SYNFLOWNET: TOWARDS MOLECULE DESIGN WITH GUARANTEED SYNTHESIS PATHWAYS

Miruna Cretu
University of Cambridge
mtc49@cam.ac.uk

Charles Harris
University of Cambridge
cch57@cam.ac.uk

Julien Roy
Valence Labs
julien.roy@recursionpharma.com

Emmanuel Bengio
Valence Labs
emmanuel.bengio@recursionpharma.com

Pietro Liò
University of Cambridge
pl219@cam.ac.uk

ABSTRACT

Recent breakthroughs in generative modelling have led to a number of works proposing molecular generation models for drug discovery. While these models perform well at capturing drug-like motifs, they are known to often produce synthetically inaccessible molecules. This is because they are trained to compose atoms or fragments in a way that approximates the training distribution, but they are not explicitly aware of the synthesis constraints that come with making molecules in the lab. To address this issue, we introduce SynFlowNet, a GFlowNet model whose action space uses chemically validated reactions and reactants to sequentially build new molecules. We evaluate our approach using synthetic accessibility scores and an independent retrosynthesis tool. SynFlowNet consistently samples synthetically feasible molecules, while still being able to find diverse and high-utility candidates. Furthermore, we compare molecules designed with SynFlowNet to experimentally validated actives, and find that they show comparable properties of interest, such as molecular weight, SA score and predicted protein binding affinity.

1 INTRODUCTION

Designing molecules with targeted biochemical properties stands as a critical challenge in pharmaceutical discovery. There has been increased attention towards employing generative models as a substitute to traditional molecular design methodologies, with the goal of enhancing chemical space exploration and streamlining the design process. Such works include variational autoencoders (Gómez-Bombarelli et al., 2018; Alperstein et al., 2019), deep reinforcement learning (Olivecrona et al., 2017), generative adversarial networks (Guimaraes et al., 2018; Cao & Kipf, 2022), normalizing flows (Zang & Wang, 2020; Shi et al., 2020), and diffusion models (Hoogeboom et al., 2022; Schneuing et al., 2023).

However, most of the *de novo* design models today do not account for important constraints, such as physical plausibility (Harris et al., 2023) or synthetic accessibility, thus leaving no guarantee that the sampled molecules can be created in the real world (Gao & Coley, 2020). Synthetic complexity scores (Ertl & Schuffenhauer, 2009; Coley et al., 2018) have been proposed as a method to assess molecules and complement generative models with knowledge on synthetic accessibility, however these heuristics and learned metrics are often oversimplified. There also have been concurrent efforts to achieve computer-aided synthesis planning, as a subsequent step to molecule generation (Coley et al., 2017; Schwaller et al., 2020). However, these approaches encounter obstacles when tasked with devising synthetic pathways for molecules inherently challenging or impossible to synthesize, thereby sometimes rendering the proposed synthesis routes impractical.

Synthetically-accessible chemical spaces can be assembled from combinations of reactions involving readily available reactants, resulting in libraries of tens of billions of molecules (Klarich et al., 2024). However, virtual screening on these datasets is impractical. It is estimated that screening 10 billion compounds using a single central processing unit can take more than 3,000 years (Sadybekov et al., 2022). The size of such spaces grows exponentially with the number of available reactants, and with

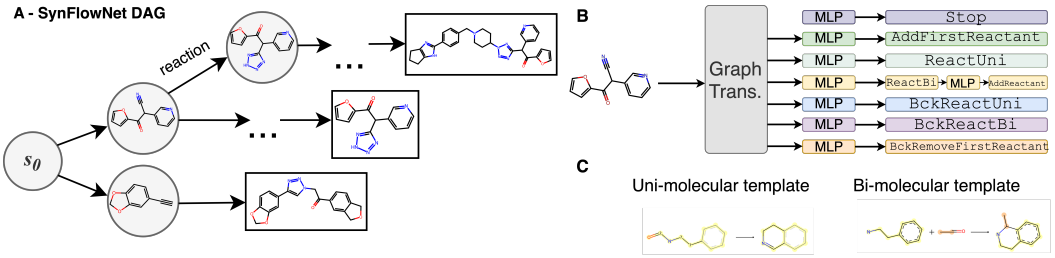

Figure 1: SynFlowNet allows for synthesis-aware molecule generation. (A) The state space of SynFlowNet generates molecules from purchasable building blocks and chemical reactions. Every final molecule is associated with a reward. Training trajectories are constructed by sampling the model forward. (B) Our policy $\pi(a|s)$ is parameterised as a graph transformer which outputs embeddings that are passed into 8 separate MLPs to predict the action logits for different action types. (C) Transitions from a state to the next happens via uni- or bi-molecular reaction templates.

thousands of novel building blocks released each year, new methods to navigate the synthesizable space are needed.

Given the challenge of synthesisability for atom- and fragment-based generative models, we propose to formulate the problem of molecule generation in an action space of chemical reactions. A reaction-based action space guarantees synthesisability, and we show that even a random sampler in this space is able to produce molecules with competitive scores compared to a fragment-based model.

Secondly, we would like our sampler to maximise desired properties, and at the same time retrieve molecules from distinct modes of the available synthetic space. Generative Flow Networks (GFlowNets) (Bengio et al., 2021) emerged as a framework that, by design, is able to generate diverse samples. GFlowNets were shown to require fewer evaluation steps of the reward function to generate samples with high reward and diversity compared to alternatives such as Markov Chain Monte Carlo or Proximal Policy Optimisation (Bengio et al., 2021).

In this work, we introduce SynFlowNet, a GFlowNet specifically trained to generate molecules from available chemical reactions and purchasable compounds. Our approach thus constrains the exploration of targets to a synthetically-accessible chemical space, sampling not only target compounds, but also synthetic routes that lead to them. In summary, our main contributions are:

- We train a GFlowNet using an action space defined by documented chemical reactions and purchasable starting materials to generate molecules that are synthesisable.
- We find that SynFlowNet is able to generate molecules with overall better scores (diversity, drug-likeness, protein binding affinity) compared to a GFlowNet with an action space composed of molecular fragments, and comparable scores to experimentally validated molecules.

