Supplementary Information

RetroSynFormer: Planning multi-step chemical synthesis routes via a Decision Transformer

Authors: Emma Granqvist, Rocío Mercado, and Samuel Genheden

A Optimization and Reward Settings

In Tables S1–S2 we describe the search space for the DT and reward parameters, which were optimized using Optuna. In Algorithm 1, we describe the reward function in detail. In Table S3 we provide details on the reward parameters for the alternative reward functions described in Table 5.

Table S1: Search space for the DT parameters and the optimal values used for training the DT model.

Model parameter	Search range	Optimal value
`activation_function`	{`relu`, `silu`, `gelu`, `tanh`, `gelu_new`}	`relu`
`action_tanh`	{`true`, `false`}	`false`
`attn_pdrop`	$\{0.01x \mid x \in \mathbb{Z}, 1 \leq x \leq 20\}$	0.02
`embd_pdrop`	$\{0.01x \mid x \in \mathbb{Z}, 1 \leq x \leq 20\}$	0.2
`hidden_size`	$\{32, 64, 128, 256, 512, 1024, 2048\}$	256
`n_heads`	$\{2x \mid x \in \mathbb{Z}, 0 \leq x \leq 32\}$	4
`n_layers`	$\{2x \mid x \in \mathbb{Z}, 0 \leq x \leq 32\}$	26
`resid_pdrop`	$\{0.01x \mid x \in \mathbb{Z}, 1 \leq x \leq 20\}$	0.08
Settings		Value
# epochs		300
Maximizing		Success rate

Algorithm S1 Pseudo code explaining how the RetroSynFormer reward function works.

$action_t \leftarrow$ retrosynformer.predict(**states**, **rewards**, **actions**)
$state_{t+1} \leftarrow$ env.step($action_t$)
$reward_{t+1} \leftarrow 0$

for s in $state_{t+1}$ **do**
  $reward_{t+1} \leftarrow reward_{t+1} + reward\_factor(s) \times scale\_depth(s) \times depth(s)$
end for
$reward_{t+1} \leftarrow \frac{reward_t+1}{len(state_t+1)}$

where

$$reward\_factor \leftarrow \begin{cases} 0, & \text{if } s \text{ is building block} \\ -2, & \text{if } s \text{ is intermediate} \\ 2, & \text{if } s \text{ is dead end} \end{cases}$$

and

$$scale\_depth \leftarrow \begin{cases} 2, & \text{if } s \text{ is building block} \\ 1, & \text{if } s \text{ is intermediate} \\ -2, & \text{if } s \text{ is dead end} \end{cases}$$

Table S2: Search space for the reward parameters and the optimal values used for training the DT model.

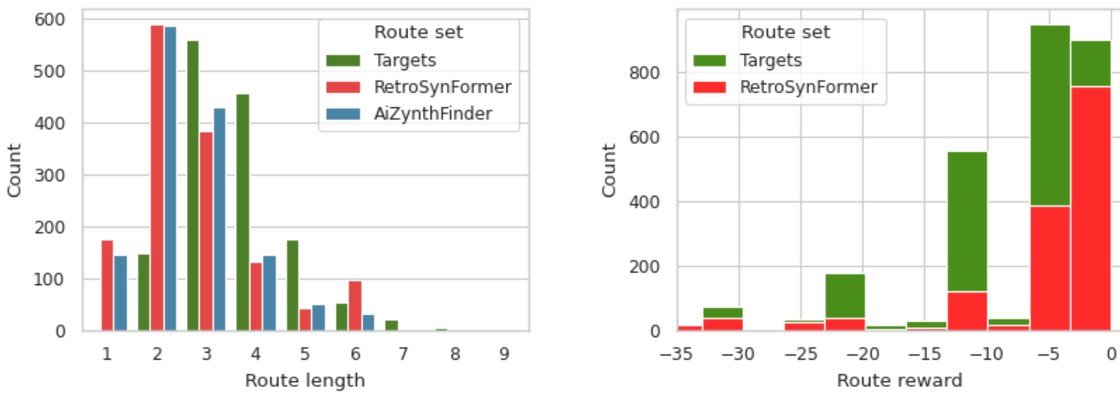
Hyperparameters	Search range	Optimal value
Building block reward	$\{0, 0.001, 0.01, 0.1, 0.25, 0.5, 1, 2, 4\}$	0
Building block scale with depth	$\{0, 0.01, 0.1, 0.5, 1, 2\}$	2
Intermediate reward	$\{0, -0.001, -0.01, -0.1, -0.25, -0.5, -1, -2, -4\}$	-2
Intermediate scale with depth	$\{0, 0.01, 0.1, 0.5, 1, 2\}$	1
Dead-end reward	$\{0, -0.001, -0.01, -0.1, -0.25, -0.5, -1, -2, -4\}$	-2
Dead-end depth	$\{0, 0.01, 0.1, 0.5, 1, 2\}$	2
Settings		**Value**
# epochs		100
Maximizing		Success rate

