Quantum Machine Learning in De Novo Drug Discovery

Vinamr Jain

Prof. Abhishek Dixit
Prof. Kedar Bhalchandra Khare

IIT Delhi

November, 2023

Challenges and Motivation

1. **Complex and Resource-Intensive Process**
   Low probability of success, with only approximately 4% of preclinical drugs eventually granted licenses.
2. **Lengthy Journey to Market** New medicine development takes a minimum of ten years to go from initial discovery to market availability.
3. **High Research and Development Costs** Estimated median capitalized R&D cost per new drug (2009-2018): $985 million, accounting for the cost of failures.
4. **Computational Challenges** Huge Hilbert Space for drug molecules due to numerous atoms. Approximately 10^{47} protein structures predicted from just 100 amino acids. Searching this vast Hilbert Space is an NP-Hard problem for traditional computers.

Promise of Quantum Advantage

1. Several studies have demonstrated that variational quantum circuit (VQC) performs the advantages in expression power, learnability, and robustness.
2. QuGAN's exponential advantages over classical GANs directly result from the ability of quantum information processors to represent N-dimensional features using log N qubits with time complexity of O(poly(log N)).
3. Recent studies showed that generative models implemented by quantum circuits with fewer architectural complexities could easily bypass their classical counterparts [1][2][4].
4. Quantum GANs can offer several opportunities e.g., (i) quantum speedup in the runtime making it possible to learn richer representation of molecules via deeper models due to the amplitude amplification property; (ii) ability to search exponentially large chemical space with few qubits and sample from distributions that may be difficult to model classically.

Classical Approach - MolGAN

1. MolGAN's a GAN for graph structure molecules trained on the QM9 database and demonstrated to generate close to 100% valid molecules
2. Generative learning with graph-structured molecules is invariant to the orderings of atoms and automates the navigation to a chemical region abundant in desired molecules

Value function-

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}}[logD(x)] + E_{z\sim p_z}[log(1-D(G(z))]$$

$$L(\omega) = \alpha \cdot L_{\text{WGAN}}(\omega) + (1-\alpha) \cdot L_{\text{GAN RL}}(\omega)$$

Classical Approach - MolGAN

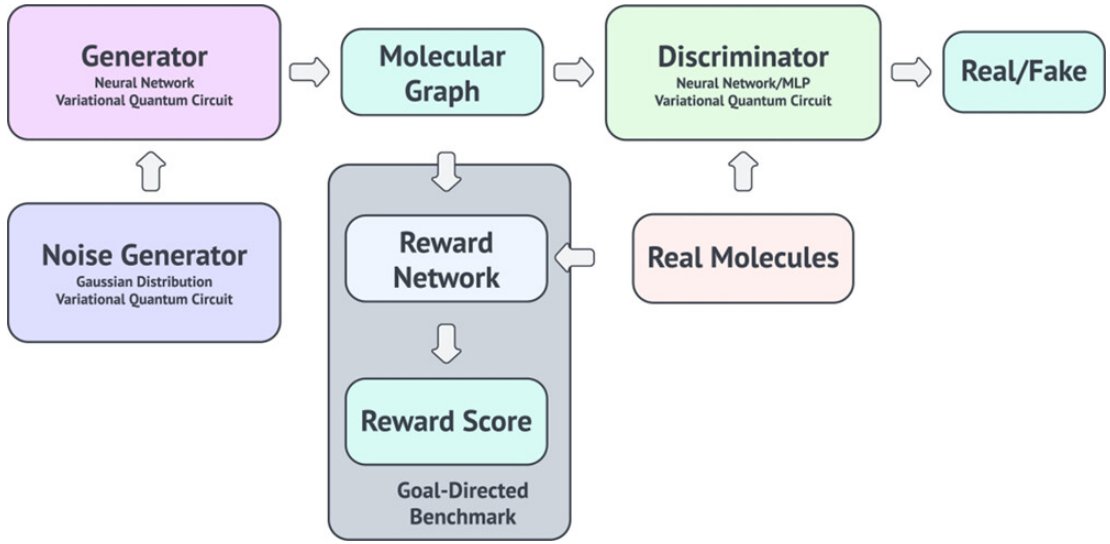


Figure: The overall pipeline of MolGAN. The reward neural network branch is enabled in the goal- directed benchmark. The classical noise generator samples from the Gaussian distribution. The classical generator is built by neural networks to generate the molecular graph. The molecular graph is represented by a bond matrix and atom vector. The classical discriminator is built by a graph-based neural network[4]

