THE POWER OF MOTIFS AS INDUCTIVE BIAS FOR LEARNING MOLECULAR DISTRIBUTIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Machine learning for molecules holds great potential for efficiently exploring the vast chemical space and thus streamlining the drug discovery process by facilitating the design of new therapeutic molecules. Deep generative models have shown promising results for molecule generation, but the benefits of specific inductive biases for learning distributions over small graphs are unclear. Our study aims to investigate the impact of subgraph structures and vocabulary design on distribution learning, using small drug molecules as a case study. To this end, we introduce Subcover, a new subgraph-based fragmentation scheme, and evaluate it through a two-step variational auto-encoder. Our results show that Subcover's improved identification of chemically meaningful subgraphs leads to a relative improvement of the FCD score by 30%, outperforming previous methods. Our findings highlight the potential of Subcover to enhance the performance and scalability of existing methods, contributing to the advancement of drug discovery.

1 INTRODUCTION

Generative models for molecules offer a way to create new compounds with specific properties, which can be useful in various fields, including drug discovery, material science, and chemistry (Bian & Xie, 2021; Choudhary et al., 2022; Hetzel et al., 2022; Zhu et al., 2022; Du et al., 2022). The ability to navigate and explore the vast chemical space more efficiently and generate novel molecules in an automated fashion can save time and resources compared to traditional laboratory methods (Reymond et al., 2012; Polishchuk et al., 2013). This way, generative models can help discover molecules with unique properties that may not have been found otherwise. Additionally, generative models can assist in optimising existing structures and provide candidate compounds with improved properties (Gao et al., 2022).

To be useful for such applications, any model must be able to abstract molecules in a way that enables it to generate new structures representative of the underlying distribution. A common approach is to learn a continuous latent distribution that captures the discrete structure of molecules (Jin et al., 2018; Liu et al., 2018; Kusner et al., 2017). This involves learning the present (i) patterns, such as rings and functional groups, (ii) relationships, such as particular bond types, and (iii) structures, the complex combinations of such geometries, and using that knowledge to generate new, diverse, and yet meaningful molecular structures.

There are two predominant approaches for learning distributions of molecules and decoding latent representations back to molecular graphs: atom-based and motif-based approaches (Yang et al., 2022). Atom-based models utilise individual atoms as the building blocks of the molecular graph, allowing them, in principle, to model any molecular structure and create highly diverse compounds. However, these models struggle to generate complex and highly symmetric patterns, such as rings (Yang et al., 2022). Motif-based models, on the other hand, extend the available building blocks with a vocabulary of common fragments or "motifs", such as a carbon ring. While a fragment can refer to any molecular substructure, motifs should represent complex geometries that affect molecular properties and provide a chemically informed inductive bias for learning molecular distributions, see Figure 1.

Yet, the expressive power of motif-based models is determined by the size and quality of the motif vocabulary, as well as the particular decomposition of molecules into motifs. We refer to the combination of the two, vocabulary and decomposition, as a fragmentation scheme. Jin et al. (2018) introduce a scheme that decomposes molecules exclusively into motifs, resulting in a vocabulary that

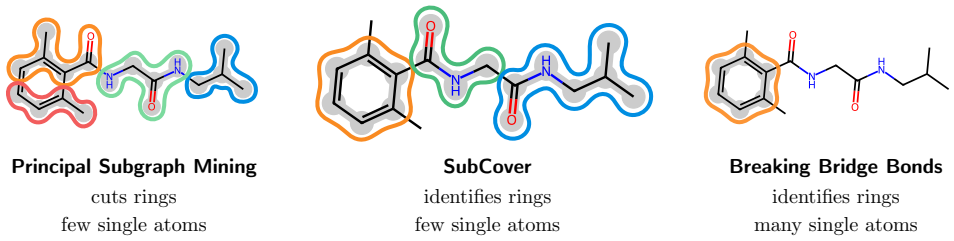

Figure 1: Example of a molecule's decomposition according to different fragmentation schemes. Subcover captures cyclic structures while using only few single atoms.

increases with the dataset size and requires the inclusion of many similar motifs. Other approaches overcome the limitation of dataset-dependent vocabulary size by including single atoms in their decomposition (Maziarz et al., 2022; Kong et al., 2022). These fragmentation schemes, however, fail to provide chemically distinct motifs and low numbers of single atoms simultaneously. Note that the latter is crucial, as a decomposition with too many single atoms hinders learning the molecules' structure. We believe that a well-designed fragmentation scheme provides an inductive bias that balances the ease of learning the molecules' decompositions and their structural properties. As a consequence, learning molecular distributions consists of two main challenges: Identifying the building blocks of an individual compound, which are determined by the respective fragmentation scheme, and learning the structure between those.

In this study, we propose Subcover, a new approach for fragmenting molecules, and examine its benefits for learning distributions of small molecules. To this end, we rely on a two-step variational auto-encoder (VAE), that extends the work by Kong et al. (2022), as this model type provides good insights into the impact of the fragmentation schemes' inductive biases.

Our main contributions can be summarised as follows:

- We investigate the impact of fragmentation methods on learning molecular graph distributions.
- We introduce Subcover, which consistently identifies large, chemically relevant motifs within a molecular graph.
- We show that Subcover improves the learning of molecular graph distributions.

2 Inductive Biases through Motif Vocabularies

Inductive biases are essential for distribution learning in high-dimensional discrete spaces as they help to reduce their combinatorial complexity. In the case of molecules, identifying motifs—frequently occurring patterns across a molecule dataset—has proven to be a strong prior facilitating the learning of graph distributions. In the following, we introduce the formalities required for building motif vocabularies. Based on this, we present two representative fragmentation schemes, Principled Subgraph Mining (PSM) (Kong et al., 2022) and Breaking Bridge Bonds (BBB) (Maziarz et al., 2022), and highlight their benefits and drawbacks. On this basis, we introduce a new fragmentation scheme: Subcover.

Preliminaries A molecular graph G can be described by the tuple $\{\mathcal{A}, \mathbf{B}\}$, where $\mathcal{A}$ is the multiset of atoms of size $a = |\mathcal{A}|$ and $\mathbf{B} \in \{0,1\}^{a \times a \times 3}$ the bonds tensor, indicating both the existence and type of a bond: single-, double-, or triple-bond. Atoms and larger subgraphs can occur multiple times within one molecule and are consequently represented by a multiset. The goal of graph generation is to learn the distribution $\mathbb{P}(\mathcal{G})$ over the set of graphs $\mathcal{G} = \{G_i\}$ and subsequently sample new graphs $G_{\text{new}} \sim \mathbb{P}(\mathcal{G})$ (Zhu et al., 2022).

