UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Improving Latent Space Diversity in SMILES Generation Using Transformer VAE

Ainin Sofia Jusoh[1,2], Muhammad Akmal Remli[1,2*], Mohd Saberi Mohamad[3,4,5], Tristan Cazenave[6] & Chin Siok Fong[7]

[1]Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, Kota Bharu, 16100, Kelantan, Malaysia
[2]Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Kota Bharu, 16100, Kelantan, Malaysia
[3]Health Data Science Lab, Department of Genetics and Genomics, College of Medical and Health Sciences, United Arab Emirates University, Al Ain, 15551, United Arab Emirates
[4]Faculty of Engineering and Technology, Multimedia University, 75450, Melaka, Malaysia
[5]Department of Biosystems Engineering, Faculty of Agricultural Technology, Universitas Brawijaya, 65145, Malang, East Java, Indonesia
[6]LAMSADE, Université Paris Dauphine- PSL, Paris, France
[7]UKM Medical Molecular Biology Institute (UMBI), 56000, Kuala Lumpur, Malaysia
Email: tgaininsofia98@gmail.com (A.S. Jusoh), akmal@umk.edu.my (M.A. Remli), saberi@uaeu.ac.ae (M.S. Mohamad), cazenave@lamsade.dauphine. fr (T. Cazenave), chinsiokfong@ppukm.ukm.edu.my (C.S. Fong).

Submitted: 15/9/2025. Revised edition: 28/11/2025. Accepted: 5/1/2026. Published online: 28/7/2026
DOI: https://doi.org/10.11113/ijic.v16n1-2.691

Abstract*—This work presents TransMolVAE, a Transformer-based variational autoencoder for molecular generation using SMILES. RNN-based models such as SmilesVAE often face posterior collapse. This happens when the latent space is not well used. To solve this, TransMolVAE applies self-attention. This allows the model to capture long-range patterns in SMILES strings. It improves reconstruction accuracy, keeps KL divergence stable, and makes better use of the latent space. The experiments showed strong results. Token accuracy increased steadily, reaching 0.72 by the fifth epoch. In contrast, SmilesVAE stayed low at 0.25. Validity, uniqueness, and novelty of molecules generated by TransMolVAE also reached above 70%. For SmilesVAE, these values were only between 50% and 55%. These findings show that TransMolVAE can generate more valid, novel, and diverse molecules compared to the baseline. The number of active units also stayed high, proving that the latent space was fully used. There was a trade-off in computation. TransMolVAE needed longer CPU time per epoch compared to SmilesVAE. This is because Transformer models are more complex and require higher computation. Even so, the improved quality of molecule generation makes TransMolVAE more suitable for drug discovery, where novelty and diversity are more important than speed. In summary, TransMolVAE combines the strengths of VAEs and Transformers. It improves reconstruction, avoids posterior collapse, and produces molecules with higher biological relevance. Future work will aim to reduce the high computational cost. This can be done with more efficient Transformer designs, larger chemical datasets, and integration with biological features. These improvements will make the model faster and even more useful for real-world drug development.

***Keywords*—Artificial Intelligence, Latent Space Diversity, Transformer, Variational Autoencoder, SMILES, Drug Discovery**

I. INTRODUCTION

Drug discovery is a long and expensive process [1,2] Developing a new drug can take more than ten years. The cost often reaches billions of dollars [3]. Many candidate molecules fail during testing, which makes the process risky and uncertain [4]. Traditional methods of searching chemical space are also slow and limited. The chemical space itself is extremely large. It is estimated to contain more than 10^60 possible molecules [5]. Exploring this space with wet lab experiments alone within a shorter time is impossible.

Computational methods have become an important alternative to wet lab experiments [6]. They can screen and analyze large numbers of compounds much faster and at a lower cost. Compared to laboratory testing, computational models can significantly reduce both time and financial expenses in the early stages of drug discovery [7]. However, while speed is important, the quality of generated molecules is even more critical [8]. Low-quality outputs can cause problems in later validation

stages, leading to wasted resources [9]. Therefore, improving molecular quality is more valuable than focusing only on faster generation, since computational methods are already much quicker than wet lab experimental approaches. To address these challenges, Machine Learning (ML) and Artificial Intelligence (AI) have emerged as strong computational solutions. In particular, deep generative models are powerful tools for molecular design [4]. They are widely used in drug discovery to create new chemical compounds. These models can explore chemical space more efficiently than traditional methods [10], and help reduce the time and cost of early drug development [11].