1.1 Background and Related Work

GFlowNets GFlowNets (Bengio et al., 2021) are a class of probabilistic models that learn a stochastic policy to generate objects x through a sequence of actions, with probability proportional to a reward $R(x)$. The sequential construction of objects x can be described as a trajectory $\tau \in \mathcal{T}$ in a directed acyclic graph (DAG) $G = (S, \mathcal{E})$, starting from an initial state s_0 and using actions a to transition from a state to the next: $s \xrightarrow{a} s'$. A GFlowNet uses a forward policy $P_F(-|s)$, which is a distribution over the children of state s, to sample a sequence of actions based on the current states. Similarly, a backward policy $P_B(-|s)$ is the distribution over the parents of state s, and can be used to calculate probabilities of backward actions, leading from terminal to initial states. The training objective which we adopt in this paper is trajectory balance (Malkin et al., 2022). This learns the initial state flow Z_θ, and the policies $P_F(-|s;\theta)$ and $P_B(-|s;\theta)$ parameterized by θ, where $Z = F(s_0) = \sum_{\tau \in \mathcal{T}} F(\tau)$.

Bengio et al. (2021) have used GFlowNets to generate molecules with high binding affinity to a protein target by linking fragments to form a junction tree (Jin et al., 2019). The framework was extended to multi-objective optimisation (Jain et al., 2023; Roy et al., 2023), where the model was trained to simultaneously optimise for binding affinity to the protein target, Synthetic Accessibility (SA), drug likeness (QED) and molecular weight. Recently, Shen et al. (2023) extended the framework to pocket-conditioned molecular generation, employing the same action space to generate molecules conditioned on different protein targets.

Synthesis-aware molecule generation The idea of tackling molecule generation and synthesis simultaneously has been addressed before, first by SYNOPSIS (Vinkers et al., 2003), which generates molecules from a starting dataset of available compounds, relying on applying chemical modifications to functional groups and assessing the value of the product with a fitness function. Works such as Bradshaw et al. (2019); Korovina et al. (2020) followed, which use neural models for one-step synthetic pathways. Gottipati et al. (2020) used a reinforcement learning setup based on policy gradient to generate compounds from reactions and commercially available reactants. Gao et al. (2021) formulate a Markov decision process to model the generation of synthesis trees, which can be optimized with respect to the desired properties of a product molecule.

Our work resembles a reinforcement learning setting for synthesis-aware molecular generation (Gottipati et al., 2020; Horwood & Noutahi, 2020). The key difference lies in the sampling distribution of the learned model. Contrary to RL, the GFlownet objective is not to generate single highest-return sequence of actions, but rather to maximise both performance and diversity by sampling terminal states proportionally to their reward. This is especially useful in the context of molecule generation, where we want to explore different modes of the distribution of interest.

2 METHODS

Problem definition We model synthetic pathways as trajectories in a GFlowNet, starting from purchasable compounds and ending with molecules that are optimized for some desired properties. At a given timestep t, the state s_t represents the current molecule and stepping forward in the environment consists in sequentially building up on that molecule by applying new pairs of reactions and reactants until either a termination action is chosen or the path reaches a maximum length. We encode chemical reactions using SMARTS templates, which are patterns that capture the structural changes occurring in a reaction by specifying which atoms and bonds are involved and how they are modified. Our method systematically constructs a synthetic pathway, applying one reaction at a time, while also optimizing for desired molecular properties captured by a reward function.

Our model is trained in an online fashion, meaning that it learns exclusively from trajectories sampled from the GFlowNet policy, without relying on an external dataset of trajectories or a set of target molecules. Note however that this framework allows for training on such datasets (such a scheme is regarded as offline training).

Action space We define five types of forward actions divided into two hierarchical levels (see Figure 1). The first level includes the actions `Stop`, `AddFirstReactant`, `ReactUni` and `ReactBi`, where `ReactUni` and `ReactBi` respectively represent uni-molecular and bi-molecular reactions. The second action level consists of the `AddReactant` action, which is sampled only if a `ReactBi` was sampled in the first level. More precisely, each trajectory starts from an empty molecular graph which is followed by a building block sampled from `AddFirstReactant`. We then continue based on the sampled action type as follows: (a) if the action type is `Stop`, we reach a terminal state and end the trajectory; (b) if a `ReactUni` action is sampled, we apply the uni-molecular reaction template to the molecule in state s to yield product molecule in state s'; (c) if the action type is `ReactBi`, the sampled reaction is used as input to an MLP, together with the state embedding, to sample a subsequent action of type `AddReactant`. This action introduces a second molecule which reacts with the molecule in state s according to the bi-molecular reaction template. Pre-computed masks ensure that only compatible reaction templates and reactants are sampled.

The backward actions types that we use are `BckReactUni`, `BckReactBi` and `BckRemoveFirstReactant`. To unfold a reverse trajectory we proceed similarly: (a) if the action-type is a `BckReactUni`, the action yields the reactant molecule directly; (b) if the

