

## 688 A Experimental details

689 In this section, we elaborate on our experimental setup. We view a set of molecules as a single  
 690 potentially disconnected graph. We follow the procedure outlined in Igashov et al. [17] and introduce  
 691 a “dummy” node as an atom type. The graph of reactant molecules has at least as many nodes as  
 692 the product molecule graph. During training and inference for RETRO PRODFLOW, we append ten  
 693 dummy nodes to each product molecule. This covers 99.4% of the reactions in the USPTO-50k test  
 694 dataset. Following Igashov et al. [17], the remaining reactions are removed from the test data. During  
 695 inference, these dummy nodes are potentially transformed into true atom nodes. Similarly, we also  
 696 append ten dummy nodes to the synthon molecule graphs when using RETRO SYNFLOW. There  
 697 are 16 atom types (not including dummy atoms) and 4 bond types (not including no bond). Our  
 698 methods are implemented in PyTorch [25], and we also use an open-source software RDKit [20], for  
 699 operations involving chemical reactions and molecular graphs.

### 700 A.1 Neural Network Model

701 We use a graph transformer network [10, 45] also used by Igashov et al. [17] to model the denoiser  
 702  $p_\theta$  of the flow matching process. The denoiser model takes a noisy graph  $(\mathbf{v}, \mathbf{E})$  and graph-level  
 703 features  $\mathbf{y}$  as input and outputs probabilities of graphs over the data distribution. In the case of RETRO  
 704 PRODFLOW, the product molecule graph is also provided as input to the denoiser model. This is done  
 705 by appending the product molecule graph’s node feature vector and adjacency matrix to the node  
 706 feature vector and adjacency matrix of the noisy graph. For RETRO SYNFLOW, both the product  
 707 molecule graph and synthon molecule graph are provided as input.

708 The graph transformer network is similar to the standard transformer architecture and con-  
 709 sists of a graph attention module depicted in Figure 6. The graph attention module takes in  
 710 input node features  $\mathbf{v}$ , edge features  $\mathbf{E}$ , and graph-level features  $\mathbf{y}$ . The FiLM is defined as  
 711  $\text{FiLM}(\mathbf{M}_1, \mathbf{M}_2) = \mathbf{M}_1 \mathbf{W}_1 + (\mathbf{M}_1 \mathbf{W}_2) \odot \mathbf{M}_2 + \mathbf{M}_2$  where  $\mathbf{W}_1, \mathbf{W}_2$  are learnable weights. Also,  
 712 PNA is defined as  $\text{PNA}(\mathbf{v}) = \text{cat}(\max(\mathbf{v}), \min(\mathbf{v}), \text{mean}(\mathbf{v}), \text{std}(\mathbf{v})) \mathbf{W}$  where  $\mathbf{W}$  is a learnable  
 713 weight.

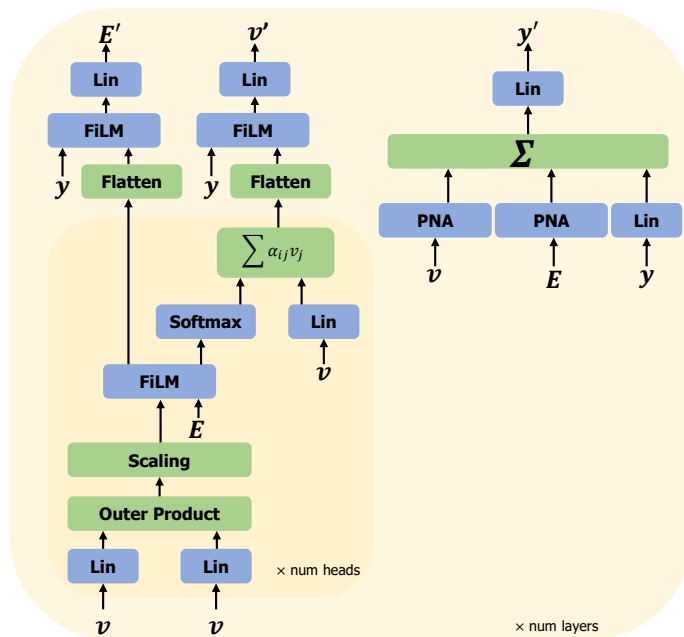


Figure 6: An overview of the graph attention module used in the graph transformer network. The output features are passed through a normalization layer and a fully connected layer at the end.

## 714 A.2 Training

715 Our training runs are done on either an NVIDIA RTX 3090 (24 GB of memory) or V100 (32 GB of  
 716 memory). We train all of our models up to 600 epochs which can take up to 32 hours. We compute  
 717 top- $k$  accuracy metrics on a portion of the validation set every fixed number of epochs and select the  
 718 checkpoint that has the highest top-1 accuracy. The models are trained using a batch size of 32. We  
 719 use AdamW [23] with a learning rate of 0.0002.

## 720 A.3 Additional Features

721 We utilize the additional features proposed by [45] and used in Igashov et al. [17] as input to our  
 722 models. We briefly state these features here for completeness.