Variational Autoencoders (VAEs) are one of the most common generative models. They are trained on molecular representations such as SMILES strings [12]. VAEs learn to encode molecules into a latent space and then decode them back into valid molecules [13]. This makes them suitable for generating novel structures with desired properties. Models such as SmilesVAE [14], GXVAE [15], GX2Mol [16] and DRAGONET [17] have shown promising results. They demonstrate how generative learning can be applied to molecular design. However, these models are based on recurrent neural networks (RNNs) [15]. RNNs are effective for short sequences but fail when the sequence becomes long [18]. RNNs suffer from several weaknesses. They face vanishing gradients, which make training unstable [19]. They also model long-term dependencies poorly, which leads to limited representation power [20]. Additionally, VAEs with RNN decoders often encounter posterior collapse [15]. This reduces diversity in the latent space and limits the ability to explore chemical space.

Transformers provide a strong alternative to RNNs. They use self-attention to capture long-range dependencies without recurrence [21]. This allows them to process sequences in parallel and learn global relationships more effectively [4]. Transformers have already transformed many areas of Natural Language Processing (NLP). Their success shows the potential for other sequence-based tasks such as molecular generation [22]. Several works have explored Transformers in chemistry. They have been applied in molecular property prediction [23], reaction prediction [24] and molecule generation [15]. These studies confirm that self-attention can handle complex molecular structures better than recurrence. Thus, integrating Transformers into generative models is a natural step forward.

In this paper, we introduce TransMolVAE. It is a Transformer-based VAE designed for SMILES sequence modeling. TransMolVAE replaces recurrent layers with self-attention. The aim is to improve latent space usage and reduce posterior collapse. This approach combines the strengths of VAEs and Transformers to advance molecular design. The main contributions of this paper are as follows. First, we propose a new architecture that integrates a Transformer with a VAE for SMILES generation. Second, we provide a detailed evaluation of latent space quality and molecular diversity. Finally, we compare our approach with baseline models such as SmilesVAE.

II. Related literature

Generative models are widely studied in molecular discovery [4,11]. They provide a way to explore chemical space and design new compounds. Several architectures have been applied, including VAEs, Generative Adversarial Networks (GANs), Transformer, flow-based models, and diffusion models. Each of these methods offers different advantages and limitations. VAEs remain one of the most popular choices in molecular generation [25,26]. They map discrete molecular representations, such as SMILES strings into continuous latent spaces [13]. This enables interpolation, optimization and smooth exploration of molecules with similar structures. SmilesVAE [14], GxVAE [15], GX2Mol [16] and DRAGONET [17] are among the well-known VAE-based models used for molecular generation. All these models were trained using the same dataset and for the same purpose, which is to learn molecular representations and generate new valid molecules. Therefore, they are suitable for direct performance comparison.

Although each model introduces different improvements, they all share the same fundamental VAE-based encoder and decoder structure. SmilesVAE was the earliest to generate molecules directly from SMILES strings [14]. Using the same idea, DRAGONET [17] improved this approach by adopting a deep graph generative network to improve novelty and structural consistency. However, the output still contains some invalid sequences during generation. Later, GxVAE [15] introduced graph-based representations to capture molecular structures more effectively, while Gx2Mol [16] combined SMILES and graph features to enhance validity and diversity. All of these models use a VAE encoder to encode the data into the latent space. However, they still produce some similar outputs because the latent space is not fully explored. These models highlight how different extensions can be built upon the same core VAE framework but still share the same limitation. In addition, these models also used the same evaluation metrics to assess performance such as validity, uniqueness and novelty of the generated molecules. Using the same dataset and metrics ensures that the model comparison is fair and consistent.

