Scientific Journal of Informatics
Vol. 13, No. 3, August 2026
p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

A Hybrid Quantum-Classical Workflow for Early-Stage Molecular Discovery: A Workflow-Level Proof-of-Concept

Charnelle Razo[1], Belinda Ndlovu[2]*
[1,2]Department of Informatics and Analytics, National University of Science and Technology, Zimbabwe

Abstract.
Purpose: Classical computational approaches struggle to model molecular quantum interactions, making drug discovery costly and time-consuming. Quantum computing shows promise in molecular modelling, but much of the existing work focuses on isolated proof-of-concept tasks. These studies usually do not examine the full molecular discovery process. This study demonstrates the feasibility of a unified hybrid quantum–classical workflow for early-stage molecular discovery and examines how quantum-based molecular representations influence later optimisation and prediction.
Methods/Study design/approach: The study developed and evaluated a five-stage hybrid quantum-classical workflow integrating query interpretation, molecular screening with quantum re-ranking, VQE-based energy estimation, property prediction and report generation. The system was implemented using Qiskit, Qiskit Nature, RDKit, and spaCy on the QM9 dataset using noiseless state-vector simulation.
Result/Findings: The NLP component was evaluated across four tiers: 180 template-based queries, 160 independently written queries, 1,000 independently written queries, and 1,000 externally sourced BioASQ Task B questions. F1-scores were 89.9%, 82.1%, 73.0%, and 71.3%. Screening achieved constraint satisfaction rates of 90%-100%. Quantum-informed ranking differed substantially from cosine similarity and evaluated the utility of this through top-k enrichment and multi-property hit rates. MIFM achieved a mean VQE absolute error of 0.0156 Ha compared with 0.0218 Ha for ZZFeatureMap and 0.0248 Ha for cosine similarity. VQE and property-prediction evaluations, scalability experiments, and simulated NISQ tests provided broader evidence of workflow behaviour.
Novelty/Originality/Value: This study presents a reproducible hybrid quantum–classical workflow for early-stage molecular discovery, extended with a chemistry-aware MolecularInteractionFeatureMap (MIFM). The study does not claim universal quantum advantage or physical quantum-hardware validation.

Keywords: Hybrid Quantum-classical Systems, Molecular Screening, VQE Optimisation, Quantum Feature Encoding, Workflow integration, MolecularInteractionFeatureMap (MIFM)
Received May 2026/ **Revised** August 2026 / **Accepted** August 2026

INTRODUCTION

Drug discovery is a task that is complex, time-consuming, costly and has high failure rates [1]. Developing a single approved therapeutic typically takes around 10 to 15 years, with costs ranging from one to three billion US dollars [2]. Only 10% to 20% of compounds entering clinical trials make it to approval [3]. To address these challenges, classical computational approaches such as molecular dynamics simulations and machine learning have been used to speed up early-stage screening and improve property prediction [4], [5]. This has made the process faster and more targeted, but classical methods have limitations, such as relying on approximations that cannot fully capture molecular interactions [6]. They also struggle to model the electronic interactions involved in reactivity, stability, and binding affinity, and they are less reliable as molecules increase in complexity [7], [8]. These limitations result in higher developmental costs, longer timelines and higher failure rates at late-stage development because inaccuracies introduced during early-stage molecular screening and optimisation pass down throughout the drug discovery process [9], motivating computational methods that represent molecular interactions more accurately, with quantum computing identified as one such solution [10].

Quantum computing offers a fundamentally different approach for molecular modelling. By utilising the properties of superposition and entanglement, quantum processors can represent molecular Hamiltonians directly in Hilbert space, making them suitable for molecular modelling tasks [11], [12]. This makes it

