Rethinking Molecular Design: Integrating Latent Variable and Auto-Regressive Models for Enhanced Goal Directed Generation

Anonymous Authors[1]

Abstract

De novo molecule design has become a highly active research area, advanced significantly through the use of state-of-the-art generative models. Despite these advances, several fundamental questions remain unanswered as the field increasingly focuses on more complex generative models and sophisticated molecular representations as an answer to the challenges of drug design. In this paper, we return to the simplest representation of molecules, and investigate overlooked limitations of classical generative approaches, particularly Variational Autoencoders (VAEs) and auto-regressive models. We propose a hybrid model in the form of a novel regularizer that leverages the strengths of both to improve validity, conditional generation, and style transfer of molecular sequences. Additionally, we provide an in depth discussion of overlooked assumptions of these models' behaviour.

1. Introduction

Molecule discovery tasks can be divided into two types. Global optimization aims to find molecules with a specific target property. Local optimization starts with an initial molecule and searches for similar molecules that have the desired property without deviating too much from the original. Past advances in molecular design have primarily leveraged either latent variable models or auto-regressive models, each with distinct advantages and shortcomings. While auto-regressive models excel in capturing conditional distributions and generating valid molecules with desired properties, they tend to overfit the training data, leading to a limited number of novel molecules when trained on small datasets (Mollaysa et al., 2020). Additionally, they lack the flexibility of latent variable models, which facilitate local optimization and complex transformations such as style transfer. On the other hand, latent variable models often struggle to generate valid molecules unless more restrictive and complex representations are applied, due to the challenges of generating valid molecules from a compressed latent space (Gómez-Bombarelli et al., 2016; Kusner et al., 2017; Dai et al., 2018; Jin et al., 2018).

In this work, instead of relying on increasingly complex generative models and sophisticated representations as solutions for every challenge, we return to the simple string based representation SMILES. By combining the strength of auto-regressive models and latent variable models, we aim to resolve the performance gap in conditional generation and enable effective style transfer without sacrificing the generative model's validity. We introduce a theoretical framework that combines the robust conditional distribution modeling capability of auto-regressive models with the flexible representational power of latent variable models. Our approach involves a dual-training mechanism where the auto-regressive model informs the training of the latent variable model, ensuring that the generative process respects both the desired property of the generated samples as captured by the auto-regressive model and the local structural coherence enforced by the latent variable.

The core of our methodology is a hybrid generative model that employs a conditional VAEs architecture with an embedded auto-regressive model to guide the decoder (generative distribution). The model is trained using a novel objective function that incorporates a regularizer to ensure the generation of valid molecules with desired properties. The regularizer is implemented in two forms: a calibration regularizer, which is a Kullback-Leibler divergence term between the marginalized generative distribution and a target distribution defined by the auto-regressive model; and a reward-based regularizer, which rewards the generative model for producing molecules with high probability under the target distribution. Each regularizer guides the generative model, enhancing its ability to produce valid and conditionally appropriate molecules. Our contribution includes:

- We present a comprehensive analysis of the limitations

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the ICML 2024 Workshop on Accessible and Efficient Foundation Models for Biological Discovery. Do not distribute.

associated with latent generative models in molecular design, particularly focusing on their ability to perform conditional generation and style transfer.

- We propose a hybrid generative framework that effectively combines the strengths of auto-regressive and latent variable models, offering a balanced approach to molecular design.

2. Methods

Suppose we are given a training set of pairs $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}; i = 1, \ldots, N$, where $\mathbf{x}_i$ corresponds to a molecule and $\mathbf{y}_i$ represents its corresponding molecular properties. Goal directed molecule design can be formulated as learning the conditional distribution $p(\mathbf{x}|\mathbf{y})$ which allow us to set a value for $\mathbf{y}$ and generate a diverse set of molecules exhibiting the specified properties. Additionally, this task can be framed as local optimization, where we start with a prototype molecule and modify it to achieve the intended property without deviating (structurally) too far from the original molecule. We refer to this as style transfer over molecules, where we aim to modify $\mathbf{x}$ to $\mathbf{x}'$ such that $\mathbf{x}'$ exhibits property $\mathbf{y}'$. This can be achieved with a latent variable model $p(\mathbf{x}|\mathbf{y}, \mathbf{z})$ that allows us to control latent factors corresponding to certain molecular structures. The style transfer task can then be defined as follows:

$$p_\theta(\mathbf{x}'|\mathbf{y}', \mathbf{x}) = \int p_\theta(\mathbf{x}'|\mathbf{y}', \mathbf{z}) q_\phi(\mathbf{z}|\mathbf{x}) \mathrm{d}\mathbf{z}. \tag{1}$$

Concretely, this involves fitting a joint generative model of the form $p_\theta(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p_\theta(\mathbf{x}|\mathbf{y}, \mathbf{z})p(\mathbf{y})p(\mathbf{z})$ and a variational distribution $q_\phi(\mathbf{z}|\mathbf{x})$ to infer the latent variable $\mathbf{z}$. This model can be trained with the standard ELBO objective function:

$$\mathcal{L}_{ELBO}(\theta, \phi) = \sum_{i=1}^{N} \Big\{ \mathbb{E}_{q_\phi(\mathbf{z}_i|\mathbf{x}_i)}[\log p_\theta(\mathbf{x}_i|\mathbf{y}_i, \mathbf{z}_i)] - D_{KL}(q_\phi(\mathbf{z}_i|\mathbf{x}_i)||p(\mathbf{z}_i)) \Big\} \tag{2}$$

which corresponds to learning a supervised (conditional) VAE (Kingma et al., 2014).

Such models are often trained to generate molecules in one shot (Kusner et al., 2017; Dai et al., 2018), meaning the decoder assumes $p(\mathbf{x}|\mathbf{y}, \mathbf{z}) = \prod_{t=0}^{T} p(x_t|\mathbf{y}, \mathbf{z})$. This was set to avoid using a strong decoder(auto-regressive network trained with teacher forcing) that assumes $p(\mathbf{x}|\mathbf{y}, \mathbf{z}) = \prod_{t=0}^{T} p(x_t|\mathbf{y}, \mathbf{z}, x_{0,1\ldots,t-1})$ as it leads to generative model $p(\mathbf{x}|\mathbf{z}, \mathbf{y})$ ignoring the latent variable $\mathbf{z}$ and defaulting to $p(\mathbf{x}|\mathbf{y})$ (Gómez-Bombarelli et al., 2016). As a result, often the generative model $p(\mathbf{x}|\mathbf{z}, \mathbf{y})$ is trained in a non-auto regressive fashion, which makes the task very challenging and often under performs compared to auto-regressive models that designed to learn to model $p(\mathbf{x}|\mathbf{y})$ which is trained to predict the next token.

Additionally, the task of conditional generation becomes challenging when using standard conditional VAEs to model the generative distribution $p(\mathbf{x}|\mathbf{z}, \mathbf{y})$. The standard objective (Eq.2) focuses only on reconstruction loss and a KL term, and has the potential to lead to scenarios where $p(\mathbf{x}|\mathbf{z}, \mathbf{y})$ may disregard $\mathbf{y}$, relying solely on $\mathbf{z}$ for molecule reconstruction. This often results in poor conditional generation performance, as the model effectively converges to $p(\mathbf{x}|\mathbf{z})$ and loses control over the properties of the generated molecules. Furthermore, the style transfer task also tends to perform poorly due to the lack of supervision during the training. The models are not exposed to examples where a latent $\mathbf{z}$ from a prototype molecule combines with a new property $\mathbf{y}$ to predict what a molecule with structure $\mathbf{z}$ and property $\mathbf{y}$ would look like. This gap in training results in the model's inability to effectively handle style transfer.

