Towards DNA-Encoded Library Generation with GFlowNets

Michał Koziarski[*,1,2]**, Mohammed Abukalam**[1,2]**, Vedant Shah**[1,2]**, Louis Vaillancourt**[2,3]**,
Doris Alexandra Schuetz**[2,3]**, Moksh Jain**[1,2]**, Almer van der Sloot**[1,2]**, Mathieu Bourgey**[1,2]**,
Anne Marinier**[2,3]**, Yoshua Bengio**[1,2]

Abstract

DNA-encoded libraries (DELs) are a powerful approach for rapidly screening large numbers of diverse compounds. One of the key challenges in using DELs is library design, which involves choosing the building blocks that will be combinatorially combined to produce the final library. In this paper we consider the task of protein-protein interaction (PPI) biased DEL design. To this end, we evaluate several machine learning algorithms on the PPI modulation task and use them as a reward for the proposed GFlowNet-based generative approach. We additionally investigate the possibility of using structural information about building blocks to design a hierarchical action space for the GFlowNet. The observed results indicate that GFlowNets are a promising approach for generating diverse combinatorial library candidates.

1 Introduction

Combinatorial libraries represent a powerful approach in chemistry for rapidly generating and screening large numbers of diverse compounds. These libraries are typically constructed by systematically combining building blocks or functional groups in various permutations. One example of combinatorial libraries is DNA-encoded libraries (DEL) (Goodnow Jr et al., 2017; Gironda-Martínez et al., 2021; Satz et al., 2022). In DELs, unique DNA tags are attached to individual small molecule compounds, creating a vast library of DNA-tagged molecules (Figure 1). Pulldown and next-generation sequencing (NGS) then allow for rapid identification of molecules with desired properties and potential drug candidates. While powerful, DELs are difficult to synthesize (the process can easily take several months), and might present challenges with respect to training machine learning models on the produced data (McCloskey et al., 2020; Lim et al., 2022). Furthermore, a given set of building blocks may represent a large number of molecules resulting from their different permutations. This makes it difficult to evaluate a given set.
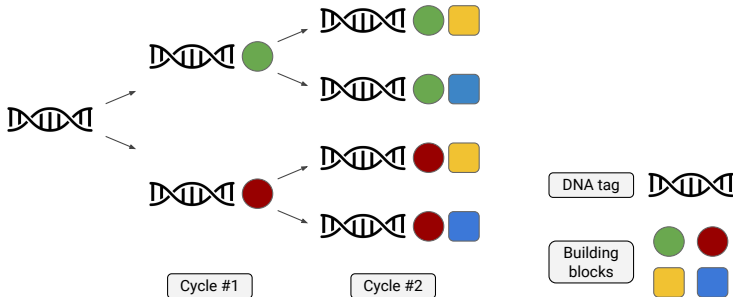

Figure 1: Illustration of DEL generation. Building blocks are first attached to a DNA tag used for identification, and later, across several consecutive cycles, combined in a combinatorial manner to produce a library of molecules built by joining together a sequence of building blocks.

[*] Corresponding author: michal.koziarski@mila.quebec, [1] Mila, [2] Université de Montréal, [3] IRIC

Due to the significant cost of synthesizing these libraries, it is desirable to construct versatile DELs, with high potential hit ratio (proportion of screened molecules with desired properties) across multiple targets, which are not known prior to the screen. This enables re-using the library in a multitude of screens. Moreover, biasing the libraries towards a specific family of targets can also be beneficial, as distinct regions of the chemical space are likely to produce higher hit rates for specific targets. One possible solution to fulfilling the above desiderata is the design of libraries biased towards modulating protein-protein interactions (PPIs) (Morelli et al., 2011; Bosc et al., 2020). Dysfunctions in PPIs are known to be connected to various disease states. Despite this, PPI-targeting drugs are rare, mainly due to challenging "druggability" (Morelli et al., 2011) and poor quality of existing libraries. This makes the design of PPI-biased DELs potentially highly impactful.