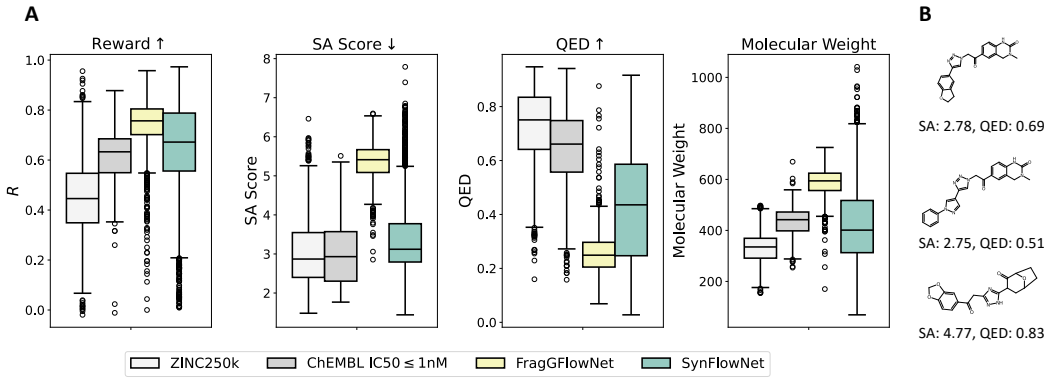

Figure 2: (A) Distributions of metrics calculated for molecules generated using SynFlowNet and the fragment-based GFlowNet. The reward reflects the predicted binding affinity to the protein target we optimise for. ZINC250k molecules are used as control and a subset of 313 ChEMBL molecules has been selected for high binding affinity against the SeH target using an IC_{50} threshold of 1nM. The model was trained to maximise the reward and the other metrics are used for analysis only. (B) Three of the highest-reward molecules sampled from SynFlowNet.

action type is `BckReactBi`, we obtain two reactants, and the molecule that is **not** a building block becomes the *next* state (or *previous* state in the DAG). If the two resulting reactants are both building blocks (which happens at the beginning of the forward trajectory), the molecule that populates the next state is picked at random from the two building blocks. The last action in a backward trajectory is `BckRemoveFirstReactant`, leading to the empty molecular graph s_0.

Model A graph neural network based on a graph transformer architecture (Yun et al., 2019) is used to parameterize the forward and backward policies. The model's action space is defined using separate MLPs for each action type (see Figure 1). Model hyperparameters and training details are discussed in the Appendix A.

Reward The reward function is defined as the normalized negative binding energy as predicted by a pretrained proxy model, available from Bengio et al. (2021) and trained on molecules docked with AutoDockVina (Trott & Olson, 2010) for the sEH (soluble epoxide hydrolase) protein target, a well studied protein which plays part in respiratory and heart disease (Imig & Hammock, 2009).

Data and Implementation We use reaction templates from the two publicly available template libraries (Button et al., 2019; Hartenfeller M, 2012). These give a total of 71 reaction templates: 13 uni- and 58 bi-molecular reactions respectively. We use 6000 randomly selected molecular building blocks from the Enamine Building Blocks (Global stock, accessed November 28, 2023) (Enamine). All reactions were run using RDKit's `RunReactants` function.

3 Experiments

We set out to answer the following questions: (1) Is a GFlowNet trained with an action space of chemical reactions capable to generate molecules of similar reward and diversity compared to GFlowNets using the previously-defined molecular-fragment action spaces? (2) Does an action space of chemical reactions succeed in addressing the challenge of synthesisability in molecular design?

We train the GFlowNet with a reaction templates action space to generate molecules with rewards proportional to sEH binding affinity and compare the results to a GFlowNet using an action space defined by the manipulation of molecular fragments (Bengio et al., 2021; Jin et al., 2019). To achieve comparable vocabularies in the state spaces, we replace the molecular fragments used in the original GFlowNet paper with fragments generated from the same library of reactants used by our model. The final evaluation consits of 1600 compounds sampled for each models. We also compare these

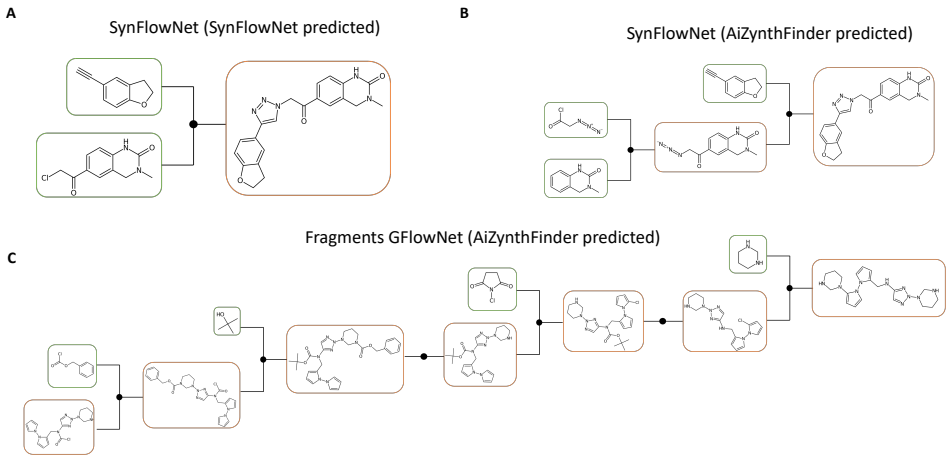

Figure 3: SynFlowNet generates molecules as well as their synthetic pathways. AiZynthFinder retrosynthetic analysis was used to further assess synthesisability. Fragment GFlowNet syntheses are longer on average. (A) SynFlowNet synthesis for the highest reward molecule. Note that an additional azide reagent is involved in the reaction; (B) AiZynthFinder synthesis (top score) for the same molecule. (C) AiZynthFinder synthesis (top score) for the top molecule sampled from the fragments GFlowNet. Molecules with a green frame are purchasable (found in stock), contrary to those with orange frames.

generated molecules against experimentally validated bioactive molecules from ChEMBL (Zdrazil et al., 2023).

Evaluation metrics To evaluate the quality of generated molecules, we used the *QED* (Bickerton et al., 2012) metric, which estimates drug-likeness, the *SA Score* and *SCScore* for synthesisability (Ertl & Schuffenhauer, 2009; Coley et al., 2018) and relative *ligand efficiency*, which is a measure of binding affinity per atom of a ligand to its binding partner. We estimate ligand efficiency by dividing the predicted reward (which is proportional to binding affinity) by atom count. Additionally, we report the *chemical validity*, *uniqueness* and *diversity* of the generated molecules. The latter was estimated using pairwise Tanimoto distances between Morgan fingerprints.