Our observations indicate that an auto-regressive model, which directly learns the conditional distribution $p(\mathbf{x}|\mathbf{y})$, demonstrates superior performance in conditional generation. This enhancement is attributed to its effective learning of the conditional distribution $\tilde{p}(\mathbf{x}|\mathbf{y}) \approx p(\mathbf{x}|\mathbf{y})$. By using this learned distribution as an approximation to the true conditional distribution, we can effectively guide our VAE model. This guidance enables the VAE to generate molecules that have a high probability under the learned distribution $\tilde{p}(\mathbf{x}|\mathbf{y})$.

2.1. Proposed Regularizer

We propose the use of a pretrained auto-regressive model $\tilde{p}(\mathbf{x}|\mathbf{y})$ as a surrogate for the true conditional distribution. This surrogate distribution can then be employed to direct the generative network, encouraging the generative model generate molecules that have a high probability under the surrogate model. To achieve this, we augment the supervised VAE objective with one of the regularizer terms outlined below.

Calibration Regularizer To achieve this constraint, we introduce a Kullback-Leibler (KL) divergence term that aligns the marginalized generative distribution $p_\theta(\mathbf{x}|\mathbf{y}_i)$ of a conditional VAE with the surrogate distribution $\tilde{p}(\mathbf{x}|\mathbf{y}_i)$.

$$\mathcal{R}_1(\theta) = \min_\theta D_{KL}(p_\theta(\mathbf{x}|\mathbf{y}_i)||\tilde{p}(\mathbf{x}|\mathbf{y}_i)) \tag{3}$$

where the marginalized conditional distribution is given by:

$$p_\theta(\mathbf{x}|\mathbf{y}_i) = \int p_\theta(\mathbf{x}|\mathbf{y}_i, \mathbf{z}) p(\mathbf{z})\, d\mathbf{z} = \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})} p_\theta(\mathbf{x}|\mathbf{y}_i, \mathbf{z})$$

Given we do not have direct access to the marginalized distribution $p_\theta(\mathbf{x}|\mathbf{y}_i)$, we estimate it using Monte Carlo integration approach: $\hat{p}_\theta(\mathbf{x}|\mathbf{y}_i) \approx \hat{p}_\theta(\mathbf{x}|\mathbf{y}_i) =$

$\frac{1}{N}\sum_{n=1}^{N} p_\theta(\mathbf{x}|\mathbf{y}_i, \mathbf{z}_n)$ and then apply the Kullback-Leibler (KL) divergence on the approximated distribution:

$$D_{KL}(\hat{p}_\theta(\mathbf{x}|\mathbf{y}_i) \parallel \tilde{p}(\mathbf{x}|\mathbf{y}_i)) = \sum_{\mathbf{x}} \hat{p}_\theta(\mathbf{x}|\mathbf{y}_i) \log \frac{\hat{p}_\theta(\mathbf{x}|\mathbf{y}_i)}{\tilde{p}(\mathbf{x}|\mathbf{y}_i)}$$

Reward-based regularizer The supervision of the generative distribution using the surrogate distribution can also be formulated using reward formulation:

$$\begin{aligned}\mathcal{R}_2(\theta) &= \max_\theta \sum_{i=1}^{N} \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})} \mathbb{E}_{\mathbf{x}\sim p_\theta(\mathbf{x}|\mathbf{y}_i,\mathbf{z})} \log \tilde{p}(\mathbf{x}|\mathbf{y}_i) \\ &= \max_\theta \sum_{i=1}^{N}\sum_{j=1}^{M} \mathbb{E}_{\mathbf{x}\sim p_\theta(\mathbf{x}|\mathbf{y}_i,\mathbf{z}_j)} \log \tilde{p}(\mathbf{x}|\mathbf{y}_i) \quad (4)\end{aligned}$$

This regularizer can be seen as expected reward objective where $\log \tilde{p}(\mathbf{x}|\mathbf{y}_i)$ can be seen as reward function and $p_\theta(\mathbf{x}|\mathbf{y}_i, \mathbf{z}_j)$ is the policy of the action, where we want to learn a policy model that maximize our reward. The regularzier can be optimized using the score function gradient estimate, i.e., $\mathbb{E}_{\mathbf{x}\sim p_\theta(\mathbf{x}|\mathbf{y}_i,\mathbf{z}_j)} [\log \tilde{p}(\mathbf{x}|\mathbf{y}_i) \nabla_\theta \log p_\theta(\mathbf{x}|\mathbf{y}_i, \mathbf{z}_j)]$ However, the score function gradient estimator is known to be noisy and unstable unless wen use large samples and control variate techniques. To avoid excessive computational complexity, we propose an alternative approach to achieve the same regularization. We reformulate our regularizer as $\max_\theta \mathbb{E}_{\mathbf{x}\sim p_\theta(\mathbf{x}|\mathbf{y}_i,\mathbf{z}_j)} \tilde{p}(\mathbf{x}|\mathbf{y}_i)$, using $\tilde{p}(\mathbf{x}|\mathbf{y}_i)$ as a positive reward function. This allows us to express the regularizer as:

$$\begin{aligned}\max_\theta \mathbb{E}_{\mathbf{x}\sim p_\theta(\mathbf{x}|\mathbf{y}_i,\mathbf{z}_j)} \tilde{p}(\mathbf{x}|\mathbf{y}_i) &= \max_\theta \int p_\theta(\mathbf{x}|\mathbf{y}_i, \mathbf{z}_j) \tilde{p}(\mathbf{x}|\mathbf{y}_i) d\mathbf{x} \\ &= \max_\theta \mathbb{E}_{\mathbf{x}\sim \tilde{p}(\mathbf{x}|\mathbf{y}_i)} p_\theta(\mathbf{x}|\mathbf{y}_i, \mathbf{z}_j) \quad (5)\end{aligned}$$

2.2. Incorporating the regularizer

We incorporate our regularzier in the standard conditional VAE objective (Eq.2). To achieve this, we first train an LSTM-based model to learn $\hat{p}(\mathbf{x}|\mathbf{y})$. This pre-trained model is then integrated into the training process of the conditional VAE (cVAE). Our final training objective consists of the standard ELBO (Eq. 2) and an additional regularizer, $\mathcal{L}(\theta, \phi) = \mathcal{L}_{ELBO}(\theta, \phi) + \lambda \mathcal{R}(\theta)$. where $\mathcal{R}(\theta)$ refers to either the calibration regularizer or the reward-based regularizer. Depending on which regularizer is used, it is either minimized together with the negative ELBO or its negative form is used when applying the reward-based regularizer. It is important to note that during training of our model, the regularizer updates only the parameters of the generative model (decoder). The parameters of the surrogate distribution $\hat{p}(\mathbf{x}|\mathbf{y})$ remain fixed.

3. Experiments

Dataset We conducted a preliminary investigation of the regularizers' behaviour on the QM9 dataset (Blum & Reymond, 2009). We use the octanol-water partition coefficient (LogP) as the conditioning property. Preprocessing details and data analysis are available in Appendix C .

Models We implemented a baseline conditional VAE along with three versions of the regularizer: the Calibration regularizer, which uses the KL divergence and is referred to as cVAE-KLD, and the two implementations of the reward-based regularizer (referred to as cVAE-Pol1 and cVAE-Pol2). For each of these we trained three versions with different decoding strategies (details provided in Appendix D.1): a one-shot decoder, where the sampling of the current token does not depend on the previous token (denoted as cVAE); an auto-regressive decoder, which uses an auto-regressive structure and explicitly samples next tokens at each generation step trained without teacher forcing (EXP-cVAE); and an auto-regressive decoder that employs teacher forcing during training to provide input tokens for the sequence (EXP-cVAE-TF). We used an LSTM model as the surrogate. Full details on model architecture and training parameters are in Appendix D.