In this paper we consider the task of computational DEL design. We utilize GFlowNets (Bengio et al., 2023) - a generative method recently used in multiple scientific discovery tasks (Bengio et al., 2021; Jain et al., 2023; Mila AI4Science et al., 2023; Volokhova et al., 2023) due to its ability to produce highly diverse samples. We postulate that diversity will play an important role in DEL design, since it allows us to propose multiple library candidates with a wide range of chemical properties, out of which the best one can be determined experimentally. We first evaluate multiple machine learning algorithms in a PPI modulation classification task, and then use the best-performing method as a reward for the proposed GFlowNet. We demonstrate that the proposed approach achieves high PPI likelihood, while keeping the sample diversity high.

2 PROBLEM FORMULATION

In this paper we consider the problem of constructing DELs biased towards high likelihood of being PPI modulators. The motivation behind this is to develop general screening libraries that could eventually be used for targeting various proteins. Ultimately, we are interested in the problem of constructing very large (> 1 million molecules) libraries, consisting of hundreds of building blocks selected out of thousands of viable ones, and possibly multiple chemical reactions assembling them. As a crucial stepping stone, in this paper we consider the simpler task of selecting a subset of already chosen building blocks, which translate to a specific library. This simplifies the problem, not only by significantly reducing the search space and required computational resources, but also by removing some of the biological constraints, for instance synthesizability of candidate building blocks.
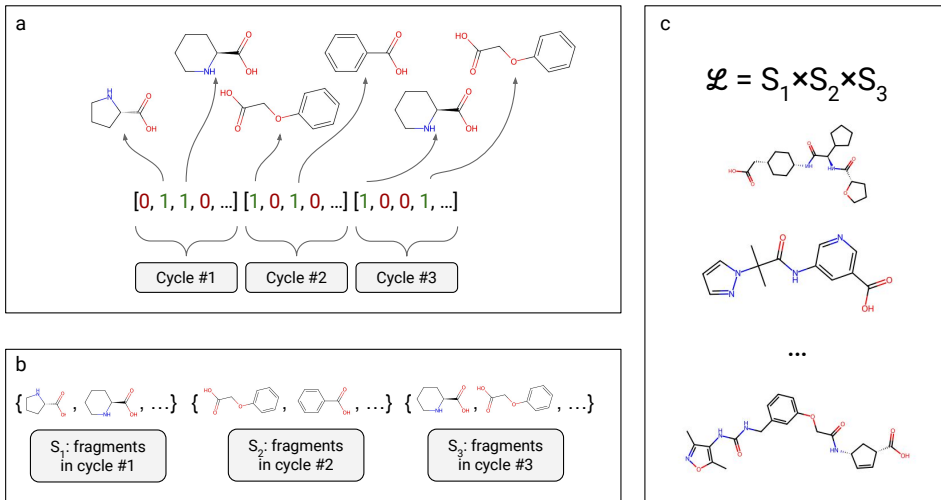

Figure 2: Illustration of the used state representation. a) Information about which building blocks are picked for a given cycle is represented as a binary vector, where each entry corresponds to a specific building block. b) For each cycle, a collection of selected building blocks is decoded. c) Resulting library is a Cartesian product of all the possible building blocks from different cycles, where each triplet of building blocks is combined to produce a final molecule.

The way in which we represent states in presented in Figure 2. Specifically, the libraries we consider consist of three synthetic cycles, having 90, 89 and 197 possible building blocks, respectively, and a single way of combining the building blocks to generate a final molecule. This corresponds to a total of approximately 1.58 million molecules being present in the library, which corresponds to the Cartesian product of all the possible building blocks from different cycles. The task is finding specific sub-libraries (constructed by selecting subsets of these building blocks) that 1) produce a library of a specified size, 2) maximize the average likelihood of a molecule in the mentioned sub-library being a PPI modulator, and 3) the sub-libraries to be mutually diverse, which enables the users to choose particular ones depending on a specific target of interest. Point 3 incorporates the choice of libraries which tend to be biased towards specific chemical properties, such as molecular weight, number of rotatable bonds, polarity, etc.

We observe that the above problem can be stated as a binary vector search, with i-th element of the vector denoting whether i-th building block will be present in the corresponding library subset. More formally, we can represent a library by $x = x_1|x_2|x_3$, with x_i denoting a sub-vector representing i-th cycle, and $x_1 \in \{0,1\}^{90}$, $x_2 \in \{0,1\}^{89}$ and $x_3 \in \{0,1\}^{197}$ (corresponding to the number of building blocks in different cycles). Then, if we denote the set of all available building blocks for i-th cycle as $\mathcal{B}_i$, and selected blocks as $\mathcal{S}_i = \{\mathcal{B}_{i,j} : \mathbb{1}[x_{i,j} = 1]\}$, the complete library subset that can be generated using x is $\mathcal{L}(x) = \mathcal{S}_1 \times \mathcal{S}_2 \times \mathcal{S}_3$.

