JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

IF: 12.35

pubs.acs.org/jacsau

Article

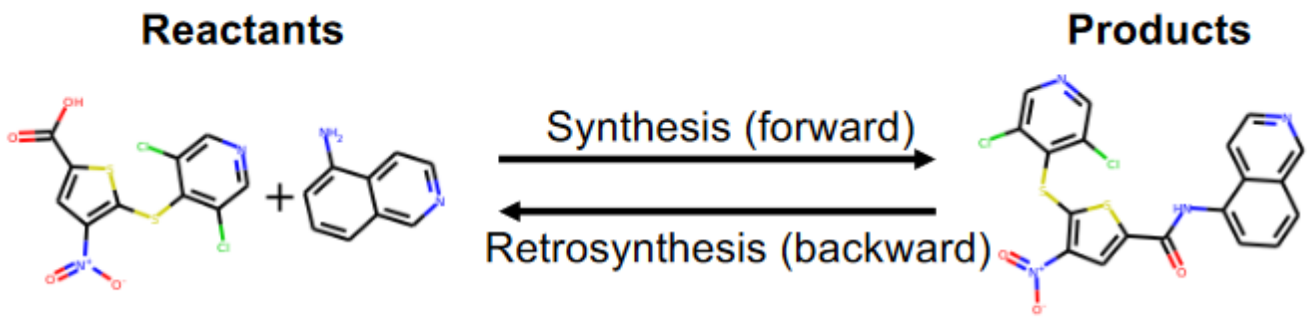
Deep Retrosynthetic Reaction Prediction using Local Reactivity and Global Attention

Shuan Chen and Yousung Jung*

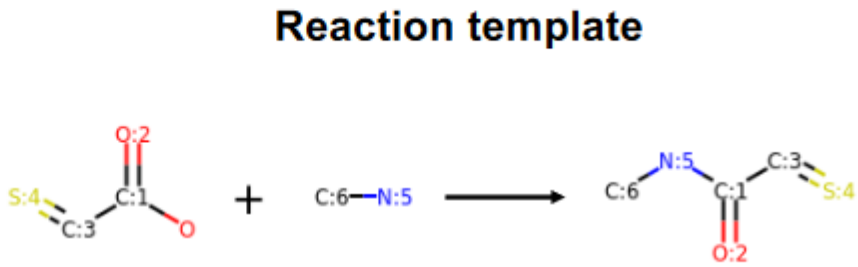
Cite This: https://doi.org/10.1021/jacsau.1c00246

