Table 1. Results on polymer generative modeling. The first row reports the oracle performance using real data as generated samples. The last row (small motif) is a variant of our model where we restrict the motif vocabulary to contain only single rings and bonds (similar to JT-VAE). "Recon." means reconstruction accuracy; "Div." means diversity; SNN means nearest neighbor similarity; "Frag / Scaf" means fragment and scaffold similarity. Except property statistics, all metrics are the higher the better.

Method	Reconstruction / Sample Quality (↑)				Property Statistics (↓)				Structural Statistics (↑)		
	Recon.	Valid	Unique	Div.	logP	SA	QED	MW	SNN	Frag.	Scaf.
Real data	-	100%	100%	0.823	0.094	6.7e-5	1.7e-5	82.3	0.706	0.995	0.462
SMILES	21.5%	93.1%	**97.3%**	0.821	1.471	0.011	**5.4e-4**	4963	0.704	0.981	0.385
CG-VAE	42.4%	**100%**	96.2%	**0.879**	3.958	2.600	0.0030	3944	0.204	0.372	0.001
JT-VAE	58.5%	**100%**	94.1%	0.864	2.645	0.157	0.0075	10867	0.522	0.925	0.297
HierVAE	**79.9%**	**100%**	97.0%	0.817	**0.525**	**0.007**	5.7e-4	**1928**	**0.708**	**0.984**	**0.390**
· Small motif	71.0%	**100%**	97.2%	0.835	0.872	0.042	0.0019	5320	0.575	0.949	0.191

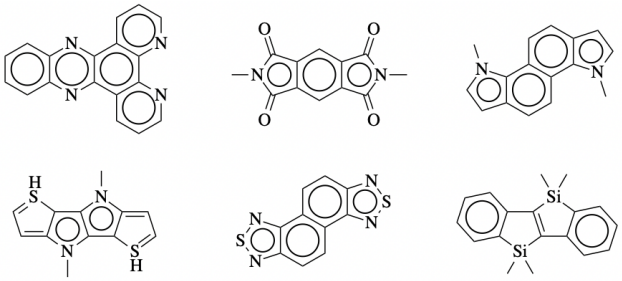

Figure 4. Examples of motif structures utilized by our model. These motifs consist of multiple rings and bonds, which are substantially more complex than previous methods (Jin et al., 2018).

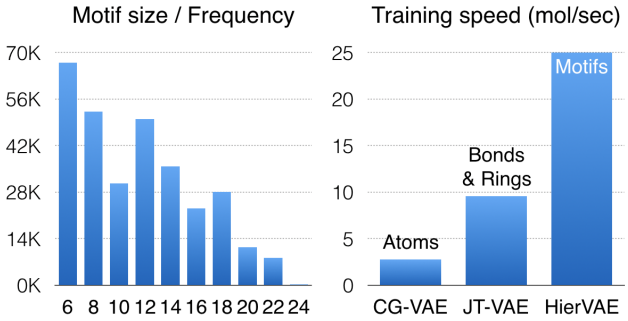

Figure 5. **Left**: Histogram of motif frequencies with respect to their sizes (i.e., number of atoms). **Right**: Training speed comparison between our method and baselines (on the same hardware).

- **Validity**: Percentage of chemically valid compounds.
- **Uniqueness**: Percentage of unique compounds.
- **Diversity**: We compute the pairwise molecular distance among generated compounds. The molecular distance $\text{dist}(X, Y)$ is defined as the Tanimoto distance over Morgan fingerprints (Rogers & Hahn, 2010) of two molecules.
- **Property statistics**: We compare property statistics between generated molecules and real data. Our properties include *partition coefficient* (logP), *synthetic accessibility* (SA), *drug-likeness* (QED) and *molecular weight* (MW). To quantitatively evaluate the distance between two distributions, we compute Frechet distance between property distributions of molecules in the generated and test sets (Polykovskiy et al., 2018).
- **Structural statistics**: We also compute structural statistics between generated molecules and real data. *Nearest neighbor similarity* (SNN) is the average similarity of generated molecules to the nearest molecule from the test set. *Fragment similarity* (Frag) and *scaffold similarity* (Scaf) are cosine distances between vectors of fragment or scaffold frequencies of the generated and the test set.

Baselines We compare our method against three state-of-the-art variational autoencoders for molecular graphs. SMILES VAE (Gómez-Bombarelli et al., 2018) is a sequence to sequence VAE that generates molecules based on their SMILES strings (Weininger, 1988). CG-VAE (Liu et al., 2018) is a graph-based VAE generating molecules atom by atom. JT-VAE (Jin et al., 2018) is also a graph-based VAE generating molecules based on simple substructures restricted to rings and bonds. Finally, we report the oracle performance of distributional statistics by using real molecules in the training set as our generated samples.

4.1.1. Results

The performance of different methods are summarized in Table 1, Our method (HierVAE) significantly outperforms all previous methods in terms of reconstruction accuracy (79.9% vs 58.5%). This validates the advantage of utilizing large structural motifs, which reduces the number of generation steps. In terms of distributional statistics, our method achieves state-of-the-art results on logP (0.525 vs 1.471), molecular weight Frechet distance (1928 vs 4863) and all the structural similarity metrics. Since our model requires fewer generation steps, our training speed is much faster than other graph-based methods (see Figure 5).