To further assess the synthetic accessibility of our compunds, we use the AiZynthFinder (Genheden et al., 2020) retrosynthesis tool for validation. This tool attempts to predict reactions 'backwards' to find synthesis pathways for a given molecule and returns a score, success rate and a synthesis tree of length n. We augment the building blocks library used by AiZynthFinder with the full Enamine Building Blocks stock (Global stock, accessed November 28, 2023) (Enamine). The reactions available to the retrosynthesis tool are the US patent office (USPTO) set (Lowe). This means that any molecule inputted to the retrosynthesis tool will be deemed synthesisable if and only if it can be synthesised using USPTO reactions and the available building blocks.

Results Figure 2 and Table A.2 summarise our results. SynFlowNet generates molecules with considerably better SA and SC scores compared to the molecular fragments GFlowNet, and better QED scores. While the average reward is lower than for our fragment counterpart, this is likely due to the abundance of nonsensical high-reward molecules in the fragment-tree space. SynFlowNet also achieves higher ligand efficiency scores on average. The computation of pairwise Tanimoto distances between molecular fingerprints of molecules generated by the two models yields the same score of 0.81, meaning that SynFlowNet is still able to generate diverse candidates even with a significantly more constrained action space.

We also benchmark against ZINC250k (Gómez-Bombarelli et al., 2016) molecules (as a control) and a curated set of compounds sourced from the ChEMBL database, specifically targeting the sEH protein (ChEMBL2409). Ligand interaction potency to the sEH target is quantified though IC_{50} values, which indicate the concentration of compound required to inhibit the protein's biological activity by 50%. We retrieve the molecules with an IC_{50} value smaller than 1nM and benchmark

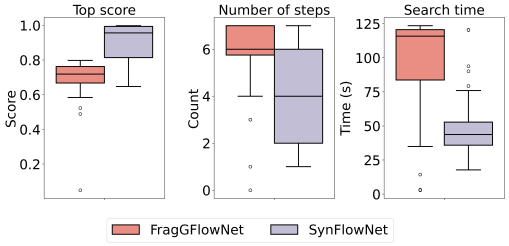

Figure 4: AiZynthFinder results. Statistics calculated for molecules generated using SynFlowNet and the fragments GFlowNet. Top score is the maximum synthesis score from all suggested routes per molecule.

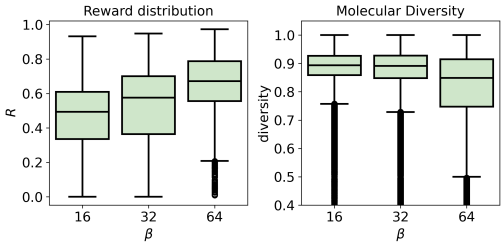

Figure 5: Effect of reward exponent (R^β) on distributions of reward and molecular diversity of sampled molecules (where diversity is quantified as the pairwise Tanimoto distance between the molecules' Morgan fingerprints).

the two GFlowNets against them. SynFlowNet molecules exhibit a promising alignment with those catalogued in ChEBML in all four metrics reported in Figure 2 but QED.

In Figure 4, we report some statistics obtained from running AiZynthFinder on the top 100 molecules. The score in the first box plot reflects the fraction of solved precursors and the number of reactions required to synthesise the target compound, and is close to 1.0 for solved compounds and typically less than 0.8 for unsolved ones (Genheden et al., 2020). 47% of the molecules sampled from SynFlowNet registered successful syntheses, while none of those generated with the fragments GFlowNet could be synthesised with the libraries of templates and building blocks used by AiZynthFinder. The number of steps and search times are also considerably higher on average for the fragments GFlowNet. In Figure 3 we show the synthetic routes for highest-reward molecules, as derived from a SynFlowNet trajectory or from AiZynthFinder. The synthesis for the molecule sampled from the fragments GFlowNet is rendered unsuccessful because one of the leaf molecules cannot be purchased.

Figure 2 showcases 3 high-reward molecules sampled from SynFlowNet, and their associated QED and SA scores. We note that two of them are derived from the same starting building block, meaning that one of the modes of high reward contains this substructure. Looking at other molecules of high reward, however, we see more structurally diverse molecules (see Figure A.9). Exploration and diversity can be controlled by reward exponent β, and the trade-off between diversity and reward is highlighted in Figure 5. This parameter's adjustability suggests we can further optimise the model's performance to achieve desired trade-offs between rewards and diversity, and such an adjustment mitigates the model's tendency to overly exploit rewards.

Limitations The main limitation of our work is the use of the pretrained reward proxy to approximate binding to sEH. This is both due to the inaccuracy of AutoDock Vina (Trott & Olson, 2010), and the uncertainty around whether the proxy model would perform well out-of-distribution for novel molecules. However, retrospective evaluation between our reward function and $\log(\text{IC}_{50})$ values gave a correlation coefficent of -0.3 and an ROC-AUC of 0.69 (Appendix Figures A.7 and A.8).

The available action space is a limiting factor in the performance of SynFlowNet, and this work can be improved with a broader library of chemical reactions. As drug design aims to simultaneously optimise for a series of pharmacological features, we aim to adapt our framework to the optimisation of multiple objectives. See Appendix A.2 for further discussions.

4 Conclusion

In this work, we introduce the first application of GFlowNets for de novo molecular design paired with forward synthesis. We have demonstrated promising results by generating molecules with high scores in terms of quality and properties, and significantly enhanced synthesisability, by defining an action space of chemical reactions and commercially available reactants. When comparing against GFlowNets defined with a molecular fragments action space, we preserve the high diversity score of generated candidates, despite significantly constraining the action space.

ACKNOWLEDGEMENTS

The authors would like to thank Alexei Lapkin, Arne Schneuing, Ilia Igashov and Arian Jamasb for insightful discussions that improved the quality of this work. Miruna Cretu is supported by the SynTech CDT at the University of Cambridge.

REFERENCES

Zaccary Alperstein, Artem Cherkasov, and Jason Tyler Rolfe. All SMILES VAE. *CoRR*, abs/1905.13343, 2019.

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow Network based Generative Models for Non-Iterative Diverse Candidate Generation. *arXiv preprint arXiv:2106.04399*, 2021.