However, the major limitation of these models lies in their use of RNNs as sequence encoders and decoders. RNN-based models are known to struggle with long sequences [18]. They often suffer from vanishing gradients, poor modeling of long-term dependencies and posterior collapse. These issues restrict the diversity of generated molecules and reduce the quality of the latent space [4]. Beyond VAEs, other generative models have been explored, such as GANs flow-based model, diffusion model and Transformers. GANs have been applied to molecule generation, but they are difficult to train and often unstable [27]. Flow-based models provide exact likelihood estimation, but they require careful design and large computational resources [28]. Additionally, diffusion models have gained attention for molecule generation [27] They achieve strong results in image synthesis, and early works suggest their potential for molecular design. Still, their use remains at an early stage compared to VAEs.

Transformers have become a breakthrough in sequence modeling [29]. Their self-attention mechanism allows them to capture long-range dependencies without recurrence [4]. This makes them more effective and scalable than RNNs. In chemistry, Transformers have been applied to tasks such as property prediction [30], molecule classification [31] and

reaction prediction [32]. Several Transformer-based molecular language models have been proposed. ChemBERTa [33] and MolBERT [34] showed that pretraining on large molecular datasets improves chemical property prediction. In addition, MolGPT [35] also applied the Transformer decoder for autoregressive molecular generation, showing competitive performance against RNN-based models. These studies highlight the strong potential of attention-based models in handling molecular representations.

Despite this progress, the integration of Transformers within VAE frameworks has received limited attention. While some works explored Transformers for direct molecule generation [17] fewer studies have combined them with latent variable models. This leaves open challenges in balancing reconstruction quality, latent space organization and molecular diversity. Our work, TransMolVAE, is motivated by this research gap. It combines the representational power of Transformers with the generative capacity of VAEs. The aim is to reduce posterior collapse, improve latent space utilization and generate molecules that are valid, unique and novel.

III. Methodology

We propose TransMolVAE, a modified version of the MolVAE in GxVAEs. Unlike the original MolVAE, which employs bidirectional GRUs, TransMolVAE replaces both the encoder and decoder with Transformer architectures. Let S = [s1, s2, ⋯ , sT] be a SMILES string, where st denotes the t-th atom. Similar to the original formulation, we assume that S is reconstructed from the latent variable Z and conditioned on the extracted gene expression feature vectors C. This process can be expressed as follows:

$$p_{\emptyset}(S|Z,C) = \prod_{t=1}^{T} p_{\emptyset}(s_t|s_{1:t-1},Z,C) \quad (1)$$

where φ denotes the parameters of the Transformer decoder. The encoder of TransMolVAE maps the input SMILES sequence into a latent representation Z using the self-attention mechanism. This mechanism enables the model to capture long-range dependencies across atoms that may be far apart in the sequence, allowing better modeling of chemical structures compared to GRU-based architectures.

The loss function of TransMolVAE follows the standard VAE objective:

$$L_T(\emptyset,\psi,S,Z,C) = E_{q\emptyset(Z|S)}[logp_{\emptyset}(S|Z,C)\} - D_{KL}(q_{\psi}(Z|S)||p_{\emptyset}(Z)) \quad (2)$$

where qψ(Z|S) is the posterior distribution parameterized by the Transformer encoder, and pφ(Z) is the prior distribution of the latent space.

IV. Experimental setup

A. Data Collection and Preprocessing

The dataset for this study was obtained from the Library of Integrated Network-Based Cellular Signatures (LINCS) database. It contains target perturbation profiles represented by gene expression data. These profiles were measured across 77 human cell lines in response to genetic perturbations, such as gene knockdown and gene overexpression. For evaluation, we used the same target proteins as in previous works (SmilesVAE [14], DRAGONET [17], GxVAE [15] and Gx2MOL [16]) with similar goals. The selected proteins include RAC-α serine/threonine-protein kinase (AKT1), RAC-β serine/threonine-protein kinase (AKT2), Aurora B kinase (AURKB), cysteine synthase A (CTSK), epidermal growth factor receptor (EGFR), histone deacetylase 1 (HDAC1), mammalian target of rapamycin (MTOR), phosphatidylinositol 3-kinase catalytic subunit (PIK3CA), decapentaplegic homolog 3 (SMAD3), and tumor protein (TP53). All molecules in the dataset were represented using the Simplified Molecular Input Line Entry System (SMILES).