3 DEL-GFlowNet

3.1 Flat variant

GFlowNets (Bengio et al., 2021; 2023) are a family of generative methods designed to learn a sampling policy π for constructing objects $x \in \mathcal{X}$ based on their non-negative reward function $R(x)$, such that $\pi(x) \propto R(x)$. This generation is done sequentially: starting from the initial state s_0, transitions $s_t \rightarrow s_{t+1} \in \mathbb{A}$ are applied between states $s \in \mathcal{S}$, forming trajectories $\tau = (s_0 \rightarrow s_1 \rightarrow \ldots \rightarrow x)$, where $\mathbb{A}$ is a predefined action space and $\mathcal{S}$ the state space. The policy π, e.g. a neural network, can then be optimized using one of the existing loss functions, such as Trajectory Balance (Malkin et al., 2022).

We utilize the GFlowNet framework in DEL construction task by operating in the previously described state space $\mathcal{S}$ of binary vectors x that can be mapped to corresponding libraries $\mathcal{L}(x)$, defining the action space $\mathbb{A} = \{1, 2, \ldots, |x| + 1\}$, with action $a = i$ denoting that i-th bit of the vector x should be flipped from 0 to 1 (plus one additional action denoting termination of sequence), and with a reward

$$R(x) = \exp\left(\frac{\beta}{|\mathcal{L}(x)|} \sum_{i=1}^{|\mathcal{L}(x)|} p(\mathcal{L}(x)_i)\right), \tag{1}$$

with $p(m)$ denoting a function predicting the likelihood of the molecule m representing a PPI modulator, and β being a predefined parameter. We enforce the size constraints of the library by masking out the end of sequence action unless the minimum specified library size is reached, in addition to masking out every action that would increase the library size above the specified maximum. We refer to this approach as DEL-GFlowNet.

3.2 Hierarchical variant

Secondly, we observe that while the selected building blocks can be represented as a simple binary vector, this leads to loss of information associated with the specific molecular structure of the block (recall that building blocks themselves are small molecular fragments). Additionally, we note that in the basic variant, the action space of DEL-GFlowNet is relatively large (while to the best of our knowledge, there are no rigorous studies on the limits of GFlowNets, our own experience indicates that in problems with action spaces larger than one or two hundred, GFlowNets become difficult to train), and would have to be further enlarged to transition to bigger library sizes. This motivated us to consider the possibility of leveraging the structural information about building blocks to construct alternative, hierarchical action space that would be more compact.
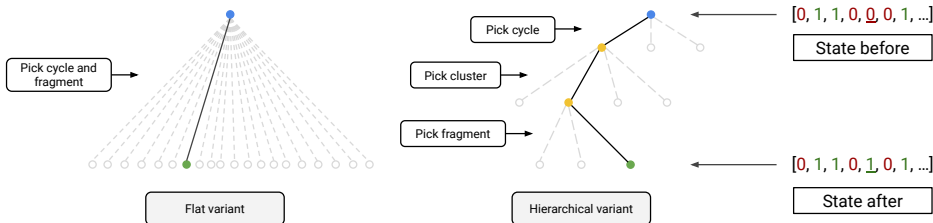
Figure 3: Illustration of action spaces of flat and hierarchical DEL-GFlowNet. Lines indicate actions (black: chosen, gray: other possible), and dots indicate states (blue: starting, yellow: intermediate, green: final). End effect in both cases is selection of one additional building block. Hierarchical variant allows us to significantly reduce the number of valid actions at any given stage.

To this end we first group the building blocks into clusters based on their molecular structure, and secondly split the action of picking a building block into three separate steps: 1) picking the cycle, 2) picking one of the clusters for the given cycle, and 3) picking one of the building blocks belonging to that cluster (Figure 3). Additionally, to make the policy learning feasible, we augment the state space with one-hot encodings of the current cycle and cluster, as well as two additional binary values indicating whether any cycle and cluster were already picked. We perform clustering using Agglomerative Clustering on 2048-dimensional ECFPs (Rogers & Hahn, 2010), with Jaccard similarity as a distance measure, setting the number of clusters to 10 for cycles 1 and 2, and to 20 for cycle 3. We refer to this hierarchical variant as H-DEL-GFlowNet.

