Deterministic global optimization for sample-efficient molecular design with generative machine learning

Jan G. Rittig[1,*], **Malte Franke**[1,*], **Alexander Mitsos**[3,1,2]
[1]RWTH Aachen University
[2]Forschungszentrum Jülich GmbH
[3]JARA Center for Simulation and Data Science (CSD)
{jan.rittig, malte.franke}@rwth-aachen.de, amitsos@alum.mit.edu
∗Contributed equally

Abstract

Generative machine learning (ML) models such as variational autoencoders (VAEs) learn continuous molecular latent spaces that can facilitate the exploration of novel molecules and materials. However, such latent spaces are typically high-dimensional, making targeted molecular optimization challenging. We therefore propose deterministic global optimization of molecular property prediction models in the form of artificial neural networks (ANNs) trained on VAEs' latent spaces. By using ANNs with ReLU activations, we formulate molecular design as a mixed-integer linear program (MILP) guaranteeing optimal molecular properties, as predicted by the ANN. Our results show superiority of the identified molecules with global optimal predicted property values compared to those found with frequently-used optimization strategies such as Bayesian optimization. Our approach thus enables finding the most promising molecules/materials according to the ANN predictions for subsequent investigation in simulations/experiments, thereby increasing the sample efficiency of ML-guided molecular design.

1 Introduction

Machine learning (ML) has shown great potential to accelerate the design of molecules and materials with desired properties in many applications such as drugs [1, 2, 3], fuels [4, 5], and catalysts [6, 7]. Various generative ML approaches have been utilized for molecular design including reinforcement learning (RL), generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion- and flow-based models, cf. overviews in [8, 9, 10]. As part of this design, predictive ML approaches such as graph neural networks (GNNs) and transformers have been employed to predict properties of molecules and materials [11, 12]. These approaches facilitate the exploration of the vast molecular space for novel molecules and materials, which can then be further investigated in so-called oracle calls, i.e., accurate computational simulations or laboratory experiments, both resource-intensive and time-consuming.

A major goal of ML-guided molecular design is to identify the most promising molecules and materials with a low number of oracle calls, i.e., a high sample efficiency. However, as of today, many generative ML approaches used for molecular design exceed realistic practical limits for the number of oracle calls [13, 14]. Gao et al. [13] have recently shown that even a limit of 10,000 oracle calls is not sufficient for many proposed generative ML approaches to identify suitable molecules in certain design tasks. In many real-world cases, especially for costly and labor-intensive experiments, the limit of oracle calls may be orders of magnitude lower, hence increasing sample efficiency of generative ML models is highly relevant and actively investigated, e.g., in [15, 16, 17].

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

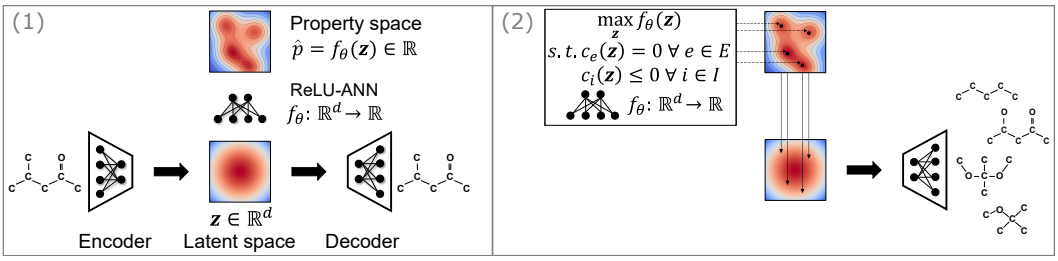

Figure 1: Deterministic global optimization for molecular design: (1) Jointly-trained VAE for molecular generation and ANN with ReLU activation for property prediction. (2) Optimization of property ANN on VAE latent space and decoding of optimal latent vectors to molecules.