QGAN-HG

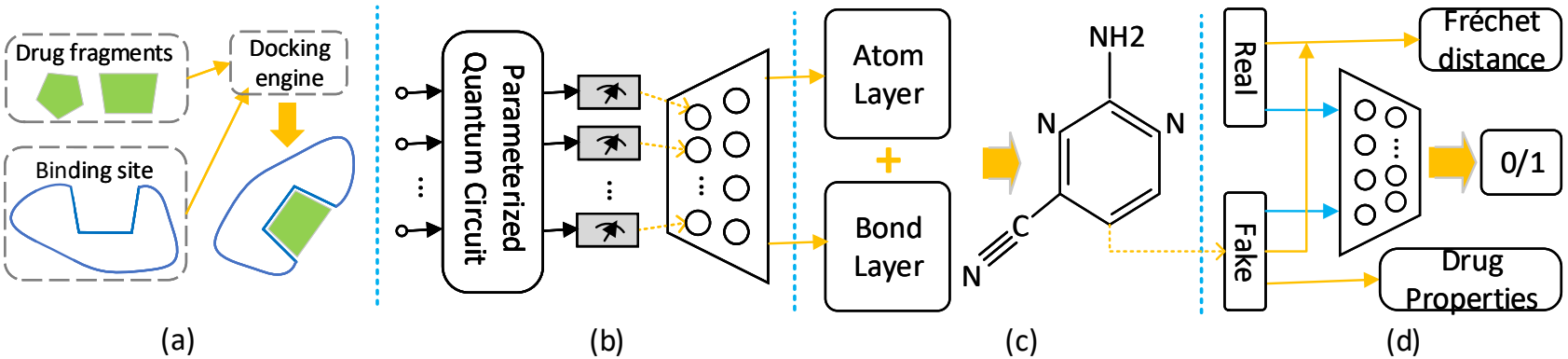


Figure: (a) Only generated molecules that have high affinity towards the receptor binding sites are considered as valid; (b) quantum stage (which is a parameterized quantum circuit with last-layer N measuring the expectation values) and classical stage (neural network with last-layer out-feature dimension of 512) separated by blue dotted line; (c) application of atom layer and bond layer for generating synthetic molecular graphs (one example synthetic molecule is given); (d) a batch of real molecules from training dataset (QM9 in this case) and a batch of synthetic molecules generated from (c) are fed into classical discriminator for real/synthetic prediction and FD score calculation, and drug properties for synthetic molecules are evaluated using RDKit package. The prediction losses from discriminator are back-forwarded to two neural networks as well as quantum circuit for updating all parameters simultaneously in each training epoch [2].