4 EXPERIMENTAL STUDY

4.1 PROXY TRAINING

Since the reward function is an essential component of the GFlowNet training, we begin by examining the possibility of training a proxy model $p(m)$ capable of reliable prediction of PPI modulation likelihood. To this end we construct a dataset consisting of 2583 compounds from the 2P2Idb database (Basse et al., 2012) with confirmed orthosteric binding to the target (treated as positives) and 1541 FDA approved drugs (treated as negatives). Note that we exclude FDA approved drugs known to modulate PPI. We make a key assumption that remaining known drugs are unlikely to modulate PPI, since that would lower their specificity. We divide this dataset into train, validation and test partitions using scaffold splitting into a 0.8:0.1:0.1 proportion.

We evaluate the performance of several classification algorithms: Random Forest (RF) trained on 2048-dimensional ECFPs, a pretrained Molformer (Wu et al., 2023), and a graph neural network (GNN) - both trained from scratch and pretrained. The details of the training can be found in Appendix A.1. The results of this comparison were presented in Table 1. As can be seen, the best performance was achieved by a pretrained GNN. While neural network-based methods outperformed RF by a margin, in general all of the classification algorithms achieved reasonable performance, indicating that they would perform well in the PPI modulator discrimination.

Table 1: Comparison of different classification algorithms for PPI modulation prediction. Best performance denoted in bold.

Method	Accuracy	Precision	Recall	AUC	F1-score	AP
RF	0.748	0.707	**0.962**	0.865	0.815	0.893
Molformer (pretrained)	0.799	0.819	0.836	0.872	0.827	0.898
GNN	0.818	0.856	0.824	0.887	0.839	0.907
GNN (pretrained)	**0.823**	**0.857**	0.832	**0.898**	**0.844**	**0.917**

This claim was further tested in an additional experiment, in which we compared the individual quality of building blocks belonging to one of the cycles. We did so by comparing average probabilities of molecules containing given building block being a PPI modulator, outputted by different

models. The results were presented in Figure 4. As can be seen, while some correlation between the models was observed, it was not perfect. Our conclusion is that while all of the models achieved comparable results on the small dataset of known PPI modulators, this only partially translates to the results on the whole library, making the choice of the optimal model difficult in practice.
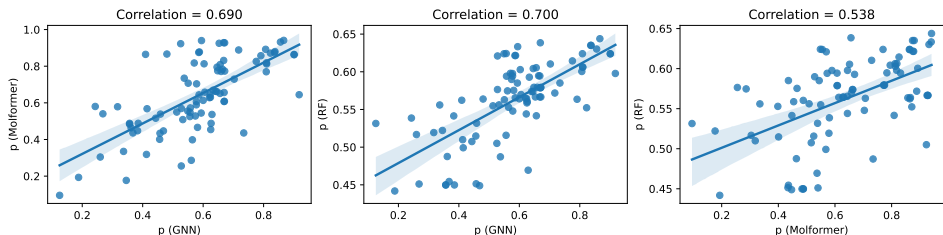

Figure 4: Comparison of model similarity. Individual points represent specific building blocks from first cycle, with average probability values of molecules containing that building block.

4.2 LIBRARY SAMPLING

Using the best performing model on the dataset of known PPI modulators (pretrained GNN) as the reward model, we evaluate the proposed DEL-GFlowNet method in a library subset sampling task. We compare it with four baselines: random sampling, a greedy approach in which 25/25/40 individually best performing blocks were selected per cycle, Markov chain Monte Carlo (MCMC) (Robert et al., 2004), and proximal policy optimization (PPO) (Schulman et al., 2017). We consider the task of sampling a library with a size between 20k and 25k molecules. Note that partial results for sampling a 90k to 100k molecules library can be found in Appendix B. Training details can be found in Appendix A. In total, we sample 5,000 samples for each method, and averaged the results across 3 random runs. We report both the average PPI likelihood of the sampled libraries, as well as their diversity:

$$\text{Diversity}(\mathcal{D}) = \frac{\sum_{(x_i, y_i) \in \mathcal{D}} \sum_{(x_j, y_j) \in \mathcal{D} \setminus \{(x_i, y_i)\}} d\left(x_i, x_j\right)}{|\mathcal{D}|(|\mathcal{D}| - 1)}, \tag{2}$$

directly on the binary vectors, with hamming distance used as distance measure.