Read Online

Background

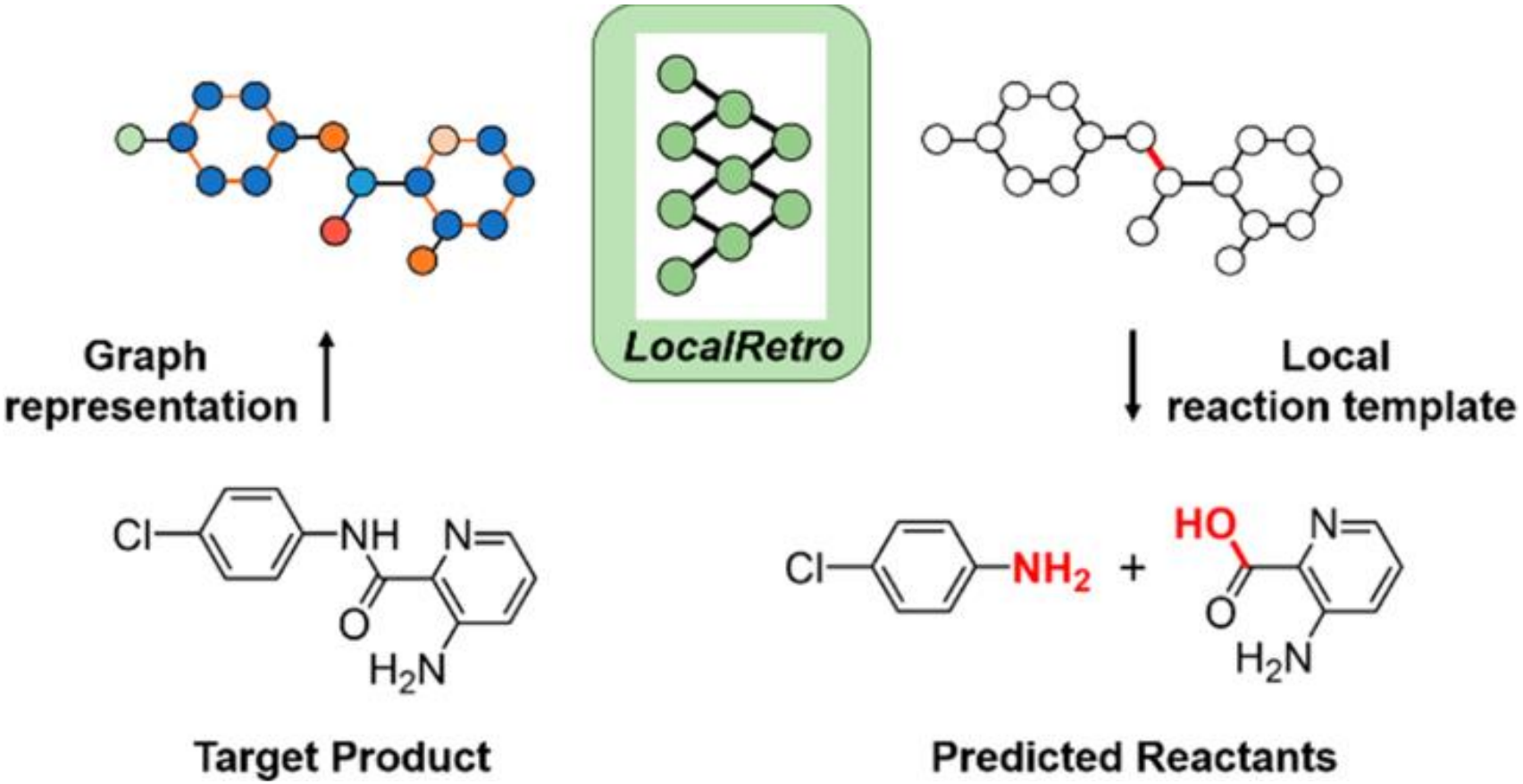


(a) Chemical reaction



(b) Reaction template

Overview

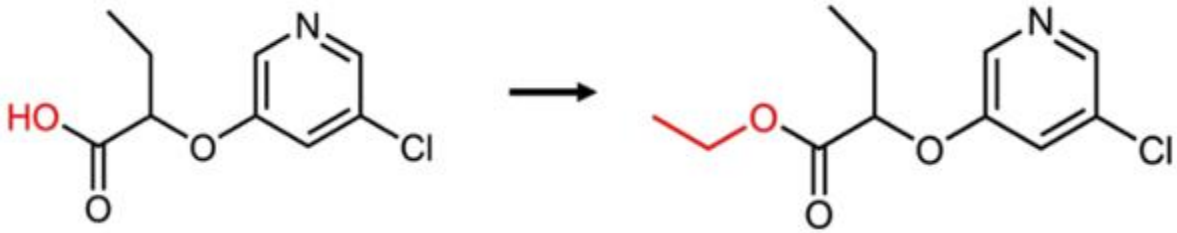

LocalRetro
Graph representation
Local reaction template
Cl
NH
N
O
H_2N
Target Product
Cl
NH_2
+
HO
N
O
H_2N
Predicted Reactants

Motivation

- All existing methods suggest reactants based on the global structures of the molecules.
- However, only local structure are edited in reaction.