QGAN-HG

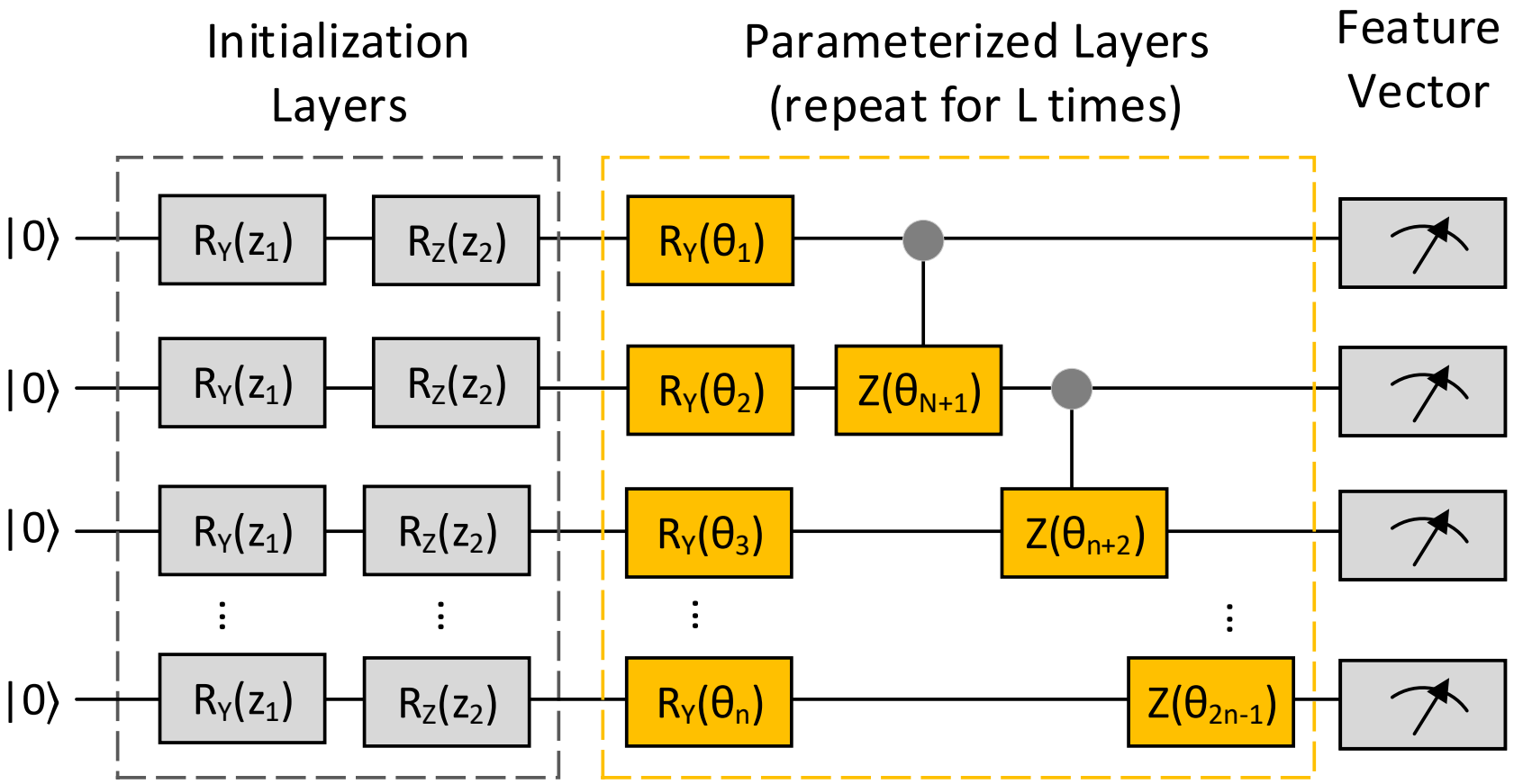
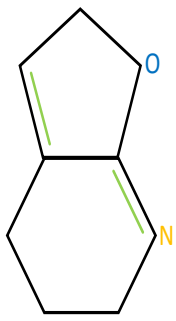


Figure: Parameterized quantum circuit to obtain feature vector of N dimensions. The circuit is composed of initialization layers, repeatable parameterized layers and measurement layer. Two CNOT gates for each ZZ interaction for creating entanglement are not shown here[2]

QGAN-HG



$$B = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 2 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \end{bmatrix}$$

$$A = [1 \quad 3 \quad 1 \quad 2 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1]$$

$$R_y(\theta) = \begin{bmatrix} cos\left(\frac{\theta}{2}\right) & -sin\left(\frac{\theta}{2}\right) \\ sin\left(\frac{\theta}{2}\right) & cos\left(\frac{\theta}{2}\right) \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} e^{-i\frac{\theta}{2}} & 0 \\ 0 & e^{i\frac{\theta}{2}} \end{bmatrix}$$

$$\text{CNOT} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

(a) (b) (c)

Figure: (a-b) A sample molecular graph from QM9 denoted by its corresponding atom vector A and bond matrix B; (c) Example Quantum Gates