For any given molecule i, fragmentation schemes partition the multiset of atoms $\mathcal{A}_i$ and their connectivity structure $\mathbf{B}$ into a multiset of fragments $\mathcal{F}_i$ whose union of elements f corresponds again to $\mathcal{A}$, $\mathcal{A} = \bigcup_{f \in \mathcal{F}} f$. Here, a fragment f can be any substructure of G_i, i.e. a single atom s or the elements of a connected subgraph that includes more than one atom. Together with a motif

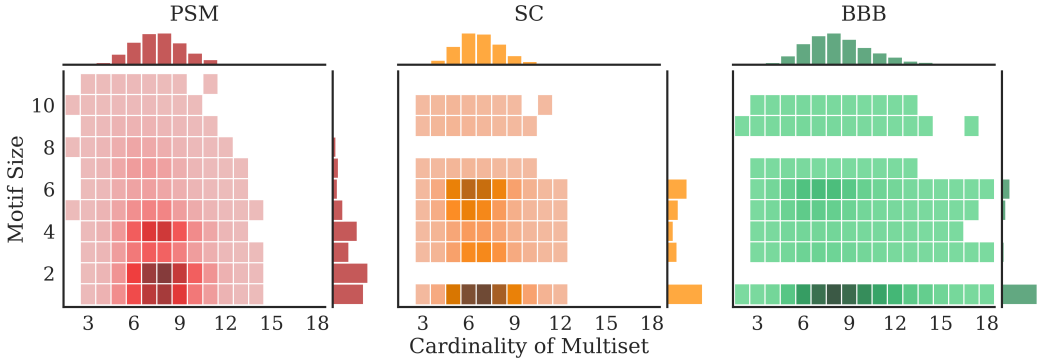

Figure 2: Different fragmentation schemes lead to different decomposition sizes $|\mathcal{F}|$. BBB leads to larger decompositions due to the increased number of single atoms $\mathcal{S}$, while PSM mostly decomposes into small motifs. Subcover combines the strength of the two: large motifs and the lowest cardinality among fragmentation schemes.

vocabulary $\mathcal{V}$, such a decomposition $\mathcal{F}$ can further be represented as $\mathcal{F} = \mathcal{M} \cup \mathcal{S}$, where $\mathcal{M}$ and $\mathcal{S}$ correspond to the multisets of motifs m—a motif m has to include more than one atom—and single atoms s, respectively. Note that we refer to those fragments f that are contained in the vocabulary $\mathcal{V}$ as motifs $m := f \in \mathcal{V}$. Further, the size of the vocab $|\mathcal{V}| = k$ is often variable, and the decision about which fragments f_i to include is usually based on a heuristic, for example, the frequency of f_i in the dataset. Some fragmentation schemes achieve decompositions of $\mathcal{A}_i$ that are completely described by motifs, $\mathcal{S}_i = \emptyset$ (Jin et al., 2018), while others use a combination of the two, $\mathcal{M}_i \neq \emptyset$ and $\mathcal{S}_i \neq \emptyset$. We associate a connectivity structure with each motif m, which allows to decompose the bond tensor $\mathbf{B}$ of a molecule G_i into $\mathbf{B} = \mathbf{B}_{\mathcal{M}_i} \cup \mathbf{B}_{\overline{\mathcal{M}_i}}$, where $\mathbf{B}_{\mathcal{M}_i}$ defines all intra-motif bonds and $\mathbf{B}_{\overline{\mathcal{M}_i}}$ all other bonds between motifs and single atoms.

On the spectrum of fragmentation schemes, we choose two representative approaches to investigate the inductive bias they provide. While Kong et al. (2022) identify motifs in a bottom-up manner by starting from single atoms and building up larger structures through merging, the approach by Maziarz et al. (2022) is top-down and leaves all cyclic structures intact. Both methods decompose molecules G into single atoms $\mathcal{S}$ and motifs $\mathcal{M}$.

Principled Subgraph Mining Kong et al. (2022) initialise the vocabulary generation process from the multiset of single atoms. Using these as initial fragments, the vocabulary $\mathcal{V}$ is built up one motif at a time. During each generation step, the current vocabulary is used to decompose all molecules into fragments. The resulting fragments are combined with their neighbours to form new candidate motifs and, among those, the most frequent one is added as a motif to the vocabulary. This way, the vocabulary $\mathcal{V}$ also includes small motifs, such as a CC chain, which reduces the number of single atoms $\mathcal{S}_i$ within a molecule's fragmentation. To identify the $\mathcal{M}$ and $\mathcal{S}$ multisets of a molecule given a fixed motif vocabulary, the same steps of iterative merging of adjacent fragments must be performed. Starting from single atoms, two fragments are merged only if the resulting fragment is part of the vocabulary $\mathcal{V}$ and its frequency is the highest among all merged fragments. This process is repeated until no merged fragment is included in the vocabulary anymore.

Breaking Bridge Bonds Unlike Principled Subgraph Mining, which is a data-driven approach, the idea of breaking bridge bonds (Maziarz et al., 2022) relies on chemical knowledge. During vocabulary construction, a molecule is fragmented by breaking all acyclic bonds adjacent to cyclic structures. For example, the connecting bond between two ring structures will be cut, and likewise, any dangling atom attached to a ring. The final set of k motifs is selected according to their frequency after applying this fragmentation to all molecules in the dataset. Following Maziarz et al. (2022), only motifs with a size of at least three atoms are considered. Once the vocabulary $\mathcal{V}$ is fixed, the fragmentation procedure is applied to a graph G_i and those fragments included in $\mathcal{V}$ are represented as motifs $\mathcal{M}_i$. The remaining fragments are represented as single atoms $\mathcal{S}_i$.

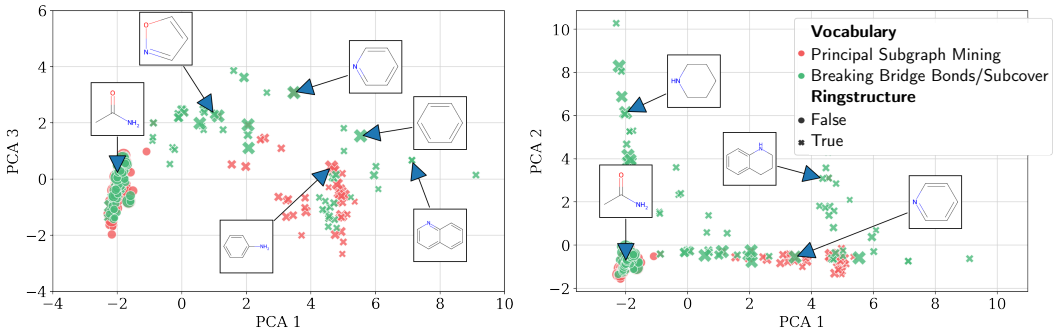

Figure 3: First three principal components for motif fingerprints for PSM and BBB / Subcover vocabularies of size $k = 128$. PSM identifies fewer rings but many chain-like structures instead. These show little variety in the first three principle components. The size of the markers indicates the frequency of a particular motif, for the BBB / Subcover vocabulary we show the counts of Subcover. In total, the vocabulary of PSM consists of 36% ring-based motifs, while BBB / Subcover has 55% of ring-like structures.