Before model training, the SMILES data were cleaned and prepared to ensure its quality. Firstly, invalid and incomplete SMILES strings were removed using RDKit. Then, the duplicate molecules were deleted to maintain data uniqueness and reduce bias. The valid SMILES strings were then canonicalized to provide a consistent molecular representation. Next, character-level tokenization was applied, where each symbol in the SMILES string was treated as an individual token. This approach was chosen because it preserves the complete molecular syntax and simplifies the tokenization process [36]. The tokenized SMILES were then converted into numerical vectors suitable for input into the generative model. These steps ensured that the dataset was clean, standardized, and ready for training.

B. Training Details

In TransMolVAE, the encoder and decoder were implemented using Transformer layers with a hidden size of 256 and a latent dimensionality of 64. The learning rate and dropout probability were set to 1e−4 and 0.1 respectively. The maximum length of generated SMILES strings was determined by the dataset and fixed at 100. The batch size of the latent vectors was set to 128. TransMolVAE was trained for 300 epochs using the Adam optimizer. To monitor latent space utilization, the variance of latent dimensions and the number of active units were recorded at each epoch.

C. Evaluation Metrics

Several metrics were used to evaluate the performance of TransMolVAE, as shown in Table I. CPU time was measured to evaluate the computational cost of training and inference, where lower values indicate higher efficiency [37]. Reconstruction accuracy was used to assess how well the model could reconstruct SMILES sequences from the latent space, with low accuracy suggesting poor learning of molecular structures [4]. The total loss was calculated as the sum of reconstruction loss and KL divergence loss. Cross-entropy loss evaluated the accuracy of token prediction during sequence reconstruction, while KL divergence measured the distance between the approximate posterior and the prior distribution of the latent space. Active units were also examined to determine how many latent dimensions contributed useful information, since a low number of active units indicates posterior collapse [38].

For the biological evaluation, validity was defined as the proportion of chemically valid molecules, which was verified using RDKit. In addition, uniqueness referred to the fraction of non-repeated molecules among the valid set, where low uniqueness reflects mode collapse [10]. While, novelty measured the fraction of unique molecules that did not appear in the training set, with low novelty suggesting overfitting [39]. Together, these computational and biological measures provide a comprehensive assessment of the efficiency, accuracy, and chemical quality of the generated molecules.

TABLE I. LIST OF PERFORMANCE EVALUATION METRICS USED IN THIS STUDY

List of Performance Evaluation Metrics	
Computational Metrics: To evaluate the model performance.	***Biological Metrics: To evaluate the quality of the generated molecules/outputs.***
CPU Time Reconstruction Accuracy Total Loss Reconstruction Loss KL Loss Active Units	Validity Uniqueness Novelty

V. Results and Discussion

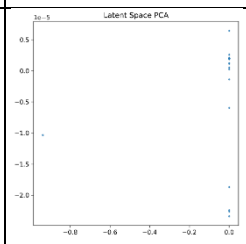
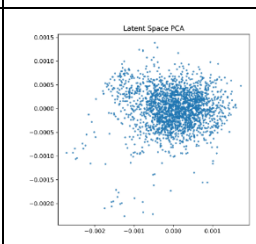
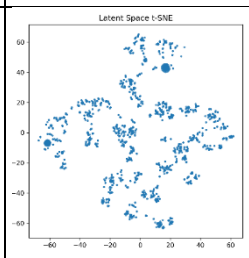
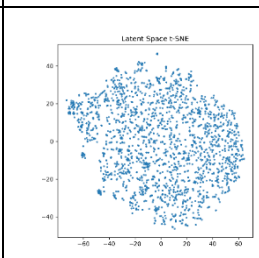
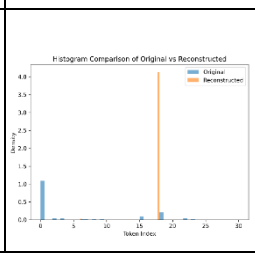
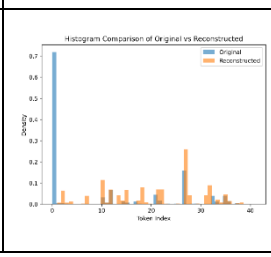
A. Latent Space Analysis

We compared the latent spaces learned by TransMolVAE and SmilesVAE using PCA and t-SNE visualizations, as shown in Table II to investigate the quality of the latent representations. In the PCA results, SmilesVAE showed a collapsed space where most points were aligned near a single axis with very little spread. This indicates that the model failed to use the latent space effectively, leading to limited molecular diversity. In contrast, TransMolVAE displayed a wider spread across both PCA axes, suggesting that its attention-based architecture captured richer and more informative chemical features. The ability of the Transformer encoder to model long-range dependencies in SMILES strings likely contributed to this better latent organization.