Table 2. Results on graph translation tasks from Jin et al. (2019). We report average improvement for continuous properties (logP), and success rate for binary properties (e.g., DRD2).

Method	logP (sim ≥ 0.6)		logP (sim ≥ 0.4)		Drug likeness		DRD2	
	Improvement	Diversity	Improvement	Diversity	Success	Diversity	Success	Diversity
JT-VAE	0.28 ± 0.79	-	1.03 ± 1.39	-	8.8%	-	3.4%	-
CG-VAE	0.25 ± 0.74	-	0.61 ± 1.09	-	4.8%	-	2.3%	-
GCPN	0.79 ± 0.63	-	2.49 ± 1.30	-	9.4%	0.216	4.4%	0.152
MMPA	1.65 ± 1.44	0.329	3.29 ± 1.12	0.496	32.9%	0.236	46.4%	**0.275**
Seq2Seq	2.33 ± 1.17	0.331	3.37 ± 1.75	0.471	58.5%	0.331	75.9%	0.176
JTNN	2.33 ± 1.24	0.333	3.55 ± 1.67	0.480	59.9%	0.373	77.8%	0.156
AtomG2G	2.41 ± 1.19	0.379	$\mathbf{3.98 \pm 1.54}$	0.563	73.6%	0.421	75.8%	0.128
HierG2G	$\mathbf{2.49 \pm 1.09}$	**0.381**	$\mathbf{3.98 \pm 1.46}$	**0.564**	**76.9%**	**0.477**	**85.9%**	0.192

Table 3. Ablation study: the importance of hierarchical graph encoding, LSTM MPN architecture and structure-based decoding.

Method	QED	DRD2
HierG2G	**76.9%**	**85.9%**
· atom-based decoder	76.1%	75.0%
· two-layer encoder	75.8%	83.5%
· one-layer encoder	67.8%	74.1%

and 2.4%. When we further remove the attachment layer (*one-layer encoder*), the performance degrades significantly on both datasets. This is because all the motif information is lost and the model needs to infer what motifs are and how motif layers are constructed for each molecule. This shows the importance of the hierarchical representation.

5. Related Work

Graph Generation Previous work have adopted various approaches for generating molecular graphs. Gómez-Bombarelli et al. (2018); Segler et al. (2017); Kusner et al. (2017); Dai et al. (2018); Guimaraes et al. (2017); Olivecrona et al. (2017); Popova et al. (2018); Kang & Cho (2018) generated molecules based on their SMILES strings (Weininger, 1988). Simonovsky & Komodakis (2018); De Cao & Kipf (2018); Ma et al. (2018) developed generative models which output the adjacency matrices and node labels of the graphs at once. You et al. (2018b); Li et al. (2018); Samanta et al. (2018); Liu et al. (2018); Zhou et al. (2018) proposed generative models which decode molecules sequentially node by node. Seff et al. (2019) developed a edit-based model which generates molecules based on insertions and deletions.

Our model is closely related to Liao et al. (2019) which generate graphs one block of nodes and edges at a time. While their encoder operates on original graphs, our encoder operates on multiple hierarchies and learns multi-resolution representations of input graphs. Our work is also closely related to Jin et al. (2018; 2019) that generate molecules based on substructures. Their decoder first generates a junction tree with substructures as nodes, and then predicts how the substructures should be attached to each other. Their substructure attachment process involves combinatorial enumeration and therefore their model cannot scale to substructures more complex than simple rings and bonds. In contrast, our model allows the motif to have flexible structures.

Graph Encoders Graph neural networks have been extensively studied for graph encoding (Scarselli et al., 2009; Bruna et al., 2013; Li et al., 2015; Niepert et al., 2016; Kipf & Welling, 2017; Hamilton et al., 2017; Lei et al., 2017; Velickovic et al., 2017; Xu et al., 2018). Our method is related to graph encoders for molecules (Duvenaud et al., 2015; Kearnes et al., 2016; Dai et al., 2016; Gilmer et al., 2017; Schütt et al., 2017). Different to these approaches, our method represents molecules as hierarchical graphs spanning from atom-level to motif-level graphs.

Our work is most closely related to (Defferrard et al., 2016; Ying et al., 2018; Gao & Ji, 2019) that learn to represent graphs in a hierarchical manner. In particular, Defferrard et al. (2016) utilized graph coarsening algorithms to construct multiple layers of graph hierarchy and Ying et al. (2018); Gao & Ji (2019) proposed to learn the graph hierarchy jointly with the encoding process. Despite some differences, all of these methods learns the hierarchy for regression or classification tasks. In contrast, our hierarchy is constructed for efficient graph generation.

6. Conclusion

In this paper, we developed a hierarchical encoder-decoder architecture generating molecular graphs using structural motifs as building blocks. The experimental results show our model outperforms prior atom and substructure based methods in both small molecule and polymer domains.