Table S3: The reward parameters for alternative rewards.

Label	Building block		Intermediate		Dead End	
	Reward	Scale	Reward	Scale	Reward	Scale
Default	0	-2	1	-2	-2	2
Increasing Building Block Reward	2	-2	1	-2	-2	2
Decreasing Building Block Reward	-2	-2	1	-2	-2	2
Remove Scaling Building Block Reward	2	0	1	-2	-2	2
Flipping Sign Intermediate Score	0	-2	1	2	-2	2
Remove Scaling Intermediate	0	-2	0	-2	-2	2
Flipping Sign Dead End Score	0	-2	1	-2	2	2
Remove Scaling Dead End	0	-2	1	-2	-2	0

B Complimentary Results

In Figure S1 we describe the route characteristics for the routes generated by the RetroSynFormer and AiZynthFinder on the N5 set. In Figures S2 and S3 we present histograms for the most common reaction templates in the N1 and N5 sets, respectively. Finally, in Table S4 we compare the performance of the RetroSynFormer and AiZynthFinder on the targets that are shared across the N1 and N5 test sets, respectively.



(a) Histograms of N5 route lengths, measured as the number of actions per route, for each route set.

(b) Stacked bar plot showing the distribution of N5 route rewards in the target versus RetroSynFormer-predicted routes.

Figure S1: Route characteristics for routes for the N5 test set target routes compared to the RetroSynFormer and AiZynthFinder solved predictions. a) The median reward for the route length is 4 for the target routes and 2 for the RetroSynFormer and AiZynthFinder routes. b) The median reward for the route reward is -8 for the target routes and -2 for the RetroSynFormer and AiZynthFinder routes.

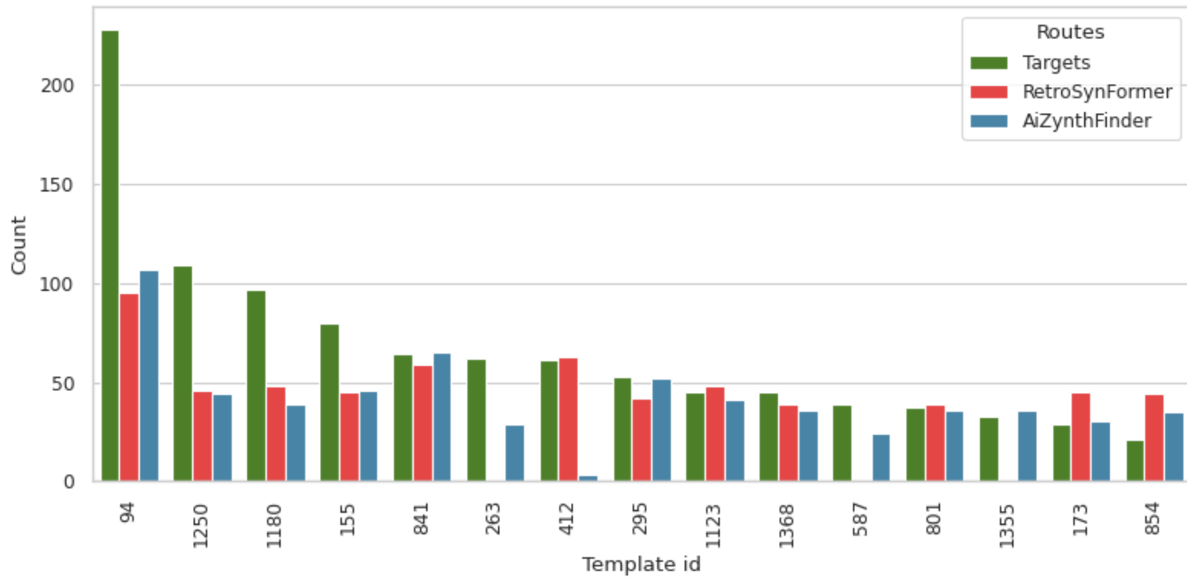
Table S4: Performance of retrosynthesis models for different template sets on the 232 N1 targets and 248 N5 targets that are common among across the test sets in Table 6.