*Corresponding author.
Email addresses: n022177311@students.nust.ac.zw (Razo), belinda.ndlovu@nust.ac.zw (Ndlovu)
DOI: https://doi.org/10.15294/sji.v13i3.50262

possible to compute properties such as ground-state energies, electron correlation, and molecular similarity using potentially fewer resources than classical methods [12]. In recent years, several hybrid quantum-classical approaches have been studied in molecular discovery and design. For instance, in [12], they showed that hybrid quantum machine learning could be used to predict binding affinity and reported improvements in prediction speed, but the evaluation was limited to a single pipeline stage. [13] showed that quantum-enhanced feature representations can improve drug–target interaction prediction on benchmark datasets, but their analysis focused only on prediction performance. Similarly, the Variational Quantum Eigensolver has been applied in molecular energy estimation within NISQ constraints as a standalone task without downstream pipeline integrations [14]. [15] explored small molecule generation in drug discovery using variational quantum circuits, but this was also a single subtask. Reviews such as [16] summarise techniques, including VQE and perturbation theory, but do not consider full pipeline integration. [16] designed a hybrid pipeline for real-world drug discovery using VQE-based QM/MM simulations and demonstrated functional feasibility for specific drug-target cases, but lacked integration with upstream screening and ranking stages.

All these studies follow a similar pattern, in which quantum methods are usually presented as isolated subtasks rather than as part of a connected workflow. This fragmentation creates a methodological gap because in practical molecular discovery workflows, the stages are dependent on each other. Screening affects optimisation difficulty, optimisation influences property prediction, and representation quality carries through the entire process. Understanding this interaction requires looking at system-level integration rather than standalone performance. To the best of our knowledge, this is one of the early controlled demonstrations of the pipeline integration of quantum-enhanced techniques across the four key stages: query interpretation and molecular constraint extraction, molecular screening, candidate optimisation, and property prediction within a single executable computational framework. This gap is not due to a lack of interest but is largely shaped by current hardware limitations. Noisy Intermediate-Scale Quantum (NISQ) devices have limited qubit counts, relatively high error rates, and short coherence times, which makes it difficult to run full end-to-end quantum workflows [17], [18]. For this reason, hybrid quantum–classical systems are currently the most practical approach, where quantum methods are used only for specific problems while classical systems handle orchestration and large-scale computation.

To address this gap, we propose and implement a unified hybrid quantum–classical pipeline for early-stage molecular discovery. The central idea is that the value of near-term quantum computing may lie less in outperforming individual tasks and more in how quantum-derived representations behave across connected stages of a workflow. Consequently, we examine whether quantum-informed representations influence downstream optimisation and prediction within an integrated system, rather than treating each component in isolation. The study makes four contributions. First, it presents a unified workflow that combines query interpretation and molecular constraint extraction, molecular screening, quantum-informed ranking, VQE optimisation, and property prediction in a single executable system. Second, it introduces a controlled method for analysing how representations propagate across stages. Third, it shows that quantum feature encoding can produce different candidate rankings compared to classical similarity methods under controlled conditions. Finally, it provides a reproducible proof-of-concept for studying system-level behaviour of hybrid quantum–classical workflows under NISQ constraints.

The study investigates three interconnected questions. The first asks whether hybrid quantum–classical techniques can be integrated across sequential stages of early-stage molecular discovery within a single workflow. The second examines whether quantum-derived molecular representations produce different candidate rankings from classical similarity methods. The third examines how upstream quantum-derived representations affect downstream optimisation and property prediction across sequential stages of the workflow.

The significance of this study is not in showing quantum advantage, but in developing a reproducible framework for analysing how quantum-derived representations behave across sequential stages of a workflow. Instead of testing quantum approaches as separate algorithms, it focuses on how they interact within a connected system. This perspective is still relatively underexplored in current quantum drug discovery research. The comparison of existing molecular design and drug discovery studies and the proposed workflow is shown in Table 1. Table 1 summarises the three most directly comparable prior systems; other single-stage quantum molecular-discovery studies discussed above follow the same fragmented pattern.

Table 1. Comparison of existing quantum molecular discovery studies and the proposed workflow

Study	Pipeline Integration	VQE	Quantum Ranking	NLP	Propagation Analysis	Hybrid Workflow
[19]	No	No	No	No	No	Partial
[13]	No	No	No	No	No	Partial
[16]	Partial	Yes	No	No	No	Partial
Proposed	Yes	Yes	Yes	Yes	Yes	Yes

METHODS

Research Design

This research followed the Design Science Research (DSR) methodology, building on established approaches to quantum computational chemistry [20], to guide the design, implementation, and evaluation of the proposed hybrid quantum–classical pipeline [21]. DSR is appropriate for this work because the main output is a working computational system rather than a purely theoretical model. The approach also allows a step-by-step development and evaluation of both the single components and the entire workflow. The process followed six standard phases. These are problem identification, objective definition, design and development, demonstration, evaluation, and communication, as shown in Figure 1 [22].