Quantum Noise, Generator and Discriminator

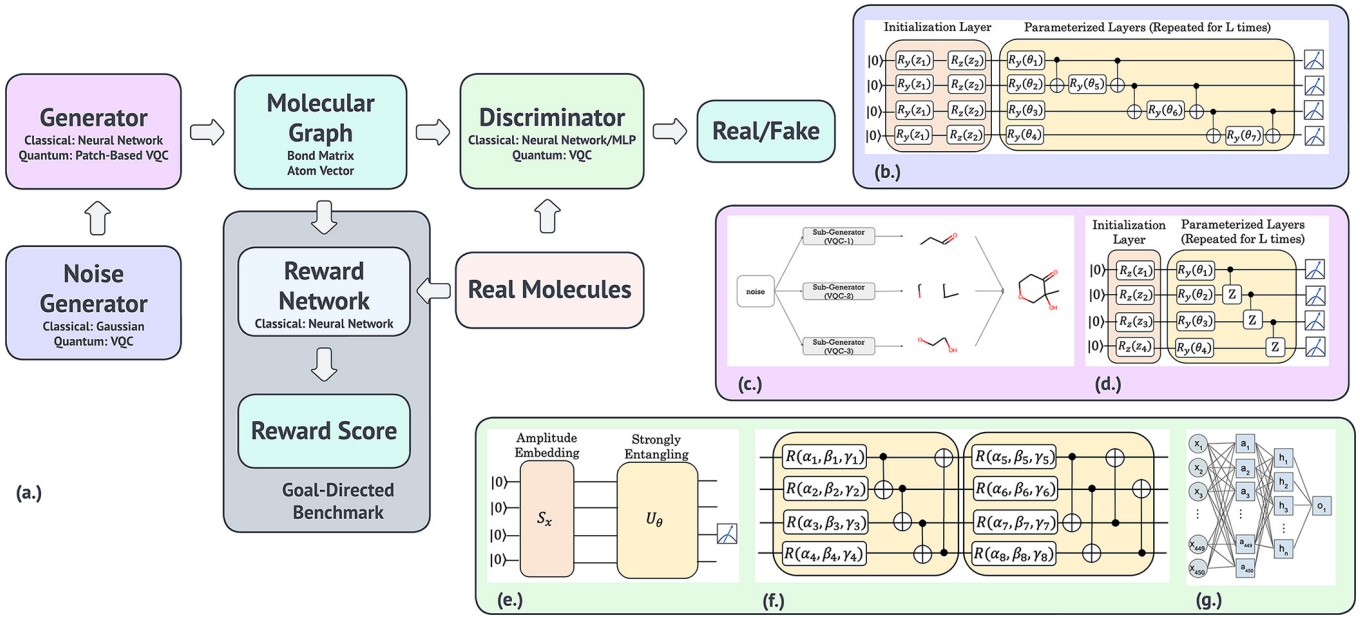


Figure: Overall pipeline[4]. (a) The overall pipeline of MolGAN with different combinations of classical/quantum components. The reward neural network branch is enabled in the goal-directed benchmark. The classical noise generator samples from the Gaussian distribution, and the quantum one uses the variational quantum circuit (VQC). The classical generator is built by neural networks, and the quantum one uses the patch-based VQC to generate the molecular graph. The molecular graph is represented by a bond matrix and atom vector. The classical discriminator is built by a graph-based neural network or multilayer perceptron (MLP), and the VQC is used in the quantum one. (b) The example of VQC in the noise generator. (c) The patch method uses multiple VQCs as subgenerators. Each subgenerator takes noise as input and outputs a partial part of the final molecular graph. The final molecular graph is constructed by concatenating all the partial patches together. (d) The example of VQC in the quantum generator. (e) The VQC of the quantum discriminator consists of the amplitude embedding circuit (Sx), the strong entanglement layers (U_θ), and the measurement. (f) The VQC of strongly entanglement layers contains multiple CNOT gates and parametrized rotational gates (R). (g) MLP-based discriminator architecture in MolGAN-CC.