We hypothesize that deterministic global optimization (DGO) in combination with ML is highly promising to increase the sample efficiency in molecular design: given an ML model, DGO can guarantee finding molecules with optimal predicted properties, thereby focusing further investigations on the most promising candidates. For instance, McDonald et al. [18] and Zhang et al. [19, 20] recently proposed a molecular design problem formulation with trained GNNs embedded as constraints, allowing to identify globally optimal molecules as a function of GNNs. However, this approach relies on expert-designed constraints to ensure chemical validity of the designed molecules and incorporates (nonlinear) graph convolutions in the optimization formulation, resulting in extensive formulation efforts and computational costs. This limits its practical application to the design of small molecules with a low number of heavy atoms or fragments, cf. [18, 19].

We propose to rather use VAEs for molecular generation in combination with DGO of artificial neural networks (ANNs) for property prediction (1) trained on the VAE's latent spaces and (2) embedded into a molecular design formulation, see Figure 1, also cf. Wang et al. [21]. In this way, DGO finds optimal predicted properties of the ANN, and the VAE can be used subsequently to decode molecules from the corresponding latent vectors. Thus, in contrast to DGO with GNNs embedded, we circumvent the need to define chemistry rules for designing molecules and to include graph convolutions in the optimization formulation. In fact, by using ReLU activations within the ANN, we can formulate molecular design as a MILP.

Our **main contribution** is the use of mixed-integer linear programming to guarantee the discovery of molecules with optimal properties according to ANN prediction models, thereby increasing sample efficiency in ML-guided molecular design.

2 Methods

2.1 VAEs for molecular design

VAEs have been extensively developed and applied to molecular design in recent years, e.g., [22, 23, 24, 25]. Utilizing their encoder-decoder architecture, VAEs learn a latent representation of molecules from which the molecules can be reconstructed. The encoder maps molecules from their input representation, e.g., in the form of text or graphs, to a continuous vector representation, referred to as latent space vectors $\mathbf{z} \in \mathbb{R}^d$ of dimensionality d, which typically follow a multivariate Gaussian distribution $\mathcal{N}(\mu, \Sigma)$. The decoder then reconstructs the molecules from the latent space. By sampling new points from the latent space and decoding them, molecular structures can be generated, enabling exploration of the molecular space. Here, multiple challenges arise, such as the chemical validity and out-of-distribution sampling of molecules, which has led to various adaptions of VAEs for molecular design, e.g., [23, 26, 27, 28]

VAEs can also be utilized for molecular property prediction [8]. Specifically, the latent space can serve as input domain to train molecular property prediction models, e.g., ANNs. Here, an ANN with parameters Θ learns to predict a molecular property p from the VAE's latent space: $f_\Theta(\mathbf{z}) = \hat{p}$. The property prediction models can either be trained after or simultaneously to the VAE training; also known as conditional VAEs. In molecular design, joint training has shown superior performance to subsequent ANN training, leading to higher accuracy and a latent space ordered by property [22, 29].

Optimization guides the sampling process of molecules from the VAEs' latent spaces. Frequently, derivative-free black-box global optimization approaches like Bayesian optimization, particle-swarm algorithms, or genetic algorithms are employed to search for promising molecules in the latent space, cf. [5, 8, 10, 30]. The properties of the identified molecules can then be evaluated with the oracle and/or by using a property prediction model. Since the latent space typically is high-dimensional and the mapping from latent space to property can be highly nonlinear, a high number of oracle/prediction model calls are expected until near-optimal points are identified. Further, gradient-based optimization approaches can be used to sample from the latent space [31]; however, they may get stuck in local optima. Therefore, we focus on DGO based on property prediction models in the form of ANNs. That is, DGO based on ANNs for property prediction that are trained on VAE latent spaces allows to identify global optimal properties and corresponding latent vectors, as recently discussed by Wang et al. [21].