Similarly, t-SNE visualizations showed a tightly clustered pattern for SmilesVAE, reflecting poor latent separation and posterior collapse. TransMolVAE, however, produced a more continuous and structured latent space, indicating that the model successfully learned smooth transitions between chemical structures. It is a property essential for generating novel and valid molecules. The improved latent organization suggests that TransMolVAE generalizes better and can explore the chemical space more effectively.

The histogram comparison between original and reconstructed molecules further supports this finding. TransMolVAE achieved a strong overlap, showing that it retained the molecular structure and statistical distribution of the input data, while SmilesVAE diverged significantly. These differences imply that the Transformer's self-attention mechanism provides more stable learning dynamics and prevents latent collapse. Overall, the results demonstrate that TransMolVAE not only achieves better quantitative performance but also produces semantically meaningful latent representations, leading to higher-quality molecular generation.

TABLE II. EVALUATION OF LATENT SPACE AND RECONSTRUCTION DISTRIBUTION FOR SMILESVAE AND TRANSMOLVAE

Type of Graph	***SmilesVAE***	***TransMolVAE***
PCA		
t-SNE		
Histogram		

B. Evaluation of TransMolVAE and SmilesVAE

The comparison between TransMolVAE and SmilesVAE shows clear differences in efficiency and biological relevance, as shown in Table III and Table IV. TransMolVAE improved latent space exploration and reconstruction quality. The KL divergence stayed stable and did not collapse. The number of active units remained high. This shows that the Transformer-based encoder avoided posterior collapse. SmilesVAE, which is RNN-based, often suffers from this problem. In addition, token accuracy improved steadily, reaching 0.72 by the fifth epoch. SmilesVAE stayed much lower, at around 0.25. Besides, validity, uniqueness, and novelty also improved to more than 70% for TransMolVAE. This indicates that the latent space learned by the Transformer supports diverse and realistic molecule generation. These results confirm that the self-attention mechanism captures long-range patterns in SMILES better than RNNs.

TABLE III. EVALUATION METRICS FOR TRANSMOLVAE

Epoch	*CPU Time (hours)*	*Recon Accuracy*	*Total Loss*	*Recon Loss*	*KL Loss*	*Active Units*	*Validity*	*Uniqueness*	*Novelty*
1	72.40	0.41	0.2480	0.2168	0.0312	60	52.10	55.30	54.80
2	71.80	0.51	0.0036	0.0034	0.0007	58	60.40	61.10	59.70
3	71.30	0.61	0.0019	0.0016	0.0011	57	66.80	67.50	65.20
4	70.90	0.69	0.0013	0.0011	0.0009	55	71.20	72.80	70.60
5	70.60	0.72	0.0010	0.0008	0.0008	54	74.50	75.00	73.80

TABLE IV. EVALUATION METRICS FOR SMILESVAE

Epoch	*CPU Time (hours)*	*Recon Accuracy*	*Total Loss*	*Recon Loss*	*KL Loss*	*Active Units*	*Validity*	*Uniqueness*	*Novelty*
1	64.22	0.23	0.3051	0.2972	0.0079	8	50.2	52.3	51.7
2	63.90	0.25	0.0055	0.0053	0.0002	2	51.1	53.0	52.2
3	63.87	0.25	0.0048	0.0047	0.0001	1	52.4	54.0	53.1
4	63.71	0.25	0.0046	0.0045	0.0001	0	54.0	55.1	54.4
5	63.68	0.25	0.0045	0.0044	0.0001	0	55.3	56.0	55.0

However, the improvement comes with higher computational cost. CPU time per epoch stayed stable for both models. TransMolVAE required much longer time than SmilesVAE. This reflects the higher complexity of Transformer models. The quadratic attention mechanism makes them more powerful but also slower. In settings with limited resources, this higher demand may reduce scalability. Still, the quality of generated molecules is better with the Transformer.