Atom reaction template

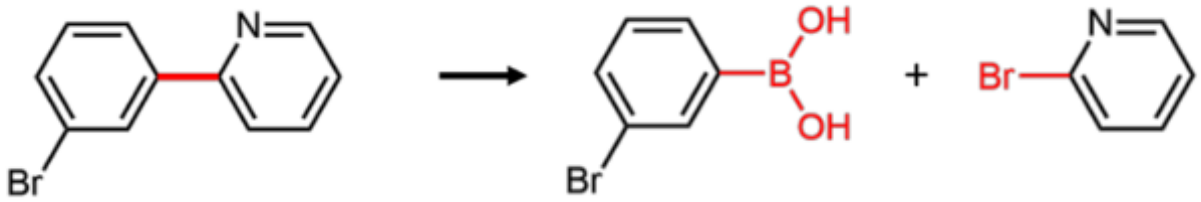
Retro-reaction



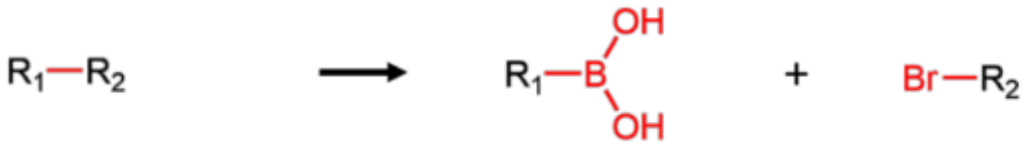
Local reaction template

Bond reaction template

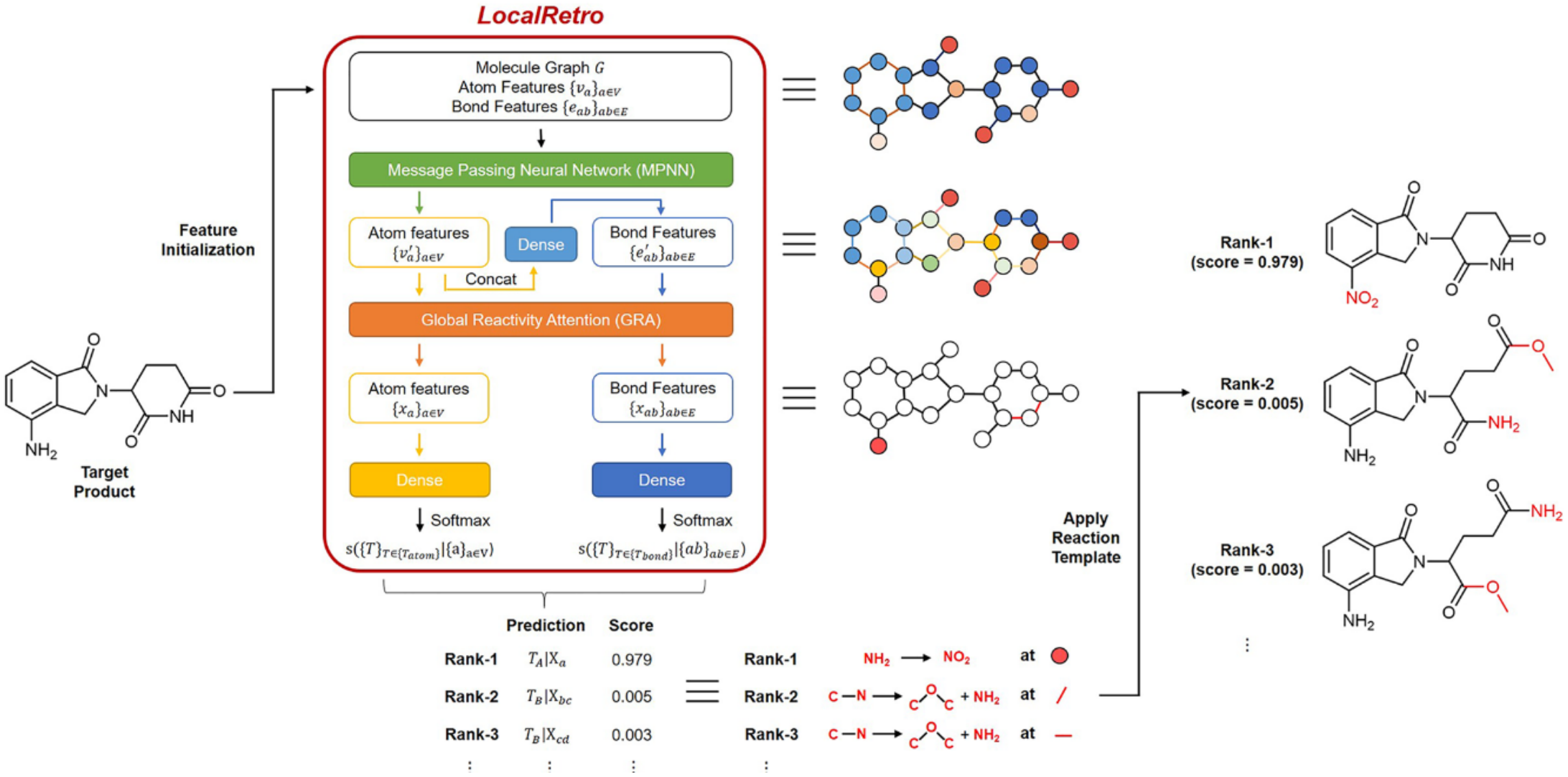
Retro-reaction



Local reaction template



System


LocalRetro
Molecule Graph G
Atom Features $\{v_a\}_{a\in V}$
Bond Features $\{e_{ab}\}_{ab\in E}$
Message Passing Neural Network (MPNN)
Atom features $\{v'_a\}_{a\in V}$
Dense
Bond Features $\{e'_{ab}\}_{ab\in E}$
Concat
Global Reactivity Attention (GRA)
Atom features $\{x_a\}_{a\in V}$
Bond Features $\{x_{ab}\}_{ab\in E}$
Dense
Dense
Softmax
Softmax
$s(\{T\}_{T\in\{T_{atom}\}}|\{a\}_{a\in V})$
$s(\{T\}_{T\in\{T_{bond}\}}|\{ab\}_{ab\in E})$
Feature Initialization
Target Product
Prediction
Score
Rank-1 $T_A|X_a$ 0.979
Rank-2 $T_B|X_{bc}$ 0.005
Rank-3 $T_B|X_{cd}$ 0.003
Rank-1 $NH_2 \rightarrow NO_2$ at
Rank-2 C–N → + NH_2 at /
Rank-3 C–N → + NH_2 at —
Apply Reaction Template
Rank-1 (score = 0.979)
Rank-2 (score = 0.005)
Rank-3 (score = 0.003)

Implementation details

- MPNN:

$$v'_a = \mathrm{MPNN}(v_a, \{v_b\}, \{e_{ab}\})$$

$$e'_{ab} = w(v'_a \| v'_b) + c$$

- Graph attention (GRA):

$$x_a = \mathrm{GRA}(v_a', \{v_a'\}_{a \in V}, \{e_{ab}'\}_{ab \in E})$$

$$x_{ab} = \mathrm{GRA}(e_{ab}', \{v_a'\}_{a \in V}, \{e_{ab}'\}_{ab \in E})$$

- Output

$$o_a = u_{\mathrm{A}}^{T}(\sigma(w_{\mathrm{A}} x_a + c_{\mathrm{A}}))$$

$$o_{ab} = u_{\mathrm{B}}^{T}(\sigma(w_{\mathrm{B}} x_{ab} + c_{\mathrm{B}}))$$

$$s(T|a) = \mathrm{Softmax}(o_a),\ T \in \{T_{\mathrm{atom}}\}$$

$$s(T|ab) = \mathrm{Softmax}(o_{ab}),\ T \in \{T_{\mathrm{bond}}\}$$

Results

Table 1. Top-*k* Exact Match Accuracy on USPTO-50K Dataset and USPTO-MIT Dataset without Given Reaction Class[a]

dataset	model	top-*k* accuracy (%)				
		K = 1	3	5	10	50
USPTO-50K	*GLN*[14]	52.5	69.0	75.6	83.7	92.4
	G2G[15]	48.9	67.6	72.5	75.5	
	GraphRetro[16]	53.7	68.3	72.2	75.5	
	Augmented Transformer[25]	**53.5**	69.4	81.0	85.7	
	MEGAN[17]	48.1	70.7	78.4	86.1	93.2
	LocalRetro wo/ GRA (this work)	49.8	75.8	84.0	91.3	**97.7**
	LocalRetro (this work)	53.4	**77.5**	**85.9**	**92.4**	**97.7**
USPTO-MIT	*LocalRetro* wo/ GRA (this work)	49.9	70.7	77.0	83.1	89.8
	LocalRetro (this work)	54.1	73.7	79.4	84.4	90.4

[a]The best top-*k* exact match accuracy is highlighted with bold font.