Dataset	Model	Test set	Success rate	Top-1 Accuracy	TED	Avg. route length
Small	RetroSynFormer50	N1	0.805	0.098	6.033	2.609
		N5	0.839	0.094	5.294	2.37
	AiZynthFinder	N1	0.914	0.142	4.746	2.590
		N5	0.907	0.093	6.423	2.902
Standard	RetroSynFormer50	N1	0.953	0.102	5.314	2.309
		N5	0.898	0.052	6.901	2.397
	AiZynthFinder	N1	0.935	0.121	5.031	2.396
		N5	0.911	0.073	7.059	2.726
Large	RetroSynFormer50	N1	0.958	0.091	5.499	2.147
		N5	0.917	0.031	7.136	2.296
	AiZynthFinder	N1	0.940	0.099	5.293	2.362
		N5	0.927	0.065	7.262	2.748

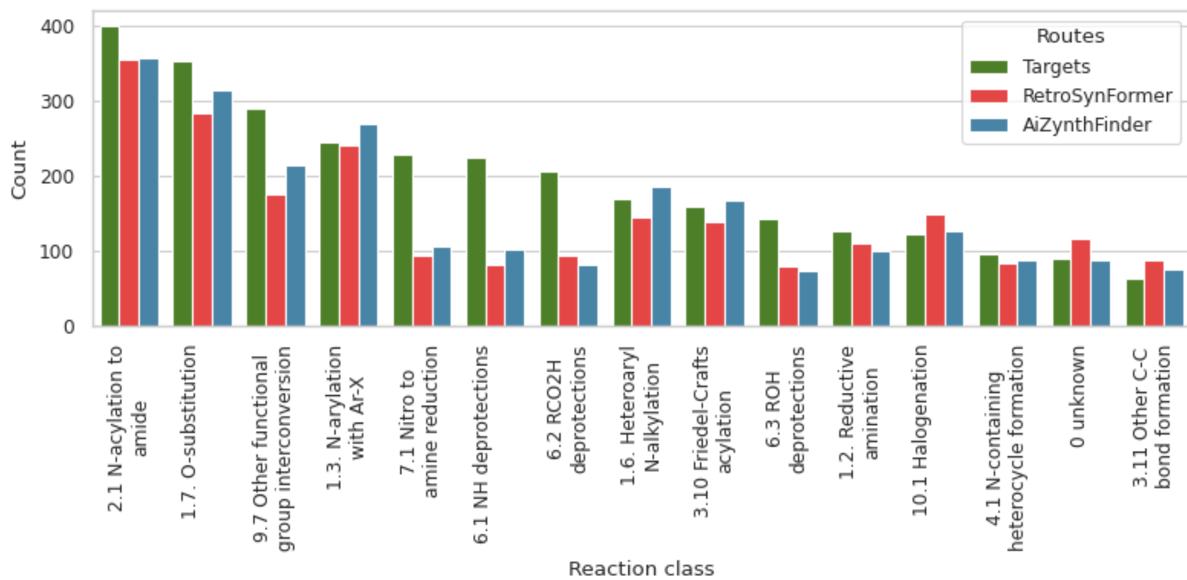
* RetroSynFormer results show averages over three runs. Standard deviations for all RetroSynFormer models are ≤ 0.02 for success rate, ≤ 0.02 for top-1 accuracy, ≤ 0.15 for TED, and ≤ 0.05 for route length.

C Reaction Templates

In Table S5, we present an illustration of the most commonly occurring templates in the predicted reactions from the RetroSynFormer.