Evaluation Conditional generation performance was evaluated using validity, novelty, and uniqueness, while conditional generation was assessed with the MAE between target and generated properties. Style transfer performance was measured by the proportion of valid style transferred molecules (style transfer validity), the proportion of transfers where the source molecule was reproduced (PCT Fail), Tanimoto similarity of Morgan fingerprints of source and target molecule (Tan) compared with two baselines as lower bounds (Tan-B1 and Tan-B2), and the MAE between target and generated properties. Examples of successful style transfer are provided in Appendix B. Details on metric computation and sampling are provided in Appendix E.

3.1. Experimental results

Performance of One-Shot Decoder Models The results of our experiments for each regularizer type (KLD, Pol1, Pol2) are presented in Tables 1-6. Notably, as table 1 shows the conditional performance of the versions of the models trained with the one-shot version of the decoder perform poorly, with the KLD, and Pol2 versions failing to generate a large proportion of valid molecules. Upon further inspection, the large property error scores can be attributed to long model outputs without termination tokens that were valid by chance. This aligns with the findings of (Gómez-Bombarelli et al., 2016). Despite this poor performance, the one-shot versions of the models performed better at the style transfer task (see Table 2), implying latent points and their neighbourhoods belonging to the posterior distribution learned by $q_\phi(z|x)$ are meaningful under the decoder.

Model	Recon	Valid	Uniq	Novelty	Prop MAE
LSTM	-	0.97	1.0	0.3024	0.3211
cVAE	0.6605	0.018	0.7111	0.8444	20.3636
cVAE-KLD	0.8418	0.0036	1.0	0.8888	1.9087
cVAE-Pol1	0.894	0.1644	0.2993	1.0	38.5748
cVAE-Pol2	0.8585	0.0056	1.0	0.8571	0.8061
(Gómez-Bombarelli et al., 2016)	0.0361	0.1030	-	0.9000	-

Table 1. Conditional generation performance of models trained with one-shot decoders

Model	PCT Valid	PCT Fail	MAE	Tan	Tan-B1	Tan-B2
cVAE	0.854	0.4272	0.6021	0.3592	0.01646	0.06603
cVAE-KLD	0.8916	0.6218	0.6899	0.3632	0.03470	0.06531
cVAE-Pol1	0.894	0.6307	0.6792	0.3627	0.004535	0.06475
cVAE-Pol2	0.9056	0.6806	0.7388	0.3636	0.04366	0.06463

Table 2. Style transfer performance of cVAE with one-shot decoders

Auto-regressive decoder models To ensure alignment with the surrogate target distribution used in the regularizer, which is trained auto-regressively, we also investigated the behavior of the regularizer and the learned latent representation when using auto-regressive decoder models. To prevent latent representation degradation we abstained from using teacher forcing. Instead, at each step, we sampled from the model's own prediction, forcing the decoder to rely on the latent representation during generation. The results are presented in Tables 3, 4, 5, and 6. Table 3 demonstrates promising results for the calibration regularizer in the EXP-cVAE-KLD model. It successfully improves both the validity and uniqueness compared to the baseline model (EXP-cVAE). The lower novelty of this model can be attributed to the low novelty of the LSTM model (Table 1).

Overall, the regularizer has clearly guided the generative/decoder distribution to better approximate the distribution induced by the auto-regressive model. Comparing the models' style transfer performance (Table 4), it shows both models are able to successfully perform style transfer (low PCT Fail). The overall low similarity of style transfer molecules can be attributed to the small molecules, as a change in one or two atoms can significantly lower the similarity score. Surprisingly, the calibration regularizer is significantly less likely to fail to change the source molecule.

Model	Recon	Valid	Uniq	Novelty	Prop MAE
EXP-cVAE	0.5535	0.0856	0.4813	0.8598	17.4750
EXP-cVAE-KLD	0.1782	0.9268	0.9918	0.3949	0.1655
EXP-cVAE-Pol1	0.1936	0.0456	0.6140	0.8772	15.6166
EXP-cVAE-Pol2	0.1908	0.9784	0.01145	1.0	34.7762

Table 3. Conditional generation performance of models trained with auto-regressive decoders without teacher forcing

Model	PCT Valid.	PCT Fail	MAE	Tan	Tan-B1	Tan-B2
EXP-cVAE	0.8376	0.3505	0.5547	0.3529	0.003819	0.06651
EXP-cVAE-KLD	0.7296	0.0948	0.5521	0.2862	0.06644	0.06401
EXP-cVAE-Pol1	0.7122	0.0901	0.5785	0.2822	0.008178	0.06596
EXP-cVAE-Pol2	0.7512	0.07641	0.5499	0.2916	0.002795	0.06517

Table 4. Style transfer performance of cVAE with auto-regressive decoder trained without teacher forcing

In our final exploration, we extended the decoder to use full teacher forcing (Tables 5, 6). As expected, the introduction of teacher forcing significantly improves the decoder's ability to generate valid molecules. However, its conditional generative performance remains low. The introduction of the calibration regularize greatly improves both the proportion of valid molecules and the conditional performance of the model. In terms of style transfer performance we see a similar trend to the non teacher forcing case, with the regularized version of the model having a much lower proportion of failed style transfer attempts. To illustrate this we have included a plot of the style transfer performance of both models (EXP-cVAE-KLD-TF and EXP-cVAE-TF) in Appendix B.

Model	Recon	Valid	Uniq	Novelty	Prop MAE
EXP-cVAE-TF	0.8281	0.1228	0.3290	0.8111	2.9578
EXP-cVAE-TF-KLD	0.3843	0.5836	0.9965	0.6292	0.4582
EXP-cVAE-TF-Pol1	0.3495	0.034	0.5647	0.8941	9.1231
EXP-cVAE-TF-Pol2	0.4198	0.9832	0.005288	1.0	35.0501
Lim et. all.-TF	0.4381	0.5700	0.9863	0.9520	0.4510

Table 5. Conditional generation performance of models trained with auto-regressive decoders with teacher forcing

Model	PCT Valid	PCT Fail	MAE	Tan	Tan B1	Tan B2
EXP-cVAE-TF	0.9188	0.1441	0.1951	0.2732	0.006704	0.06701
EXP-cVAE-TF-KLD	0.91	0.0567	0.2002	0.2167	0.06823	0.06688
EXP-cVAE-TF-Pol1	0.9004	0.06821	0.9187	0.2175	0.01447	0.06635
EXP-cVAE-TF-Pol2	0.9056	0.07641	0.1966	0.2171	0.002696	0.06708
Lim et all-TF	-	0.0210	0.3439	0.1424	0.08160	0.11450

Table 6. Style transfer performance of cVAE with auto-regressive decoder trained with teacher forcing

Note that the results of the auto-regressive decoder (trained with teacher forcing) contradict the common assumption that a "strong decoder" could degrade the latent space. Therefore, we conducted a more in-depth analysis, which we provide in Appendix Section A.4. We conclude when trained in an auto-regressive fashion with teacher forcing, the latent representation is utilized until a certain length of the sequence, beyond which the decoder does not rely on the latent representation to generate the rest of the sequence.

4. Conclusions

Our framework addresses key limitations in classic generative methods, particularly enhancing conditional generation and style transfer. Novel regularizers, such as the calibration and reward-based regularizers, guide the generative process, improving the fidelity and validity of generated molecules. This work is intended to return to fundamental models based on simple SMILES representations to investigate commonly accepted hypotheses and provide in-depth analysis. Experimental results demonstrated the potential of the proposed regularizers, however, further investigation and testing are still needed. Our work also provided insights about the models and assumption used in molecular design, paving the way for future research in more complex settings.

References

Deep reinforcement learning for de novo drug design. *Science Advances*, 4, 2017. URL https://api.semanticscholar.org/CorpusID:38125055.

Bayer, J. and Osendorfer, C. Learning stochastic recurrent networks. *arXiv preprint arXiv:1411.7610*, 2014.

Bird, A. and Williams, C. K. Customizing sequence generation with multi-task dynamical systems. *arXiv preprint arXiv:1910.05026*, 2019.