2.2 Deterministic global optimization of ANNs

ANNs can be embedded into optimization problem formulations, which can be solved to global optimality by using DGO [32]. This allows finding the inputs to the ANN that correspond to an optimal predicted target value. In contrast to aforementioned derivative-free or gradient-descent methods, DGO with an embedded ANN results in provable optimality, within a given tolerance. Due to the nonlinear activation functions, problem formulations with embedded ANNs generally result in nonlinear programs, and solving them to global optimality with dedicated techniques, cf. [32], can be computationally expensive. By using piecewise linear ReLU activation functions, the optimization problem can be formulated as MILP [33], which can facilitate solving.

2.3 Deterministic global optimization of ReLU-ANNs with VAEs for molecule generation

We propose to use DGO of property ReLU-ANNs in combination with VAEs for molecular design, cf. Figure 1. That is, we jointly train a VAE on molecular generation and an ANN with ReLU activation for molecular property prediction. After training, we embed the ReLU-ANN into a MILP to find optimal predicted property values with respect to the latent space: $\max_{\mathbf{z}} f_\Theta(\mathbf{z})$. Our formulation of molecular design as a MILP, instead of a nonlinear program used by Wang et al. [21], therefore avoids nonlinearities. Solving this MILP yields the best property value $\hat{p}^*$ as well as a corresponding latent point $\mathbf{z}^*$. The decoder then reconstructs a molecule from this latent point. We thereby guarantee to find the molecule with the best predicted property value within a given optimality tolerance.

3 Molecular design case studies

We consider two molecular design tasks. First, we aim at maximizing the **molecular weight** as an illustrative example with a simple oracle. For model training and testing, we use the ZINC dataset of 250 k small organic molecules [34, 35]. Second, based on the Tartarus benchmark [14], we consider minimizing the **singlet-triplet gap**, which can improve emitter efficiency [36]. Thus, we evaluate our approach in a practical setting with quantum mechanics-based oracles. Here, we focus on the emitter design task using a filtered portion of the GDB13 database, containing about 380 k organic molecules with up to 13 heavy atoms [37]. For details on the oracle and dataset, we refer to [14].

Models: We consider two alternative VAEs for molecule generation, namely the string-based SMILES-VAE [22, 38] and the graph-based Junction Tree VAE [23]. The VAEs are trained jointly with ReLU-ANNs for predicting molecular properties. For the training, validation, and testing of the VAEs and ANNs, we split the respective datasets randomly into 90%/5%/5%.

Optimization strategies: We benchmark our DGO approach against the following strategies for sampling molecules from VAEs' latent spaces: Random search, BO based on [39], and GAs using the package by [40]. For DGO with embedded ReLU-ANNs, we use the tool from [41] with Gurobi [42].

Optimization setup: We set the bounds of the search space to the two-sigma (Σ) region of the latent space, i.e., the VAE should be able to capture about 95% of the training distribution. We run random search, BO, and GA until 1000 molecules are identified and repeat each run with 5 different seeds. We run DGO with an optimality tolerance of 10^{-6}, using the *PoolSearchMode* functionality of Gurobi (cf. [42]) to systematically search for the n best solutions. We choose $n = 1000$ to ensure comparability with the other optimization strategies. In practical applications, we would likely use a