We report the numerical results in Table 2. We compute both distance and average probability in three settings: across all samples, for top-100 candidate libraries with highest reward, and for the best sampled candidate. Additionally, to further illustrate the diversity, in Figure 5 we visualize the distributions of average values of several chemical properties of the generated libraries. Several observations can be made based on the results. First of all, it's worth noting that the simple greedy approach outperforms all of the other methods in a single library design, finding a library candidate with near-perfect probability. However, we argue that taking into account the uncertainty associated with the quality of the proxy, in practice it is unlikely to produce the "optimal" (with respect to real, unknown proportion of possible PPI modulators) library. Instead, what is required is proposal of multiple diverse library candidates which can be evaluated experimentally, which is something the greedy approach is incapable of. PPO behaves similarly: while it achieves very good performance of an individual library, it collapses to a single mode, and has low diversity of solutions. GFlowNets outperform remaining methods, producing libraries with higher estimated PPI likelihood and slightly higher diversity. The second point is further illustrated in Figure 5, where we can observe GFlowNet-produced libraries to have wider ranges of selected average properties, including molecular weight, cLogP and the number of non-hydrogen atoms. Finally, we note that this is more pronounced for H-DEL-GFlowNet than DEL-GFlowNet, which also achieves slightly higher performance as shown in Table 2.

5 DISCUSSION & FUTURE DIRECTIONS

In this paper we frame the combinatorial library design problem as a binary vector search task and design DEL-GFlowNet - a generative method for sampling diverse library candidates biased towards

Table 2: Comparison of different DEL sampling strategies. Best performance denoted in bold.

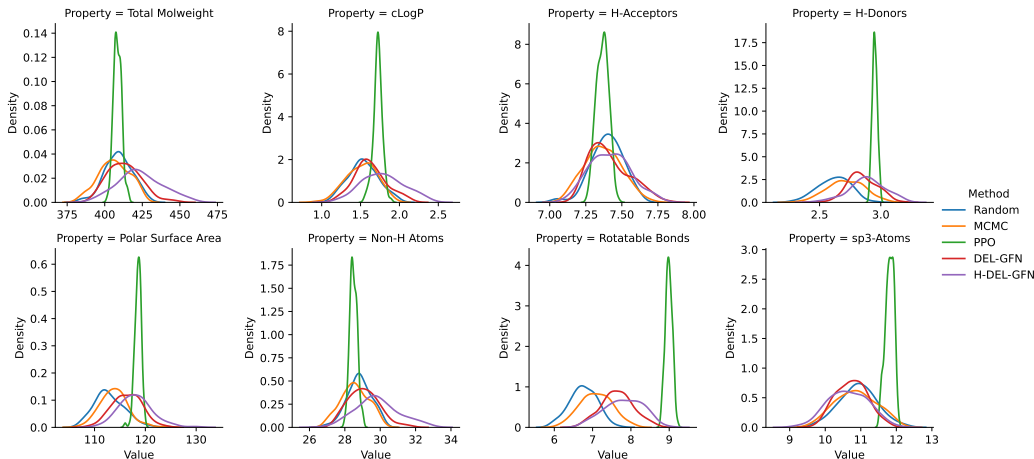| Method	Prob.	Div.	Top-100 prob.	Top-100 div.	Top-1 prob.
Random	0.582 ± 0.000	0.378 ± 0.000	0.706 ± 0.002	0.374 ± 0.001	0.754 ± 0.011
Greedy	-	-	-	-	**0.992**
MCMC	0.696 ± 0.000	0.393 ± 0.000	0.809 ± 0.000	0.385 ± 0.001	0.850 ± 0.002
PPO	$\mathbf{0.974 \pm 0.006}$	0.105 ± 0.015	$\mathbf{0.983 \pm 0.005}$	0.090 ± 0.012	0.985 ± 0.004
DEL-GFN	0.767 ± 0.003	0.397 ± 0.001	0.873 ± 0.003	0.383 ± 0.001	0.912 ± 0.011
H-DEL-GFN	0.783 ± 0.013	$\mathbf{0.402 \pm 0.001}$	0.885 ± 0.010	$\mathbf{0.390 \pm 0.001}$	0.918 ± 0.008

Figure 5: Distributions of average chemical library properties of top-100 generated libraries.