Additionally, SmilesVAE showed clear signs of posterior collapse. The KL divergence dropped close to zero. The number of active units reached zero after the fourth epoch. This means the latent space was not used well. Besides, token accuracy stayed very low, around 0.25. While, validity, uniqueness, and novelty were only between 50% and 55%. These results show that RNN-based VAEs are not strong enough for molecular generation. They are efficient in computation but weaker in outcome. Overall, there is a trade-off between performance and efficiency. TransMolVAE improves reconstruction, biological quality, and latent space stability. But it takes longer to train because of higher computation. SmilesVAE is faster but produces weak and less diverse molecules.

These results highlight the importance of balancing computational efficiency and molecular quality in generative modeling, as discussed earlier in the Section 1. Drug discovery needs faster approaches, especially for diseases like cancer that require immediate treatment [4]. Computational methods help speed up this process compared to traditional wet-lab experiments [40]. However, speed alone is not enough because the generated molecules must also be high in quality to ensure their safety and effectiveness. Therefore, even though computational models improve efficiency, their results should still be judged by the quality of the generated compounds. This is also why exploring the unexplored chemical or latent space is important, as it allows the discovery of novel and more effective molecules [41,42]). In this case, TransMolVAE is a better choice as it produces higher-quality molecular representations, even though it needs more computational resources.

VI. CONCLUSION

In this work, we introduced TransMolVAE, a Transformer-based VAE for molecular generation using SMILES. Unlike RNN-based models such as SmilesVAE, TransMolVAE uses self-attention to learn long-range patterns in SMILES strings. This helps the model achieve better reconstruction accuracy, stable KL divergence, and more effective use of the latent space. The results showed that TransMolVAE achieved higher token accuracy than SmilesVAE. The model also produced molecules with higher validity, uniqueness, and novelty. These results prove that TransMolVAE can generate more diverse and valid molecules compared to the baseline model. However, the model required longer CPU time because Transformer models are more complex and need higher computation. This shows a trade-off between performance and computational cost. Future work can focus on reducing this cost by using GPU acceleration, model pruning, or mixed-precision training. Another possible improvement is to replace SMILES with SELFIES (Self-Referencing Embedded Strings), an upgraded version of SMILES that automatically removes invalid sequences. Since SMILES can still contain invalid molecular strings, SELFIES offers a more robust molecular representation and ensures that all generated sequences correspond to valid chemical structures. Thus,

making the generation process more reliable. Besides improving computation, TransMolVAE can also be integrated into the drug discovery pipeline, specifically during the research and development (R&D) phase. The R&D phase is the first and most time-consuming stage of drug discovery. By generating high-quality and diverse molecules early in this stage, TransMolVAE can help shorten the initial screening and compound design process, accelerating the overall drug discovery timeline. Overall, TransMolVAE shows strong ability in generating diverse and valid molecules and can serve as a useful model for future AI-based drug discovery research.

ACKNOWLEDGMENT

This research was financially supported by the Sultan Mizan Ant arctic Research Foundation-Yayasan Penyelidikan Antartika Sultan Mizan (YPASM) under grant number R/YPASM/B1500/01850A/003/ 2024/01332, and the France-Malaysia Collaboration Programme for Joint Research (MyTIGER 2023).

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] Berdigaliyev, N., & Aljofan, M. (2020). An overview of drug discovery and development. *Future Medicinal Chemistry, 12*(10), 939–947.

[2] Ocana, A., Pandiella, A., Privat, C., Bravo, I., Luengo-Oroz, M., Amir, E., & Gyorffy, B. (2025). Integrating artificial intelligence in drug discovery and early drug development: A transformative approach. *Biomarker Research, 13*(1).

[3] Bikku, T., Malligunta, K. K., Thota, S., & Surapaneni, P. P. (2025). Improved quantum algorithm: A crucial stepping stone in quantum-powered drug discovery. *Journal of Electronic Materials*.

[4] Jusoh, A. S., Remli, M. A., Mohamad, M. S., Cazenave, T., & Fong, C. S. (2025). How generative artificial intelligence can transform drug discovery? *European Journal of Medicinal Chemistry, 295*, 117825.

[5] Pracht, P., Bohle, F., & Grimme, S. (2020). Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Physical Chemistry Chemical Physics, 22*(14), 7169–7192.