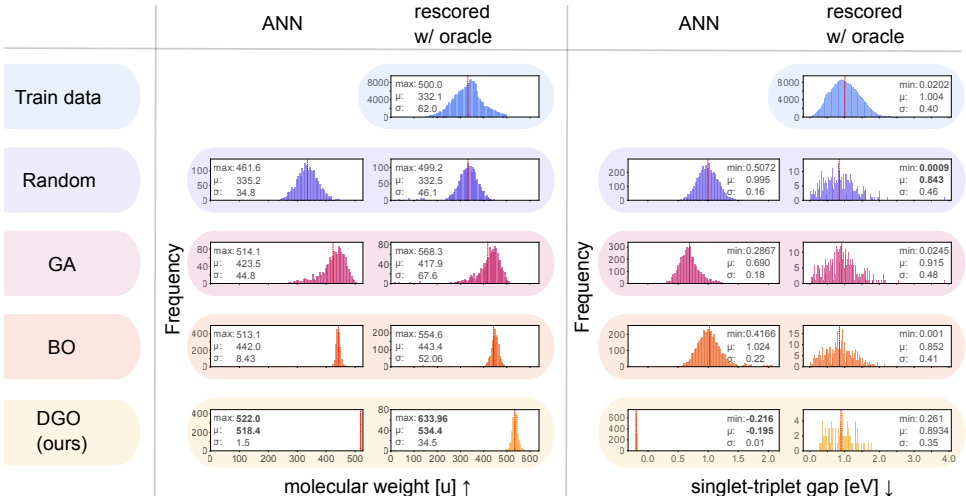

Figure 2: Molecular design results for different sampling/optimization strategies using SMILES-VAE.

much looser tolerance and only one or a few solutions, i.e., a smaller value for n. An overview of additional hyperparameters and settings is provided in the Appendix.

4 Results

The results of the two molecular design case studies using the SMILES-VAE with the different optimization strategies are shown in Figure 2; the results for the Junction Tree VAE are in the Appendix. We report the distribution of property values predicted by the ANN for the identified molecules. The ANNs exhibit a coefficient of determination (R^2) on the test set of 0.99 and 0.42 for the molecular weight and singlet-triplet gap, respectively. The significantly lower accuracy highlights the difficulty in predicting the singlet-triplet gap. Future work should address improving the accuracy, e.g., by adapting the VAE architecture and training to learn a molecular representation that facilitates the prediction step, which is out of the scope of our work since we focus on optimization. Further, we report the rescored property values that are obtained by applying the oracles to the identified molecules. For the singlet-triplet, we rescore only the best 100 molecules due to the high oracle cost.

Frequently employed optimization strategies can fail to find optimal molecules. Considering the property values predicted by the ANN, BO and GAs find better molecules than random search in both case studies, as expected. The global optima found by DGO exceed the best points identified by BO and GAs. In case of limited oracle calls, DGO is in fact the only method that guarantees finding the most promising molecules according to the property ANN. This guarantee typically entails higher computational cost than the other strategies; yet, we found DGO solution times of less than 24 hours. Future work could investigate the balance between proving optimality, run time, and solution quality.

DGO exploits weak spots of inaccurate property ANNs. For the singlet-triplet gap minimization, the DGO reveals that the lowest value of the ANN prediction is -0.216 eV. Although negative singlet-triplet gaps have recently been reported, e.g., in [43], such low values are highly unlikely. Hence, we rather attribute this prediction to the inaccuracy of the ANN – which is confirmed in the subsequent oracle evaluation. By the formulation of the molecular optimization problem, DGO exploits such weak spots, i.e., regions with high predictive inaccuracies and/or uncertainties, which can correspond to molecules with physically unreasonable or overly optimistic predictions. This can be accounted for by adapting the problem formulation and refining the ANN, e.g, by considering uncertainty quantification of molecular property predictions [44, 45, 46].

DGO can increase sample efficiency in molecular design. When applying the oracle functions to rescore the best molecules based on the ANN, we find that, for the molecular weight, DGO outperforms the other methods. DGO yields significantly higher values for both the molecule with the highest molecular weight and the mean molecular weight of the identified molecules beyond the training data, therefore increasing sample efficiency.

Inaccurate property ANNs limit improvements on sample efficiency. For the singlet-triplet gap, we find that the distributions of the rescored property values do not match those of the property ANN. In fact, one molecule identified in random search has the lowest singlet-triplet gap of 0.0009 eV according to the oracle, followed by 0.001 eV of a molecule found with BO, cf. Appendix. Interestingly both molecules are superior to the best candidates found in the Tartarus benchmark with 0.008 eV [14]. Yet, the results reveal that optimization based on less accurate ANNs limits targeted ML-guided molecular design. Future work could constrict the latent space to areas of high certainty, cf. [47].

