer chromatography (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.

10.2 Synthesis



(*R*)-2-Allyl-2-methylcyclohexan-1-one [20]. For the synthesis of (*R*)-2-allyl-2-methylcyclohexan-1-one, procedure reported by Pupo et. al was applied [21]. 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.

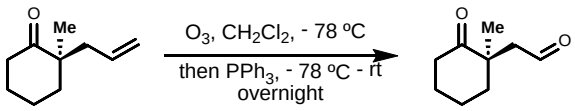
R_f: 0.50 (EtOAc/Hex= 1:9)

$[\alpha]_D^{22}$: 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)

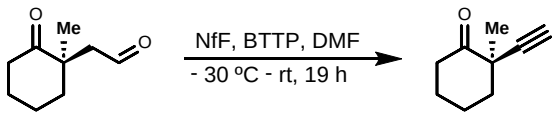
$[\alpha]_D^{22}$: -64.52 (c = 0.155, CH_2Cl_2)

^{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 [22]. 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)

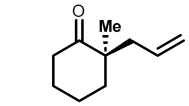
$[\alpha]_D^{22}$: 274.14 (c = 0.116, CH_2Cl_2)

^{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)

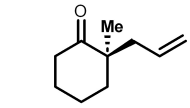
^{13}C NMR (100 MHz, $CDCl_3$): δ 208.8, 86.4, 72.7, 45.8, 42.1, 38.6, 28.2, 23.3, 22.4

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

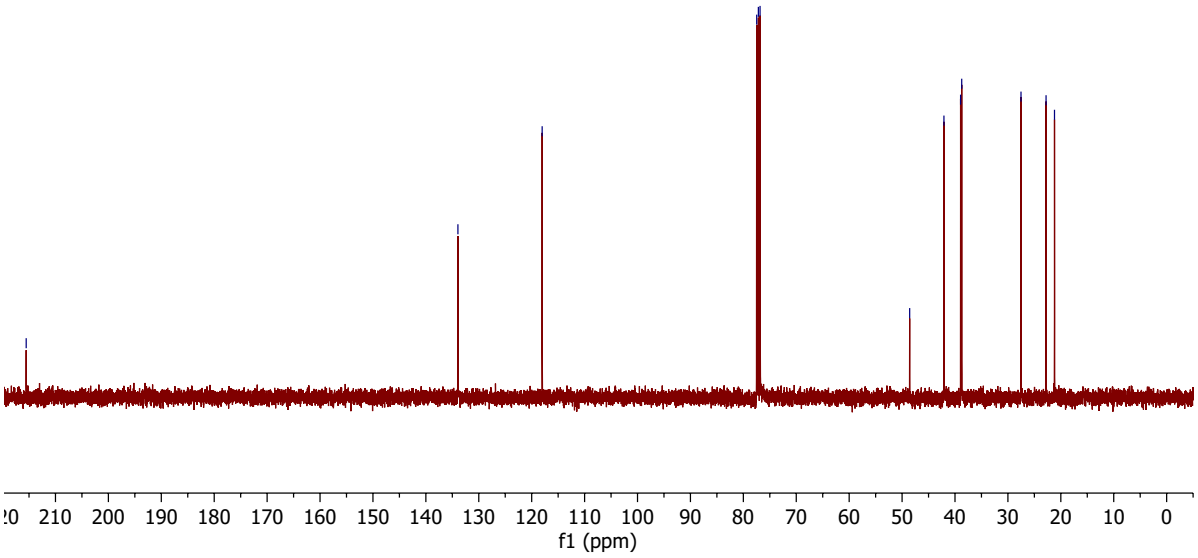
HRMS(*m/z*): calc'd. for $C_9H_{13}O^+$: 137.0961; detected: 137.0964.