Quantum Noise, Generator and Discriminator

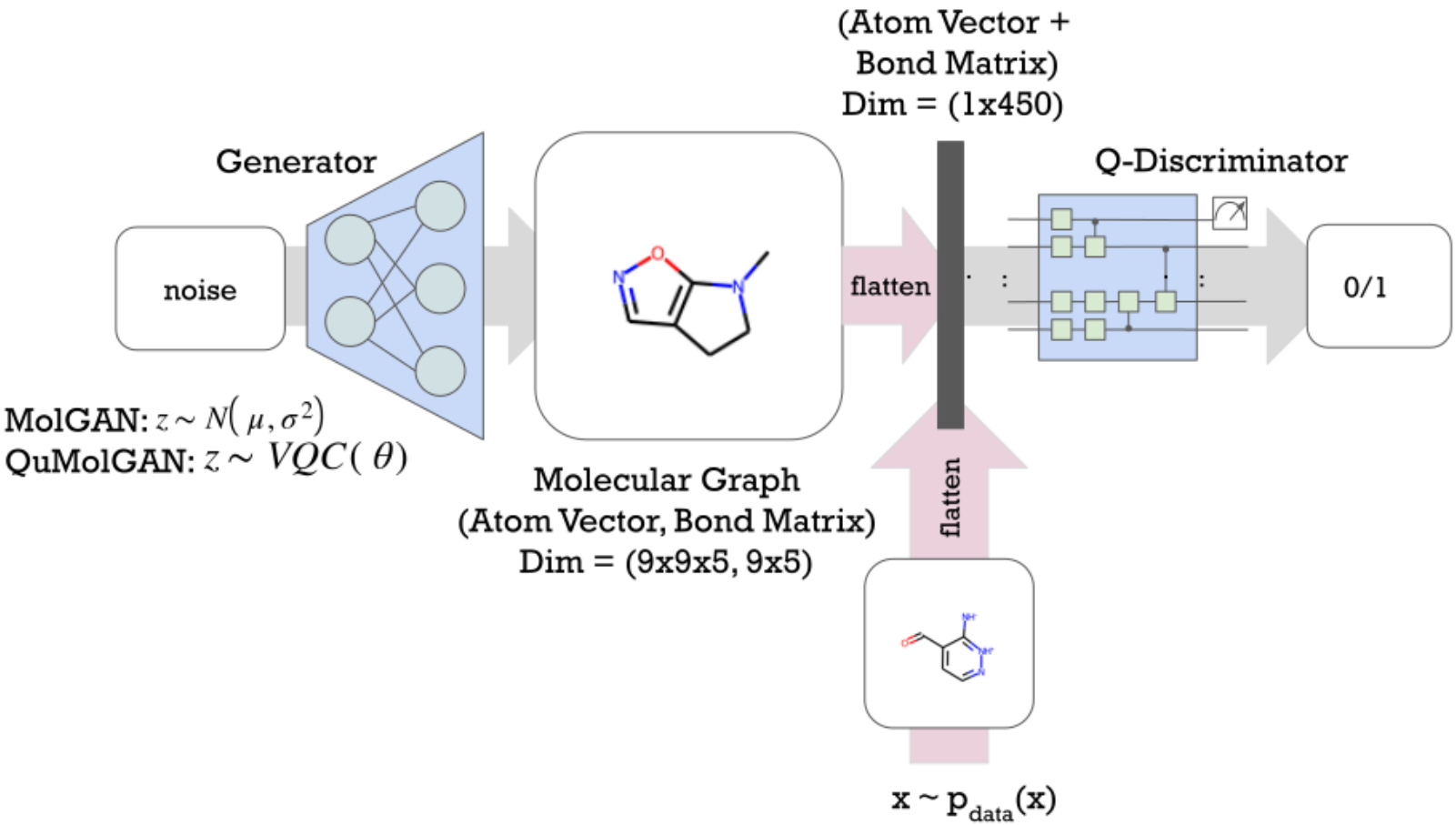


Figure: MolGAN Quantum Discriminator Schema[4]