5 Conclusion

We introduced a new optimization strategy for ML-guided molecular design: DGO of ANNs trained on VAE latent spaces for molecular property prediction. Our approach guarantees finding the best molecules with respect to the predicted properties by the ANNs, surpassing established methods such as BO and GAs. We find that candidates identified with DGO can also transfer to the most promising molecules according to the oracle function, e.g., simulations or experiments, thereby increasing sample efficiency in molecular design. The transferability depends on the predictive accuracy of the ANN, as in some cases DGO uncovers the ANN's weak spots. Overall, DGO of ANNs in combination with VAEs is highly promising for molecular design and can also be transferred to other applications.

Acknowledgments and Disclosure of Funding

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 466417970 – within the Priority Programme "SPP 2331: Machine Learning in Chemical Engineering". This work was also performed as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE). Simulations were performed with computing resources granted by RWTH Aachen University under project "rwth1232". The authors thank the anonymous reviewers for their helpful comments and feedback.

References

[1] Xiaochu Tong, Xiaohong Liu, Xiaoqin Tan, Xutong Li, Jiaxin Jiang, Zhaoping Xiong, Tingyang Xu, Hualiang Jiang, Nan Qiao, and Mingyue Zheng. Generative models for de novo drug design. *Journal of Medicinal Chemistry*, 64(19):14011–14027, 2021.

[2] Tiago Sousa, João Correia, Vítor Pereira, and Miguel Rocha. Generative deep learning for targeted compound design. *Journal of Chemical Information and Modeling*, 61(11):5343–5361, 2021.

[3] Yuanqi Du, Arian R Jamasb, Jeff Guo, Tianfan Fu, Charles Harris, Yingheng Wang, Chenru Duan, Pietro Liò, Philippe Schwaller, and Tom L Blundell. Machine learning-aided generative molecular design. *Nature Machine Intelligence*, pages 1–16, 2024.

[4] Nursulu Kuzhagaliyeva, Samuel Horváth, John Williams, Andre Nicolle, and S Mani Sarathy. Artificial intelligence-driven design of fuel mixtures. *Communications Chemistry*, 5(1):111, 2022.

[5] Jan G Rittig, Martin Ritzert, Artur M Schweidtmann, Stefanie Winkler, Jana M Weber, Philipp Morsch, Karl Alexander Heufer, Martin Grohe, Alexander Mitsos, and Manuel Dahmen. Graph machine learning for design of high-octane fuels. *AIChE Journal*, 69(4):e17971, 2023.

[6] Gabriel dos Passos Gomes, Robert Pollice, and Alán Aspuru-Guzik. Navigating through the maze of homogeneous catalyst design with machine learning. *Trends in Chemistry*, 3(2):96–110, 2021.

[7] Julia Westermayr, Joe Gilkes, Rhyan Barrett, and Reinhard J Maurer. High-throughput property-driven generative design of functional organic molecules. *Nature Computational Science*, 3(2):139–148, 2023.

[8] Benjamin Sanchez-Lengeling and Alán Aspuru-Guzik. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365, 2018.

[9] Camille Bilodeau, Wengong Jin, Tommi Jaakkola, Regina Barzilay, and Klavs F Jensen. Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(5):e1608, 2022.

[10] Dylan M Anstine and Olexandr Isayev. Generative models as an emerging paradigm in the chemical sciences. *Journal of the American Chemical Society*, 145(16):8736–8750, 2023.

[11] Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93, 2022.

[12] Afnan Sultan, Jochen Sieg, Miriam Mathea, and Andrea Volkamer. Transformers for molecular property prediction: Lessons learned from the past five years. *Journal of Chemical Information and Modeling*, 64(16):6259–6280, 2024.

[13] Wenhao Gao, Tianfan Fu, Jimeng Sun, and Connor Coley. Sample efficiency matters: a benchmark for practical molecular optimization. *Advances in Neural Information Processing Systems, Track Datasets and Benchmarks*, 35:21342–21357, 2022.

