

Learning to Extend Molecular Scaffolds with Structural Motifs

Presentation for ICLR 2022

Krzysztof Maziarz¹, Henry Jackson-Flux¹, Pashmina Cameron¹, Finton Sirockin²,
Nadine Schneider², Nikolaus Stiefl², Marwin Segler¹, Marc Brockschmidt¹

Collaboration between ¹Microsoft Research and ²Novartis

Research Questions

Observation 1: in practice, molecule generation constrained by presence of a given **scaffold** is required

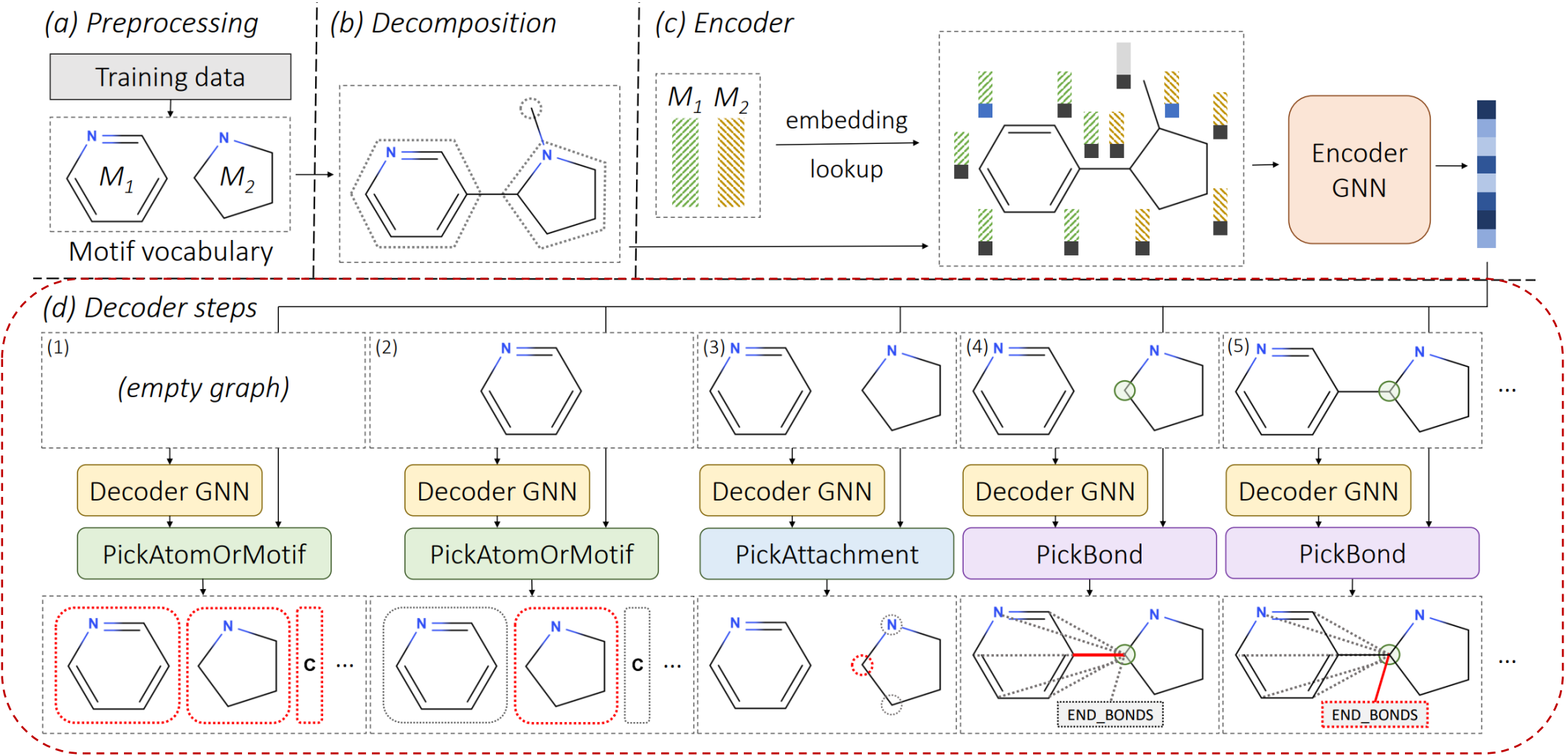
Q1: Can one model succeed at both scaffold-based and unconstrained generation?

Observation 2: **motifs** (common molecule fragments) crucial for success of recent graph-based models (JT-VAE, HierVAE), but limit expressivity

Q2: Can we get the benefits of fragment-based generation without its restrictions?

Our new model **MoLeR** answers “yes” to both.