Results- QuGAN-HG

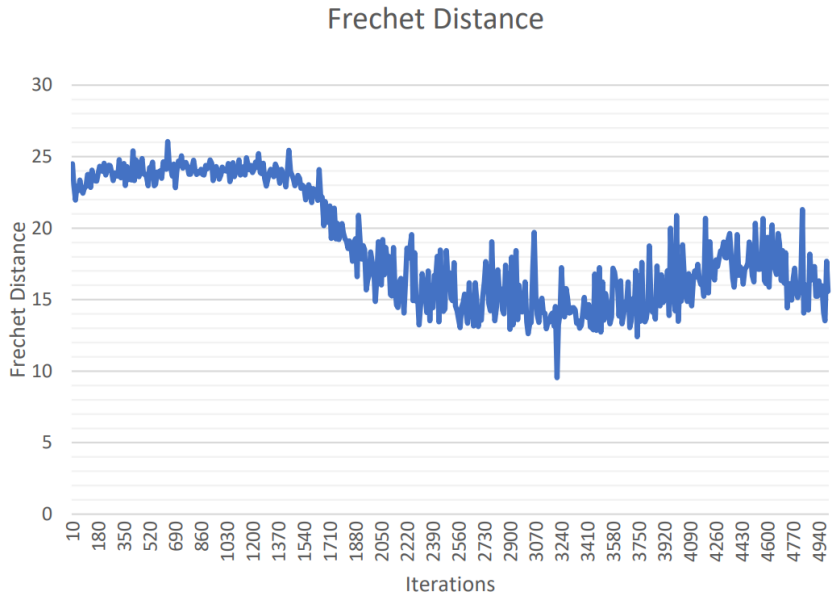


Figure: Frechet Distance Bonds and Atoms (Reproduced) for QGAN-HG

REPRODUCED DRUG PROPERTIES-

Diversity	Drugcandidate	logP	Novel	NP	QED	SA	Unique	Valid
1.00	0.4593	0.7101	1.00	0.7538	0.5081	0.2263	0.7857	0.875