[14] AkshatKumar Nigam, Robert Pollice, Gary Tom, Kjell Jorner, John Willes, Luca Thiede, Anshul Kundaje, and Alán Aspuru-Guzik. Tartarus: A benchmarking platform for realistic and practical inverse molecular design. *Advances in Neural Information Processing Systems*, 36:3263–3306, 2023.

[15] Jeff Guo and Philippe Schwaller. Augmented memory: Capitalizing on experience replay to accelerate de novo molecular design. *arXiv preprint arXiv:2305.16160*, 2023.

[16] Jeff Guo and Philippe Schwaller. Beam enumeration: probabilistic explainability for sample efficient self-conditioned molecular design. *arXiv preprint arXiv:2309.13957*, 2023.

[17] Michael Dodds, Jeff Guo, Thomas Löhr, Alessandro Tibo, Ola Engkvist, and Jon Paul Janet. Sample efficient reinforcement learning with active learning for molecular design. *Chemical Science*, 15(11):4146–4160, 2024.

[18] Tom McDonald, Calvin Tsay, Artur M Schweidtmann, and Neil Yorke-Smith. Mixed-integer optimisation of graph neural networks for computer-aided molecular design. *Computers & Chemical Engineering*, 185:108660, 2024.

[19] Shiqiang Zhang, Juan Campos, Christian Feldmann, David Walz, Frederik Sandfort, Miriam Mathea, Calvin Tsay, and Ruth Misener. Optimizing over trained GNNs via symmetry breaking. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 44898–44924, 2023.

[20] Shiqiang Zhang, Juan S. Campos, Christian Feldmann, Frederik Sandfort, Miriam Mathea, and Ruth Misener. Augmenting optimization-based molecular design with graph neural networks. *Computers & Chemical Engineering*, 186:108684, 2024.

[21] Zihao Wang, Teng Zhou, and Kai Sundmacher. A novel machine learning-based optimization approach for the molecular design of solvents. In *Computer Aided Chemical Engineering*, volume 51, pages 1477–1482. Elsevier, 2022.

[22] 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.

[23] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Junction tree variational autoencoder for molecular graph generation. In *International Conference on Machine Learning*, pages 2323–2332. PMLR, 2018.

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

[25] Wengong Jin, Regina Barzilay, and Tommi Jaakkola. Hierarchical generation of molecular graphs using structural motifs. In *International Conference on Machine Learning*, pages 4839–4848. PMLR, 2020.

[26] Jaechang Lim, Sang-Yeon Hwang, Seokhyun Moon, Seungsu Kim, and Woo Youn Kim. Scaffold-based molecular design with a graph generative model. *Chemical science*, 11(4):1153–1164, 2020.

[27] Orion Dollar, Nisarg Joshi, David AC Beck, and Jim Pfaendtner. Attention-based generative models for de novo molecular design. *Chemical Science*, 12(24):8362–8372, 2021.

[28] Ryan J Richards and Austen M Groener. Conditional β-VAE for de novo molecular generation. *arXiv preprint arXiv:2205.01592*, 2022.

[29] Oliver Schilter, Alain Vaucher, Philippe Schwaller, and Teodoro Laino. Designing catalysts with deep generative models and computational data. a case study for suzuki cross coupling reactions. *Digital Discovery*, 2(3):728–735, 2023.

[30] Robin Winter, Floriane Montanari, Andreas Steffen, Hans Briem, Frank Noé, and Djork-Arné Clevert. Efficient multi-objective molecular optimization in a continuous latent space. *Chemical Science*, 10(34):8016–8024, 2019.

[31] Qi Liu, Miltiadis Allamanis, Marc Brockschmidt, and Alexander Gaunt. Constrained graph variational autoencoders for molecule design. *Advances in Neural Information Processing Systems*, 31, 2018.