Results

Table 2. Top-*k* Exact Match Accuracy on USPTO-50K Dataset with Given Reaction Class[a]

dataset	model	top-*k* accuracy (%)				
		$K = 1$	3	5	10	50
USPTO-50K	*GLN*[14]	**64.2**	79.1	85.2	90.0	93.2
	G2G[15]	61.0	81.3	86.0	88.7	
	GraphRetro[16]	63.9	81.5	85.2	88.1	
	MEGAN[17]	60.7	82.0	87.5	91.6	95.3
	LocalRetro wo/ GRA (this work)	62.3	86.1	91.8	96.0	97.9
	LocalRetro (this work)	63.9	**86.8**	**92.4**	**96.3**	**97.9**

[a]The best top-*k* exact match accuracy is highlighted with bold font.

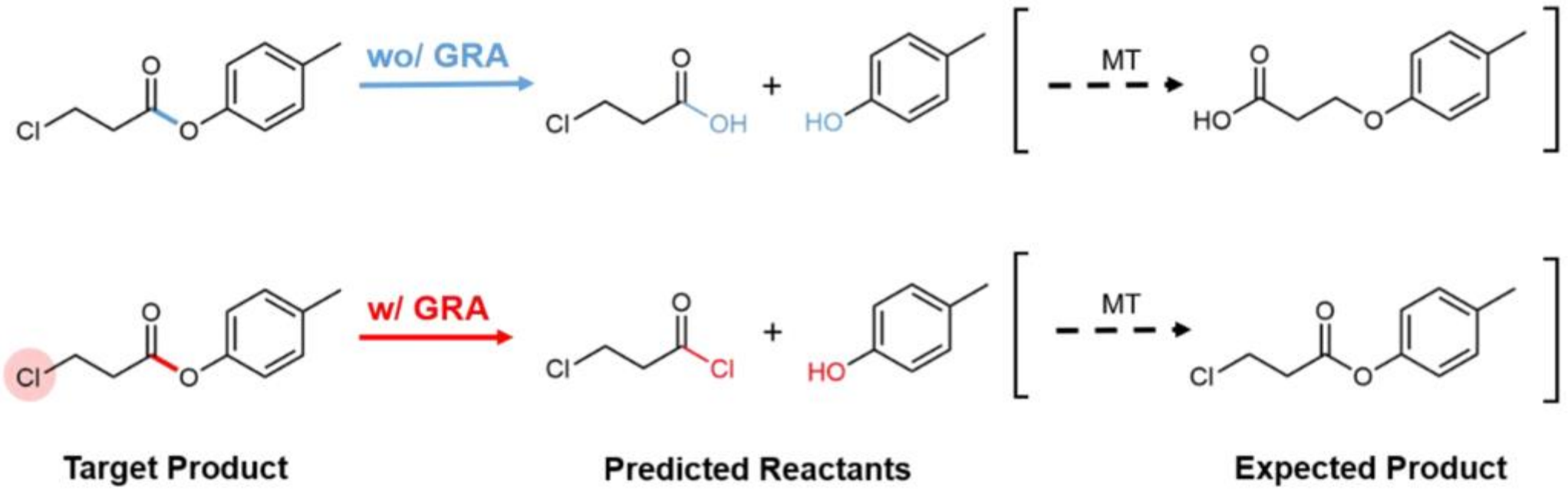
Results

Table 3. Top-*k* Round-Trip Accuracy on the USPTO-50K Dataset and USPTO-MIT Dataset[a]

dataset	model	top-*k* accuracy (%)		
		K = 1	3	5
USPTO-50K	*GLN*[14]	88.4	95.0	97.1
	LocalRetro wo/GRA (this work)	88.2	97.8	98.9
	LocalRetro (this work)	**89.5**	**97.9**	**99.2**
USPTO-MIT	*LocalRetro* wo/GRA (this work)	85.7	95.7	97.3
	LocalRetro (this work)	87.0	95.9	97.4

[a]The best top-*k* round-trip accuracy is highlighted with bold font.