Figure: Reprodued Drug property results for QGAN-HG

Results- QuMolGAN-HR

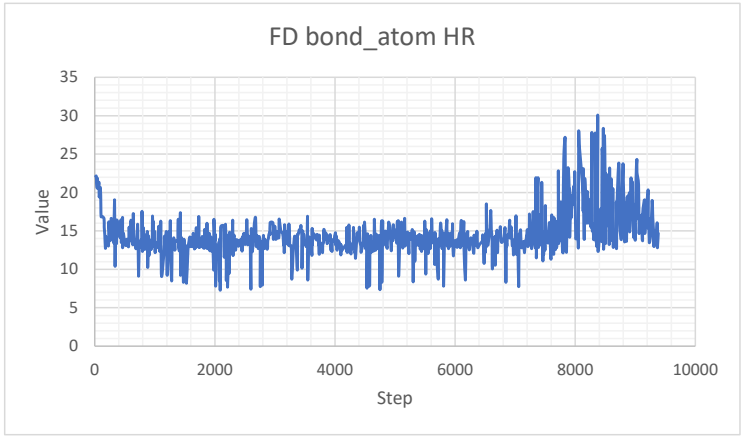


Figure: Frechet Distance Bonds and Atoms (Reproduced) for QuMolGAN-HR

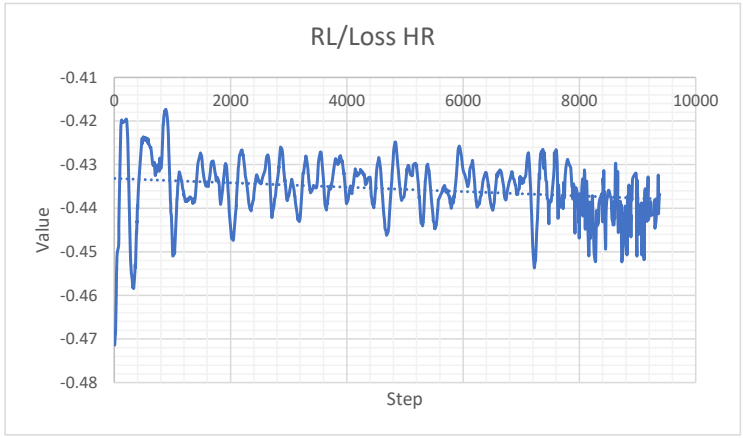


Figure: Reinforcement Loss (Reproduced) for QuMolGAN-HR

REPRODUCED DRUG SCORES HR-

Valid	Unique	Novel	NP	QED	Solute	SA	Diverse	Drug candidate
0.8472	0.1861	1.00	0.95	0.54	0.36	0.42	0.65	0.53

Figure: Reprodued Drug property results for QuMolGAN-HR

Results- QuMolGAN-NR

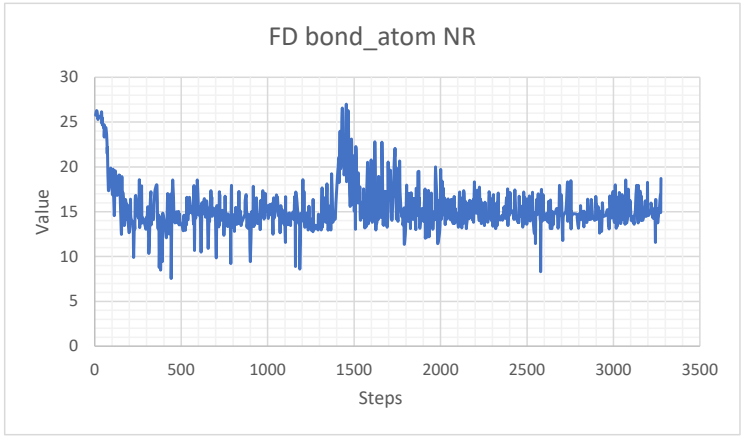


Figure: Frechet Distance Bonds and Atoms (Reproduced) for QuMolGAN-NR

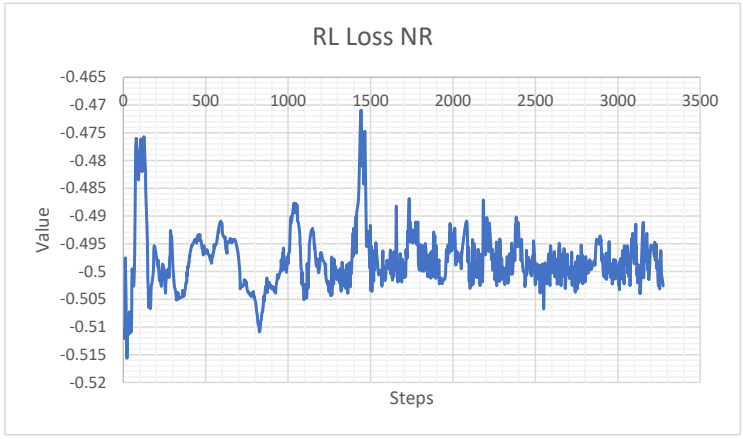


Figure: Reinforcement Loss (Reproduced) for QuMolGAN-NR

REPRODUCED DRUG SCORES NR (0 epochs)-

Valid	Unique	Novel	NP	QED	Solute	SA	Diverse	Drug candidate
0.5842	0.2319	nan	nan	nan	nan	nan	nan	0.40