[32] Artur M Schweidtmann and Alexander Mitsos. Deterministic global optimization with artificial neural networks embedded. *Journal of Optimization Theory and Applications*, 180(3):925–948, 2019.

[33] Bjarne Grimstad and Henrik Andersson. ReLU networks as surrogate models in mixed-integer linear programs. *Computers & Chemical Engineering*, 131:106580, 2019.

[34] Teague Sterling and John J Irwin. Zinc 15–ligand discovery for everyone. *Journal of Chemical Information and Modeling*, 55(11):2324–2337, 2015.

[35] Matt J Kusner, Brooks Paige, and José Miguel Hernández-Lobato. Grammar variational autoencoder. In *International Conference on Machine Learning*, pages 1945–1954. PMLR, 2017.

[36] Rafael Gómez-Bombarelli, Jorge Aguilera-Iparraguirre, Timothy D Hirzel, David Duvenaud, Dougal Maclaurin, Martin A Blood-Forsythe, Hyun Sik Chae, Markus Einzinger, Dong-Gwang Ha, Tony Wu, et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nature Materials*, 15(10):1120–1127, 2016.

[37] Lorenz C Blum and Jean-Louis Reymond. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *Journal of the American Chemical Society*, 131(25):8732–8733, 2009.

[38] Daniil Polykovskiy, Alexander Zhebrak, Benjamin Sanchez-Lengeling, Sergey Golovanov, Oktai Tatanov, Stanislav Belyaev, Rauf Kurbanov, Aleksey Artamonov, Vladimir Aladinskiy, Mark Veselov, et al. Molecular sets (MOSES): a benchmarking platform for molecular generation models. *Frontiers in Pharmacology*, 11:565644, 2020.

[39] The GPyOpt authors. GPyOpt: A Bayesian optimization framework in python. `http://github.com/SheffieldML/GPyOpt`, 2016. Accessed on 01.08.2024.

[40] Ryan Mohammad Solgi. geneticalgorithm (python package). `https://github.com/rmsolgi/geneticalgorithm`, 2020. Accessed on 01.08.2024.

[41] Laurens Lueg, Bjarne Grimstad, Alexander Mitsos, and Artur M. Schweidtmann. reluMIP: Open source tool for MILP optimization of ReLU neural networks. `https://github.com/ChemEngAI/ReLU_ANN_MILP`, 2021. Accessed on 01.08.2024.

[42] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2024.

[43] Naoya Aizawa, Yong-Jin Pu, Yu Harabuchi, Atsuko Nihonyanagi, Ryotaro Ibuka, Hiroyuki Inuzuka, Barun Dhara, Yuki Koyama, Ken-ichi Nakayama, Satoshi Maeda, et al. Delayed fluorescence from inverted singlet and triplet excited states. *Nature*, 609(7927):502–506, 2022.

[44] Lior Hirschfeld, Kyle Swanson, Kevin Yang, Regina Barzilay, and Connor W. Coley. Uncertainty quantification using neural networks for molecular property prediction. *Journal of Chemical Information and Modeling*, 60(8):3770–3780, 2020.

[45] AkshatKumar Nigam, Robert Pollice, Matthew F. D. Hurley, Riley J. Hickman, Matteo Aldeghi, Naruki Yoshikawa, Seyone Chithrananda, Vincent A. Voelz, and Alán Aspuru-Guzik. Assigning confidence to molecular property prediction. *Expert Opinion on Drug Discovery*, 16(9):1009–1023, 2021.

[46] Tom Wollschläger, Nicholas Gao, Bertrand Charpentier, Mohamed Amine Ketata, and Stephan Günnemann. Uncertainty estimation for molecules: Desiderata and methods. In *Proceedings of the 40th International Conference on Machine Learning*, pages 37133–37156, 2023.

[47] Pascal Notin, José Miguel Hernández-Lobato, and Yarin Gal. Improving black-box optimization in VAE latent space using decoder uncertainty. *Advances in Neural Information Processing Systems*, 34:802–814, 2021.