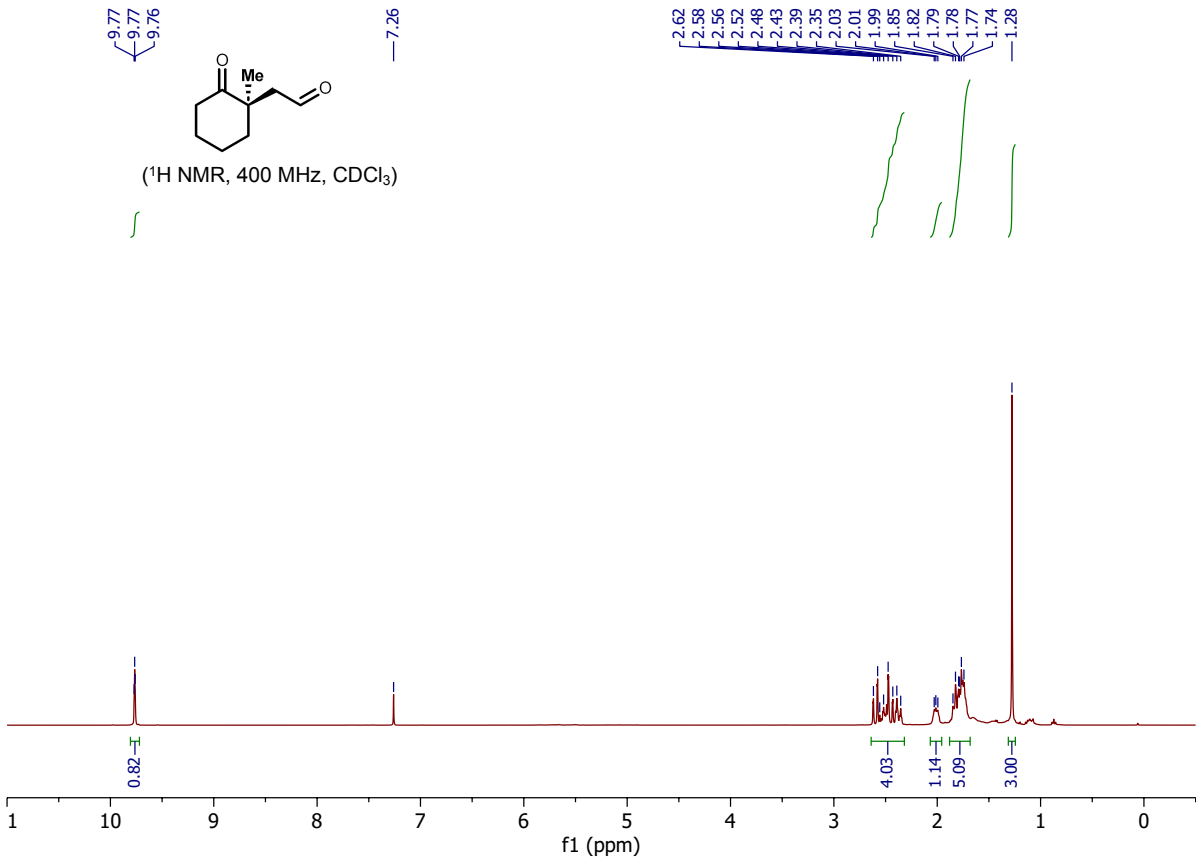


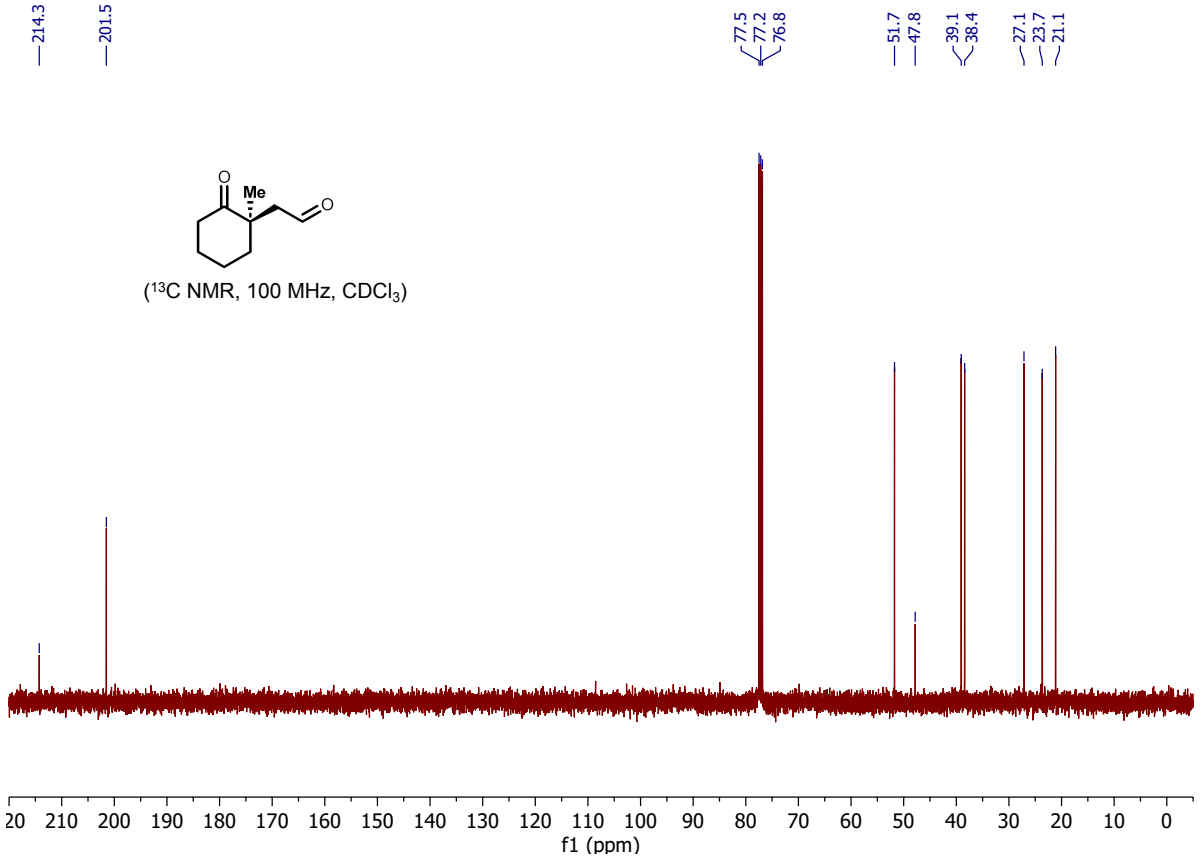
(^{1}H NMR, 400 MHz, $CDCl_3$)