Bjerrum, E. J. and Threlfall, R. Molecular generation with recurrent neural networks (rnns). *ArXiv*, abs/1705.04612, 2017. URL https://api.semanticscholar.org/CorpusID:8280382.

Blum, L. C. and Reymond, J.-L. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.*, 131:8732, 2009.

Boulanger-Lewandowski, N., Bengio, Y., and Vincent, P. Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription. *arXiv preprint arXiv:1206.6392*, 2012.

Chung, J., Kastner, K., Dinh, L., Goel, K., Courville, A. C., and Bengio, Y. A recurrent latent variable model for sequential data. In *Advances in neural information processing systems*, pp. 2980–2988, 2015.

Dai, H., Tian, Y., Dai, B., Skiena, S., and Song, L. Syntax-directed variational autoencoder for structured data. *arXiv preprint arXiv:1802.08786*, 2018.

Fabius, O. and van Amersfoort, J. R. Variational recurrent auto-encoders. *arXiv preprint arXiv:1412.6581*, 2014.

Gómez-Bombarelli, R., Duvenaud, D. K., Hernández-Lobato, J. M., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *CoRR*, abs/1610.02415, 2016. URL http://arxiv.org/abs/1610.02415.

Grechishnikova, D. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Scientific Reports*, 11, 2019. URL https://api.semanticscholar.org/CorpusID:213566533.

Jin, W., Barzilay, R., and Jaakkola, T. Junction tree variational autoencoder for molecular graph generation. *arXiv preprint arXiv:1802.04364*, 2018.

Kingma, D. P., Mohamed, S., Rezende, D. J., and Welling, M. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems*, pp. 3581–3589, 2014.

Kotsias, P.-C., Arús-Pous, J., Chen, H., Engkvist, O., Tyrchan, C., and Bjerrum, E. J. Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks. *Nature Machine Intelligence*, 2:254 – 265, 2020. URL https://api.semanticscholar.org/CorpusID:219457600.

Kusner, M. J., Paige, B., and Hernández-Lobato, J. M. Grammar variational autoencoder. *arXiv preprint arXiv:1703.01925*, 2017.

Lim, J., Ryu, S., Kim, J. W., and Kim, W. Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of Cheminformatics*, 10, 2018a. URL https://api.semanticscholar.org/CorpusID:49273733.

Lim, J., Ryu, S., Kim, J. W., and Kim, W. Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *Journal of cheminformatics*, 10(1):1–9, 2018b.

Liu, Q., Allamanis, M., Brockschmidt, M., and Gaunt, A. Constrained graph variational autoencoders for molecule design. In *Advances in Neural Information Processing Systems 31*, pp. 7806–7815. 2018.

Mokaya, M., Imrie, F., van Hoorn, W. P., Kalisz, A., Bradley, A. R., and Deane, C. M. Testing the limits of smiles-based de novo molecular generation with curriculum and deep reinforcement learning. *Nature Machine Intelligence*, 5:386–394, 2022. URL https://api.semanticscholar.org/CorpusID:250926481.

Mollaysa, A., Paige, B., and Kalousis, A. Goal-directed generation of discrete structures with conditional generative models. *Advances in Neural Information Processing Systems*, 33, 2020.

Olivecrona, M., Blaschke, T., Engkvist, O., and Chen, H. Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9, 2017. URL https://api.semanticscholar.org/CorpusID:2978311.

Segler, M. H., Kogej, T., Tyrchan, C., and Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS central science*, 4 (1):120–131, 2017.

Simonovsky, M. and Komodakis, N. Graphvae: Towards generation of small graphs using variational autoencoders. *arXiv preprint arXiv:1802.03480*, 2018.

Thiede, L. A., Krenn, M., Nigam, A., and Aspuru-Guzik, A. Curiosity in exploring chemical spaces: intrinsic rewards for molecular reinforcement learning. *Machine Learning: Science and Technology*, 3, 2020. URL https://api.semanticscholar.org/CorpusID:229339732.

Wang, J., Hsieh, C.-Y., Wang, M., Wang, X., Wu, Z., Jiang, D., Liao, B., Zhang, X., Yang, B., He, Q., Cao, D., Chen, X., and Hou, T. Multi-constraint molecular generation based on conditional transformer, knowledge distillation and reinforcement learning. *Nature Machine Intelligence*, 3:914 – 922, 2021. URL https://api.semanticscholar.org/CorpusID:244575754.

Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., and Chao, L. S. Learning deep transformer models for machine translation. In *Annual Meeting of the Association for Computational Linguistics*, 2019. URL https://api.semanticscholar.org/CorpusID:174799399.

You, J., Liu, B., Ying, R., Pande, V., and Leskovec, J. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems 31*, pp. 6412–6422. 2018.

A. Literature review

A.1. Auto regressive models on Smiles

Early auto-regressive approaches to molecule generation (Segler et al., 2017) (Bjerrum & Threlfall, 2017) showed that auto-regressive model can successfully learn a distribution of SMILES strings and generate realistic candidates with high accuracy. Furthermore, Segler et al. (2017) proposed conditional generation be achieved by filtering the generated molecules for candidates with desirable properties or by fine tuning the model on a curated dataset of desirable molecules. Methods have also been developed to directly control the conditioning and avoid the need for further fine-tuning. For example, concatenating conditional information to inputs at the first or every generation step and encoding conditional information into the model's initial hidden state (Kotsias et al., 2020).

A parallel approach to guiding auto-regressive models for conditional generation uses policy optimization and a set of objective functions to fine-tune auto retrogressive models. The seminal examples in this direction are REINVENT (Olivecrona et al., 2017) and ReLeaSe (201, 2017). More recent has incorporated various other innovations from the field of reinforcement learning such as (Thiede et al., 2020) and curriculum learning (Mokaya et al., 2022).

Following their success in the field of natural language processing, the majority of recent publications have adopted various transformer based architectures. Wang et al. (2019) and Grechishnikova (2019) propose transformer based approaches to structure based drug design. Finally, Wang et al. (2021) propose a multi-step pipeline where knowledge is distilled from a transformer model to an RNN-based model which is then further fine-tuned using reinforcement learning.

Although these have been found to outperform older architectures such as RNN and LSTM, the difference is minimal in simpler settings. For this reason and to maintain a similar architecture between the VAE decoder and auto-regressive model and prior VAE based models, we have opted to employ a conditional LSTM model in this study.

A.2. latent variable models on Smiles

The seminal implementation of a latent variable model for SMILES (Gómez-Bombarelli et al., 2016) is based on a conditional VAE that encodes SMILES as continuous latent vectors. It also uses a network to predict molecular properties from latent vectors, allowing for efficient property optimization. Lim et al. (2018a) propose an approach based on a conditional VAE for simultaneous control of multiple properties. Although successful, the smiles generated by the model are prone to syntactic and semantic errors. Following works aimed to address this by operating on grammar rules for generating SMILES (Kusner et al., 2017), and later on attribute grammars (Dai et al., 2018) (to address errors that are still possible when using a regular CFG). These models were able to generate a much higher proportion of valid smiles at the cost of computational and implementation complexity.

Others works such as Simonovsky & Komodakis (2018), Jin et al. (2018), You et al. (2018) and Liu et al. (2018) explored models that operate on graph representations, largely improving the ability to generate valid molecules. A notable example is JT-VAE (Jin et al., 2018) witch is able to guarantee validity of generated molecules by constructing molecules from a library of fragments.