A VAE and optimization hyperparameter

Parameter	Value
q_cell	gru
q_bidir	False
q_d_h	1024
q_n_layers	3
q_dropout	0.1
d_cell	gru
d_n_layers	3
d_dropout	0.1
d_z	32
d_d_h	512
freeze_embeddings	False
emb_size	128
n_batch	512
clip_grad	50
kl_start	0
kl_w_start	0
kl_w_end	0.001
lr_start	1e-4
lr_n_period	30
lr_n_restarts	1
lr_n_mult	1
lr_end	1e-4
n_last	1000
ANN_dropout	0.1
ANN_n_layers	3
ANN_hidden_dim	64

Table 1: Model hyperparameters for SMILES-VAE training

Parameter	Value
model_type	GP
kernel	Matern52
acquisition_type	expected improvement
acquisition_optimizer_type	lbfgs
normalize_Y	True
evaluator_type	Thompson sampling
batch_size	10

Table 2: Hyperparameters for Bayesian optimization

Parameter	Value
population_size	50
mutation_probability	0.1
elit_ratio	0.01
crossover_probability	0.5
parents_portion	0.3
crossover_type	uniform
max_iteration_without_improv	None

Table 3: Genetic Algorithm Hyperparameters

B Top molecules identified for singlet-triplet gap

Random search identifies C(O)=CC2=CS1C=CS2N=COC1=O with a singlet-triplet gap of 0.0009 eV calculated by the oracle. Further, BO finds O=C1OC2=CN=C(C1=CO)C=C2 with a calculated singlet-triplet gap of 0.001 eV.

C Additional results for JT-VAE

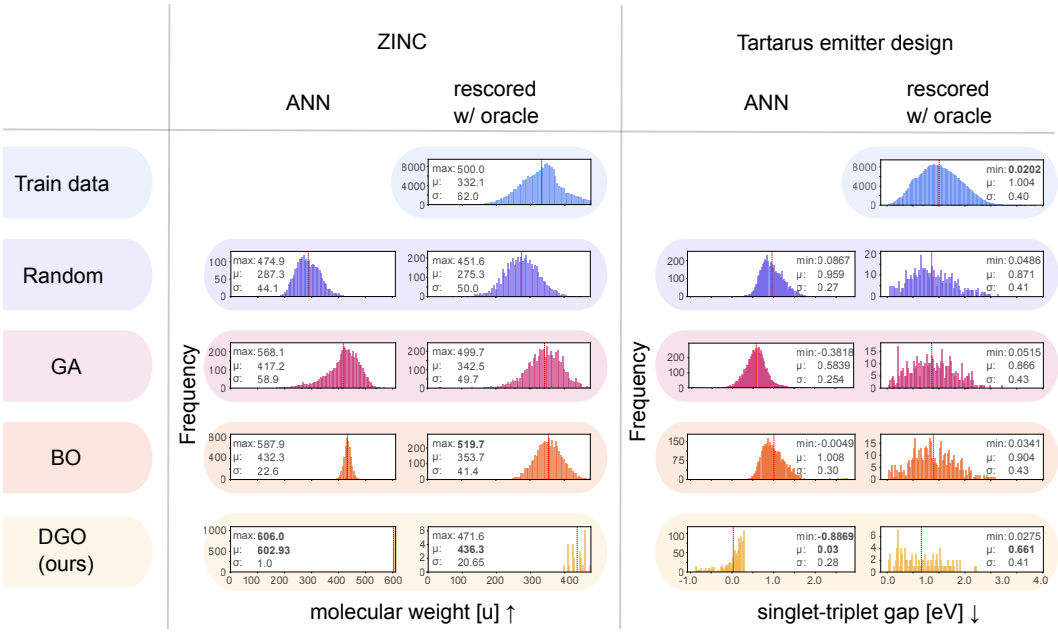

Figure 3: Molecular design results for different sampling/optimization strategies using Junction Tree VAE.