G. Richard Bickerton, Gaia V. Paolini, Jérémy Besnard, Sorel Muresan, and Andrew L. Hopkins. Quantifying the chemical beauty of drugs. *Nature Chemistry*, 4(2):90–98, 2012. doi: 10.1038/nchem.1243.

John Bradshaw, Brooks Paige, Matt J. Kusner, Marwin H. S. Segler, and José Miguel Hernández-Lobato. A model to search for synthesizable molecules. *CoRR*, abs/1906.05221, 2019.

Jacob Buckman, Aurko Roy, Colin Raffel, and Ian Goodfellow. Thermometer encoding: One hot way to resist adversarial examples. In *International Conference on Learning Representations*, 2018.

Alexander Button, Daniel Merk, Jan A Hiss, and Schneider Gisbert. Automated de novo molecular design by hybrid machine intelligence and rule-driven chemical synthesis. *Nature Machine Intelligence*, 1(7):307–315, 2019.

Nicola De Cao and Thomas Kipf. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2022.

ChEMBL2409. Chembl2409 Target Report Card. `https://www.ebi.ac.uk/chembl/target_report_card/CHEMBL2409/`. Accessed: 28.03.2024.

Connor W. Coley, Luke Rogers, William H. Green, and Klavs F. Jensen. Computer-assisted retrosynthesis based on molecular similarity. *ACS Central Science*, 3(12):1237–1245, 2017. doi: 10.1021/acscentsci.7b00355.

Connor W. Coley, Luke Rogers, William H. Green, and Klavs F. Jensen. Scscore: Synthetic complexity learned from a reaction corpus. *Journal of Chemical Information and Modeling*, 58(2):252–261, 2018. doi: 10.1021/acs.jcim.7b00622.

Enamine. Enamine. `https://enamine.net`. Accessed: 28.11.2023.

Peter Ertl and Ansgar Schuffenhauer. Estimation of synthetic acces- sibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics*, (1):1–11, 2009.

Wenhao Gao and Connor W. Coley. The synthesizability of molecules proposed by generative models. *Journal of Chemical Information and Modeling*, 60(12):5714–5723, 2020. doi: 10.1021/acs.jcim.0c00174.

Wenhao Gao, Rocío Mercado, and Connor W. Coley. Amortized tree generation for bottom-up synthesis planning and synthesizable molecular design. *CoRR*, abs/2110.06389, 2021.

S. Genheden, A. Thakkar, V. Chadimová, J. L. Reymond, O Engkvist, and Esben Bjerrum. Aizynthfinder: a fast, robust and flexible open-source software for retrosynthetic planning. *J Cheminform*, 12:70, 2020. doi: 10.1186/s13321-020-00472-1.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 1263–1272, 2017.

Rafael Gómez-Bombarelli, David Duvenaud, José Miguel Hernández-Lobato, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *CoRR*, abs/1610.02415, 2016.

Sai Krishna Gottipati, Boris Sattarov, Sufeng Niu, Yashaswi Pathak, Haoran Wei, Shengchao Liu, Shengchao Liu, Simon Blackburn, Karam Thomas, Connor Coley, Jian Tang, Sarath Chandar, and Yoshua Bengio. Learning to navigate the synthetically accessible chemical space using reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 3668–3679, 2020.

Gabriel Lima Guimaraes, Benjamin Sanchez-Lengeling, Carlos Outeiral, Pedro Luis Cunha Farias, and Alán Aspuru-Guzik. Objective-reinforced generative adversarial networks (organ) for sequence generation models. *arXiv preprint arXiv:1705.10843*, 2018.

Rafael Gómez-Bombarelli, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, 4(2):268–276, 2018. doi: 10.1021/acscentsci.7b00572.

Charles Harris, Kieran Didi, Arian Jamasb, Chaitanya Joshi, Simon Mathis, Pietro Lio, and Tom Blundell. Posecheck: Generative models for 3d structure-based drug design produce unrealistic poses. In *NeurIPS 2023 Generative AI and Biology (GenBio) Workshop*, 2023.

Walter M Rupp M Reisen F Proschak E Weggen S Stark H Schneider G. Hartenfeller M, Zettl H. DOGS: reaction-driven de novo design of bioactive. *PLOS Computational Biology*, 8(2), 2012. doi: 10.1371/journal.pcbi.1002380.

Emiel Hoogeboom, Víctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 8867–8887, 2022.

Julien Horwood and Emmanuel Noutahi. Molecular design in synthetically accessible chemical space via deep reinforcement learning. *ACS Omega*, 5(51):32984–32994, December 2020. doi: 10.1021/acsomega.0c04153.

John D. Imig and Bruce D. Hammock. Soluble epoxide hydrolase as a therapeutic target for cardiovascular diseases. *Nat Rev Drug Discov*, 8(2):794–805, 2009. doi: 10.1038/nrd2875.

Moksh Jain, Sharath Chandra Raparthy, Alex Hernandez-Garcia, Jarrid Rector-Brooks, Yoshua Bengio, Santiago Miret, and Emmanuel Bengio. Multi-objective GFlowNets. *arXiv preprint arXiv:2210.12765*, 2023.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv preprint arXiv:1802.04364*, 2019.

Kathryn Klarich, Brian Goldman, Trevor Kramer, Patrick Riley, and W. Patrick Walters. Thompson samplingan efficient method for searching ultralarge synthesis on demand databases. *Journal of Chemical Information and Modeling*, 64(4):1158–1171, 2024. doi: 10.1021/acs.jcim.3c01790.

Ksenia Korovina, Sailun Xu, Kirthevasan Kandasamy, Willie Neiswanger, Barnabas Poczos, Jeff Schneider, and Eric Xing. Chembo: Bayesian optimization of small organic molecules with synthesizable recommendations. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108, pp. 3393–3403, 2020.

D. Lowe. Chemical reactions from US patents. `https://figshare.com/articles/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873`. Accessed: 31.04.2018.