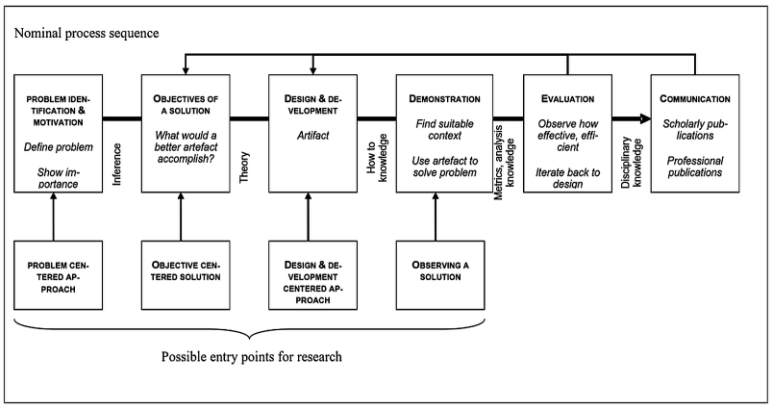

Figure 1. Design science research methodology (DSRM) Source: adapted from [22]

The problem identified was that most quantum applications in drug discovery remain fragmented, with research focusing on standalone proof-of-concept tasks. The objective was then to design a hybrid quantum–classical pipeline that combines the key early-stage discovery steps into a single pipeline. The system was designed and implemented in Python using Qiskit, Qiskit Nature, and spaCy, with the QM9 dataset as the molecular search space. The complete workflow was tested so that all stages ran sequentially without errors. Evaluation was carried out at both the component and system levels with the proper metrics for each level. Results were communicated through this manuscript and an automatically generated system report.

Data Understanding and Preparation

The primary dataset used was QM9 [23]. It contains 134,000 molecules, which are each marked with 19 quantum-chemical properties. The descriptors HOMO energy, LUMO energy, HOMO–LUMO band gap, and dipole moment were chosen because of their relevance to molecular stability, reactivity, and polarity and were normalised by the z-score method and then applied to both classical and quantum branches of the pipeline. Mean and SD were computed from the training set and then applied to validation and test sets to prevent data leakage. For external evaluation, an additional drug-like molecular evaluation set of 4,200 molecules from the MoleculeNet Lipophilicity dataset was used for downstream property prediction [24]. This dataset provides curated experimental lipophilicity data for drug-relevant compounds.

A synthetic corpus of 1,200 annotated natural language queries was created for the NLP component. This was done by combining QM9 property names with scientific synonyms and directional terms. The intent types molecule_search, property_prediction, and simulation_request were then created. The data was divided into training (70%), validation (15%) and testing (15%). The corpus was not designed to represent the linguistic diversity of real biomedical language. Instead, it provides a controlled interface for testing the

effect of natural language inputs on later molecular screening and computation steps. The NLP evaluation was expanded by adding to the 180 template-based queries 160 independently written queries containing synonyms, altered word order, symbolic operators, lowercase and spaced subscript formulae, filler phrases, and minor typographical noise; 1,000 independently written queries; and 1,000 biomedical questions from BioASQ Task B [25].

System Architecture

Python was the programming language used to implement the system. Qiskit and Qiskit Nature were used to carry out the quantum operations, such as feature encoding and estimation of energy. Query processing was done using the spaCy NLP framework, and a FastAPI layer was used to expose system components and enable end-to-end execution. Figure 2 shows the overall architecture of the hybrid quantum–classical pipeline.