723 **Cycles.** Message Passing Neural Networks cannot detect graph cycles, so we add them as features  
 724 using formulas up to cycles of size 6. We compute node-level features (how many cycles does this  
 725 node belong to) up to size 5 and graph-level features (how many cycles does this graph have) up  
 726 to size 6. Fortunately, we can use formulas to compute the graph-level features  $\mathbf{y}_i$  and node-level  
 727 features  $\mathbf{X}_i$ , which can be efficiently computed on the GPU. In the following formulas,  $d$  denotes the  
 728 vector containing node degrees and  $\|\cdot\|_F$  denotes the Frobenius norm:

$$\begin{aligned}\mathbf{X}_3 &= \text{diag}(\mathbf{A}^3)/2 \\ \mathbf{X}_4 &= (\text{diag}(\mathbf{A}^4) - d(d-1) - \mathbf{A}(d\mathbf{1}_n^T)\mathbf{1}_n)/2 \\ \mathbf{X}_5 &= (\text{diag}(\mathbf{A}^5) - 2\text{diag}(\mathbf{A}^3) \odot d - \mathbf{A}((\text{diag}(\mathbf{A}^3)\mathbf{1}_n^T)\mathbf{1}_n) + \text{diag}(\mathbf{A}^3))/2 \\ \mathbf{y}_3 &= \mathbf{X}_3^T \mathbf{1}_n/3 \\ \mathbf{y}_4 &= \mathbf{X}_4^T \mathbf{1}_n/4 \\ \mathbf{y}_5 &= \mathbf{X}_5^T \mathbf{1}_n/5 \\ \mathbf{y}_6 &= \text{Tr}(\mathbf{A}^6) - 3\text{Tr}(\mathbf{A}^3 \odot \mathbf{A}^3) + 9\|\mathbf{A}(\mathbf{A}^2 \odot \mathbf{A}^2)\|_F \\ &\quad - 6\langle \text{diag}(\mathbf{A}^2), \text{diag}(\mathbf{A}^4) \rangle + 6\text{Tr}(\mathbf{A}^4) - 4\text{Tr}(\mathbf{A}^3) \\ &\quad + 4\text{Tr}(\mathbf{A}^2 \mathbf{A}^2 \odot \mathbf{A}^2) + 3\|\mathbf{A}^3\|_F - 12\text{Tr}(\mathbf{A}^2 \odot \mathbf{A}^2) + 4\text{Tr}(\mathbf{A}^2).\end{aligned}$$

729 **Spectral Features.** We compute graph-level features: the number of connected components (which  
 730 is the multiplicity of the 0 eigenvalue), and the first 5 non-zero eigenvalues of the graph Laplacian.  
 731 We also compute node-level features: an estimate of the biggest connected component and the first  
 732 two eigenvectors associated with the first two non-zero eigenvalues. Since molecular graphs in  
 733 USPTO-50k have fewer than 100 nodes, the computation of these spectral features is not a concern.

## 734 B Additional Ablation Studies

735 In this section, we provide some additional ablation studies examining the performance of our  
 736 methods. In Table 5, we evaluate the performance of RETRO PRODFLOW-RS when sampling  
 737  $N = 100$  reactants with  $M = 2$  synthon predictions. We vary  $N_1$ , the number of reactant predictions  
 738 generated for the highest ranking synthon prediction from the reaction center identification model.  
 739 We verify that we need to generate more reactants for the highest-scoring synthon prediction to obtain  
 740 competitive top- $k$  accuracy. Next, we conduct a study to understand how the number of sampling  
 741 steps affects the performance of flow matching compared to RetroBridge. We find that  $T = 50$   
 742 sampling steps is sufficient for RETRO PRODFLOW to obtain SOTA results. Although RetroBridge  
 743 achieves a higher accuracy at  $T = 5$  or  $T = 10$  sampling steps compared to flow matching, both  
 744 methods fail to reach competitive performance. At  $T = 50$  steps, RETRO PRODFLOW achieves some  
 745 improvement over RetroBridge.

746 We also investigate the inference time required by RETRO SYNFLOW to sample  $N = 100$  reactants  
 747 with  $T = 50$  sampling steps. We run this test on an RTX 3090 with 24 GB of memory. The  
 748 average time required to sample 100 reactants for each product molecule in USPTO-50k test dataset  
 749 is  $5.46 \pm 2.95$  seconds.

Table 5: Top- $k$  accuracy (exact match) on the USPTO-50k validation dataset of RETRO SYNFLOW with  $M = 2$  synthon predictions, sampling  $N = 100$  reactants and varying the split. Given  $N_1$ , we have  $N_2 = 100 - N_1$ .

$N_1$	Top- $k$ Accuracy			
	$k = 1$	$k = 3$	$k = 5$	$k = 10$
90	58.2	77.4	81.9	84.4
80	58.3	78.0	82.3	84.7
70	58.1	77.5	82.0	84.6
60	56.6	77.0	81.6	84.5
50	48.5	76.1	81.2	84.2

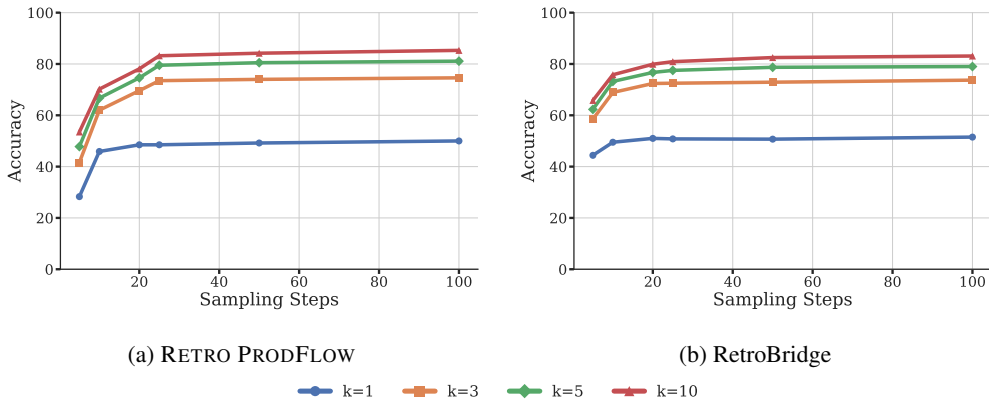


Figure 7: The performance of RETRO PRODFLOW and RetroBridge as we vary the number of sampling steps.