Study of GRA

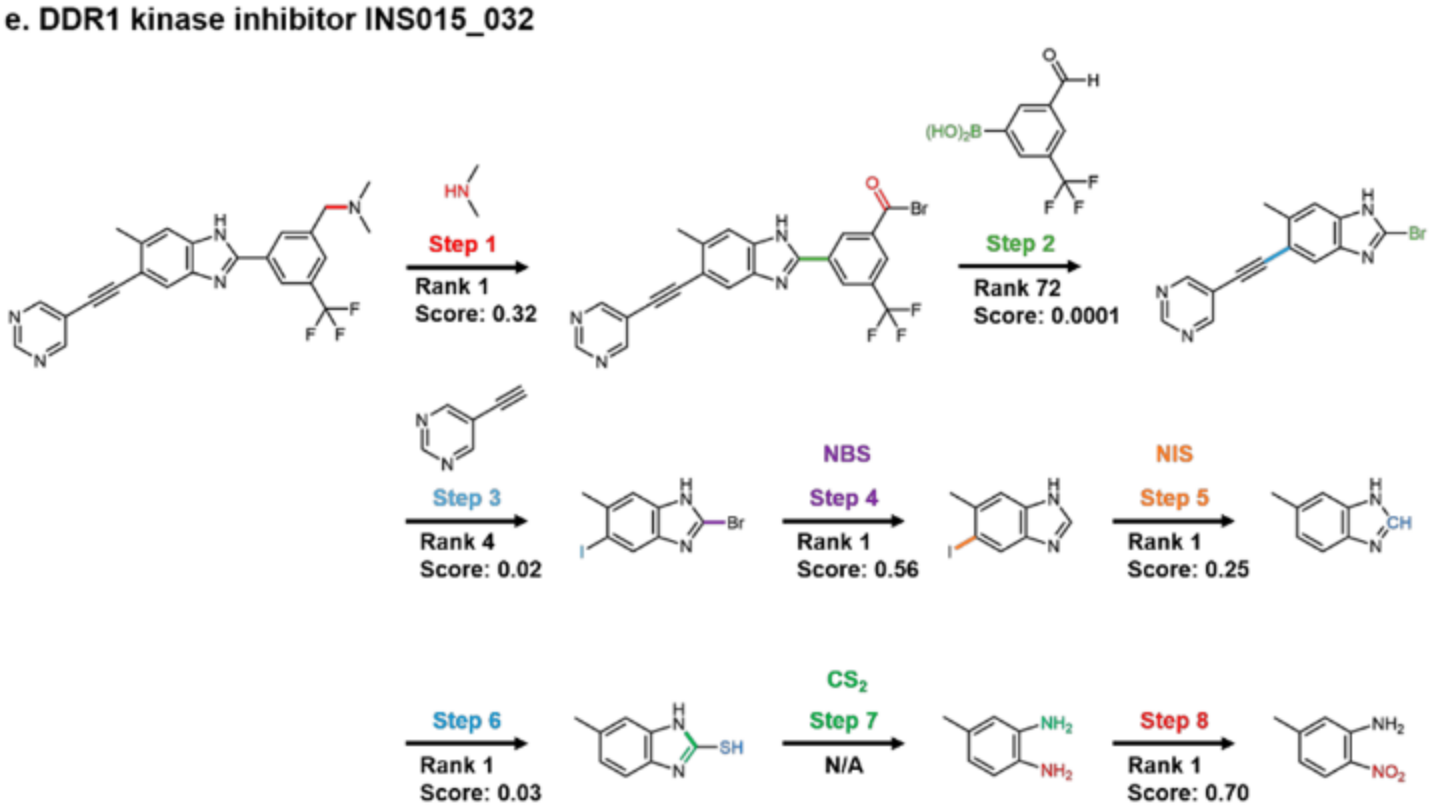

wo/ GRA
w/ GRA
MT
MT
Target Product
Predicted Reactants
Expected Product

Case study

Case study on considered in various retrosynthesis and inverse design works in the literature: lenalidomide, salmeterol, kinase inhibitors, etc.

Identical as reported in literature

Case study



Identical as reported in literature

Take home message

- Local reactivity is essential for retrosynthesis.

RetCL: A Selection-based Approach for Retrosynthesis via Contrastive Learning

Hankook Lee[1*] Sungsoo Ahn[2] Seung-Woo Seo[3] You Young Song[4]
Eunho Yang[15] Sung-Ju Hwang[15] Jinwoo Shin[1]

[1]Korea Advanced Institute of Science and Technology (KAIST)
[2]Mohamed bin Zaeyed University of Artificial Intelligence
[3]Standigm [4]Samsung Electronics [5]AITRICS

IJCAI21

Motivation

- Formulate the retrosynthesis as a retrieval task rather than generation task, therefore:
  - Can consider the availability.
  - Generalize to unseen templates.

Compare different paradigms

- Template free methods: choose from whole space of (possibly unavailable) molecules.
- Template based methods: choose from reactants extracted from known reaction templates.
- This work: choose from expert defined template library, which can be updated during inference.