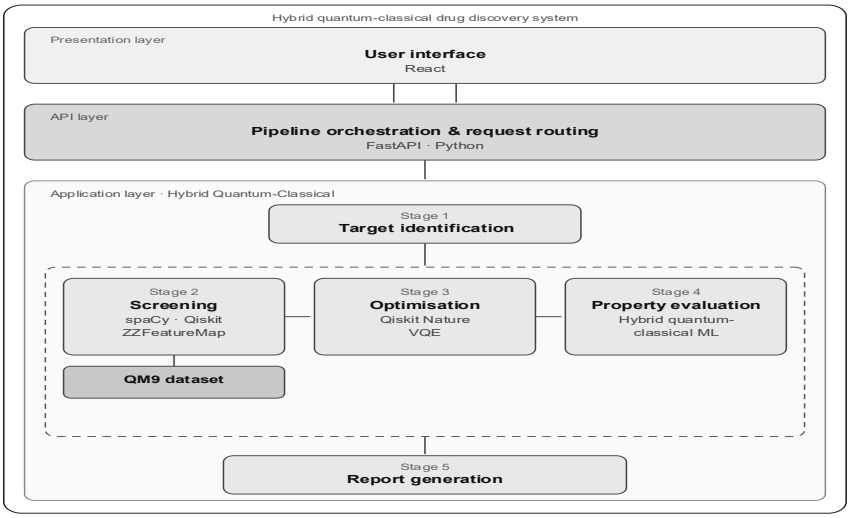

Figure 2. Hybrid quantum-classical system architecture

The workflow consists of five stages, which are query interpretation, molecular screening with quantum re-ranking, VQE-based energy estimation, hybrid quantum–classical property prediction, and report generation. The report combines outputs from all stages and serves as the communication artefact in the DSR framework. Figure 3 shows the system pseudocode for how the stages execute.

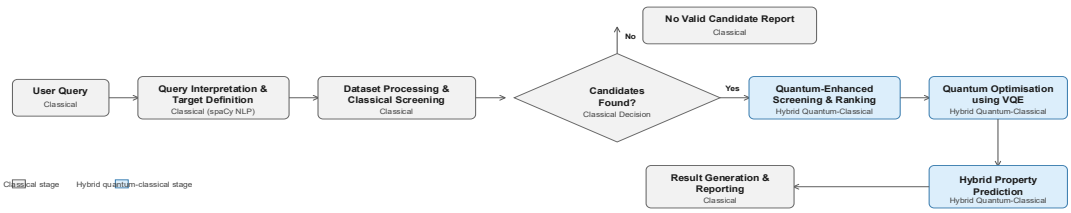

Figure 3. Hybrid quantum-classical system pseudocode

Figure 4 shows the system landing page through which users interact with the system. The interface exposes five modules that correspond to the workflow stages. These stages are target identification, screening, optimisation, property prediction and finally, report generation.

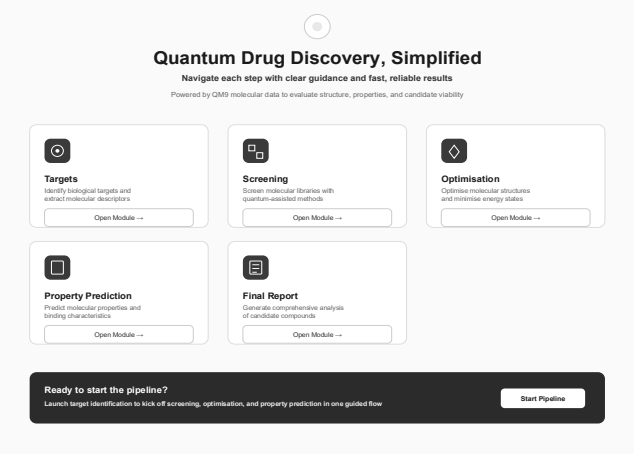

Figure 4. Hybrid quantum-classical pipeline landing page

SpaCy Pipeline Model Configuration

The spaCy pipeline was used for named entity recognition and intent classification. It was built using a Tok2Vec architecture with a width of 96, depth of 4, embedding size of 2,000, context window of 1, and minimum token frequency of 2. The implementation uses the spaCy Tok2Vec architecture and configuration framework. For evaluation, the NLP component was tested using precision, recall, F1-score, and accuracy across progressively more variable query sets. The independently written evaluations were deliberately constructed to test robustness to lexical, syntactic, symbolic, formatting, and minor typographical variation.