## 750 C Round-trip Visualization

751 This section provides a short case study analyzing the outputs of RETRO PRODFLOW and RETRO  
 752 PRODFLOW-RS with  $K = 2$  particles. We aim to understand how top- $k$  accuracy can decrease  
 753 when applying inference-time steering to guide generations towards outputs that optimize round-trip  
 754 accuracy. We look at the top-1 accuracy results on the USPTO-50k test dataset for simplicity. Our  
 755 main finding from this ablation is that it is still possible for RETRO PRODFLOW to generate reactants  
 756 that are incorrect, *i.e.*, do not match the true reactants and are not feasible, *i.e.*, the forward synthesis  
 757 model prediction does not match the ground-truth product. Table 6 shows how steering based on  
 758 a round-trip reward affects the incorrect/correct prediction made by RETRO PRODFLOW. In total,  
 759 257 correct examples in the test dataset get converted to incorrect examples when applying reward  
 760 steering. On the other hand, 251 incorrect examples are converted to correct examples when applying  
 761 steering. As we increase the number of particles, *i.e.*, increase the strength of steering, this gap  
 762 widens. This results in an overall decrease in exact-match accuracy as we force reactants towards  
 763 more diverse and feasible predictions. Figures 8 and 9 show the visualizations between the outputs  
 764 of RETRO PRODFLOW and RETRO PRODFLOW-RS. The predicted product column refers to the  
 765 prediction of the forward-synthesis model given the predicted reactants as input.

## 766 D Additional Sampling Scheme

767 Recently, there has been increasing interest in developing advanced adaptive sampling schemes for  
 768 discrete diffusion and flow matching models [15, 27, 19, 26]. These developments aim to reduce  
 769 errors in the generation process while improving inference speed. As explained in Section 2.1, we  
 770 update the intermediate sample  $\mathbf{x}_t$  using the following transition kernel,  $\mathbf{x}_{t+h}^i \sim \text{Cat}(\mathbf{x}_{t+h}^i; \delta(\mathbf{x}_t^i) +$   
 771  $hu_t^i(\mathbf{x}_{t+h}^i, \mathbf{x}_t))$ , which is analogous to the Euler update step in continuous flow matching. Inspired

Table 6: Top-1 predicted reactants from RPF and RPF-RS quantified into four categories. The round-trip match column indicates whether the prediction made by RPF-RS is a round-trip match.

RPF	RPF-RS	Round-Trip Match	Count	Percentage
Correct	Incorrect	T	225	4.5
Correct	Incorrect	F	32	0.64
Incorrect	Correct	T	226	4.5
Incorrect	Correct	F	25	0.50
Correct	Correct	T	1848	36.9
Correct	Correct	F	388	7.7
Incorrect	Incorrect	T	1810	36.1
Incorrect	Incorrect	F	453	9.0

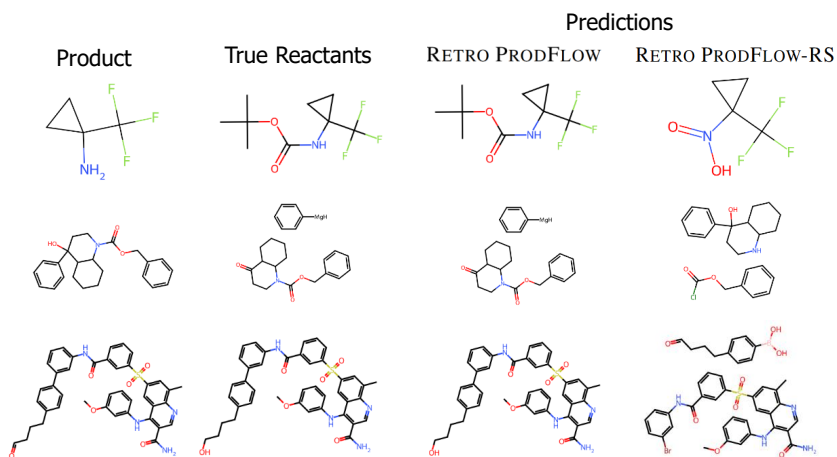


Figure 8: A visualization of reactions where RETRO PRODFLOW-RS generates an incorrect reactant prediction that is still feasible, while the RETRO PRODFLOW generates the correct reactant prediction. There are 225 examples in the test set that correspond to this case.