Both methods, PSM and BBB, reduce the number of single atoms $\mathcal{S}$ through the construction of $\mathcal{V}$, which by itself provides a good inductive bias as it decreases the degrees of freedom required to represent molecules. Yet, simply reducing $|\mathcal{S}|$ is insufficient. While the bottom-up PSM approach leads to many molecule decompositions that are completely described by motifs, $G_i = \mathcal{M}_i$ and $\mathcal{S}_i = \emptyset$, the identified motifs often contain few atoms, cf. Figure 2. In addition, Figure 3 shows the qualitative difference between identified motifs revealing an additional drawback of PSM: The motifs in a PSM vocabulary are often chain-like and chemically indistinct. This implies that rings are cut and defining properties of G_i are not reflected by its motif set $\mathcal{M}_i$. In contrast, the top-down BBB approach focuses on leaving cyclic structures intact and considers everything else, i.e. chain-like structures, as "remainder" fragments. While this maintains chemical integrity, it also leads to many fragments $f \notin \mathcal{V}$ being represented as single atoms. This happens even though a motif $m \in \mathcal{V}$ exists which is contained in f, $m \subset f$.

Algorithm 1 Subcover

Require: Molecule G, vocabulary $\mathcal{V}$
  $\{f_i\}_{i=0}^{j} \leftarrow$ BreakBridgeBonds(G)
  $\mathcal{F} \leftarrow \{\}$
  for $i \in \{0, \ldots, j\}$ **do**
    if $f_i \in \mathcal{V}$ **then**
      $\mathcal{F} \leftarrow \mathcal{F} \cup f_i$
    else
      $\mathcal{F} \leftarrow \mathcal{F} \cup$ FindMInF$(f_i, \{\})$
    end if
  end for
  procedure FINDMINF$(f, \mathcal{F})$
    $m^* \leftarrow$ None
    for m in $\mathcal{V}$ **do**
      if $m \subset f$ and $|m| \geq |m_*|$ **then**
        if and $c(m) > c(m_*)$ **then**
          $m^* \leftarrow m$
        end if
      end if
    end for
    if m^* is not None **then**
      $f \leftarrow$ DeleteMotifFromFrag(f, m^*)
      $\mathcal{F} \leftarrow \mathcal{F} \cup m^*$
      $\mathcal{F} \leftarrow$ FindMInF$(f, \mathcal{F})$
    end if
    return $\mathcal{F}$
  end procedure

Subcover To combine the strengths of both approaches, we use the same fragmentation as BBB to construct the motif vocabulary $\mathcal{V}$. For the decomposition, however, we go beyond the initial comparison of fragments f and motifs $m \in \mathcal{V}$. If $f \notin \mathcal{V}$, we recursively search for matching subgraphs $m \subset f$. By identifying substructures, we aim to reduce the number of single atoms as PSM does. Should multiple motifs m_j be contained in the fragment f, which is often the case, the largest and most frequent motif m_* is selected and the fragment further decomposed, $f = m_* \cup \hat{f}$. This process is repeated with $\hat{f}$ until no more motifs can be identified and the resulting decomposition $f = \mathcal{M}_f \cup \mathcal{S}_f$ is added to the multisets $\mathcal{M}_i$ and $\mathcal{S}_i$ of the molecule G_i. This reduces the number of single atoms $\mathcal{S}_i$ significantly and achieves a more concise, yet chemically meaningful fragmentation, see Figure 2. Note that our way of subgraph identification in Subcover

allows mapping structures that are identical up to a charge. While we are not explicitly decoding ions with our model yet, this facilitates the capturing of the molecule's geometry and topology, even for small vocabularies. A formalisation of Subcover is presented in Algorithm 1.

3 TWO-STEP VAE

Fully autoregressive approaches are common for learning distributions of molecules. Yet, the learnt distribution can be difficult to interpret and evaluate independently of the applied corrections and postprocessing. We introduce a two-step VAE approach, which first predicts the multiset of atoms and motifs in a recurrent fashion and subsequently infers their bonds. We build on top of the architecture of (Kong et al., 2022) and equip our model with multiplicity and motif features to fully harness the provided vocabulary. Our generative model follows a variational encoder-decoder architecture (Kingma & Welling, 2013). In doing so, we train an encoder to map the structural graph G_i of a molecule i to a d-dimensional latent representation $\mathbf{z}_i$. In a subsequent step, the molecule is reconstructed by applying a two-step decoder, that separates the reconstruction of molecule fragments and structure.

Encoder The encoder learns the variational posterior $q_\theta(\mathbf{z}_i \mid G_i)$ that defines the latent representation of each molecule i through $\mathbf{z}_i|G_i \sim \mathcal{N}\big(\mu_\theta(G_i), \sigma^2_\theta(G_i)I\big)$. We parameterise the mean and variance of the approximate posterior as:

$$\big(\mu_\theta(G_i), \sigma^2_\theta(G_i)\big) = h_\theta\big([g_\theta(\mathbf{X}_i, \mathbf{E}_i), \mathbf{f}_i]\big), \tag{1}$$

where $g_\theta(\cdot,\cdot)$ is a multi-layer graph neural network (GNN), whose message-passing operation acts on the original molecular graph with the node and edge features $\mathbf{X}_i$ and $\mathbf{E}_i$, respectively, and returns a graph representation. The initial node and edge features are learned. Each node $v \in V$ in G_i is represented by a node feature vector $\mathbf{x}_{v;i} \in \mathbf{X}_i$, that is given by a concatenation of learned embeddings of the corresponding motif and atom identifiers and multiplicity information, as well as its motif fingerprint information. The connectivity of the molecular graph is encoded by a learned edge embedding that indicates the bond type. For $g_\theta(\cdot,\cdot)$, we leverage a multi-layer transformer convolution backbone (Shi et al., 2021) and attain the graph representation of each molecule by a learned aggregation of the transformed node features, see A.1 for more details. This graph representation is concatenated with a learned embedding of a molecule's fingerprint information $\mathbf{f}_i$ before it is mapped to the mean and variance by a multilayer perceptron (MLP) h_θ.