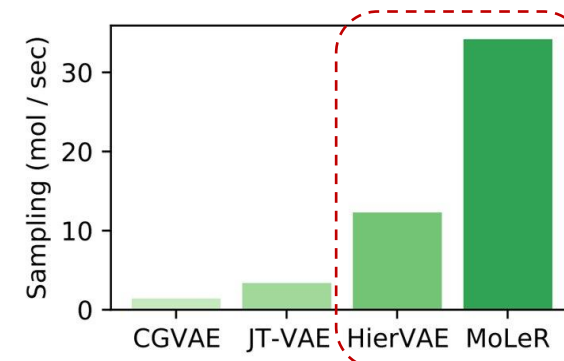
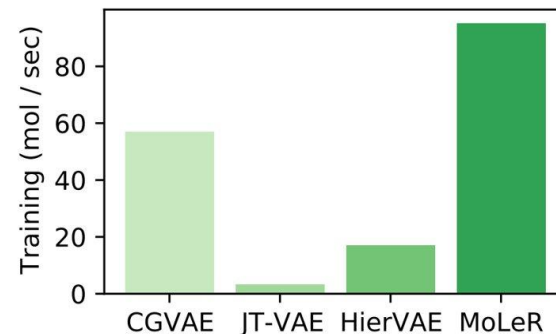
MoLeR Model: Overview



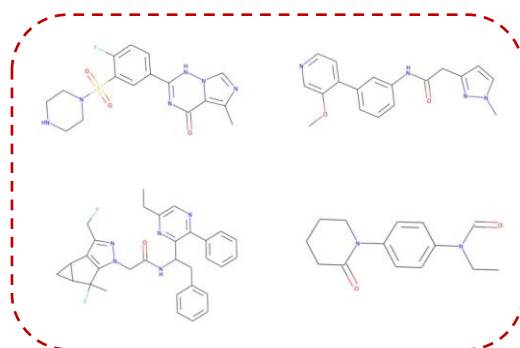
decoder steps independent given encoder output → can start from scaffold

Results

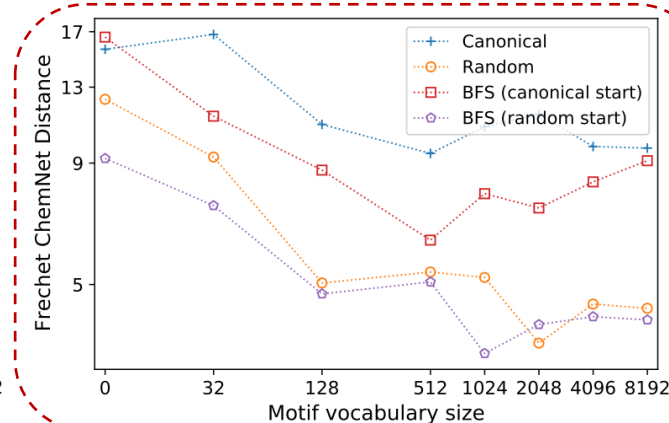
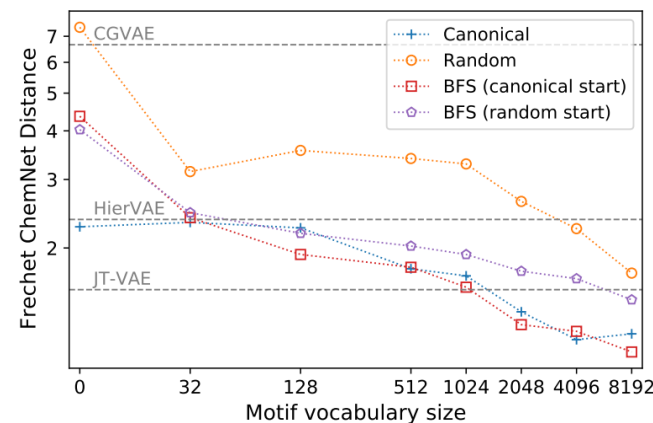
Method	GuacaMol		Scaffolds	
	Score	Quality	Score	Quality
Best of dataset	0.61	0.77	0.17	-
SMILES LSTM	0.87	0.77	0.45	-
SMILES GA	0.72	0.36	0.45	-
GRAPH MCTS	0.45	0.22	0.20	-
GRAPH GA	0.90	0.40	0.79	-
CDDD + MSO	0.90	0.58	0.92	0.59
MNCE-RL	0.92	0.54	0.95	0.47
MoLeR + MSO	0.82	0.75	0.93	0.63



sampling 2.7x faster than HierVAE



*new benchmarks
(complex scaffolds)*



randomized generation order crucial for working with scaffolds



Thank you for listening!

Poster 7209

arXiv: arxiv.org/abs/2103.03864

Code: github.com/microsoft/molecule-generation



Background: scaffold-constrained latent neighborhood learned by MoLeR