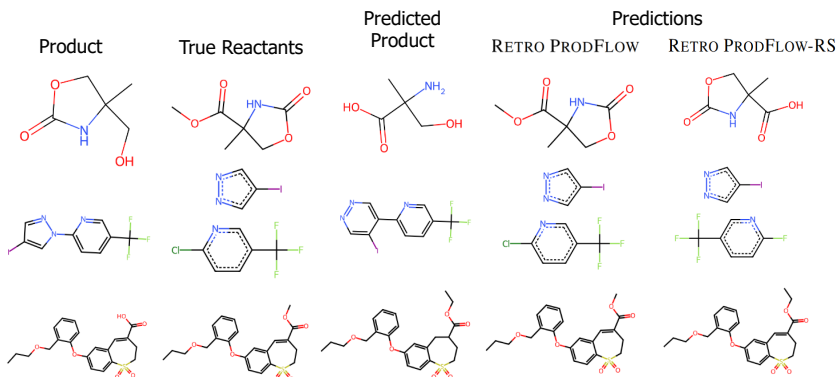


Figure 9: A visualization of reactions where RETRO PRODFLOW-RS generates an incorrect reactant prediction that is infeasible. RETRO PRODFLOW generates the correct reactant prediction. There are 32 examples in the test set that correspond to this case.

by this analogy, we explore a higher-order sampling scheme [27] based on the Runge-Kutta (RK) method for solving ODEs. The update step for the method is as follows:

$$\hat{\mathbf{x}}_{t+h}^i \sim \text{Cat}(\hat{\mathbf{x}}_{t+h}^i; \delta(\mathbf{x}_t^i) + hu_t^i(\hat{\mathbf{x}}_{t+h}^i, \mathbf{x}_t)) \quad (3)$$

$$\mathbf{x}_{t+h}^i \sim \text{Cat}\left(\mathbf{x}_{t+h}^i; \delta(\mathbf{x}_t^i) + \frac{1}{2}hu_t^i(\mathbf{x}_{t+h}^i, \mathbf{x}_t) + \frac{1}{2}hu_{t+h}^i(\mathbf{x}_{t+h}^i, \hat{\mathbf{x}}_{t+h}^i)\right). \quad (4)$$

This update step requires two model evaluations from  $p_\theta$  instead of one. Table 7 compares the performance of RETRO PRODFLOW using this sampling scheme with 25 steps against RETRO PRODFLOW using the Euler-inspired sampling scheme with 50 steps.

Table 7: Top- $k$  accuracy of RPF on the USPTO-50k test set sampling  $N = 50$  reactants per product.

Model	1	3	5	10
RPF	49.6	73.3	79.6	83.6
RPF (RK 25 steps)	49.3	71.2	76.4	80.0
RPF (RK 50 steps)	49.3	72.5	78.6	82.3

## E Predictions Visualization

We provide some additional visualizations of the generated reactants from our methods. In the following figures, an “E” represents an exact-match between the prediction reactant and an “R” indicates a round-trip match but not an exact-match. We show the top-3 reactants.

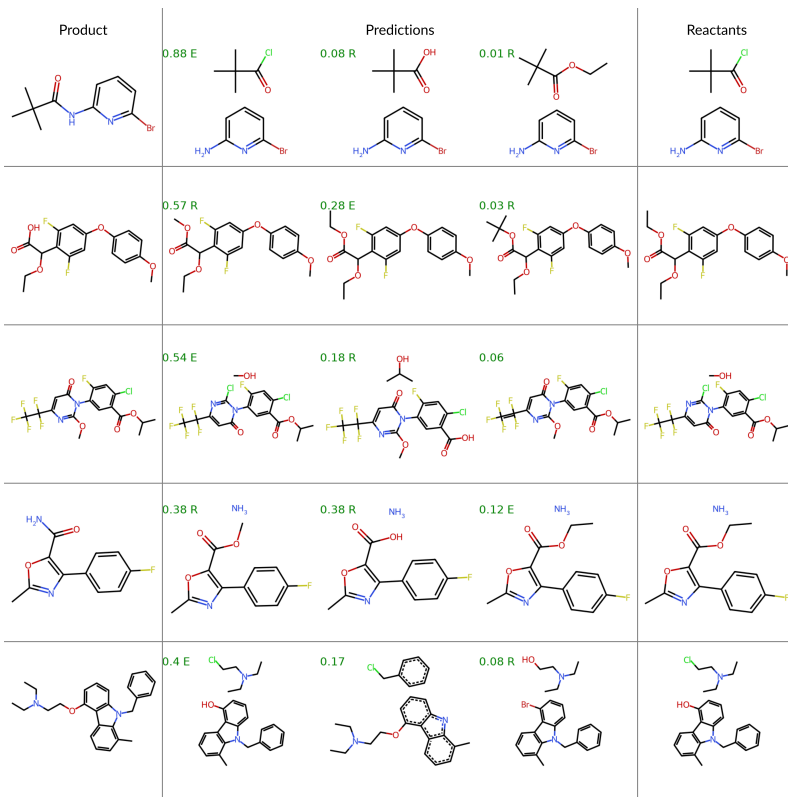


Figure 10: Visualizations of predictions made by RETRO PRODFLOW. Examples are taken from the USPTO-50k test set randomly.