[6] Degenhardt, Y., Poss, M., & Gao, X. (2025). Computer-aided design for cancer-targeted peptide drugs. In *Methods in Pharmacology and Toxicology* (pp. 45–62).

[7] Niazi, S. K., & Mariam, Z. (2025). Artificial intelligence in drug development: Reshaping the therapeutic landscape. *Therapeutic Advances in Drug Safety, 16*.

[8] Hagen, J., Hornung, L., Barham, W., Mukhopadhyay, S., Bess, A., Contrera, K., Basu, D., Sandulache, V., Spielmann, G., & Kansara, S. (2025). Artificial intelligence in head and neck cancer: Towards precision medicine. *Cancers, 17*(18), 3023.

[9] Suveena, S., Rekha, A. A., Rani, J. R., Oommen, O. V., & Ramakrishnan, R. (2025). The translational impact of bioinformatics on traditional wet lab techniques. *Advances in Pharmacology*.

[10] Vogt, M. (2023). Exploring chemical space—Generative models and their evaluation. *Artificial Intelligence in the Life Sciences, 3*, 100064.

[11] Bilodeau, C., Jin, W., Jaakkola, T., Barzilay, R., & Jensen, K. F. (2022). Generative models for molecular discovery: Recent advances and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 12*(5).

[12] Iqbal, T., & Qureshi, S. (2022). The survey: Text generation models in deep learning. *Journal of King Saud University – Computer and Information Sciences, 34*(6), 2515–2528.

[13] Ochiai, T., Inukai, T., Akiyama, M., Furui, K., Ohue, M., Matsumori, N., Inuki, S., Uesugi, M., Sunazuka, T., Kikuchi, K., Kakeya, H., & Sakakibara, Y. (2023). Variational autoencoder-based chemical latent space for large molecular structures with 3D complexity. *Communications Chemistry, 6*(1).

[14] Alperstein, Z., Cherkasov, A., & Rolfe, J. T. (2023). All SMILES variational autoencoder for molecular property prediction and optimization. In *Challenges and Advances in Computational Chemistry and Physics* (pp. 85–115).

[15] Li, C., & Yamanishi, Y. (2025). AI-driven transcriptome profile-guided hit molecule generation. *Artificial Intelligence*, 104239.

[16] Li, C., & Yamanishi, Y. (2026). Gx2Mol: De novo generation of hit-like molecules from gene expression profiles (pp. 333–349).

[17] Yamanaka, C., Uki, S., Kaitoh, K., Iwata, M., & Yamanishi, Y. (2023). De novo drug design based on patient gene expression profiles via deep learning. *Molecular Informatics, 42*(8–9).

[18] Lipton, Z. C., Berkowitz, J., & Elkan, C. (2015). A critical review of recurrent neural networks for sequence learning. *arXiv*.

[19] Yazdani, S., Singh, A., Saxena, N., Wang, Z., Palikhe, A., Pan, D., Pal, U., Yang, J., & Zhang, W. (2025). Generative AI in depth: A survey of recent advances, model variants, and real-world applications. *Journal of Big Data, 12*(1).

[20] Gangwal, A., & Lavecchia, A. (2025). Artificial intelligence in natural product drug discovery: Current applications and future perspectives. *Journal of Medicinal Chemistry*.

[21] Tiezzi, M., Casoni, M., Betti, A., Guidi, T., Gori, M., & Melacci, S. (2025). Back to recurrent processing at the crossroad of transformers and state-space models. *Nature Machine Intelligence, 7*(5), 678–688.

[22] Jiang, J., Ke, L., Chen, L., Dou, B., Zhu, Y., Liu, J., Zhang, B., Zhou, T., & Wei, G. (2024). Transformer technology in molecular science. *Wiley Interdisciplinary Reviews: Computational Molecular Science, 14*(4).

[23] Sultan, A., Sieg, J., Mathea, M., & Volkamer, A. (2024). Transformers for molecular property prediction: Lessons learned from the past five years. *Journal of Chemical Information and Modeling, 64*(16), 6259–6280.

[24] Qiu, X., Shao, S., Wang, H., & Tan, X. (2025). Bio-K-Transformer: A pre-trained transformer-based sequence-to-sequence model for adverse drug reactions prediction. *Computer Methods and Programs in Biomedicine, 260*, 108524.