Overview

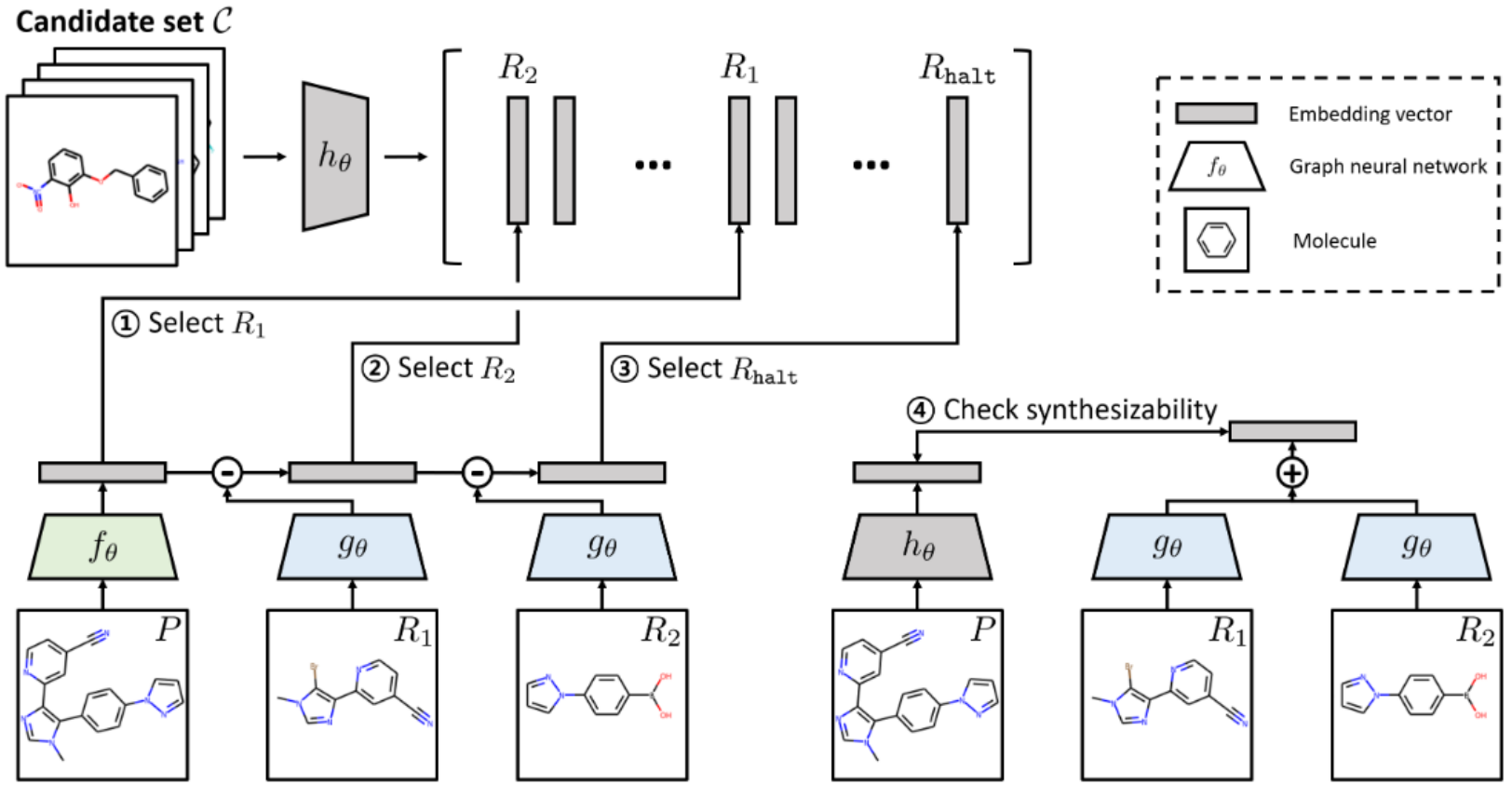
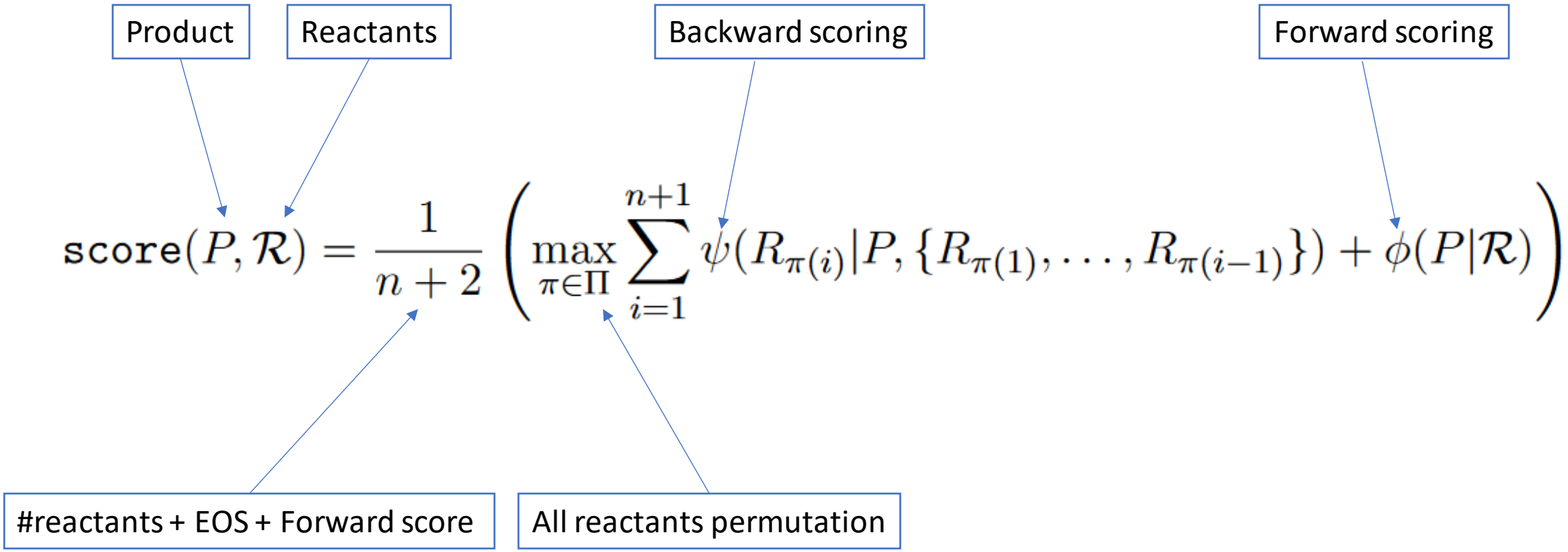


Figure 2: Illustration of the search procedure in RetCL. It first (1-3) selects reactants sequentially based on $\psi(R|P, \mathcal{R}_{\texttt{given}})$, and then (4) check the synthesizability of the selected reactant-set based on $\phi(P|\mathcal{R})$. The overall score is the average over all scores from (1) to (4).