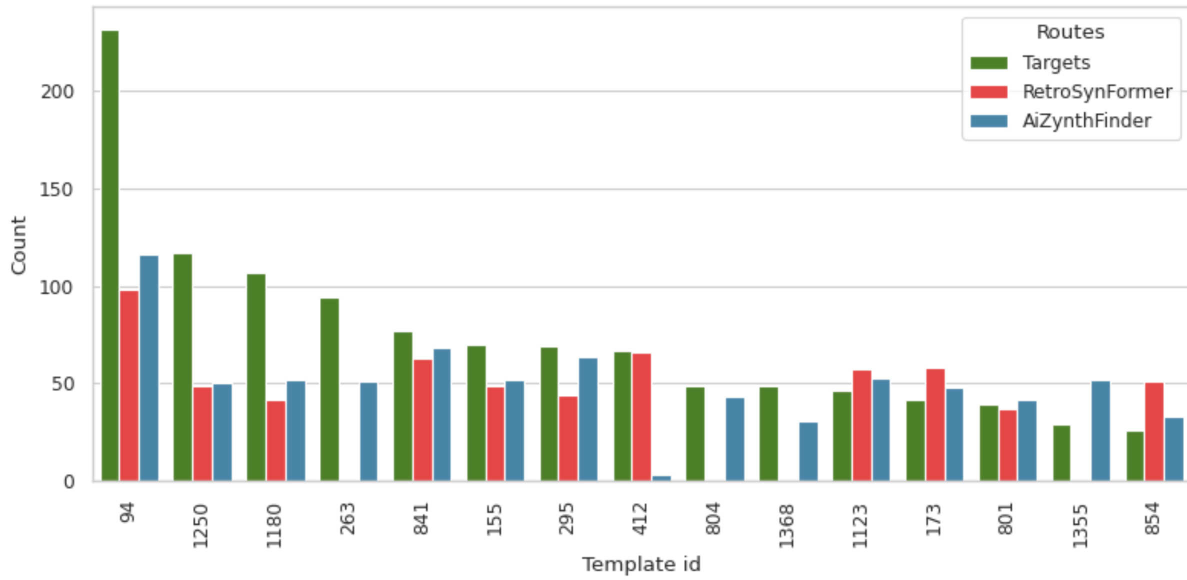


(a) Reaction template frequency distribution.

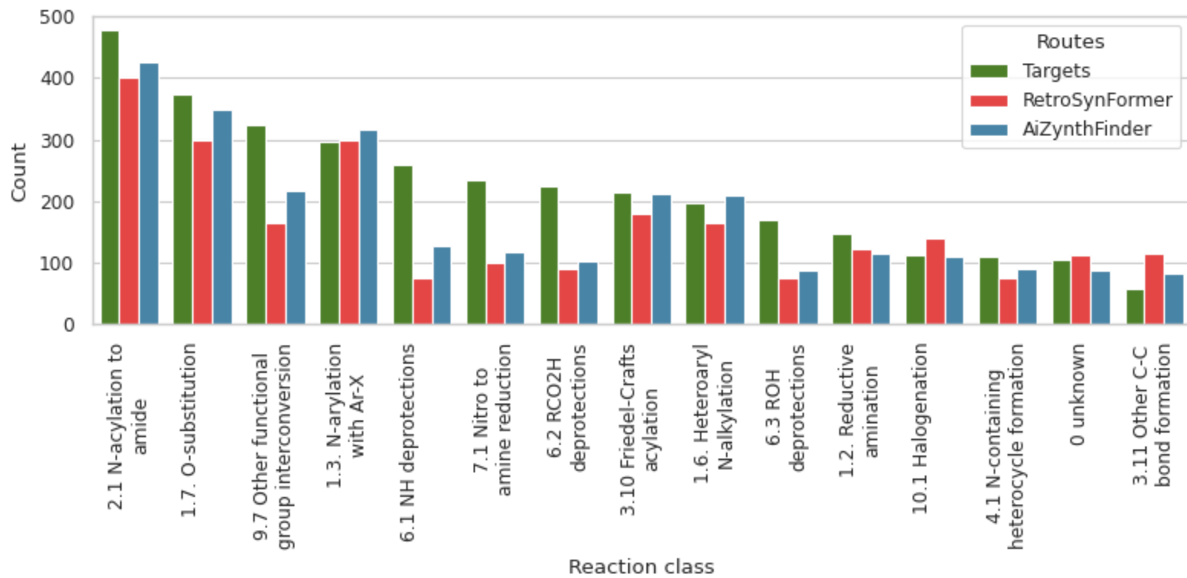


(b) Reaction classes frequency distribution.