Decoder The decoder is trained to reconstruct the molecule given the latent representation $\mathbf{z}_i$ sampled from the approximate posterior q_θ. In a first step, we autoregressively predict the multiset of a molecule's fragments. We leverage a simple one-layer recurrent neural network (RNN) for the multiset prediction by casting it as a fragment classification task. Formally, the RNN is trained to minimise the negative log-likelihood:

$$\mathcal{L}_M = -\sum_{j=1}^{N} \log \mathbb{P}_\phi(f_j \mid f_{j-1}, \ldots, f_0, \mathbf{z}), \tag{2}$$

where $\mathbb{P}_\phi(f_j|f_{j-1}, \cdots, f_0, \mathbf{z})$ is trained to decode the sequence of molecule fragments f_j starting from a start token f_0 and ending with a stop token.

In a second step, we use the multiset of molecule fragments $\mathcal{F}_i$ and latent representation $\mathbf{z}_i$ to infer the structure of the molecule, i.e. the existence of bonds and their types. To predict the molecule's structure, we train two MLPs to parameterise $\mathbb{P}(\mathbf{b}_{u,v} \mid \mathbf{z}, \mathbf{x}_u, \mathbf{x}_v)$, where the node features are attained by a multi-layer GNN on the graph of atoms, see A.2 for more details. This graph is fully connected except for the known intra-motif connections. We leverage the node embeddings from the encoder and contextualise them further with $\mathbf{z}_i$ and the multiset of molecule fragments, which is only possible due to our canonical representation of molecules. Finally, we concatenate the node representations across all GNN layers and predict the adjacencies and bond types to minimise the log-likelihood:

$$\mathcal{L}_S = -\sum_{\substack{v,u \in V \\ u \neq v}} \log \mathbb{P}(\mathbf{b}_{u,v} \mid \mathbf{z}, \mathbf{x}_u, \mathbf{x}_v).$$

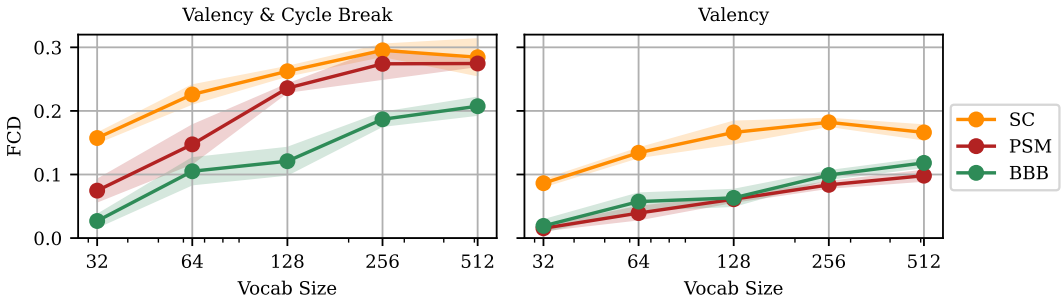

Figure 4: FCD achieved with different fragmentation schemes over increasing vocabulary size with cycle breaking postprocessing (left) and only valency correction (right).

The VAE is trained to minimise the following loss function:

$$\mathcal{L} = \beta(\lambda_S \mathcal{L}_S + \lambda_M \mathcal{L}_M + \lambda_V \mathcal{L}_V) + (1 - \beta) D_{\mathrm{KL}}, \tag{3}$$

where D_{KL} is the KL-Divergence between the variational posterior q_θ and a standard normal prior, and $\mathcal{L}_V$ is a valency regularization loss on the generated molecule.

In addition to incorporating a valency regularisation into the training process, we employ two forms of post-processing during inference. Specifically, we perform a valency correction that enforces adherence to the valid valency of each atom by restricting the number of bonds accordingly. Further correction can be done via a "Cycle Breaking" procedure, which limits cyclic structures outside of motifs to a size of either five or six (Kong et al., 2022). When applying this procedure, we allow no more than two shared atoms between such cycles. The decision on which bonds to include is based on the model's confidence about its prediction.

4 Experiments

Data We conduct all experiments on the ZINC molecule dataset (Kusner et al., 2017), which contains a total of 249,457 molecules from the ZINC database (Gómez-Bombarelli et al., 2018). The dataset is randomly split into training, validation, and test sets with a ratio of 0.8 / 0.1/ 0.1. All vocabularies are extracted from the entire dataset and the specified vocabulary size does not include single atoms. We train all models for 20 epochs and the hyperparameters used for the baselines are taken from their respective publications. Meanwhile, the hyperparameters for our model are described in Appendix B. For every model-vocabulary pair, we report the mean and standard deviation over three random model initialisations. Results for sampling new molecules from the learnt distribution are reported for a sample set of 10,000 molecules.

Comparison of Fragmentation Schemes The Fréchet Chemnet Distance (FCD) (Preuer et al., 2018) is a metric to determine the distance between two sets of molecular graphs. It takes into account both chemical and biological information, as well as the diversity within each set. The comparison of the fragmentation schemes Subcover (SC), Principled Subgraph Mining (PSM), and Breaking Bridge Bonds (BBB), combined with heavy postprocessing (Valency & Cycle Break), as well as postprocessing only to ensure chemical validity of bonds (Valency), is depicted in Figure 4. To generate new molecules, we sample latent embeddings $\mathbf{z}_i \sim \mathcal{N}(0, 1)$ and apply our decoder. Experiments beyond sampling from a standard Gaussian distribution can be found in Appendix E, where Subcover achieves FCD scores over 0.35. Our results show that SC outperforms the other fragmentation methods. Although the motifs found by BBB may have chemical significance, the high rate of single atoms hinders learning of the distribution over the multiset and, consequently, the molecular structure. The fact that PSM achieves a worse FCD score than SC, even though its single atom rate is the lowest across approaches, indicates that only reducing the number of single atoms in a molecule fragmentation is not sufficient. This is further supported by the FCD scores that the fragmentation schemes achieve without major postprocessing. These findings also suggest that due to the indistinctiveness of the PSM motifs, they are difficult to connect correctly. The fact that PSM can only achieve reasonable FCD scores with cycle correction is likely a result of breaking apart rings, as the model must then learn to create rings and may do so excessively during inference.

Table 1: Performance metrics for several graph generation models. We highlight the best results **overall** and of two-step approaches. For all metrics except SA, higher values indicate superior performance. All models are evaluated on a vocabulary of size 256, except for JTVAE, which has a fixed vocabulary size of around 780. Results for varying vocabulary size are in Appendix D.