PPI modulators. Furthermore, we extended the approach by introducing a hierarchical action space based on clustering the molecular structures of the building blocks. We evaluate the approach on a task of selecting a subset of existing library. While the results seem promising, as demonstrated by the ability to sample high-quality libraries with diverse chemical properties, there are also several limitations that would likely need to be addressed in order to scale the method to facilitate the design of significantly larger libraries.

One challenge lies in training the classification algorithm predicting the likelihood of a molecule being a PPI modulator. While we were able to train multiple classification algorithms, each achieving good performance on a small dataset of molecules with known PPI modulation activity, and their performance is roughly comparable on that dataset, non-negligible differences were observed with respect to ranking individual building blocks of the considered DEL. This indicates that the performance on the small dataset of known PPI modulators does not necessarily translate to reliable out-of-distribution prediction on DEL, and poses additional difficulty related to ambiguity of which proxy model should be used as a reward for GFlowNet training. While typically this problem can be solved by gathering additional data, in the case of general PPI modulation this becomes tricky. This is because while positives can be identified somehow reliably, this is not the case for negatives (while a molecule can be labeled a PPI modulator if it modulates at least one PPI, determining if it will not be able to modulate any PPI is difficult).

Secondly, while we observed promising results for sampling library candidates with GFlowNets, several issues remain here as well. Most importantly, the considered problem of library subset selection is still relatively small, and it is not clear how the performance would scale with a significantly increased number of possible building blocks and/or reactions combining them. This would also introduce an additional consideration: since for the proposed DEL the total possible number of generated molecules was manageable (1.58 million), making it feasible to just precompute proxy probabilities for every molecule, larger variants might require e.g. stochastically sampling smaller batches of possible rewards for the sake of reward computation. Finally, while some (quite mod-

est) improvement in performance was observed for the hierarchical approach for the smaller library subset, it wasn't observed in the case of a larger subset, making it inconclusive to what extent introducing hierarchy is beneficial. Nevertheless, we argue that since the approach itself is quite naive, in particular due to the use of ECFPs on small molecular fragments for clustering, further investigation might be warranted.

REFERENCES

Marie Jeanne Basse, Stephane Betzi, Raphael Bourgeas, Sofia Bouzidi, Bernard Chetrit, Veronique Hamon, Xavier Morelli, and Philippe Roche. 2P2Idb: a structural database dedicated to orthosteric modulation of protein–protein interactions. *Nucleic acids research*, 41(D1):D824–D827, 2012.

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. *Advances in Neural Information Processing Systems*, 34:27381–27394, 2021.

Yoshua Bengio, Salem Lahlou, Tristan Deleu, Edward J Hu, Mo Tiwari, and Emmanuel Bengio. GFlowNet foundations. *Journal of Machine Learning Research*, 24(210):1–55, 2023.

Nicolas Bosc, Christophe Muller, Laurent Hoffer, David Lagorce, Stéphane Bourg, Carine Derviaux, Marie-Edith Gourdel, Jean-Christophe Rain, Thomas W Miller, Bruno O Villoutreix, et al. Fr-PPIChem: An academic compound library dedicated to protein–protein interactions. *ACS chemical biology*, 15(6):1566–1574, 2020.

Adrián Gironda-Martínez, Etienne J Donckele, Florent Samain, and Dario Neri. DNA-encoded chemical libraries: a comprehensive review with succesful stories and future challenges. *ACS Pharmacology & Translational Science*, 4(4):1265–1279, 2021.

Robert A Goodnow Jr, Christoph E Dumelin, and Anthony D Keefe. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nature Reviews Drug Discovery*, 16(2):131–147, 2017.

Moksh Jain, Tristan Deleu, Jason Hartford, Cheng-Hao Liu, Alex Hernandez-Garcia, and Yoshua Bengio. GFlowNets for AI-driven scientific discovery. *Digital Discovery*, 2(3):557–577, 2023.