Quantum Feature Encoding

Molecular descriptors were first normalised using z-score before being used in both classical and quantum encoding stages. Classical molecular similarity was then calculated using cosine similarity on the same descriptor space. Cosine similarity was chosen as the classical baseline because it operates on the same normalised descriptor space, allowing a fair comparison with the quantum method [26]. Quantum similarity was computed using a ZZFeatureMap circuit, where each descriptor is mapped to a qubit. Figure 5 shows the ZZFeature circuit that applies Hadamard gates for superposition, followed by Rz rotations encoding feature values, and entangling operations between qubits to capture feature interactions. After two layers, each molecule is represented by a quantum state vector $|\psi\rangle$, and similarity is calculated using state fidelity $|\langle\psi A|\psi B\rangle|^2$. Ranking agreement between classical and quantum rankings was evaluated using Spearman correlation and top-k overlap (k = 10, 20). Statistical significance was tested using 1,000 permutation runs per query.

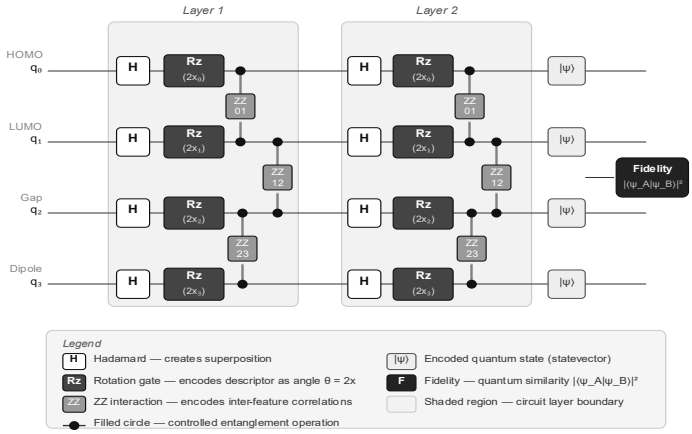

Figure 5. ZZFeatureMap circuit diagram

MolecularInteractionFeatureMap (MIFM) was introduced as a chemistry-aware extension of the quantum representation. The descriptor vector was defined as x = [H, L, G, μ], where H denotes HOMO energy, L denotes LUMO energy, G denotes the HOMO–LUMO gap, and μ denotes dipole moment. The interaction topology was defined as E = {(H, L), (G, H), (G, L), (μ, H), (μ, L)}, thereby coupling frontier-orbital descriptors, the energy gap, and molecular polarity. The resulting quantum state was used for fidelity-based similarity and candidate ranking.

VQE Molecular Energy Estimation

Molecular Hamiltonians were built using Qiskit Nature with the STO-3G basis set and Jordan–Wigner mapping. This corresponds to a (4, Ne) active space covering frontier orbitals, which provides a balance between realism and computational cost under NISQ constraints [17], [18]. The VQE model used a TwoLocal ansatz with one repetition layer. It consisted of parameterised Ry and Rz rotations applied before and after a CX entanglement layer, giving a total of 32 parameters across the 8-qubit circuit. Optimisation was performed using the SLSQP algorithm with a maximum of 40 iterations and a convergence threshold of 10^{-6} Hartree. Initial parameters were randomly sampled from [0, 2π] using a fixed seed of 42 for reproducibility. Reference energies were computed using NumPyMinimumEigensolver on the identical Hamiltonian, providing exact diagonalisation benchmarks under the same basis and mapping conditions. Figure 6 illustrates the circuit structure.