The score function



The score function

$$\psi(R|P, \mathcal{R}_{\texttt{given}}) = \texttt{CosSim}\left(f_\theta(P) - \sum\nolimits_{S \in \mathcal{R}_{\texttt{given}}} g_\theta(S),\ h_\theta(R)\right),$$

$$\phi(P|\mathcal{R}) = \texttt{CosSim}\left(\sum\nolimits_{R \in \mathcal{R}} g_\theta(R),\ h_\theta(P)\right),$$

f_θ Query of product

g_θ Query of reactant

h_θ Key of a molecule

Contrastive learning

$$p(R|P,\mathcal{R}_{\texttt{given}},\mathcal{C}) = \frac{\exp(\psi(R|P,\mathcal{R}_{\texttt{given}})/\tau)}{\sum_{R'\in\mathcal{C}\setminus\{P\}}\exp(\psi(R'|P,\mathcal{R}_{\texttt{given}})/\tau)},$$

$$q(P|\mathcal{R},\mathcal{C}) = \frac{\exp(\phi(P|\mathcal{R})/\tau)}{\sum_{P'\in\mathcal{C}\setminus\mathcal{R}}\exp(\phi(P'|\mathcal{R})/\tau)},$$

$$\mathcal{L}_{\texttt{backward}}(P,\mathcal{R}|\theta,\mathcal{C}) = -\max_{\pi\in\Pi}\sum_{i=1}^{n+1}\log p(R_{\pi(i)}|P,\{R_{\pi(1)},\ldots,R_{\pi(i-1)}\},\mathcal{C}),$$

$$\mathcal{L}_{\texttt{forward}}(P,\mathcal{R}|\theta,\mathcal{C}) = -\log q(P|\mathcal{R},\mathcal{C}),$$

Contrastive learning

It’s costly to use all (6*10^5) negative examples

$$\mathcal{C}_{\mathcal{B}} = \{M \mid \exists\, (\mathcal{R}, P) \in \mathcal{B} \text{ such that } M = P \text{ or } M \in \mathcal{R}\},$$

$$\mathcal{L}(\mathcal{B}|\theta) = \frac{1}{|\mathcal{B}|} \sum_{(\mathcal{R},P)\in\mathcal{B}} \Big(\mathcal{L}_{\texttt{backward}}(P, \mathcal{R}|\theta, \mathcal{C}_{\mathcal{B}}) + \mathcal{L}_{\texttt{forward}}(P, \mathcal{R}|\theta, \mathcal{C}_{\mathcal{B}}) \Big).$$

Hard negative mining

$$\widetilde{\mathcal{C}}_{\mathcal{B}} = \mathcal{C}_{\mathcal{B}} \cup \bigcup\nolimits_{M\in\mathcal{C}_{\mathcal{B}}} \{\text{Top-}K \text{ nearest neighbors of } M \text{ from } \mathcal{C}\},$$

Category	Method	Top-1	Top-3	Top-5	Top-10	Top-20	Top-50
Reaction type is unknown							
Template-free	Transformer (Karpov et al., 2019)	37.9	57.3	62.7	-	-	-
	SCROP (Zheng et al., 2019)	43.7	60.0	65.2	68.7	-	-
	Transformer (Chen et al., 2019)	44.8	62.6	67.7	71.1	-	-
	G2Gs (Shi et al., 2020)	**48.9**	**67.6**	**72.5**	**75.5**	-	-
Template-based	retrosim (Coley et al., 2017b)	37.3	54.7	63.3	74.1	82.0	85.3
	neuralsym (Segler & Waller, 2017)	44.4	65.3	72.4	78.9	82.2	83.1
	GLN (Dai et al., 2019)	**52.5**	**69.0**	**75.6**	**83.7**	**89.0**	**92.4**
Selection-based	Bayesian-Retro (Guo et al., 2020)	47.5	67.2	77.0	80.3	-	-
	RetCL (Ours)	**71.3**	**86.4**	**92.0**	**94.1**	**95.0**	**96.4**
Reaction type is given as prior							
Template-free	seq2seq (Liu et al., 2017)	37.4	52.4	57.0	61.7	65.9	70.7
	Transformer† (Chen et al., 2019)	54.1	70.0	74.2	77.8	80.4	83.3
	SCROP (Zheng et al., 2019)	59.0	74.8	78.1	81.1	-	-
	G2Gs (Shi et al., 2020)	**61.0**	**81.3**	**86.0**	**88.7**	-	-
Template-based	retrosim (Coley et al., 2017b)	52.9	73.8	81.2	88.1	91.8	92.9
	neuralsym (Segler & Waller, 2017)	55.3	76.0	81.4	85.1	86.5	86.9
	GLN (Dai et al., 2019)	**64.2**	**79.1**	**85.2**	**90.0**	**92.3**	**93.2**
Selection-based	Bayesian-Retro (Guo et al., 2020)	55.2	74.1	81.4	83.5	-	-
	RetCL (Ours)	**78.9**	**90.4**	**93.9**	**95.2**	**95.8**	**96.7**

Results

Table 2: The top-k exact match accuracy (%) of our RETCL, Transformer (Chen et al., 2019) and GLN (Dai et al., 2019) with discarding predictions not in the candidate set $\mathcal{C}$.

Category	Method	Top-1	Top-5	Top-10	Top-50	Top-100	Top-200
Reaction type is unknown							
Template-free	Transformer (Chen et al., 2019)	59.6	74.3	77.0	79.4	79.5	79.6
	RETCL (Ours)	71.3	**92.0**	**94.1**	**96.4**	**96.7**	**97.1**
Template-based	GLN (Dai et al., 2019)	**77.3**	90.0	92.5	93.3	93.3	93.3
Reaction type is given as prior							
Template-free	Transformer (Chen et al., 2019)	68.4	82.4	84.3	85.9	86.0	86.1
	RETCL (Ours)	78.9	**93.9**	**95.2**	**96.7**	**97.1**	**97.5**
Template-based	GLN (Dai et al., 2019)	**82.0**	91.7	92.9	93.3	93.3	93.3