Katherine S Lim, Andrew G Reidenbach, Bruce K Hua, Jeremy W Mason, Christopher J Gerry, Paul A Clemons, and Connor W Coley. Machine learning on DNA-encoded library count data using an uncertainty-aware probabilistic loss function. *Journal of Chemical Information and Modeling*, 62(10):2316–2331, 2022.

Nikolay Malkin, Moksh Jain, Emmanuel Bengio, Chen Sun, and Yoshua Bengio. Trajectory balance: Improved credit assignment in GFlowNets. *Advances in Neural Information Processing Systems*, 35:5955–5967, 2022.

Kevin McCloskey, Eric A Sigel, Steven Kearnes, Ling Xue, Xia Tian, Dennis Moccia, Diana Gikunju, Sana Bazzaz, Betty Chan, Matthew A Clark, et al. Machine learning on DNA-encoded libraries: a new paradigm for hit finding. *Journal of Medicinal Chemistry*, 63(16):8857–8866, 2020.

Mila AI4Science, Alex Hernandez-Garcia, Alexandre Duval, Alexandra Volokhova, Yoshua Bengio, Divya Sharma, Pierre Luc Carrier, Michał Koziarski, and Victor Schmidt. Crystal-GFN: sampling crystals with desirable properties and constraints. *arXiv preprint arXiv:2310.04925*, 2023.

Xavier Morelli, Raphaël Bourgeas, and Philippe Roche. Chemical and structural lessons from recent successes in protein–protein interaction inhibition (2P2I). *Current opinion in chemical biology*, 15(4):475–481, 2011.

Christian P Robert, George Casella, Christian P Robert, and George Casella. The Metropolis—Hastings algorithm. *Monte Carlo statistical methods*, pp. 267–320, 2004.

David Rogers and Mathew Hahn. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754, 2010.

Alexander L Satz, Andreas Brunschweiger, Mark E Flanagan, Andreas Gloger, Nils JV Hansen, Letian Kuai, Verena BK Kunig, Xiaojie Lu, Daniel Madsen, Lisa A Marcaurelle, et al. DNA-encoded chemical libraries. *Nature Reviews Methods Primers*, 2(1):3, 2022.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

Teague Sterling and John J Irwin. ZINC 15–ligand discovery for everyone. *Journal of chemical information and modeling*, 55(11):2324–2337, 2015.

Alexandra Volokhova, Michał Koziarski, Alex Hernández-García, Cheng-Hao Liu, Santiago Miret, Pablo Lemos, Luca Thiede, Zichao Yan, Alán Aspuru-Guzik, and Yoshua Bengio. Towards equilibrium molecular conformation generation with GFlowNets. *arXiv preprint arXiv:2310.14782*, 2023.

Fang Wu, Dragomir Radev, and Stan Z Li. Molformer: Motif-based transformer on 3D heterogeneous molecular graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 5312–5320, 2023.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018a.

Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. Representation learning on graphs with jumping knowledge networks. In *International conference on machine learning*, pp. 5453–5462. PMLR, 2018b.

A TRAINING DETAILS

A.1 PROXY

GNN model consisted of 5 GIN layers Xu et al. (2018a) with hidden dimensionality of 500, utilized Jumping Knowledge shortcuts Xu et al. (2018b) and had a single output MLP layer. Pretraining was done in an unsupervised fashion on ZINC15 dataset Sterling & Irwin (2015). Training was done for 30 epochs using Adam optimizer with learning rate of 5×10^{-5} and batch size of 50. **Molformer** used published pretrained weights, and had 12 layers with dimensionality of 768 each. Finetuning was done for 50 epochs, using learning rate of 3×10^{-5} and batch size of 16. **Random Forest** used 1000 trees.

A.2 GFLOWNET

GFlowNets were trained for 5000 training iterations, using learning rate of 1×10^{-4}, batch size of 50 forwards samples and 50 replay samples, with prioritized replay buffer of up to 1000 samples. Both forward and backward policies were 5 layer MLPs, with hidden dimensionality of 512. Training was done using Trajectory Balance loss, $\beta = 64$ and random action probability of 0.1.