A.3. Auto regressive model with latent variable

There have been several attempts to introduce latent variables into RNNs such VRNN (Chung et al., 2015), and others (Bayer & Osendorfer, 2014; Boulanger-Lewandowski et al., 2012; Fabius & van Amersfoort, 2014). However, the goal of such latent variables is to introduce more stochasticity into the model as a standard RNN has limited power to model the variability observed in highly-structured data. Integrating latent variables into the hidden state of the RNN aims to be able to model more variability that is observed in the structured data. When used to model the conditional distribution $p(\mathbf{x}|\mathbf{y})$, while such models help to learn a distribution with more diverse samples compared to standard RNN models, they do not provide sufficient control over the style of the generated samples. Recently, an approach similar to VRNN, Bird & Williams (2019) also introduces a latent variable $\mathbf{z}$ into RNNs, in addition to modelling more variability of sequences, this paper aims to capture sequence-specific features through latent the variable $\mathbf{z}$ and use it to perform style transfer. The latent variable $\mathbf{z}$ is inferred from input output sequence pairs and the parameters of the RNN are made dependent on $\mathbf{z}$. This enables each sequence to be modelled as a single dynamical system and capture instance-specific variation through $\mathbf{z}$ which later helps us to perform style transfer over sequences. Due to auto-regressive models' strong conditional generation performance, it would be interesting to perform style transfer with auto-regressive models in an unsupervised fashion without the need of

Condition	Reconstruction
Standard (using $\mathbf{z}$ throughout)	0.4351
Using $\mathbf{z}$ up to 15th step	0.4312
Using $\mathbf{z}$ up to 10th step	0.3922
Using $\mathbf{z}$ up to 1st step	0.0025
Not using $\mathbf{z}$ (prior from step zero)	0.0000

Table 7. Reconstruction Rates with Different Conditions

constructing a paired dataset.

A.4. Strong decoder

It has been assumed or stated by (Gómez-Bombarelli et al., 2016) that when trained with teacher forcing, the auto-regressive decoder tends to ignore the encoded $\mathbf{z}$ and recover the next token purely from the current token. However, this assumption was overlooked by (Lim et al., 2018b), who proceeded with a similar type of model featuring a strong decoder (an RNN trained with teacher forcing). Mathematically, it is reasonable to assume that if one can learn $p(\mathbf{x}|\mathbf{y})$ using an auto-regressive model trained to predict the next token (Segler et al., 2017), then when training a generative model $p(\mathbf{x}|\mathbf{z},\mathbf{y})$ through a VAE using the same model as a decoder, the generative model could easily converge to the auto-regressive model $p(\mathbf{x}|\mathbf{y})$ by ignoring $\mathbf{z}$. This is because it is easy to recover the next token $p(x_t|\mathbf{y}, x_1, \ldots, x_{t-1}, \mathbf{z})$ without relying on $\mathbf{z}$ at all. This assumption has been generally accepted in subsequent research papers, which have trained $p(\mathbf{x}|\mathbf{z},\mathbf{y})$ without the auto-regressive component when using latent variable models to avoid degenerated latent representations.

To investigate the impact of the latent variable $\mathbf{z}$, we took the model from (Lim et al., 2018b) and tested if their model leads to degraded latent representation. The first reliable measure would be reconstruction, but they did not report the reconstruction performance of their model. Since the model training is very slow, we retrained their model on the QM9 dataset as a proof of concept.

The model performance is reported in Tables 5 and 6. It shows that we have a reconstruction rate of 0.43. Note that $\mathbf{y}$ also carries some information about $\mathbf{x}$. Therefore, to test further, we tried the reconstruction with latent $\mathbf{z}$ that is sampled from the prior instead of sampled from the posterior. This resulted in a 0% reconstruction rate, which further validates that the latent variable $\mathbf{z}$ does carry useful information about the encoded $\mathbf{x}$.

After investigating whether the learned latent variable $\mathbf{z}$ carries meaningful information, we discovered that it does, but only up to a certain step in the auto-regressive generation process. For instance, with the QM9 dataset, where the maximum sequence length T is under 40, the decoder operates in an auto-regressive setting. During reconstruction, while decoding progressively, we observed the following results presented in Table 7:

This indicates that $\mathbf{z}$ carries information that is utilized only until the 16th step, after which it becomes ineffective. This behavior arises from the auto-regressive training with teacher forcing. Initially, when generating the next token conditioned on $\mathbf{y}$ and the current token, there are many possible characters that could be the next token, each leading to a SMILES string with the property $\mathbf{y}$. Therefore, it is crucial for $\mathbf{z}$ to guide the decoding process to ensure the generation of the specific $\mathbf{x}$ that was used to encode $\mathbf{z}$. This makes $\mathbf{z}$ essential for generating the next token at the beginning of the sequence. However, as the sequence progresses, the number of possible next tokens significantly reduces, making it clear what the next token should be based on the current token. Consequently, $\mathbf{z}$ becomes less useful, leading the model to utilize $\mathbf{z}$ primarily for encoding the initial structure of the original SMILES string, but not for the entire sequence.

A.5. Reward based Regularizer Behaviour

The two versions of the reward based regularizer (Pol1, Pol2) fail to produce a high proportion of valid or unique molecules respectively. This result is also fixed regardless of the decoder structure. A possible cause of this is high variance during training. In future work we intend to explore this by modifying the sampling strategy and greatly increasing the number of samples. Another solution we will explore is sampling latent points for the regularizer by encoding smiles from the training set, as the discrepancy between it's conditional generation and style transfer property errors implies that the effect of the regularizer is overpowering the effect of the reconstruction loss in regions of the search space outside the vicinity of the aggregate posterior of the encoder.

A.6. Analysis

1. Can we actually disentangle the latent variable $\mathbf{z}$ and property $\mathbf{y}$ First of all, unlike in the space of image where we can actually annotate the variations in the data, molecular space is much more complicated. When we apply vanilla VAE to encode molecules to a latent space, we can assume that $\mathbf{z}$ encodes the molecular structure, and previous works have shown that points that are close in the latent space result in molecules that are structurally similar. However, once we assume a molecule is generated from a property variable y and some other latent factor $\mathbf{z}$ where $\mathbf{y}$ and $\mathbf{z}$ are independent under the prior, $p(\mathbf{x}, \mathbf{y}, \mathbf{z}) = p(\mathbf{x}|\mathbf{y}, \mathbf{z})p(\mathbf{z})p(\mathbf{y})$, then we can not really say what $\mathbf{z}$ encodes. Therefore, it is not really clear what $\mathbf{z}$ encodes in this setting. Even though we still assume $\mathbf{z}$ encodes some structural features of the molecules, for instance, some functional groups that are not dependent on the property $\mathbf{y}$, it is hard to clarify what aspect of the molecule that $\mathbf{z}$ learns. However, it is clear is that $\mathbf{z}$ does not fully specify the molecule's structure since the property of the molecule is fully determined by the structure of the molecule, and we assume the learned $\mathbf{z}$ is disentangled from $\mathbf{y}$. We can not modify the property of a molecule without modifying its structure. However, note that $\mathbf{z}$ is a learned factor, assuming that it is independent of the $\mathbf{y}$. Therefore, $\mathbf{z}$ should not learn some variations in the molecule data that is independent of the conditioning properties. However, in reality, for a given molecule, can we always modify its property to the value we want and stay close to the original molecules in terms of structure.

B. Examples of Style Transfer Molecules

Figure 1 shows five examples of successful style transfer with the EXP-cVAE-TF-KLD model where the generated molecule successfully matched the target property. Figures 2 and 3 show the style transfer performance of the teacher forcing cVAE with and without the KLD based regularizer for a random sample of molecules from the test set. Finally, Figure 4 plots the properties attained from style transfer with the KLD-TF model for a range of target values. The molecules selected for this plot were the molecules with the highest, lowest and median property scores from the test set.

C. Dataset and Property Statistics

C.1. Dataset and Property Processing

Training used properties and canonical smiles extracted from the QM9 dataset file (provided as atom coordinates and types) using the RDKit chemo-informatics package (version 2023.09.6), ignoring molecules that resulted in invalid smiles. The resulting smiles were encoded as integers and augmented with start and end tokens. The remainder of each sequence was padded with 0 tokens.

C.2. Data Splits

We withheld 10000 random molecules from the dataset for the validation set (tracking progress and determining early stopping) and 10000 for the test set (performance metric computation). The remaining 113885 molecules were used as the training set. The same test-train-validation split was used to train all models, including the baseline model from (Lim et al., 2018a) and the LSTM model for the regularizers.