	FCD ↑	KL ↑	Int. Div ↑	SA ↓	QED ↑
JT-VAE	0.70 ± 0.006	0.95 ± 0.004	0.86 ± 0.001	0.89 ± 0.006	0.07 ± 0.003
MoLer	$\mathbf{0.80 \pm 0.016}$	$\mathbf{0.98 \pm 0.002}$	0.87 ± 0.001	0.56 ± 0.032	0.10 ± 0.002
PS-VAE (CB)	0.23 ± 0.001	$\underline{0.83 \pm 0.001}$	$\underline{\mathbf{0.89 \pm 0.001}}$	1.86 ± 0.001	0.14 ± 0.001
PS-VAE (V)	0.08 ± 0.001	0.67 ± 0.001	$\underline{\mathbf{0.89 \pm 0.001}}$	2.85 ± 0.010	$\underline{\mathbf{0.17 \pm 0.002}}$
Ours (CB)	$\underline{0.30 \pm 0.009}$	0.82 ± 0.010	0.86 ± 0.002	$\underline{\mathbf{0.52 \pm 0.029}}$	0.04 ± 0.009
Ours (V)	0.15 ± 0.004	0.70 ± 0.007	0.88 ± 0.002	1.94 ± 0.028	0.09 ± 0.007

As the vocabulary size increases, both Subcover and PSM stagnate in performance. Overall, the results indicate superior performance of Subcover in the small-vocabulary settings compared to all fragmentation scheme baselines.

Molecular Distribution Learning Finally, to contextualise the results of this study, we provide a comparison of our model with the best-performing fragmentation scheme, Subcover, to three state-of-the-art molecule generation methods: Junction Tree VAE (JTVAE) (Jin et al., 2018), MoLer Maziarz et al. (2022) and PS-VAE Kong et al. (2022). We defer an overview of related work for molecule generation methods to Appendix C. As per (Brown et al., 2019; Polykovskiy et al., 2020), the Fréchet Chemnet Distance, the Kullback-Leibler Divergence, and Internal Diversity are normalised to a scale of $[0, 1]$, with higher values indicating superior performance. Besides FCD, the KL metric measures the Kullback-Leibler Divergence of two sets of properties with respect to physiochemical properties and the Internal Diversity specifies distances of molecules within the generated set based on Tanimoto similarity (Benhenda, 2017). Additionally, we report results on the quantitative estimate of drug-likeness (QED) (Bickerton et al., 2012) as well as the synthetic accessability of drug-like molecules (SA) (Ertl & Schuffenhauer, 2009). Our model trained with Subcover shows improved results on the FCD metric compared to the two-step PS-VAE approach, especially without correction. All evaluated models perform consistent w.r.t. diversity of the generated molecules. Specifically, both MoLer and our method generate synthetically accessible molecules, emphasizing the significance of chemical rules for molecule fragmentation and vocabulary construction. Among the autoregressive models, MoLer demonstrates strong performance across all metrics. The findings of this study regarding the impact of inductive biases are also relevant to autoregressive models. Although they inherently possess a more effective approach to handling individual atoms through their autoregressive nature, they, too, can benefit from a consistent and meaningful fragmentation scheme.

5 Conclusion

We present Subcover, a novel fragmentation scheme that combines the benefits of both data-driven and rule-based techniques to provide an effective inductive bias for learning molecular distributions. Subcover outperforms existing fragmentation methods by producing more meaningful motifs and utilising the vocabulary more efficiently, thus reducing single atoms. These attributes are shown to be crucial for accurately learning distributions over molecular graphs. Subcover has the potential to improve the performance of various other graph generation models, including autoregressive models. We believe that the findings of this work will not only be useful for the advancement of molecular graph generation but also have broader implications for related fields, such as drug discovery, materials science, and computational biology.

REFERENCES

Sungsoo Ahn, Binghong Chen, Tianzhe Wang, and Le Song. Spanning tree-based graph generation for molecules. In *International Conference on Learning Representations*, 2021.

Rim Assouel, Mohamed Ahmed, Marwin H Segler, Amir Saffari, and Yoshua Bengio. Defactor: Differentiable edge factorization-based probabilistic graph generation. *arXiv preprint arXiv:1811.09766*, 2018.

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.

Mostapha Benhenda. ChemGAN challenge for drug discovery: can AI reproduce natural chemical diversity? *arXiv preprint arXiv:1708.08227*, 2017.

Yuemin Bian and Xiang-Qun Xie. Generative chemistry: drug discovery with deep learning generative models. *Journal of Molecular Modeling*, 27(3):71, February 2021. ISSN 0948-5023. doi: 10.1007/s00894-021-04674-8. URL https://doi.org/10.1007/s00894-021-04674-8.

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, February 2012. ISSN 1755-4349. doi: 10.1038/nchem.1243. URL https://www.nature.com/articles/nchem.1243.

Xavier Bresson and Thomas Laurent. A two-step graph convolutional decoder for molecule generation. *arXiv preprint arXiv:1906.03412*, 2019.

Nathan Brown, Marco Fiscato, Marwin H. S. Segler, and Alain C. Vaucher. GuacaMol: Benchmarking Models for De Novo Molecular Design. *Journal of Chemical Information and Modeling*, 59(3): 1096–1108, March 2019. ISSN 1549-9596, 1549-960X. doi: 10.1021/acs.jcim.8b00839. URL http://arxiv.org/abs/1811.09621.

Kamal Choudhary, Brian DeCost, Chi Chen, Anubhav Jain, Francesca Tavazza, Ryan Cohn, Cheol Woo Park, Alok Choudhary, Ankit Agrawal, Simon J. L. Billinge, Elizabeth Holm, Shyue Ping Ong, and Chris Wolverton. Recent advances and applications of deep learning methods in materials science. *npj Computational Materials*, 8(1):1–26, 2022. ISSN 2057-3960. doi: 10.1038/s41524-022-00734-6. URL https://www.nature.com/articles/s41524-022-00734-6.

Nicola De Cao and Thomas Kipf. Molgan: An implicit generative model for small molecular graphs. *arXiv preprint arXiv:1805.11973*, 2018.

Yuanqi Du, Tianfan Fu, Jimeng Sun, and Shengchao Liu. Molgensurvey: A systematic survey in machine learning models for molecule design. *arXiv preprint arXiv:2203.14500*, 2022.

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

Daniel Flam-Shepherd, Tony Wu, and Alan Aspuru-Guzik. Graph deconvolutional generation. *arXiv preprint arXiv:2002.07087*, 2020.

Wenhao Gao, Tianfan Fu, Jimeng Sun, and Connor W Coley. Sample efficiency matters: a benchmark for practical molecular optimization. *arXiv preprint arXiv:2206.12411*, 2022.

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. Publisher: ACS Publications.

Leon Hetzel, Simon Böhm, Niki Kilbertus, Stephan Günnemann, Mohammad Lotfollahi, and Fabian Theis. Predicting single-cell perturbation responses for unseen drugs. *arXiv preprint arXiv:2204.13545*, 2022.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *International conference on machine learning*, pp. 2323–2332. PMLR, 2018.

Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical Generation of Molecular Graphs using Structural Motifs. Technical report, International Conference on Machine Learning, April 2020.

Hiroshi Kajino. Molecular hypergraph grammar with its application to molecular optimization. In *International Conference on Machine Learning*, pp. 3183–3191. PMLR, 2019.

Yash Khemchandani, Stephen O'Hagan, Soumitra Samanta, Neil Swainston, Timothy J. Roberts, Danushka Bollegala, and Douglas B. Kell. DeepGraphMolGen, a multi-objective, computational strategy for generating molecules with desirable properties: a graph convolution and reinforcement learning approach. *Journal of cheminformatics*, 12(1):1–17, 2020. Publisher: BioMed Central.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Xiangzhe Kong, Wenbing Huang, Zhixing Tan, and Yang Liu. Molecule Generation by Principal Subgraph Mining and Assembling, September 2022. URL http://arxiv.org/abs/2106.15098. arXiv:2106.15098 [cs, q-bio].

Matt J. Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar Variational Autoencoder. Technical Report arXiv:1703.01925, arXiv, March 2017. URL http://arxiv.org/abs/1703.01925. arXiv:1703.01925 [stat] type: article.

G. Landrum. RDKit, 2010.

Yujia Li, Oriol Vinyals, Chris Dyer, Razvan Pascanu, and Peter Battaglia. Learning Deep Generative Models of Graphs. Technical Report arXiv:1803.03324, arXiv, March 2018. URL http://arxiv.org/abs/1803.03324. arXiv:1803.03324 [cs, stat] type: article.

Jaechang Lim, Sang-Yeon Hwang, Seungsu Kim, Seokhyun Moon, and Woo Youn Kim. Scaffold based molecular design using graph generative model. *Chemical Science*, 11(4):1153–1164, 2020. ISSN 2041-6520, 2041-6539. doi: 10.1039/C9SC04503A. URL http://arxiv.org/abs/1905.13639. arXiv:1905.13639 [cs, q-bio, stat].

Meng Liu, Keqiang Yan, Bora Oztekin, and Shuiwang Ji. GraphEBM: Molecular Graph Generation with Energy-Based Models. Technical Report arXiv:2102.00546, arXiv, April 2021. URL https://arxiv.org/abs/2102.00546.

Qi Liu, Miltiadis Allamanis, Marc Brockschmidt, and Alexander L. Gaunt. Constrained Graph Variational Autoencoders for Molecule Design. Technical report, Advances in Neural Information Processing Systems, 2018.

Youzhi Luo, Keqiang Yan, and Shuiwang Ji. GraphDF: A Discrete Flow Model for Molecular Graph Generation. Technical report, International Conference on Machine Learning, June 2021.

Tengfei Ma, Jie Chen, and Cao Xiao. Constrained Generation of Semantically Valid Graphs via Regularizing Variational Autoencoders. Technical report, Advances in Neural Information Processing Systems, 2018.

Krzysztof Maziarz, Henry Jackson-Flux, Pashmina Cameron, Finton Sirockin, Nadine Schneider, Nikolaus Stiefl, Marwin Segler, and Marc Brockschmidt. Learning to Extend Molecular Scaffolds with Structural Motifs. Technical Report arXiv:2103.03864, arXiv, April 2022. URL http://arxiv.org/abs/2103.03864. arXiv:2103.03864 [cs, q-bio] type: article.

Rocío Mercado, Tobias Rastemo, Edvard Lindelöf, Günter Klambauer, Ola Engkvist, Hongming Chen, and Esben Jannik Bjerrum. Graph Networks for Molecular Design. *Machine Learning: Science and Technology*, 2(2):025023, 2021. Publisher: IOP Publishing.

AkshatKumar Nigam, Pascal Friederich, Mario Krenn, and Alán Aspuru-Guzik. Augmenting genetic algorithms with deep neural networks for exploring the chemical space. *arXiv preprint arXiv:1909.11655*, 2019.

P. G. Polishchuk, T. I. Madzhidov, and A. Varnek. Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of Computer-Aided Molecular Design*, 27(8):675–679, August 2013. ISSN 1573-4951. doi: 10.1007/s10822-013-9672-4.

Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez-Lengeling, Sergey Golovanov, Oktai Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, Artur Kadurin, Simon Johansson, Hongming Chen, Sergey Nikolenko, Alan Aspuru-Guzik, and Alex Zhavoronkov. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models, October 2020. URL http://arxiv.org/abs/1811.12823. arXiv:1811.12823 [cs, stat].

Mariya Popova, Mykhailo Shvets, Junier Oliva, and Olexandr Isayev. MolecularRNN: Generating realistic molecular graphs with optimized properties. Technical Report arXiv:1905.13372, arXiv, May 2019. URL http://arxiv.org/abs/1905.13372. arXiv:1905.13372 [cs, q-bio, stat] type: article.

Kristina Preuer, Philipp Renz, Thomas Unterthiner, Sepp Hochreiter, and Günter Klambauer. Fréchet ChemNet Distance: A Metric for Generative Models for Molecules in Drug Discovery. *Journal of Chemical Information and Modeling*, 58(9):1736–1741, September 2018. ISSN 1549-9596. doi: 10.1021/acs.jcim.8b00234. URL https://doi.org/10.1021/acs.jcim.8b00234. Publisher: American Chemical Society.

Jean-Louis Reymond, Lars Ruddigkeit, Lorenz Blum, and Ruud van Deursen. The enumeration of chemical space. *WIREs Computational Molecular Science*, 2(5):717–733, 2012. ISSN 1759-0884. doi: 10.1002/wcms.1104. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1104. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1104.

Bidisha Samanta, Abir De, Gourhari Jana, Pratim Kumar Chattaraj, Niloy Ganguly, and Manuel Gomez-Rodriguez. NeVAE: A Deep Generative Model for Molecular Graphs, September 2019. URL http://arxiv.org/abs/1802.05283. arXiv:1802.05283 [physics, stat].

Chence Shi, Minkai Xu, Zhaocheng Zhu, Weinan Zhang, Ming Zhang, and Jian Tang. GraphAF: a Flow-based Autoregressive Model for Molecular Graph Generation. Technical Report arXiv:2001.09382, arXiv, February 2020. URL http://arxiv.org/abs/2001.09382. arXiv:2001.09382 [cs, stat] type: article.

Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification, May 2021. URL http://arxiv.org/abs/2009.03509. arXiv:2009.03509 [cs, stat].

Martin Simonovsky and Nikos Komodakis. GraphVAE: Towards Generation of Small Graphs Using Variational Autoencoders, February 2018. URL http://arxiv.org/abs/1802.03480. arXiv:1802.03480 [cs].