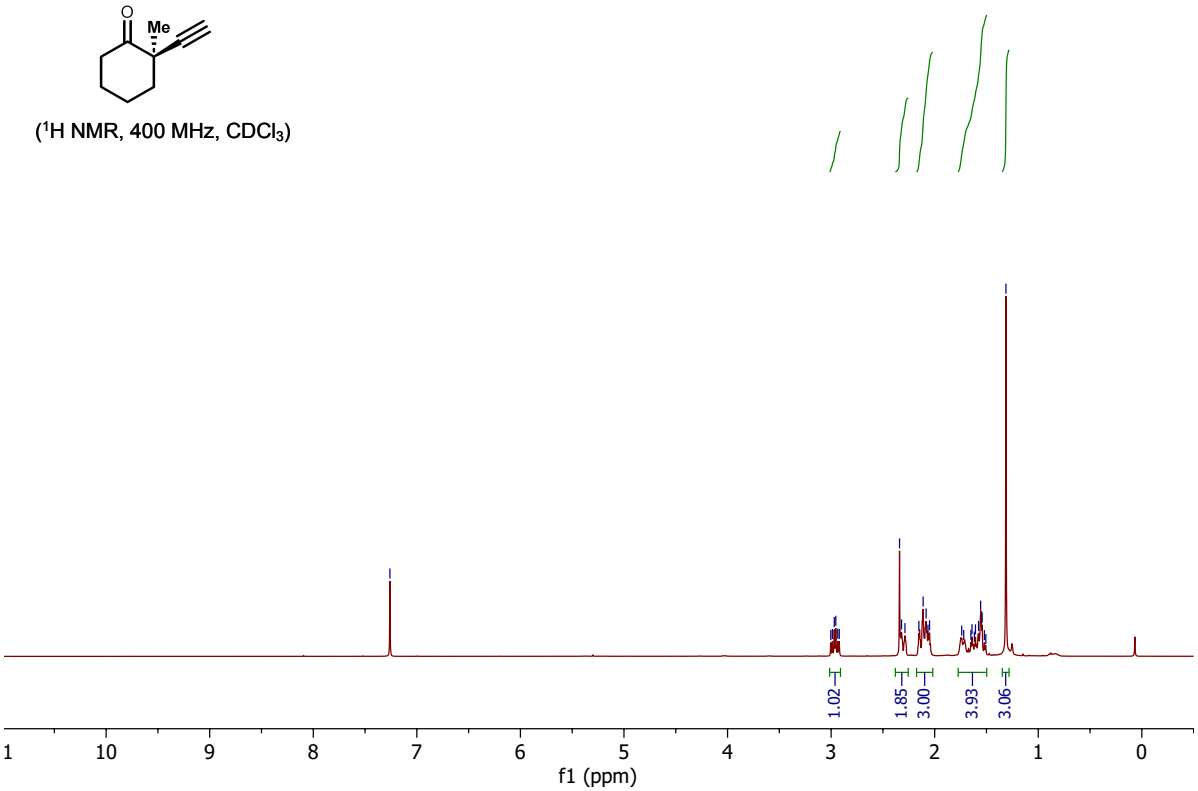


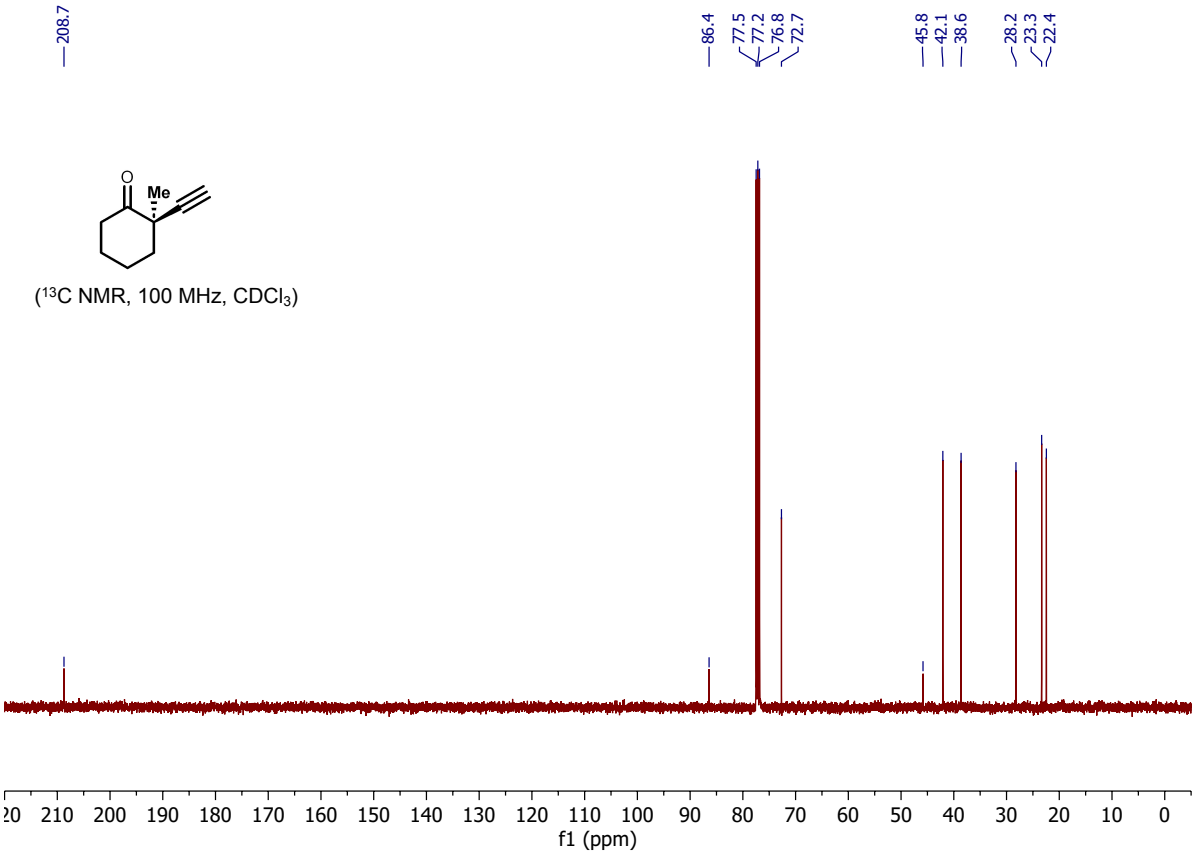
(^{13}C NMR, 100 MHz, $CDCl_3$)




9.77
9.77
9.76
7.26
2.62
2.58
2.56
2.52
2.48
2.43
2.39
2.35
2.03
2.01
1.99
1.85
1.82
1.79
1.78
1.77
1.74
1.28
Me
O
O
(1H NMR, 400 MHz, CDCl3)
0.82
4.03
1.14
5.09
3.00
1
10
9
8
7
6
5
4
3
2
1
0
f1 (ppm)


214.3
201.5
77.5
77.2
76.8
51.7
47.8
39.1
38.4
27.1
23.7
21.1
Me
O
O
(13C NMR, 100 MHz, CDCl3)
20
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)


Me
(1H NMR, 400 MHz, CDCl3)
7.26
3.00
2.99
2.97
2.95
2.94
2.92
2.34
2.32
2.29
2.15
2.11
2.08
2.06
2.05
1.74
1.65
1.64
1.61
1.58
1.56
1.54
1.02
1.85
3.00
3.93
3.06
f1 (ppm)


Me
(13C NMR, 100 MHz, CDCl3)
208.7
86.4
77.5
77.2
76.8
72.7
45.8
42.1
38.6
28.2
23.3
22.4
f1 (ppm)

References

[1] 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.