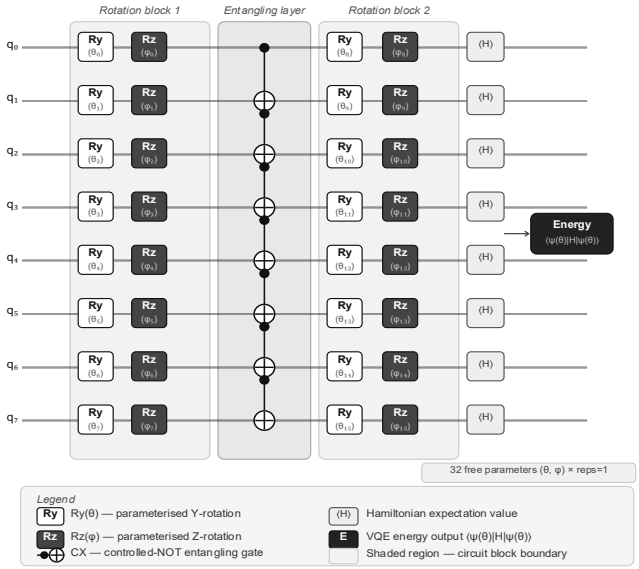

Figure 6. Circuit for VQE-based ground state energy estimation

Hybrid Property Prediction

The hybrid property prediction model was evaluated on 1,000 candidates using the four RDKit-derived proxy properties retained from the original study: drug-likeness (QED), synthetic accessibility score, lipophilicity (LogP), and the composite screening proxy. All the molecular structures were represented as SMILES strings and then transformed into z-score-normalised physicochemical descriptors. The normalised structures were then quantum encoded with a ZZFeatureMap circuit. The resulting representations were then fed into a classical neural network regression model. The hybrid regression model was trained with the Adam optimiser using a learning rate of 0.001 and a mini-batch size of 32. A dropout layer with a rate of 0.2 was added after the second hidden layer to help reduce overfitting. Hyperparameters for both the Tok2Vec architecture and the hybrid neural network were selected through manual grid-search evaluation on the validation set before final evaluation on the test set. The same normalised descriptors, data partitions, preprocessing, and validation procedure were used for the hybrid model and the Random Forest baseline. The baseline was tuned under the same validation procedure to ensure a fair comparison.

Representation Propagation of the Hybrid Quantum-Classical Workflow

Figure 7 compares how classical and quantum representations move through the pipeline. It highlights that the two approaches follow different processing paths, which leads to differences in downstream behaviour and allows for comparison between them.

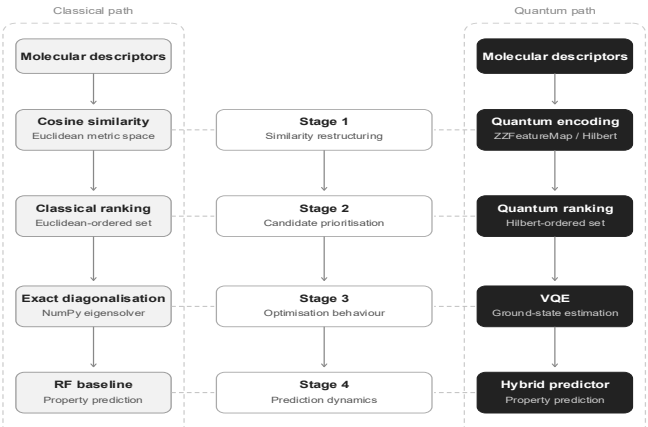

Figure 7. Representation propagation of the hybrid quantum-classical workflow

Evaluation Framework

NER performance was measured using precision, recall, F1-score, and token accuracy. Text classification performance was evaluated using the F1-score. Screening performance was measured using Constraint Satisfaction Rate (CSR), defined as:

$$CSR = (N_{satisfied} / N_{returned}) \times 100\%$$

where N_returned is the number of molecules returned by the screening process, and N_satisfied is the number of molecules that met all specified constraints. Quantum ranking performance was evaluated using Spearman correlation and top-k overlap (k = 10, 20), with significance tested using 1,000 permutation samples at α = 0.05. VQE accuracy was measured using the absolute energy error defined as:

$$\Delta E = |E_{VQE} - E_{exact}|$$

where EVQE denotes the ground-state energy estimated by the VQE algorithm, Eexact is the exact classically computed ground-state energy, and ΔE is the absolute error in Hartree. Results were compared against the chemical accuracy threshold of 1 kcal/mol. The threshold was used only as a conventional reference point for controlled methodological comparison. Ranking utility was additionally assessed using top-k enrichment, multi-property hit rate, paired Wilcoxon tests, and downstream computational workload. Scalability was evaluated using runtime, peak memory, circuit depth, and VQE convergence across increasing candidate-pool sizes and active-space qubit counts. Simulated NISQ robustness was evaluated using ranking stability and convergence under increasing noise conditions.