Nianzu Yang, Huaijin Wu, Junchi Yan, Xiaoyong Pan, Ye Yuan, and Le Song. Molecule Generation for Drug Design: a Graph Learning Perspective. Technical Report arXiv:2202.09212, arXiv, February 2022. URL http://arxiv.org/abs/2202.09212. arXiv:2202.09212 [cs] type: article.

Soojung Yang, Doyeong Hwang, Seul Lee, Seongok Ryu, and Sung Ju Hwang. Hit and Lead Discovery with Explorative RL and Fragment-based Molecule Generation. *Advances in Neural Information Processing Systems*, 34:7924–7936, 2021.

Jiaxuan You, Bowen Liu, Rex Ying, Vijay Pande, and Jure Leskovec. Graph Convolutional Policy Network for Goal-Directed Molecular Graph Generation. Technical Report arXiv:1806.02473, arXiv, February 2019. URL http://arxiv.org/abs/1806.02473. arXiv:1806.02473 [cs, stat] type: article.

Zhaoning Yu and Hongyang Gao. Molecular Representation Learning via Heterogeneous Motif Graph Neural Networks. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 25581–25594. PMLR, June 2022. URL https://proceedings.mlr.press/v162/yu22a.html. ISSN: 2640-3498.

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*, pp. 617–626, August 2020. doi: 10.1145/3394486.3403104. URL http://arxiv.org/abs/2006.10137. arXiv:2006.10137 [physics, stat].

Yanqiao Zhu, Yuanqi Du, Yinkai Wang, Yichen Xu, Jieyu Zhang, Qiang Liu, and Shu Wu. A Survey on Deep Graph Generation: Methods and Applications, August 2022. URL http://arxiv.org/abs/2203.06714. arXiv:2203.06714 [cs, q-bio].

A METHOD DETAILS

A.1 ENCODER GNN

Our encoder GNN leverages a transformer convolution backbone and computes the transformed node features at layer k as follows:

$$\mathbf{x}_i^{(k+1)} = \mathbf{W}_1 \mathbf{x}_i^{(k)} + \sum_{j \in \mathcal{N}(i)} \alpha_{i,j}^{(k)} \left(\mathbf{W}_2 \mathbf{x}_j^{(k)} + \mathbf{W}_6 \mathbf{e}_{ij} \right), \tag{4}$$

where $\mathbf{W}_n$ are learned matrices and the attention coefficients $\alpha_{i,j}^{(k)}$ are computed via multi-head dot product attention:

$$\alpha_{i,j}^{(k)} = \text{softmax} \left(\frac{(\mathbf{W}_3 \mathbf{x}_i^{(k)})^\top (\mathbf{W}_4 \mathbf{x}_j^{(k)} + \mathbf{W}_6 \mathbf{e}_{ij})}{\sqrt{h}} \right). \tag{5}$$

To attain graph features $g_i \in \mathbb{R}^{(H+1) \cdot d_{enc}}$ we aggregate the transformed node features by applying a sum aggregation and a learned weighted aggregation:

$$g_i = [\sum_{v \in V_i} \mathbf{x}_v^{(k)}; \sum_{v \in V_i} \text{MLP}_{\theta_1}(\mathbf{x}_1^{(k)}) \text{MLP}_{\sigma_1}(\mathbf{x}_1^{(k)}); \ldots; \sum_{v \in V_i} \text{MLP}_{\theta_H}(\mathbf{x}_v^{(k)}) \text{MLP}_{\sigma_H}(\mathbf{x}_v^{(k)})], \tag{6}$$

where MLP_{θ_h} and MLP_{σ_h} are two-layer MLPs of feature head h.

A.2 DECODER GNN

The decoder GNN is based on an MLP-based edgeweighter for which we employ residual connection between the layers:

$$\begin{aligned} \mathbf{m}_i^{(l+1)} &= \sum_{j \in \mathcal{N}(i)} \text{MLP}_{\phi_l^{edge}}(\mathbf{z}, \mathbf{h}_i^{(l)}) \cdot \mathbf{\Theta} \cdot \mathbf{h}_j^{(l)} \\ \mathbf{h}_i^{(l+1)} &= \text{MLP}_{\phi_l^{node}}(\mathbf{m}_i^{(l+1)}, \mathbf{h}_i^{(l)}), \end{aligned} \tag{7}$$

where $\mathbf{\Theta}$ is a weight matrix, and $\text{MLP}_{\phi_l^{edge}}$ and $\text{MLP}_{\phi_l^{node}}$ are a two- and three-layer MLP, respectively.

B MODEL HYPERPARAMETERS

Table 2 details all model and training hyperparameters used to obtain the results for our two-step VAE. These parameters are fixed for training across vocabulary sizes and types.

C RELATED WORK

Methods for molecule generation differ primarily in the chosen molecular representation. While first attempts characterise molecules through their SMILES or SELFIES strings Gómez-Bombarelli et al. (2018); Nigam et al. (2019); Kusner et al. (2017), focus soon shifted towards a 2D graph representation of molecules due to desirable properties such as permutation invariance. In a parallel line of work, 3D graph generation is concerned with predicting the geometry of a molecule from its 2D representation, called conformer. We refer the reader to (Du et al., 2022) for an overview of 1D, 2D, and 3D generation methods.

Yang et al. (2022) separate the 2D molecule generation literature into three subcategories: all-at-once, node-based and fragment-based. Many methods, both node-based (Khemchandani et al., 2020; Shi

Optimizer	Adam
Learning rate	0.000874
Learning rate decay	0.99
Batch size	64
Gradient clipping magnitude	3
Loss weights $\lambda_S, \lambda_M, \lambda_v$	1
β initalization	0
β maximum	0.2
β annealing start	0
β annealing frequency	100
β annealing step size	1.4347e-05
Valency penalty initalization	0
Valency penalty maximum	0.001
Valency penalty annealing start	0
Valency penalty annealing frequency	1000
Valency penalty annealing step size	0.00025
Encoder layers	4
Decoder layers	1
Latent representation size	128
Atom identity embedding size	50
Motif identity embedding size	75
Atom multiplicity embedding size	30
Motif multiplicity embedding size	30
Global graph feature size	115
Motif feature size	100
Edge decoder connection weight	0.3

Table 2: Hyperparameters for our two-step VAE.

et al., 2020; Popova et al., 2019; Mercado et al., 2021; Luo et al., 2021; Liu et al., 2018; Li et al., 2018; Assouel et al., 2018; Ahn et al., 2021) and fragment-based (Yu & Gao, 2022; You et al., 2019; Yang et al., 2021; Lim et al., 2020; Kajino, 2019; Jin et al., 2020; Bengio et al., 2021), rely on the autoregressive decoding of the molecular representation, either by attaching single atoms or larger fragments, such as motifs or scaffolds. In Table 3, we compare to MoLer, a molecular graph generation model trained to extend structural scaffolds using the BBB fragmentation scheme Maziarz et al. (2022). Additionally, we report results for Junction Tree Variational Autoencoder (JTVAE) (Jin et al., 2018), a model that autoregressively builds a junction tree of a molecule and extracts a molecular graph based purely on fragments.