[2] Haote Li, Yu Shee, Brandon Allen, Federica Maschietto, Anton Morgunov, and Victor Batista. Kernel-elastic autoencoder for molecular design. *PNAS nexus*, 3(4):pgae168, 2024.

[3] Connor W Coley, William H Green, and Klavs F Jensen. Rdchiral: An rdkit wrapper for handling stereochemistry in retrosynthetic template extraction and application. *Journal of chemical information and modeling*, 59(6):2529–2537, 2019.

[4] Bo Qu, Jaehee Lee, Lifen Wu, Anjan K Saha, Guisheng Li, Yibo Xu, Zhulin Tan, Jason Brazzillo, Nizar Haddad, Scott Pennino, et al. Scalable process of spiro (cyclopropane) oxazepane pyridine carboxylic acid through kulinkovich, mitsunobu, and pd-catalyzed intramolecular c–n coupling. *Organic Process Research & Development*, 26(9):2779–2787, 2022.

[5] Rogelio P Frutos, Thomas G Tampone, Jason A Mulder, Sonia Rodriguez, Nathan K Yee, Bing-Shiou Yang, and Chris H Senanayake. Development of a practical process for the synthesis of pde4 inhibitors. *Organic Process Research & Development*, 20(5):982–988, 2016.

[6] Guisheng Li, Zhulin Tan, Yibo Xu, Kanwar PS Sidhu, Bo Qu, Melissa A Herbage, Magnus C Eriksson, Xingzhong Zeng, Carl A Busacca, Jean-Nicolas Desrosiers, et al. Process development of the bace inhibitors bi 1147560 bs and bi 1181181 mz. *Organic Process Research & Development*, 26(12):3345–3372, 2022.

[7] Xiao-Jun Wang, Rogelio P Frutos, Li Zhang, Xiufeng Sun, Yibo Xu, Thomas Wirth, Thomas Nicola, Lawrence J Nummy, Dhileep Krishnamurthy, Carl A Busacca, et al. Asymmetric synthesis of lfa-1 inhibitor birt2584 on metric ton scale. *Organic Process Research & Development*, 15(5):1185–1191, 2011.

[8] Hanjun Dai, Chengtao Li, Connor Coley, Bo Dai, and Le Song. Retrosynthesis prediction with conditional graph logic network. *Advances in Neural Information Processing Systems*, 32, 2019.

[9] Shuan Chen and Yousung Jung. Deep retrosynthetic reaction prediction using local reactivity and global attention. *JACS Au*, 1(10):1612–1620, 2021.

[10] Marwin HS Segler and Mark P Waller. Neural-symbolic machine learning for retrosynthesis and reaction prediction. *Chemistry–A European Journal*, 23(25):5966–5971, 2017.

[11] Vignesh Ram Somnath, Charlotte Bunne, Connor Coley, Andreas Krause, and Regina Barzilay. Learning graph models for retrosynthesis prediction. *Advances in Neural Information Processing Systems*, 34:9405–9415, 2021.

[12] 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.

[13] 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. *Nature Communications*, 14(1):6155, 2023.

[14] 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. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 531–539, 2021.

[15] 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. *Journal of Chemical Information and Modeling*, 61(1):123–133, 2021.

[16] Mikołaj Sacha, Mikołaj Błaz, Piotr Byrski, Paweł Dabrowski-Tumanski, Mikołaj Chrominski, Rafał Loska, Paweł Włodarczyk-Pruszynski, and Stanisław Jastrzebski. Molecule edit graph attention network: modeling chemical reactions as sequences of graph edits. *Journal of Chemical Information and Modeling*, 61(7):3273–3284, 2021.

[17] 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. *Nature communications*, 11(1):5575, 2020.

[18] 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. *Chemical Science*, 13(31):9023–9034, 2022.

[19] Zipeng Zhong, Jie Song, Zunlei Feng, Tiantao Liu, Lingxiang Jia, Shaolun Yao, Tingjun Hou, and Mingli Song. Recent advances in deep learning for retrosynthesis. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 14(1):e1694, 2024.

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

[21] Gabriele Pupo, Roberta Properzi, and Benjamin List. Asymmetric catalysis with co2: The direct α-allylation of ketones. *Angewandte Chemie International Edition*, 55(20):6099–6102, 2016.

[22] Ekaterina V Boltukhina, Andrey E Sheshenev, and Ilya 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.