C.3. Data Exploration

All LogP values were normalized according to the distribution in the training-set (mean=0.2982513513338983, std=1.0008837558144368). The distribution of unnormalized LogP values for the entire dataset is presented in Figure 5, and the distribution of sequence lengths in Figure 6. The frequencies of various characters in the entire dataset can be found in Table 8.

Symbol	Frequency	Normalized Frequency
C	767830	0.38
O	177823	0.09
(	118216	0.06
)	118216	0.06
#	37027	0.02
N	98355	0.05
=	94597	0.05
1	259368	0.13
[	12523	0.01
H	10568	0.01
3	35634	0.02
+	1847	0.00
]	12523	0.01
-	1974	0.00
2	131751	0.07
c	78726	0.04
n	41409	0.02
o	10174	0.01
F	3314	0.00
4	4952	0.00
5	174	0.00

Table 8. Symbol Frequencies in Entire QM9 Dataset

D. Model Architecture and Training Parameters

D.1. Decoder Structure

Our experiments employ three different decoder structures: the one-shot decoder (cVAE), explicitly auto-regressive decoder without teacher forcing (EXP-cVAE) and explicit auto-regressive decoder with teacher-forcing:

- **One-Shot**: The decoder does not accept the tokens generated from previous steps as input and propagates sequential information solely through the hidden state. Sequences are sampled from the model's predictions by independently sampling tokens with probabilities proportionate to each decoder output for each step in the sequence.
- **Explicit/Auto-Regressive**: At each step in the sequence t, a token is sampled proportionately to the decoders outputs and provided as an input at $t + 1$.
- **Explicit/AR with Teacher Forcing**: At each step in the sequence t, a token is sampled proportionately to the decoder outputs. The token provided as input to the decoder at $t + 1$ is taken from the ground truth SMILES string. During inference the model is sampled in the same manner as the non-teacher forcing version (EXP-cVAE).

The auto-regressive model that constitutes the one-shot decoder

D.2. LSTM Parameters

For the surrogate distribution $\tilde{p}(\mathbf{x}|\mathbf{y})$, we trained a conditional LSTM model with a maximum likelihood objective until convergence. The performance of this pre-trained model is shown in Table 3. The LSTM model consisted of an initial embedding layer (47x512) three stacked LSTM layers with input feature size 513 and output feature size 512 and a final linear layer with input size 512 and output size 47. Conditional information was provided by prepending it to the input at each time step. LSTM layers were trained with a dropout probability of 0.2. Initial and final layer weights were initialized using a xavier uniform initialization strategy.

D.3. VAE Architecture and Training

All VAE models shared a common architecture and parameters (except for regularizer specific parameters and weights). Training was conducted for up to 500 epochs, with early stopping after 50 epochs without validation loss improvement. Regularizer weights were determined so that the initial regularizer loss component was approximately an order of magnitude smaller than the initial KL loss component.

The VAE model encoder consisted of an initial embedding layer followed by three 1 dimensional convolutional layers of dimensions 47x9x9, 9x9x9 and 9x10x11, each followed by a ReLu activation function. These were fed into two linear layers, one for predicting the embedding mean (435x56) and one for the log variance (435x56). The state decoder consisted an initial embedding layer (47x56) followed by three stacked GRU layers (113x501, 501x501, 501x501) each followed and a final output layer (501x47). The models were trained with gradient clipping above set to magnitude 0.2 and a learning rate scheduler set to reduce learning rate on plateaus to a minimum of 1e-6. An overview of the training parameters is presented in Table 9.

D.4. Implementation and Hardware Details

The model was implemented using the PyTorch library (version 2.2.1) and trained on a NVIDIA GeForce RTX 4090 graphics card using CUDA version 12.3.

Parameter	Value
Optimizer	Adam
Batch Size	128
Learning Rate	1e-4
Std. for Reparameterization	0.01
ELBO KLD loss component weight	1.0
KLD Regularizer Weight	0.1
Pol1 Regularizer Weight	0.01
Pol1 MC Sample Size for Latent Points	4
Pol1 MC Sample Size for Decoder	10
Pol 2 Regularizer Weight	0.1
Pol2 MC Sample Size	5

Table 9. Training parameters for VAE models

E. Metric Computation Details

E.1. Conditional Generation Metrics

Conditional generation metrics were computed by sampling 500 random latent points and equipping each with a random property value from the test set. Each pair was decoded 5 separate times to give the final sample. Validity was computed as the proportion of all generated smiles that were parsed to valid molecules by RdKit, uniqueness was the number of unique valid smiles out of all valid smiles and novelty as proportion of valid smiles who's canonical form is not present in the train set. Reconstruction was computed by encoding a molecule 5 separate times and decoding from it's encoding and property 5 separate times. The final reconstruction is the proportion of smiles that are identical to the original.

E.2. Style Transfer Metrics

Style transfer metrics were computed by encoding 2500 source SMILES from the test set and equipping each with one of 2500 target properties sampled independently from the test set. The style transfer validity of the resulting set of generated SMILES was computed as the proportion of generated SMILES that did not equal the source SMILES. The similarity was computed as the Tanimoto similarity of each pair of source and generated SMILES according to their Morgan fingerprint calculated with radius 3.

E.3. Baselines for Tanimoto Similarity

The first Tanimoto baseline was computed as the average similarity between the set SMILES generated through style transfer and a paired set of generations using random latent points instead of the encoded source molecules. The second Tanimoto baseline compared the similarity of two set of style transferred smiles with a shared set of target properties.