Nikolay Malkin, Moksh Jain, Emmanuel Bengio, Chen Sun, and Yoshua Bengio. Trajectory balance: Improved credit assignment in gflownets. *CoRR*, abs/2201.13259, 2022.

Marcus Olivecrona, Thomas Blaschke, Ola Engkvist, and Hongming Chen. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(48), 2017. doi: 10.1186/s13321-017-0235-x.

Julien Roy, Pierre-Luc Bacon, Christopher Pal, and Emmanuel Bengio. Goal-conditioned gflownets for controllable multi-objective molecular design. *arXiv preprint arXiv:2306.04620*, 2023.

Arman A. Sadybekov, Anastasiia V. Sadybekov, Yongfeng Liu, Christos Iliopoulos-Tsoutsouvas, Xi-Ping Huang, Julie Pickett, Blake Houser, Nilkanth Patel, Ngan K. Tran, Fei Tong, Nikolai Zvonok, Manish K. Jain, Olena Savych, Dmytro S. Radchenko, Spyros P. Nikas, Nicos A. Petasis, Yurii S. Moroz, Bryan L. Roth, Alexandros Makriyannis, and Vsevolod Katritch. Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature*, 601(7893):452–459, 2022. doi: 10.1038/s41586-021-04220-9.

Arne Schneuing, Yuanqi Du, Charles Harris, Arian Jamasb, Ilia Igashov, Weitao Du, Tom Blundell, Pietro Lió, Carla Gomes, Max Welling, Michael Bronstein, and Bruno Correia. Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695*, 2023.

Philippe Schwaller, Riccardo Petraglia, Valerio Zullo, Vishnu H. Nair, Rico Andreas Haeuselmann, Riccardo Pisoni, Costas Bekas, Anna Iuliano, and Teodoro Laino. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.*, 11: 3316–3325, 2020. doi: 10.1039/C9SC05704H.

Tony Shen, Mohit Pandey, Jason Smith, Artem Cherkasov, and Martin Ester. Tacogfn: Target conditioned gflownet for structure-based drug design. *arXiv preprint arXiv:2310.03223*, 2023.

Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, and Jian Tang. Graphaf: a flow-based autoregressive model for molecular graph generation. *arXiv preprint arXiv:2001.09382*, 2020.

Oleg Trott and Arthur J. Olson. Autodock Vina: improving the speed and accuracy of docking with a new scoring func- tion, efficient optimization, and multithreading. *Journal of Computational Chemistry*, 31(2):455–461, 2010. doi: 10.1002/jcc.21334.

H. Maarten Vinkers, Marc R. de Jonge, Frederik F. D. Daeyaert, Jan Heeres, Lucien M. H. Koymans, Joop H. van Lenthe, Paul J. Lewi, Henk Timmerman, Koen Van Aken, and Paul A. J. Janssen. Synopsis: SYNthesize and OPtimize System in Silico. *Journal of Medicinal Chemistry*, 46(13): 2765–2773, 2003. doi: 10.1021/jm030809x.

Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J Kim. Graph transformer networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Chengxi Zang and Fei Wang. Moflow: An invertible flow model for generating molecular graphs. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pp. 617–626, 2020. doi: 10.1145/3394486.3403104.

Barbara Zdrazil, Eloy Felix, Fiona Hunter, Emma J Manners, James Blackshaw, Sybilla Corbett, Marleen de Veij, Harris Ioannidis, David Mendez Lopez, Juan F Mosquera, Maria Paula Magarinos, Nicolas Bosc, Ricardo Arcila, Tevfik Kizilören, Anna Gaulton, A Patrícia Bento, Melissa F Adasme, Peter Monecke, Gregory A Landrum, and Andrew R Leach. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, 52(D1):D1180–D1192, 2023. doi: 10.1093/nar/gkad1004.

A Implementation Details

A.1 Training

Model and training details We adapt the framework from Bengio et al. (2021) to train a GFlowNet sampler over a space of synthesisable molecules, which are assembled from an action space of

chemical reactions and reactants. A graph neural network with a graph transformer architecture (Yun et al., 2019) is used to produce a state-conditional distribution over the actions. A state is represented as a molecular graph in which nodes contain atom features. Edge attributes are bond type and the indices of the atoms which are its attachment points. This representation is augmented with a fully-connected virtual node, which is an embedding of the conditional encoding of the desired sampling temperature, obtained using an MLP. The sampling temperature is controlled by a temperature parameter β, which also plays a role in reward modulation, allowing for exponential scaling of the rewards (by making rewards received during training equal to R^β). We experimented with sampling β from multiple distributions, and use a constant distribution in the reported results in this paper. We used a thermometer encoding of the temperature (Buckman et al., 2018).

The model is trained using the trajectory balance objective (Malkin et al., 2022) and thus is parameterised by forward and backward action distributions P_F and P_B and an estimation of the partition function $Z = \sum_{\tau \in \mathcal{T}} F(\tau)$. The hyperparameters used for training SynFlowNet and the fragment-based GFlowNet are summarised in Table A.2. A maximum trajectory length of 5 is used.

Hyperparameters	Values	
	SynFlowNet	Fragments GFlowNet
Batch size	64	64
GFN temperature parameter β	64	64
Number of training steps	1000	1000
Number of GNN layers	4	4
GNN node embedding size	128	128
Learning rate (P_F)	10^{-4}	10^{-4}
Learning rate (P_B)	10^{-4}	10^{-4}
Learning rate (Z)	10^{-3}	10^{-3}

Table A.1: Hyperparameters used in our training pipelines.

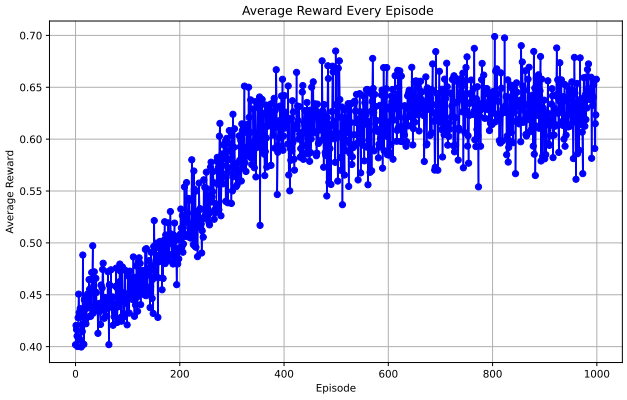

Figure A.1: Average reward recorded during training.