Figure: Reprodued Drug property results for QuMolGAN-NR

Sample Generated Molecules

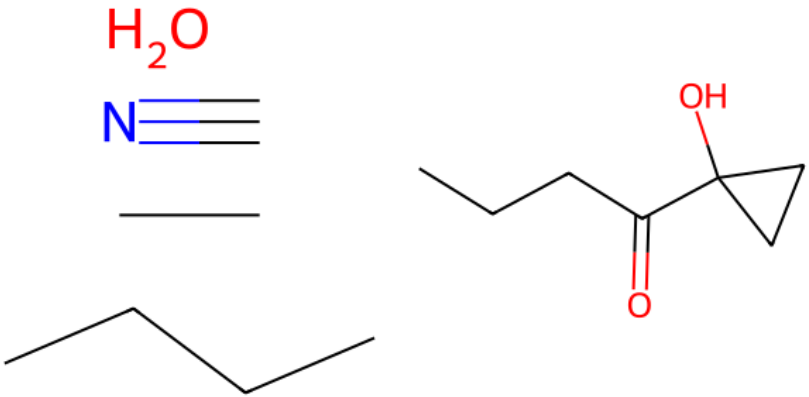


Figure: Sample generated molecules in QuMolGAN-HR

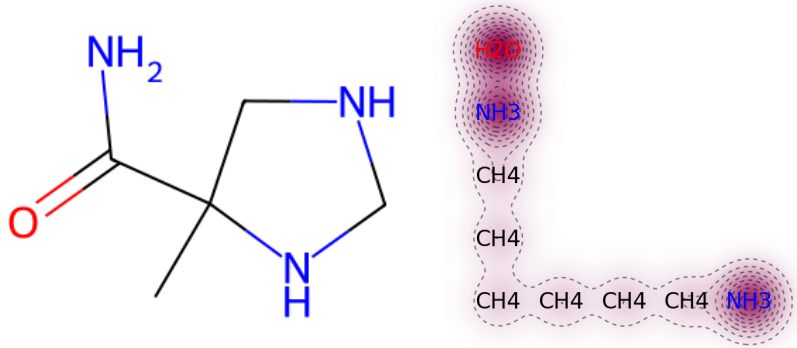


Figure: Sample generated molecule in QuMolGAN-NR and QGAN-HG respectively

Conclusion

1. QGAN-HG outperforms classical MolGAN in both drug properties and Frechet distance, even with a significant reduction in the number of parameters used.
2. QuMolGAN-HR outperforms even QGAN-HG in drug property scores like Drug candidate, QED, SA and NP scores, despite it being trained only for a single epoch.
3. However, the training processing is resource-consuming and time-consuming, even on state of the art classical computers.

Future Scope

1. Quantum Discriminator Model (MolGAN-CQ)
   - 1.1 Demonstrate Quantum Advantage
   - 1.2 Challenge: Implementing Problem-Specific Ansatz
   - 1.3 Explore Integration with Hybrid Quantum-Classical Generative Models
   - 1.4 Investigate Group Theoretic Structure in Drug Molecules (Inspiration from Glick et al. [5])
   - 1.5 Difficulty: Identifying Symmetry in Drug Molecules
2. Classical Generative Methods
   - 2.1 Explore Autoencoders
   - 2.2 Investigate Quantum Advantage in Classical Generative Methods
   - 2.3 Reference: Li & Ghosh (2022) [6]

References

[1] Pyrkov, A., Aliper, A., Bezrukov, D., Lin, Y., Polykovskiy, D., Kamya, P., Ren, F., & Zhavoronkov, A. (2023). Quantum computing for near-term applications in generative chemistry and drug discovery.

[2] J. Li, R. O. Topaloglu and S. Ghosh, "Quantum Generative Models for Small Molecule Drug Discovery,"

[3] https://pennylane.ai/qml/demos/tutorial_quantum_gans

[4] Exploring the Advantages of Quantum Generative Adversarial Networks in Generative Chemistry Po-Yu Kao, Ya-Chu Yang, Wei-Yin Chiang, Jen-Yueh Hsiao, Yudong Cao, Alex Aliper, Feng Ren, Alán Aspuru-Guzik, Alex Zhavoronkov, Min-Hsiu Hsieh, and Yen-Chu Lin.

[5] Glick, J. R., Gujarati, T. P., Corcoles, A. D., Kim, Y., Kandala, A., Gambetta, J. M., & Temme, K. (2021). Covariant quantum kernels for data with group structure.

[6] Li, J. and Ghosh, S., 2022, March. Scalable variational quantum circuits for autoencoder-based drug discovery.

[7] De Cao, N., & Kipf, T. (2018). MolGAN: An implicit generative model for small molecular graphs.