Figure S2: Histograms of predicted reaction templates and classes for the N1 test set. a) Comparison of the counts for the 15 most common predicted templates compared to ground truth and baseline. Illustrations of the template IDs can be found in Table S5. b) Comparison of the counts for the 15 most common predicted reaction classes compared to ground truth and baseline.



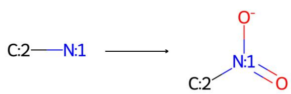
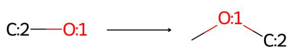
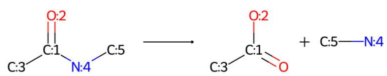
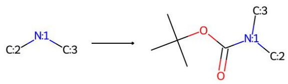
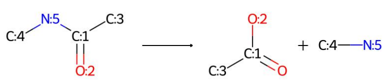
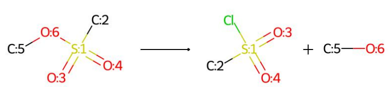
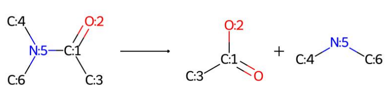
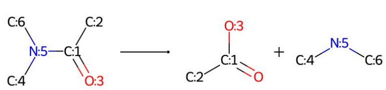
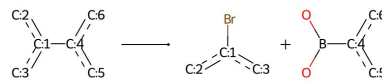
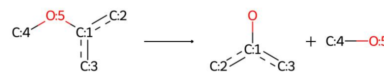
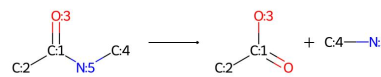
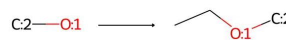
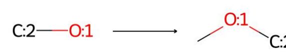
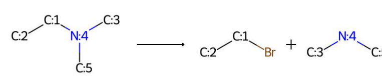
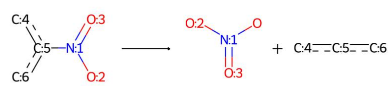
(a) Reaction template frequency distribution.



(b) Reaction classes frequency distribution.

Figure S3: Histograms of predicted reaction templates and classes for the N5 test set. a) Comparison of the counts for the 15 most common predicted templates compared to ground truth and baseline. Illustrations of the template IDs can be found in Table S5. b) Comparison of the counts for the 15 most common predicted reaction classes compared to ground truth and baseline.

Table S5: Visualization of the most common templates mentioned in Figure S2a and S3a.

Template id	Template
94	
155	
173	
263	
412	
587	
801	
804	
841	
854	
1123	
1180	
1250	
1355	
1368	

D Out of Distribution Evaluation

The standard and large model was evaluated on a set of 600 randomly sampled routes extracted from the *Journal of Medicinal Chemistry* (JMC), used in [29]. The result for the evaluation can be found in Table D.

The mean pairwise Tanimoto similarity (based on 1024 Morgan Fingerprint) between the JMC targets and the molecules in the standard training set is 0.123. We observe that JMC routes are more challenging and that both RetroSynFormer and AiZynthFinder are not generalizing well to this external dataset. However we observe that the large models with additional reaction templates performs better better compared to the standard model. The JMC target routes contains templates outside of the the ones used for training and we can see that the restricted template space could be a limiting factor for the model performance. In addition, the building block stock used is comprised of the leafs from the PaRoutes, these building blocks are likely not the optimal starting materials for the JMC routes and could be one reason why the number of solved routes are lower for both RetroSynFormer and AiZynthFinder.

Table S6: Performance of RetroSynFormer compared to AiZynthFinder evaluated on 600 randomly sampled routes from Journal of Medicinal Chemistry.

	Standard		Large	
	RetroSynFormer50	AiZynthFinder	RetroSynFormer50	AiZynthFinder
Success rate (%) ↑	0.461	0.500	0.558	0.615
Top-1 accuracy ↑	0.013	0.013	0.014	0.019
Mean TED ↓	9.018	10.016	8.87	9.840
Mean # reactions per route ↓	2.787	2.940	2.616	2.781

* RetroSynFormer results show averages over three runs using beam width 50. Standard deviation for the RetroSynFormer models is ≤ 0.008 for the success rate, ≤ 0.002 for the top-1 accuracy, ≤ 0.03 for the mean tree edit distance (TED), and ≤ 0.02 for the mean # reactions per route.