Reward details As mentioned in Section 2, we use a proxy trained on molecules docked with AutoDock Vina (Trott & Olson, 2010) for the sEH target; we use the weights provided by Bengio et al. (2021), resulted from training an MPNN (Gilmer et al., 2017) that receives an atom graph as input. Model architecture is detailed in Bengio et al. (2021). The proxy was trained with a dataset of 300k randomly generated molecules down to a test MSE of 0.6. Note that the reward scale presented in our findings deviates from that of the original GFlowNet publication. Specifically, the rewards in our analysis are adjusted by a factor of 1/8.

A.2 EVALUATION

While we report synthetic accessibility/complexity scores, this is primarily to enable comparison to baselines. Such scores become less relevant when one finds available synthetic routes for a compound, which our model achieves for every generated molecule (provided that the reaction can be carried out with high yield in a synthetic chemistry lab, which is not guaranteed). Figure 2 supports to some extent the reliability of the heuristics behind the SA score, as it shows a clear distribution in the lower range of the score's domain, as expected for molecules that are derived from chemical reactions. At the same time, high SA scores are also recorded (see Figure 2), exposing the score's limitations in assessing the real synthesisability of molecules. Correlation between synthesisability and heuristic scores have been discussed previously (Gao & Coley, 2020). We sought to further assess synthesisability using a retrosynthesis software, which validated the enhanced synthesisability of molecules generated from our model, but such a tool is still limited by the library of templates it is trained on.

Method	Validity	Uniqueness	Diversity	QED (↑)	SA (↓)	SCS (↓)	MW (↓)
SynFlowNet	100%	93.2%	0.81	0.43	3.4	3.8	419.2
Frag GFlowNet	99.8%	100%	0.81	0.26	5.4	4.9	588.9

Table A.2: Comparisons according to sample quality and property statistics of molecules generated with SynFlowNet and a fragments-based GFlowNet.

Method	AiZynthFinder success
SynFlowNet	47%
Frag GFlowNet	0%

Table A.3: Percentages of molecules (out of top 100) for which AiZynthFinder provided successful retrosynthesis routes.

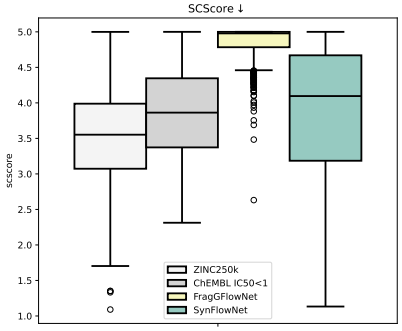

Figure A.2: SCScores.

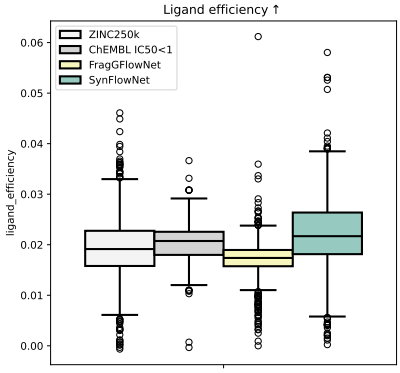

Figure A.3: Ligand efficiency.

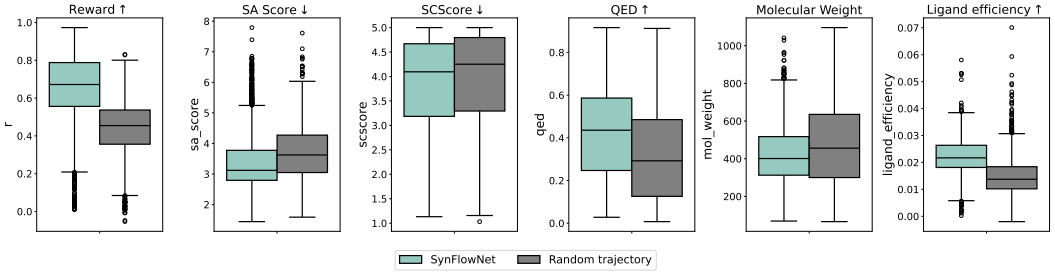

Figure A.4: Comparison to molecules generated by randomly combining reaction templates and building blocks.

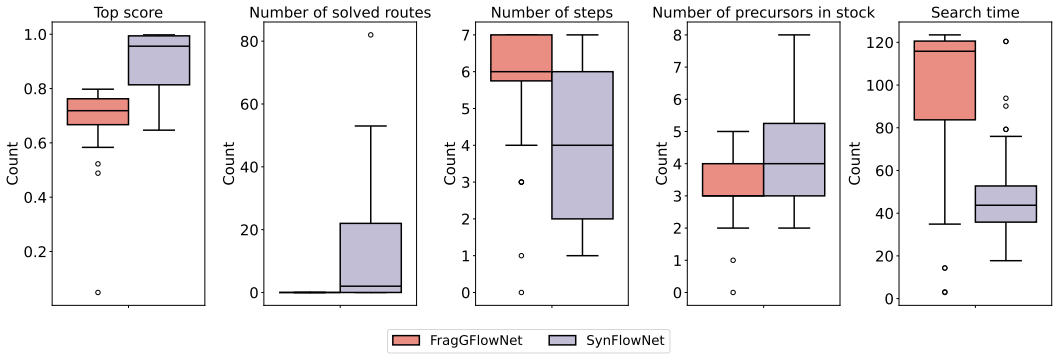

Figure A.5: AiZynthFinder statistics for SynFlowNet vs. fragments GFlowNet model. The second box plot is number of solved routes per molecule, out of all routes generated for that molecule.