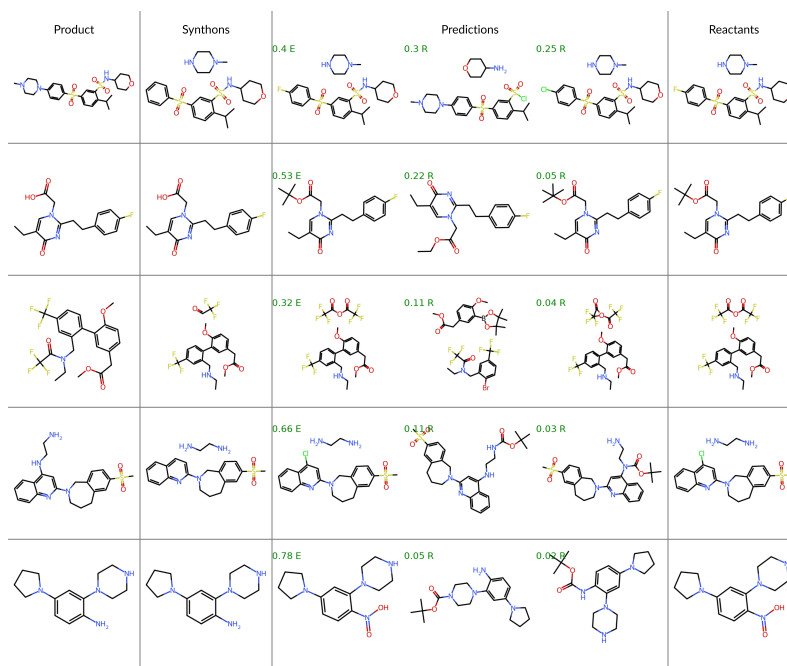


Figure 11: Visualizations of predictions made by RETRO SYNFLOW. Examples are taken from the USPTO-50k test set randomly.

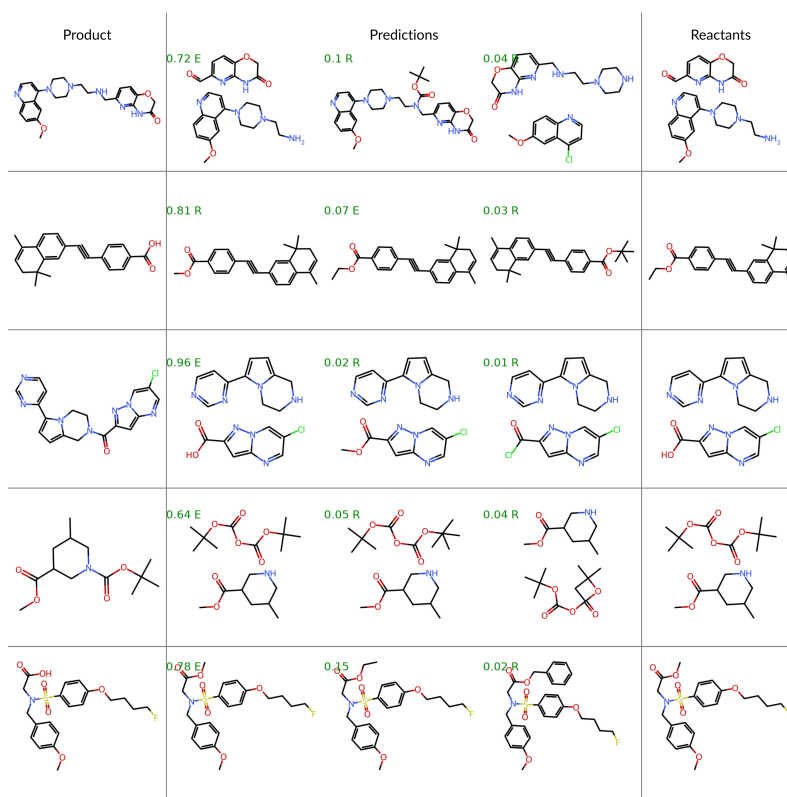


Figure 12: Visualizations of predictions made by RETRO PRODFLOW-RS. Examples are taken from the USPTO-50k test set randomly.