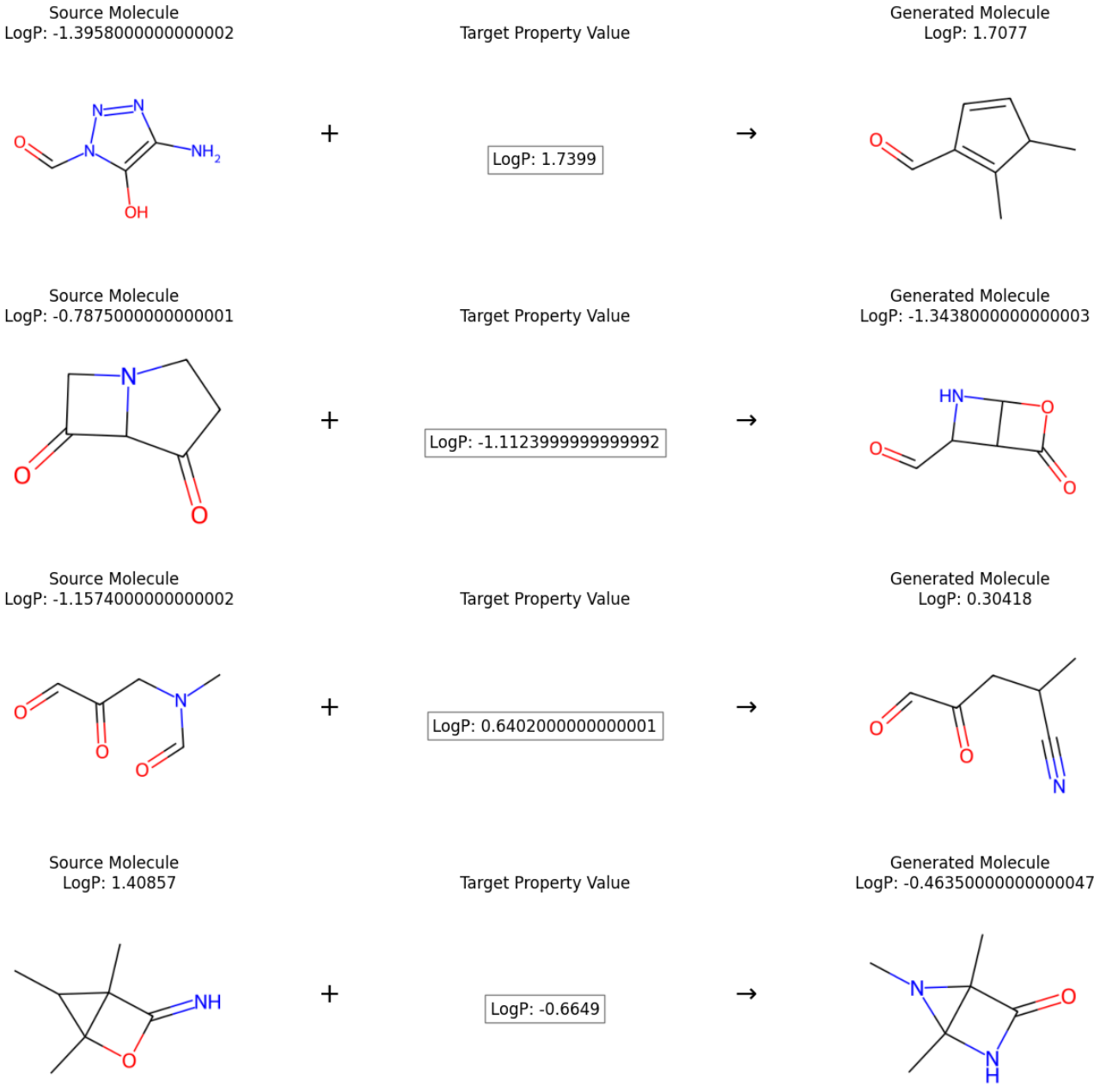

Figure 1. Examples of Successful Style Transfer

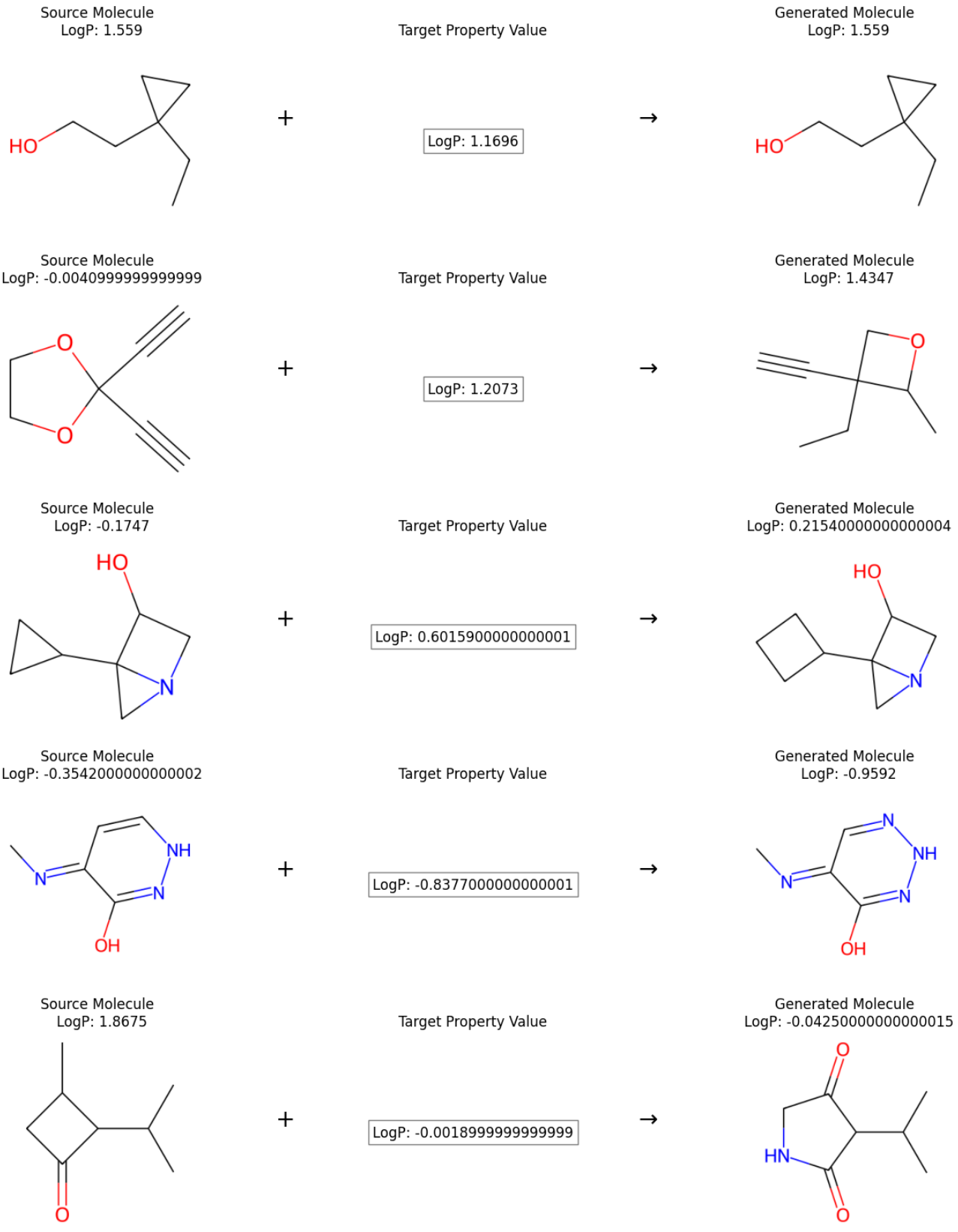

Figure 2. Random Sample of Style Transfer Attempts with EXP-CVAE-TF Model

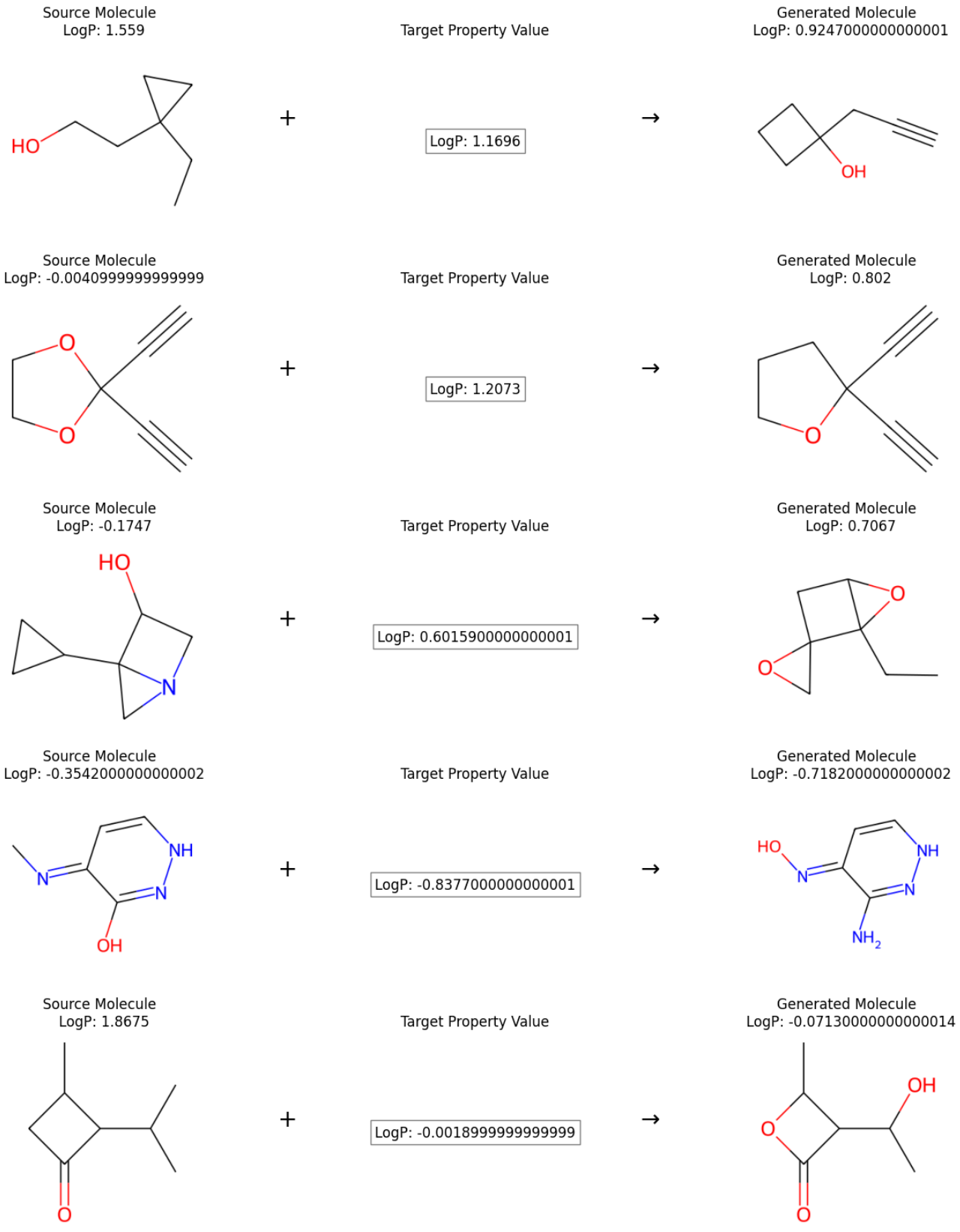

Figure 3. Random Sample of Style Transfer Attempts with EXP-cVAE-KLD Model