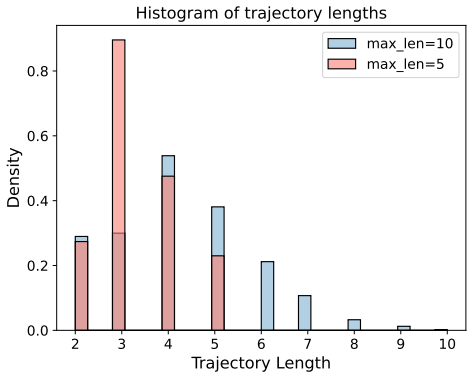

Figure A.6: Histogram for lengths of trajectories sampled from SynFlowNet trained with different maximum trajectory lengths. Note that the first state in the trajectory is an empty molecular graph, meaning that a trajectory of length equal to 2 has one building block only, a trajectory of length 3 contains one reaction etc.

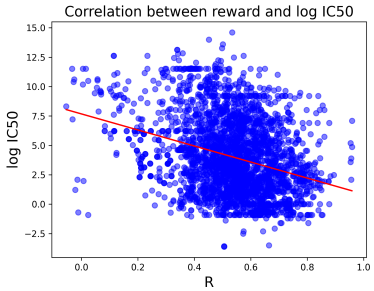

Figure A.7: Correlation between IC_{50} values and computed rewards. A Pearson correlation coefficient of -0.3 was obtained.

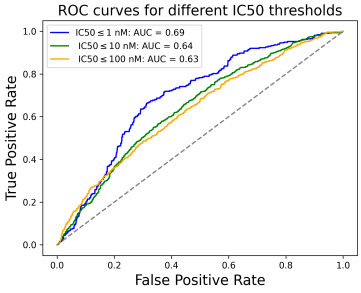

Figure A.8: ROC curves for predicted rewards, with various IC_{50} thresholds for binding. We use a threshold of IC_{50}<1nM to select molecules from ChEMBL to compare against.

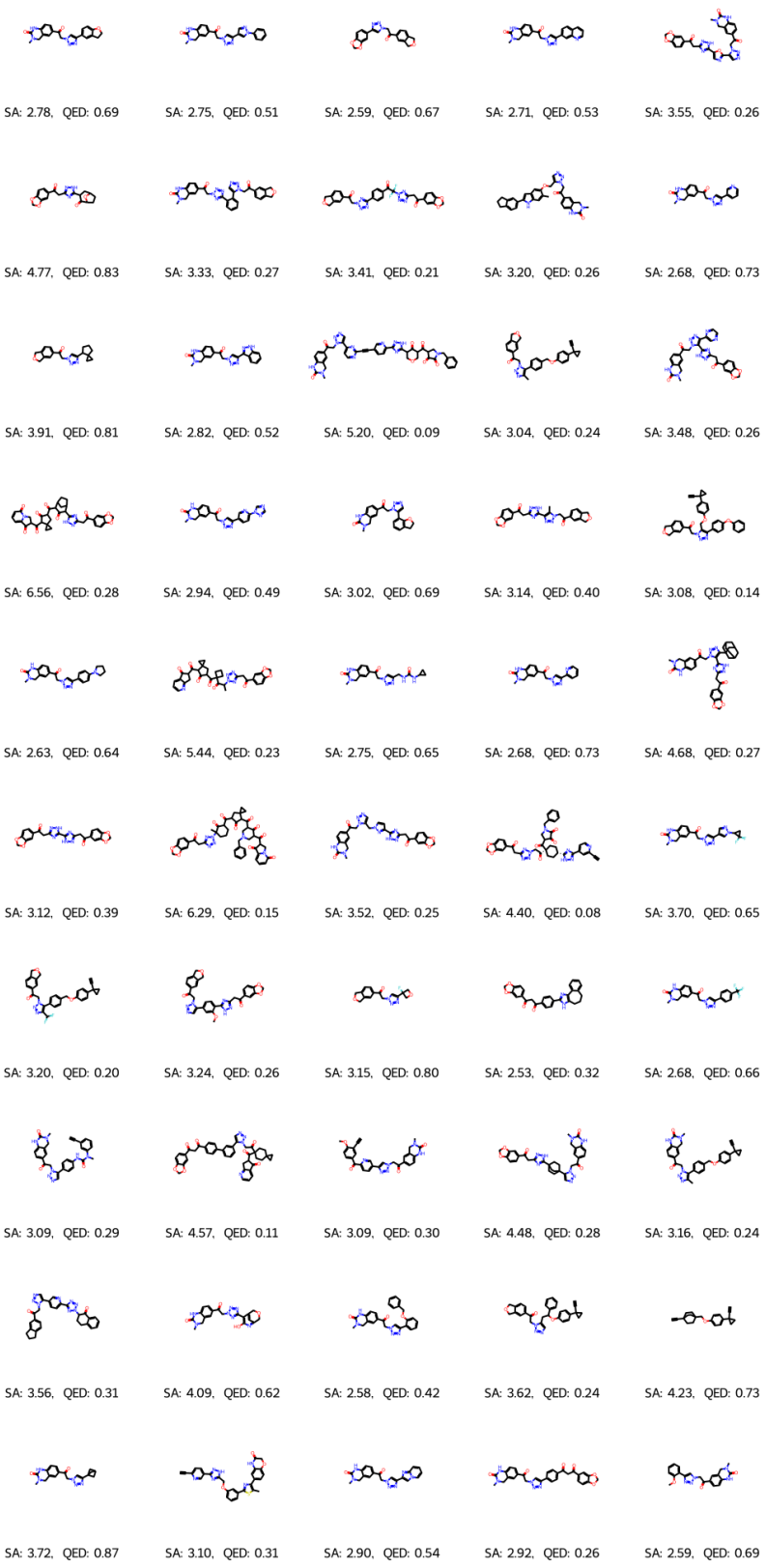

Figure A.9: Top 50 molecules sampled from SynFlowNet (by reward). Rewards for these molecules range from 0.87 to 0.97.