## References

- [1] J. Bose, T. Akhound-Sadegh, G. Huguet, K. Fatras, J. Rector-Brooks, C.-H. Liu, A. C. Nica, M. Korablyov, M. M. Bronstein, and A. Tong. SE(3)-Stochastic Flow Matching for Protein Backbone Generation. Oct. 2023. URL <https://openreview.net/forum?id=kJFIH23hXb>. (Cited on page 9)
- [2] A. Campbell, J. Yim, R. Barzilay, T. Rainforth, and T. Jaakkola. Generative flows on discrete state-spaces: Enabling multimodal flows with applications to protein co-design. *arXiv preprint arXiv:2402.04997*, 2024. (Cited on page 2)
- [3] B. Chen, T. Shen, T. S. Jaakkola, and R. Barzilay. Learning to Make Generalizable and Diverse Predictions for Retrosynthesis, Oct. 2019. URL <http://arxiv.org/abs/1910.09688>. arXiv:1910.09688 [cs]. (Cited on page 1)
- [4] S. Chen and Y. Jung. Deep Retrosynthetic Reaction Prediction using Local Reactivity and Global Attention. *JACS Au*, 1(10):1612–1620, Oct. 2021. doi: 10.1021/jacsau.1c00246. URL <https://doi.org/10.1021/jacsau.1c00246>. Publisher: American Chemical Society. (Cited on page 6)
- [5] C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, and K. F. Jensen. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Central Science*, 3(5):434–443, May 2017. ISSN 2374-7943. doi: 10.1021/acscentsci.7b00064. URL <https://doi.org/10.1021/acscentsci.7b00064>. Publisher: American Chemical Society. (Cited on page 9)
- [6] C. W. Coley, L. Rogers, W. H. Green, and K. F. Jensen. Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Central Science*, 3(12):1237–1245, Dec. 2017. ISSN 2374-7943. doi: 10.1021/acscentsci.7b00355. URL <https://doi.org/10.1021/acscentsci.7b00355>. Publisher: American Chemical Society. (Cited on page 1)
- [7] E. J. Corey and W. T. Wipke. Computer-Assisted Design of Complex Organic Syntheses. *Science*, 166(3902):178–192, Oct. 1969. doi: 10.1126/science.166.3902.178. URL <https://www.science.org/doi/10.1126/science.166.3902.178>. Publisher: American Association for the Advancement of Science. (Cited on page 1)
- [8] H. Dai, C. Li, C. Coley, B. Dai, and L. Song. Retrosynthesis Prediction with Conditional Graph Logic Network. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL <https://proceedings.neurips.cc/paper/2019/hash/0d2b2061826a5df3221116a5085a6052-Abstract.html>. (Cited on pages 1, 6, and 9)
- [9] I. Dunn and D. R. Koes. Mixed Continuous and Categorical Flow Matching for 3D De Novo Molecule Generation, Apr. 2024. URL <http://arxiv.org/abs/2404.19739>. arXiv:2404.19739 [q-bio]. (Cited on page 9)
- [10] V. P. Dwivedi and X. Bresson. A Generalization of Transformer Networks to Graphs, Jan. 2021. URL <http://arxiv.org/abs/2012.09699>. arXiv:2012.09699 [cs]. (Cited on page 17)
- [11] F. Eijkelboom, G. Bartosh, C. A. Naesseth, M. Welling, and J.-W. van de Meent. Variational Flow Matching for Graph Generation. *arXiv preprint arXiv:2406.04843*, 2024. (Cited on page 3)
- [12] P. Gaiński, M. Koziarski, K. Maziarz, M. Segler, J. Tabor, and M. Śmieja. RetroGFN: Diverse and Feasible Retrosynthesis using GFlowNets, Apr. 2025. URL <http://arxiv.org/abs/2406.18739>. arXiv:2406.18739 [cs]. (Cited on pages 6, 7, and 9)
- [13] I. Gat, T. Remez, N. Shaul, F. Kreuk, R. T. Chen, G. Synnaeve, Y. Adi, and Y. Lipman. Discrete Flow Matching. *arXiv preprint arXiv:2407.15595*, 2024. (Cited on page 2)
- [14] M. Hartenfeller, M. Eberle, P. Meier, C. Nieto-Oberhuber, K.-H. Altmann, G. Schneider, E. Jacoby, and S. Renner. A Collection of Robust Organic Synthesis Reactions for In Silico Molecule Design. *Journal of Chemical Information and Modeling*, 51(12):3093–3098, Dec. 2011. ISSN 1549-9596. doi: 10.1021/ci200379p. URL <https://doi.org/10.1021/ci200379p>. Publisher: American Chemical Society. (Cited on page 9)



- [15] P. Holderrieth, M. S. Albergo, and T. Jaakkola. LEAPS: A discrete neural sampler via locally equivariant networks, Feb. 2025. URL <http://arxiv.org/abs/2502.10843>. arXiv:2502.10843 [cs]. (Cited on page 19)
- [16] G. Huguet, J. Vuckovic, K. Fatras, E. Thibodeau-Laufer, P. Lemos, R. Islam, C.-H. Liu, J. Rector-Brooks, T. Akhound-Sadegh, M. Bronstein, A. Tong, and A. J. Bose. Sequence-Augmented SE(3)-Flow Matching For Conditional Protein Backbone Generation, Dec. 2024. URL <http://arxiv.org/abs/2405.20313>. arXiv:2405.20313 [cs]. (Cited on page 9)
- [17] I. Igashov, A. Schneuing, M. Segler, M. M. Bronstein, and B. Correia. RetroBridge: Modeling Retrosynthesis with Markov Bridges. Oct. 2023. URL <https://openreview.net/forum?id=770DetV8He>. (Cited on pages 4, 6, 7, 9, 17, and 18)
- [18] E. Kim, D. Lee, Y. Kwon, M. S. Park, and Y.-S. Choi. Valid, Plausible, and Diverse Retrosynthesis Using Tied Two-Way Transformers with Latent Variables. *Journal of Chemical Information and Modeling*, 61(1):123–133, Jan. 2021. ISSN 1549-9596. doi: 10.1021/acs.jcim.0c01074. URL <https://doi.org/10.1021/acs.jcim.0c01074>. Publisher: American Chemical Society. (Cited on page 6)
- [19] J. Kim, K. Shah, V. Kontonis, S. Kakade, and S. Chen. Train for the Worst, Plan for the Best: Understanding Token Ordering in Masked Diffusions, Mar. 2025. URL <http://arxiv.org/abs/2502.06768>. arXiv:2502.06768 [cs]. (Cited on page 19)
- [20] G. Landrum. RDKit: Open-Source Cheminformatics Software | BibSonomy. URL <https://www.bibsonomy.org/bibtex/28d01fceecd6bf2486e47d7c4207b108/salotz>. (Cited on page 17)
- [21] B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. Luu Nguyen, S. Ho, J. Sloane, P. Wender, and V. Pande. Retrosynthetic Reaction Prediction Using Neural Sequence-to-Sequence Models. *ACS Central Science*, 3(10):1103–1113, Oct. 2017. ISSN 2374-7943. doi: 10.1021/acscentsci.7b00303. URL <https://doi.org/10.1021/acscentsci.7b00303>. Publisher: American Chemical Society. (Cited on page 9)
- [22] L. Long, R. Li, and J. Zhang. Artificial Intelligence in Retrosynthesis Prediction and its Applications in Medicinal Chemistry. *Journal of Medicinal Chemistry*, 68(3):2333–2355, Feb. 2025. ISSN 0022-2623. doi: 10.1021/acs.jmedchem.4c02749. URL <https://doi.org/10.1021/acs.jmedchem.4c02749>. Publisher: American Chemical Society. (Cited on page 1)
- [23] I. Loshchilov and F. Hutter. Decoupled Weight Decay Regularization, Jan. 2019. URL <http://arxiv.org/abs/1711.05101>. arXiv:1711.05101 [cs]. (Cited on page 18)
- [24] S. Nie, F. Zhu, Z. You, X. Zhang, J. Ou, J. Hu, J. Zhou, Y. Lin, J.-R. Wen, and C. Li. Large Language Diffusion Models, Feb. 2025. URL <http://arxiv.org/abs/2502.09992>. arXiv:2502.09992 [cs]. (Cited on page 9)
- [25] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL <https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>. (Cited on page 17)
- [26] F. Z. Peng, Z. Bezemek, S. Patel, J. Rector-Brooks, S. Yao, A. Tong, and P. Chatterjee. Path Planning for Masked Diffusion Model Sampling, Feb. 2025. URL <http://arxiv.org/abs/2502.03540>. arXiv:2502.03540 [cs]. (Cited on page 19)
- [27] Y. Ren, H. Chen, Y. Zhu, W. Guo, Y. Chen, G. M. Rotskoff, M. Tao, and L. Ying. Fast Solvers for Discrete Diffusion Models: Theory and Applications of High-Order Algorithms, Feb. 2025. URL <http://arxiv.org/abs/2502.00234>. arXiv:2502.00234 [cs]. (Cited on pages 19 and 21)