Within the all-at-once category, Simonovsky & Komodakis (2018) and Ma et al. (2018) predict the graph adjacency, as well as the node and edge features in one step. Here, the number of considered nodes usually is fixed manually before inference. Liu et al. (2021) follow the same approach, but replace the VAE-based architecture with an energy-based learning technique. Similarly, De Cao & Kipf (2018) predict the entire molecular graph at once and train within a GAN framework including a permuation invariant discriminator. In contrast, Zang & Wang (2020) first generate a bond tensor through normalizing flows and subsequently assign node features to the graph whose computatiton is conditioned on the same bond tensor. Lastly, Bresson & Laurent (2019) as well as Flam-Shepherd et al. (2020) first predict the set of atoms present in the molecular graph, and in a second step compute the connection among these. Samanta et al. (2019) has a similar generation procedure, while additionally learning the 3D coordinates of the resulting molecular graph. Lastly, (Kong et al., 2022) first decode the set of nodes and fragments autoregressively via an RNN and then predict attachements of nodes and fragments.

	k	FCD	KL	Int. Div.	SA	QED
JTVAE	780	0.70 ± 0.01	0.95 ± 0.01	0.86 ± 0.01	0.89 ± 0.01	0.07 ± 0.01
Moler	32	0.71 ± 0.01	0.98 ± 0.01	0.87 ± 0.01	0.49 ± 0.03	0.09 ± 0.00
	64	0.74 ± 0.01	0.99 ± 0.01	0.87 ± 0.01	0.52 ± 0.03	0.09 ± 0.01
	128	0.76 ± 0.01	0.98 ± 0.01	0.87 ± 0.01	0.52 ± 0.01	0.09 ± 0.01
	256	0.79 ± 0.02	0.98 ± 0.01	0.87 ± 0.01	0.56 ± 0.03	0.10 ± 0.01
	512	0.82 ± 0.01	0.98 ± 0.01	0.87 ± 0.01	0.59 ± 0.04	0.09 ± 0.01
PS-VAE	32	0.09 ± 0.01	0.78 ± 0.01	0.90 ± 0.01	2.34 ± 0.00	0.16 ± 0.00
	64	0.14 ± 0.01	0.81 ± 0.01	0.90 ± 0.01	2.08 ± 0.02	0.16 ± 0.01
	128	0.18 ± 0.01	0.81 ± 0.01	0.89 ± 0.01	2.02 ± 0.00	0.15 ± 0.01
	256	0.23 ± 0.01	0.83 ± 0.01	0.89 ± 0.01	1.88 ± 0.01	0.14 ± 0.01
	512	0.30 ± 0.01	0.84 ± 0.01	0.89 ± 0.01	1.74 ± 0.00	0.13 ± 0.01
Ours	32	0.16 ± 0.01	0.73 ± 0.03	0.86 ± 0.01	0.93 ± 0.07	0.05 ± 0.02
	64	0.23 ± 0.01	0.79 ± 0.01	0.87 ± 0.01	0.81 ± 0.02	0.04 ± 0.01
	128	0.26 ± 0.01	0.82 ± 0.03	0.86 ± 0.01	0.69 ± 0.04	0.04 ± 0.01
	256	0.30 ± 0.01	0.82 ± 0.01	0.86 ± 0.01	0.52 ± 0.03	0.04 ± 0.01
	512	0.28 ± 0.03	0.83 ± 0.01	0.87 ± 0.01	0.48 ± 0.01	0.04 ± 0.01

Table 3: Results for learning distributions over molecular graphs across several graph generation models and increasing vocabulary size.

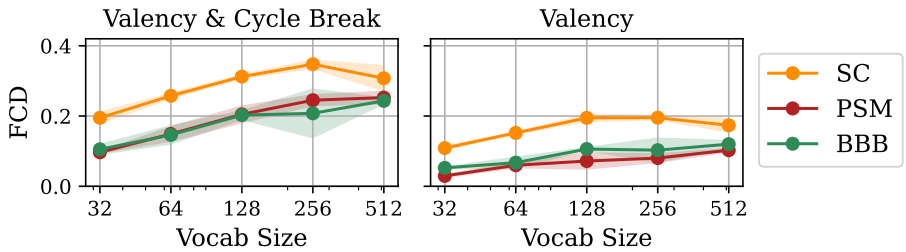

Figure 5: FCD achieved with different fragmentation schemes over increasing vocabulary size when sampling from a modified latent distribution.

D Baseline Results over varying vocabulary sizes

In addition to Table 1, we expand on these result by reporting the Fréchet Chemnet Distance, the Kullback-Leibler divergence, the Internal Diversity I, the synthetic accessability (SA) as well as the druglikeness (QED) across vocabulary sizes in Table 3. For calculation of these metrics we rely on the open-source implementations RDKit Landrum (2010), the GuacaMol benchmark (Brown et al., 2019) and the MOSES benchmark (Polykovskiy et al., 2020).

E Sampling from the Training Latent Distribution

While a Variational Autoencoder is trained to structure its latent space according to $\mathbf{z}_i \sim \mathcal{N}(0, 1)$, the latent space may be structured slightly differently even for a fully trained model. Specifically, a VAE faces a tradeoff between reconstruction performance on the training data and the KL Divergence between the approximate posterior and the prior distribution. This tradeoff is controlled via the β parameter during training. In a further evaluation, we modify the sampling procedure of our method slightly by first encoding $D = 10.000$ molecules, calculating their mean and standard deviation and then sampling new molecules according to:

$$\mathbf{z} \sim \big(\frac{1}{D}\sum_{j=0}^{D} \mu_\theta(G_j), \frac{1}{D}\sum_{j=0}^{D} \sigma^2_\theta(G_j)\big) \tag{8}$$

The results of this evaluation are shown in Figure 5. The findings reveal that with a modified distribution, PSM and BBB exhibit comparable FCD results. The results indicate that PSM does not gain significant improvements in FCD scores as a result of modifying the latent distribution compared to Figure 4. BBB on the other hand improves significantly over its performance on a standard Gaussian distribution, suggesting that the latent representations derived from BBB fragmentation make the concise structuring of the latent space more challenging. In this context, Subcover demonstrates FCD scores exceeding 0.35 with a moderately-sized vocabulary.