[25] Jørgensen, P. B., Schmidt, M. N., & Winther, O. (2018). Deep generative models for molecular science. *Molecular Informatics, 37*(1–2).

[26] Khater, T., Alkhatib, S. A., AlShehhi, A., Pitsalidis, C., Pappa, A. M., Ngo, S. T., Chan, V., & Truong, V. K. (2025). Generative artificial intelligence-based model optimization

towards molecule design enhancement. *Journal of Cheminformatics, 17*(1), 116.

[27] Elton, D. C., Boukouvalas, Z., Fuge, M. D., & Chung, P. W. (2019). Deep learning for molecular design: A review of the state of the art. *Molecular Systems Design & Engineering, 4*(4), 828–849.

[28] Özçelik, R., Brinkmann, H., Criscuolo, E., & Grisoni, F. (2025). Generative deep learning for de novo drug design—A chemical space odyssey. *Journal of Chemical Information and Modeling*.

[29] Zhang, S., Fan, R., Liu, Y., Chen, S., Liu, Q., & Zeng, W. (2023). Applications of transformer-based language models in bioinformatics: A survey. *Bioinformatics Advances, 3*(1).

[30] Khambhawala, A., Lee, C. H., Pahari, S., & Kwon, J. S.-I. (2025). Minimizing late-stage failure in drug development with transformer models: Enhancing drug screening and pharmacokinetic predictions. *Chemical Engineering Journal, 508*, 160423.

[31] Ünlü, A., Çevrim, E., Yiğit, M. G., Sarıgün, A., Çelikbilek, H., Bayram, O., Kahraman, D. C., Olğaç, A., Rifaioglu, A. S., Banoğlu, E., & Doğan, T. (2025). Target-specific de novo design of drug candidate molecules with graph-transformer-based generative adversarial networks. *Nature Machine Intelligence, 7*(9), 1524–1540.

[32] Sagawa, T., & Kojima, R. (2025). ReactionT5: A pre-trained transformer model for accurate chemical reaction prediction with limited data. *Journal of Cheminformatics, 17*(1), 126.

[33] Sadaghiyanfam, S., Kamberaj, H., & Isler, Y. (2025). Leveraging ChemBERTa and machine learning for accurate toxicity prediction of ionic liquids. *Journal of the Taiwan Institute of Chemical Engineers, 171*, 106030.

[34] Masood, M. A., Kaski, S., Ceulemans, H., Herman, D., & Heinonen, M. (2024). Balancing imbalanced toxicity models: Using MolBERT with focal loss. In *Lecture Notes in Computer Science* (pp. 82–97).

[35] Bagal, V., Aggarwal, R., Vinod, P. K., & Priyakumar, U. D. (2022). MOLGPT: Molecular generation using a transformer-decoder model. *Journal of Chemical Information and Modeling, 62*(9), 2064–2076.

[36] Liyaqat, T., Ahmad, T., & Saxena, C. (2025). Advancements in molecular property prediction: A survey of single and multimodal approaches. *Archives of Computational Methods in Engineering*.

[37] Li, D., Chen, X., Becchi, M., & Zong, Z. (2016). Evaluating the energy efficiency of deep convolutional neural networks on CPUs and GPUs. In *IEEE International Conference on Big Data and Cloud Computing (BdCloud)*.

[38] Song, T., Huang, Z., Liu, X., & Sun, J. (2025). Preventing posterior collapse with DVAE for text modeling. *Entropy, 27*(4), 423.

[39] Bhadwal, A. S., Kumari, M., & Kumar, A. (2025). PCF-VAE: Posterior collapse free variational autoencoder for de novo drug design. *Scientific Reports, 15*(1).

[40] Bakhtiary, N., Ghalandari, B., Ghorbani, F., & Liu, C. (2025). Dry-lab computational approaches for simulating bioinspired materials/structures for regenerative medicine. In *Principles of Bioinspired and Biomimetic Regenerative Medicine* (pp. 861–899).

[41] Chakraborty, R., & Hasija, Y. (2023). Utilizing deep learning to explore chemical space for drug lead optimization. *Expert Systems with Applications, 229*, 120592.

[42] Vogt, M. (2022). Using deep neural networks to explore chemical space. *Expert Opinion on Drug Discovery, 17*(3), 297–304.