- [28] M. Sacha, M. Błaż, P. Byrski, P. Dąbrowski-Tumański, M. Chromiński, R. Loska, P. Włodarczyk-Pruszyński, and S. Jastrzębski. Molecule Edit Graph Attention Network: Modeling Chemical Reactions as Sequences of Graph Edits. *Journal of Chemical Information and Modeling*, 61(7):3273–3284, July 2021. ISSN 1549-9596. doi: 10.1021/acs.jcim.1c00537. URL <https://doi.org/10.1021/acs.jcim.1c00537>. Publisher: American Chemical Society. (Cited on pages 6 and 9)
- [29] N. Schneider, N. Stiefl, and G. A. Landrum. What’s What: The (Nearly) Definitive Guide to Reaction Role Assignment. *Journal of Chemical Information and Modeling*, 56(12):2336–2346, Dec. 2016. ISSN 1549-9596. doi: 10.1021/acs.jcim.6b00564. URL <https://doi.org/10.1021/acs.jcim.6b00564>. (Cited on page 6)
- [30] P. Schwaller, R. Petraglia, V. Zullo, V. H. Nair, R. Andreas Haeuselmann, R. Pisoni, C. Bekas, A. Iuliano, and T. Laino. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chemical Science*, 11(12):3316–3325, 2020. doi: 10.1039/C9SC05704H. URL <https://pubs.rsc.org/en/content/articlelanding/2020/sc/c9sc05704h>. Publisher: Royal Society of Chemistry. (Cited on page 6)
- [31] M. H. S. Segler and M. P. Waller. Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. *Chemistry – A European Journal*, 23(25):5966–5971, 2017. ISSN 1521-3765. doi: 10.1002/chem.201605499. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/chem.201605499>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/chem.201605499>. (Cited on pages 1 and 9)
- [32] S.-W. Seo, Y. Y. Song, J. Y. Yang, S. Bae, H. Lee, J. Shin, S. J. Hwang, and E. Yang. GTA: Graph Truncated Attention for Retrosynthesis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1):531–539, May 2021. ISSN 2374-3468. doi: 10.1609/aaai.v35i1.16131. URL <https://ojs.aaai.org/index.php/AAAI/article/view/16131>. Number: 1. (Cited on pages 6 and 9)
- [33] C. Shi, M. Xu, H. Guo, M. Zhang, and J. Tang. A Graph to Graphs Framework for Retrosynthesis Prediction. In *Proceedings of the 37th International Conference on Machine Learning*, pages 8818–8827. PMLR, Nov. 2020. URL <https://proceedings.mlr.press/v119/shi20d.html>. ISSN: 2640-3498. (Cited on pages 2, 5, 6, and 9)
- [34] R. Singhal, Z. Horvitz, R. Teehan, M. Ren, Z. Yu, K. McKeown, and R. Ranganath. A General Framework for Inference-time Scaling and Steering of Diffusion Models, Jan. 2025. URL <http://arxiv.org/abs/2501.06848>. arXiv:2501.06848 [cs]. (Cited on pages 2 and 4)
- [35] V. R. Somnath, C. Bunne, C. Coley, A. Krause, and R. Barzilay. Learning Graph Models for Retrosynthesis Prediction. In *Advances in Neural Information Processing Systems*, volume 34, pages 9405–9415. Curran Associates, Inc., 2021. URL [https://proceedings.neurips.cc/paper\\_files/paper/2021/hash/4e2a6330465c8ffcaa696a5a16639176-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2021/hash/4e2a6330465c8ffcaa696a5a16639176-Abstract.html). (Cited on pages 5 and 6)
- [36] Y. Song, J. Gong, M. Xu, Z. Cao, Y. Lan, S. Ermon, H. Zhou, and W.-Y. Ma. Equivariant Flow Matching with Hybrid Probability Transport for 3D Molecule Generation. In *Neural Information Processing Systems 2023*, Nov. 2023. URL <https://openreview.net/forum?id=hHUZ5V9XFu>. (Cited on page 9)
- [37] M. Stanley and M. Segler. Fake it until you make it? Generative de novo design and virtual screening of synthesizable molecules. *Current Opinion in Structural Biology*, 82:102658, Oct. 2023. ISSN 0959-440X. doi: 10.1016/j.sbi.2023.102658. URL <https://www.sciencedirect.com/science/article/pii/S0959440X2300132X>. (Cited on page 1)
- [38] H. Stark, B. Jing, C. Wang, G. Corso, B. Berger, R. Barzilay, and T. Jaakkola. Dirichlet Flow Matching with Applications to DNA Sequence Design, May 2024. URL <http://arxiv.org/abs/2402.05841>. arXiv:2402.05841 [q-bio]. (Cited on page 9)
- [39] F. Strieth-Kalthoff, F. Sandfort, M. H. Segler, and F. Glorius. Machine learning the ropes: principles, applications and directions in synthetic chemistry. *Chemical Society Reviews*, 49(17):6154–6168, 2020. (Cited on page 1)