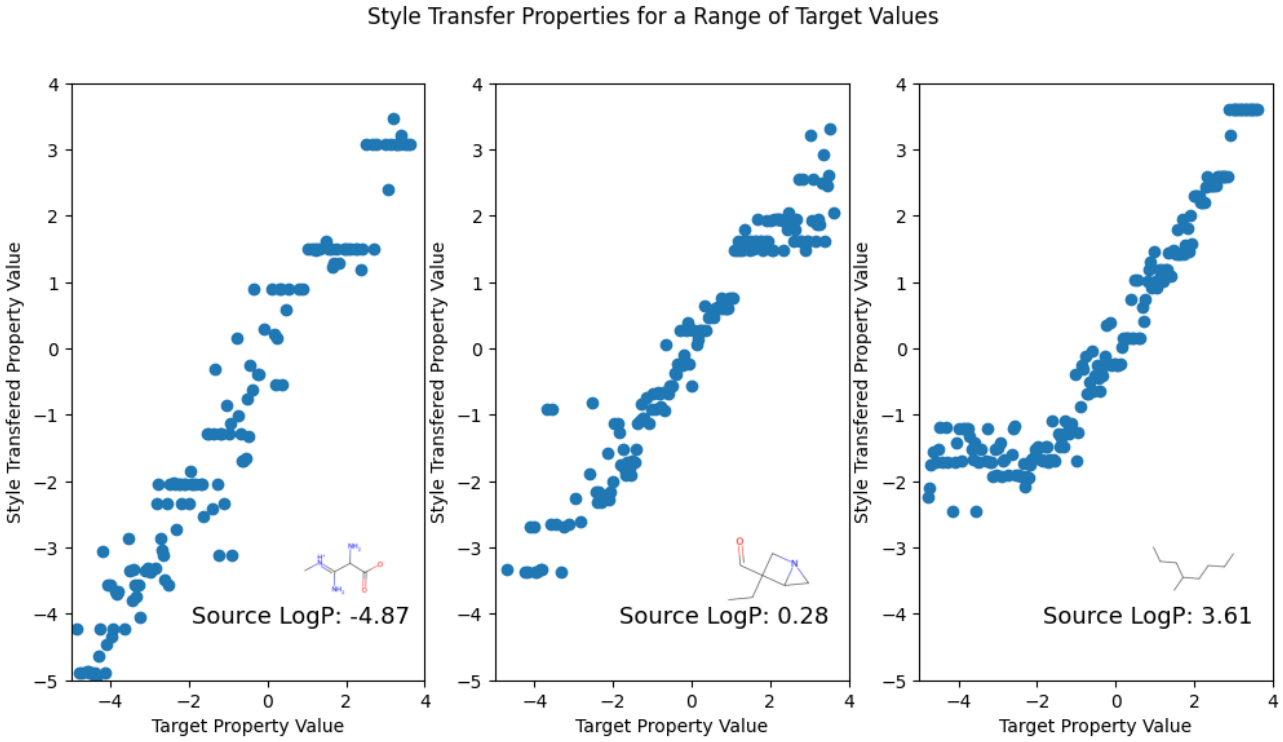

Figure 4. Style Transfer Properties for a Range of Targets of KLD-TF Model

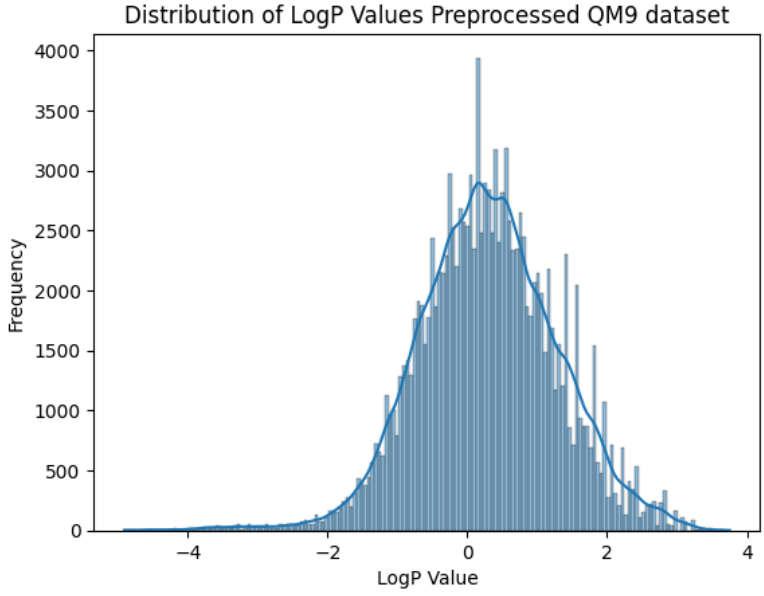

Figure 5. Distribution of Property Values

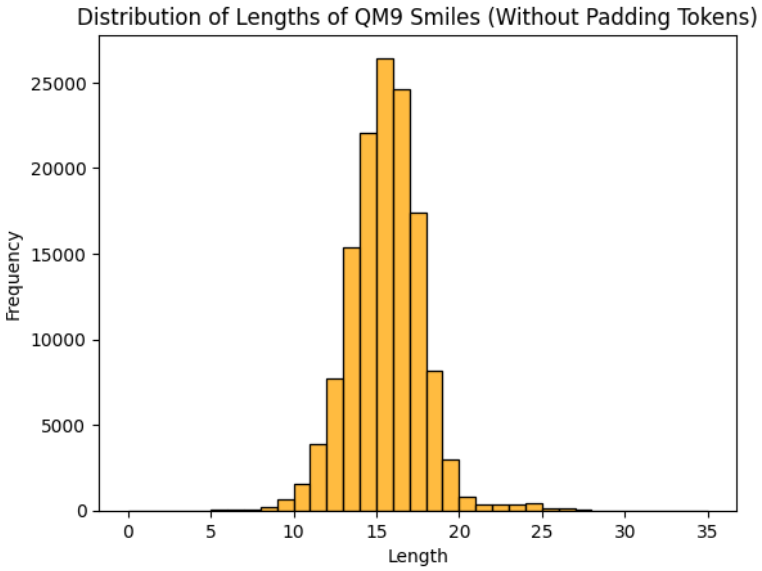

Figure 6. Distribution of Seqence Lengths