Case study

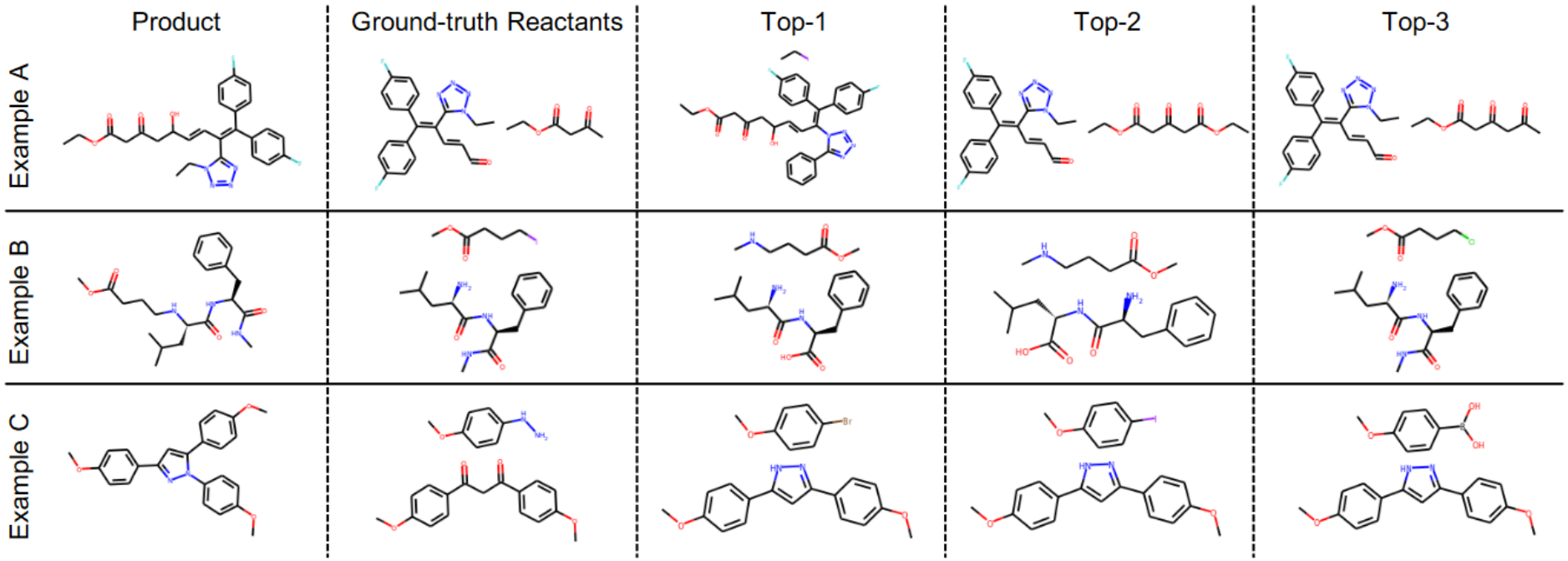


Figure 3: Failure cases of RETCL.

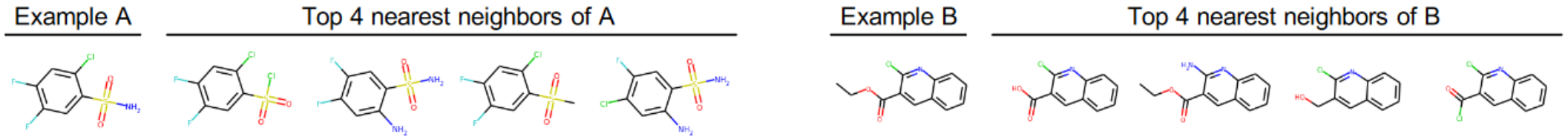


Figure 4: The top-4 nearest neighbors of two randomly sampled molecules in $\mathcal{C}$.

Ablation

Table 3: Ablation study.

$\phi(P\|\mathcal{R})$	K	`sum`	Top-1	Top-10
✓			59.5	79.8
✓	1		69.6	92.2
✓	2		70.9	92.7
✓	4		71.1	92.9
	4		69.8	90.3
✓	4	✓	71.3	94.1

Generalization to unseen templates

Table 4: The top-10 exact match accuracy (%) of our RETCL and GLN (Dai et al., 2019) trained on USPTO-50k without reaction types from 6 to 10. The average column indicates the average of class-wise accuracy for each reaction type.

		Reaction type									
Method	Average	1	2	3	4	5	6	7	8	9	10
GLN (Dai et al., 2019)	39.7	84.3	92.2	70.7	59.3	89.7	0.0	0.0	0.0	0.5	0.0
RETCL (Ours)	**55.6**	**93.9**	**97.6**	**86.4**	**67.0**	**95.6**	**59.1**	**11.9**	**18.3**	**26.1**	0.0

Generalization to larger dataset

Table 5: Generalization to USPTO-full.

Method	Top-1	Top-10	Top-50
Transformer (Chen et al., 2019)	29.9	46.6	51.0
GLN (Dai et al., 2019)	26.7	42.2	46.7
RetCL (Ours)	39.9	57.1	60.9

Generalization to unseen candidates

Table 6: Generalization to unseen candidates.

$\lvert\mathcal{C}_{\texttt{train}}\rvert$	Top-1	Top-5	Top-10	Top-20	Top-50
91,297	69.0	88.1	91.0	92.8	94.4
671,518	71.3	92.0	94.1	95.0	96.4

Multi-step retrosynthesis

Table 7: Multi-step retrosynthesis.

	Single		Hybrid	
Single-step model	MLP	TF	TF+TF	RETCL+TF
Succ. rate (%)	86.84	91.05	90.54	96.84
Avg. length	-	4.30	4.31	3.90

Take home message

- Retrieval based methods may be a new direction.