- [40] R. Sun, H. Dai, L. Li, S. Kearnes, and B. Dai. Towards understanding retrosynthesis by energy-based models. In *Advances in Neural Information Processing Systems*, volume 34, pages 10186–10194. Curran Associates, Inc., 2021. URL <https://proceedings.neurips.cc/paper/2021/hash/5470abe68052c72afb19be45bb418d02-Abstract.html>. (Cited on pages 6 and 9)
- [41] S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk, and B. A. Grzybowski. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angewandte Chemie International Edition*, 55(20):5904–5937, 2016. ISSN 1521-3773. doi: 10.1002/anie.201506101. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201506101>. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201506101>. (Cited on page 9)
- [42] I. V. Tetko, P. Karpov, R. Van Deursen, and G. Godin. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nature Communications*, 11(1): 5575, Nov. 2020. ISSN 2041-1723. doi: 10.1038/s41467-020-19266-y. URL <https://www.nature.com/articles/s41467-020-19266-y>. Publisher: Nature Publishing Group. (Cited on pages 1, 6, and 9)
- [43] Z. Tu and C. W. Coley. Permutation Invariant Graph-to-Sequence Model for Template-Free Retrosynthesis and Reaction Prediction. *Journal of Chemical Information and Modeling*, 62(15):3503–3513, Aug. 2022. ISSN 1549-9596. doi: 10.1021/acs.jcim.2c00321. URL <https://doi.org/10.1021/acs.jcim.2c00321>. Publisher: American Chemical Society. (Cited on pages 6 and 9)
- [44] U. V. Ucak, I. Ashyrmamatov, J. Ko, and J. Lee. Retrosynthetic reaction pathway prediction through neural machine translation of atomic environments. *Nature Communications*, 13(1): 1186, Mar. 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-28857-w. URL <https://www.nature.com/articles/s41467-022-28857-w>. Publisher: Nature Publishing Group. (Cited on page 2)
- [45] C. Vignac, I. Krawczuk, A. Siraudin, B. Wang, V. Cevher, and P. Frossard. Digress: Discrete denoising diffusion for graph generation. *arXiv preprint arXiv:2209.14734*, 2022. (Cited on pages 17 and 18)
- [46] Y. Wan, C.-Y. Hsieh, B. Liao, and S. Zhang. Retroformer: Pushing the Limits of End-to-end Retrosynthesis Transformer. In *Proceedings of the 39th International Conference on Machine Learning*, pages 22475–22490. PMLR, June 2022. URL <https://proceedings.mlr.press/v162/wan22a.html>. ISSN: 2640-3498. (Cited on page 9)
- [47] C. Yan, Q. Ding, P. Zhao, S. Zheng, J. YANG, Y. Yu, and J. Huang. RetroXpert: Decompose Retrosynthesis Prediction Like A Chemist. In *Advances in Neural Information Processing Systems*, volume 33, pages 11248–11258. Curran Associates, Inc., 2020. URL <https://proceedings.neurips.cc/paper/2020/hash/819f46e52c25763a55cc642422644317-Abstract.html>. (Cited on page 5)
- [48] S. Zheng, J. Rao, Z. Zhang, J. Xu, and Y. Yang. Predicting Retrosynthetic Reactions Using Self-Corrected Transformer Neural Networks. *Journal of Chemical Information and Modeling*, 60(1):47–55, Jan. 2020. ISSN 1549-9596. doi: 10.1021/acs.jcim.9b00949. URL <https://doi.org/10.1021/acs.jcim.9b00949>. Publisher: American Chemical Society. (Cited on pages 6 and 9)