A.3 MCMC

We used Metropolis–Hastings variant of MCMC. We initialize chains by sampling random binary vectors corresponding to the library size within specified size constraints. Importantly, unlike for GFlowNets, we allow bidirectional bit flipping. We use the number of random chains equal to

number of target samples (5000), and use chain length of 250. Importantly, we disallow actions that would lead to exceeding specified library size, and do not count them towards the chain length. Note that in this setting, we use $5\times$ the budget of proxy calls compared to the GFlowNet.

A.4 PPO

The PPO policy was trained for 2000 iterations. The training converges within 1500 iterations. We use a learning rate of 1×10^{-4} with a decay period of 1×10^{6} and a decay coefficient of 0.5. We use Adam optimizer, a batch size of 2, a random action probability of 0.001 and a reward scaling factor (β) of 64. The policy is a MLP with two hidden layers of size 256 each. We use a clipping factor (ϵ) of 0.1 and an entropy coefficient of 1×10^{-3}. We collect 64 trajectories per iteration and train for 16 epochs per iteration.

B LARGE LIBRARY SAMPLING

The experiments from Section 4.2 were repeated with library size constrained between 90k and 100k. Note that in contrast to the experiment with smaller library, due to computational considerations batch size was restricted to 25 forward samples and 25 replay samples. For the greedy baseline, we used 35/35/80 building blocks per cycle. Importantly, the large scale experiment did not seem to converge in the specified number of iterations, so the results should be treated as preliminary.

Table 3: Comparison of different DEL sampling strategies, with requested library size between 90k and 100k. Best performance denoted in bold.

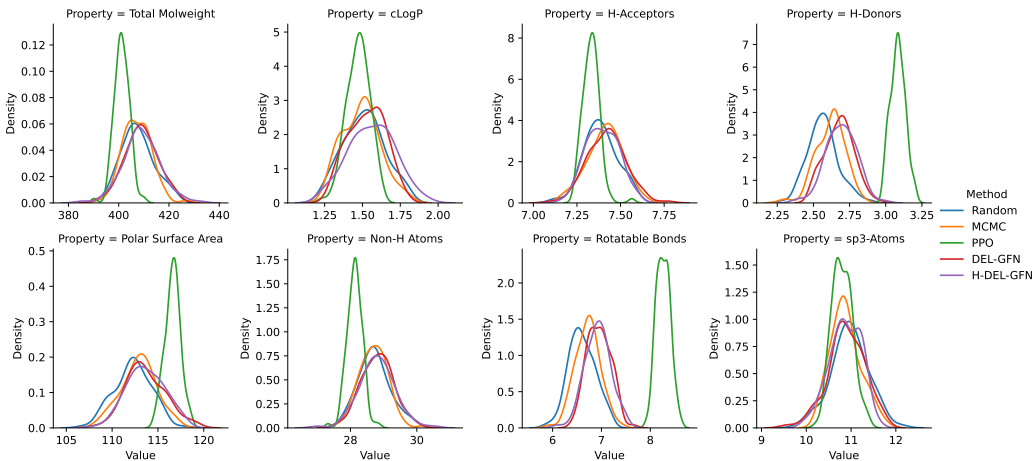| Method	Prob.	Div.	Top-100 prob.	Top-100 div.	Top-1 prob.
Random	0.582 ± 0.000	$\mathbf{0.480 \pm 0.000}$	0.671 ± 0.002	$\mathbf{0.473 \pm 0.001}$	0.726 ± 0.013
Greedy	-	-	-	-	0.975
MCMC	0.619 ± 0.000	0.478 ± 0.000	0.706 ± 0.002	0.466 ± 0.001	0.747 ± 0.012
PPO	$\mathbf{0.977 \pm 0.003}$	0.103 ± 0.010	$\mathbf{0.986 \pm 0.002}$	0.086 ± 0.008	$\mathbf{0.988 \pm 0.001}$
DEL-GFN	0.669 ± 0.001	0.475 ± 0.000	0.755 ± 0.002	0.455 ± 0.001	0.797 ± 0.003
H-DEL-GFN	0.656 ± 0.004	0.476 ± 0.001	0.749 ± 0.003	0.456 ± 0.001	0.786 ± 0.004

Figure 6: Distributions of average chemical library properties of top-100 generated libraries, with requested library size between 90k and 100k.