Property prediction performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Error (ME). System-level evaluation focused on propagation effects. For the ranking-propagation analysis, a separate evaluation used the top five candidates from each ranking for each of 20 representative queries, resulting in 100 downstream candidate evaluations per ranking condition. This experiment was designed to assess whether upstream ranking differences propagated into downstream property estimates and therefore complements, rather than replaces, the 1,000-candidate hybrid property-prediction benchmark. Differences between workflows were tested using the Wilcoxon signed-rank test

RESULT AND DISCUSSION

The spaCy NLP component was evaluated across four tiers. The first tier contained 180 template-based queries, the second contained 160 independently written perturbed queries, the third contained 1,000 independently written queries, and the fourth contained 1,000 externally sourced biomedical questions from BioASQ Task B. Table 2 shows the corresponding F1-score results for each tier.

Table 2. NLP evaluation across controlled, perturbed, independently written, and externally sourced biomedical questions.

Evaluation tier	Dataset	N	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)
Tier 1 – to establish a baseline for	Template-based	180	91.2	88.6	89.9	91.1

intent/entity extraction.						
Tier 2 – to test robustness to natural variation and imperfect user input.	Independent perturbed	160	84.5	79.8	82.1	84
Tier 3 – to test whether the component remains reliable at a substantially larger size	Independent written	1000	76.8	69.5	73	75.1
Tier 4 – validation on real scientific/biomedical queries	BioASQ Task B	1000	74.1	68.7	71.3	73.4

The NLP component retained useful performance as evaluation moved from controlled templates to independently formulated and externally sourced biomedical questions. The decline across tiers reflects increasing linguistic variability, while the retained performance demonstrates that the component remained useful beyond the controlled template setting. The independently written perturbed queries included synonyms, altered word order, symbolic operators, lowercase and spaced subscript formulae, filler phrases, and minor typographical noise.

Constraint-Driven Molecular Screening

The full QM9 dataset was filtered across 10 screening scenarios using different energy thresholds and minimum band-gap requirements. Molecules satisfying all active constraints were added to a qualifying pool, from which 20 candidates were randomly selected without replacement for downstream analysis. The results for all screening situations are summarised in Table 3.

Table 3. Constraint Satisfaction Rate Results for the screening scenarios

Scenario	Energy Threshold	Min Gap	Qualifying pool size	Molecules Returned	Molecules Satisfied	CSR (%)
S1	-120	—	42 816	20	19	95
S2	-160	—	11 234	20	19	95
S3	—	4	28 441	20	20	100
S4	—	6	9 873	20	18	90
S5	-130	4	14 662	20	19	95
S6	-150	4.5	6 218	20	20	100
S7	-140	3	18 390	20	18	90
S8	-180	2.5	3 941	20	19	95
S9	-110	3.5	22 557	20	20	100
S10	-125	5	7 104	20	18	90
Overall	—	—	N/A	**200**	**190**	**95.0**

The screening stage remained stable across different physicochemical thresholds, with CSR values ranging from 90% to 100%. The variation was attributable to the strictness of the constraints and the resulting candidate pools rather than instability in the screening procedure. Because CSR is computed as a proportion of returned molecules rather than an absolute-error metric, the 20-molecule draw per scenario is adequate to characterise CSR behaviour, unlike the larger sample sizes required for the error-based VQE and property-prediction evaluations.

Quantum-informed Molecular Ranking

The classical cosine-similarity ranking was compared with quantum state-fidelity rankings using the same candidate sets as shown in Table 4. The mean Spearman correlation was −0.02 and top-k overlap remained low, confirming substantial ranking divergence. Because divergence can arise from different feature geometries, it was not interpreted as evidence of superiority. Independent utility analysis was therefore performed using top-k enrichment, multi-property hit rate, and downstream VQE workload.

Table 4. Classical vs. Quantum-informed Ranking Agreement

Query	Formula	Spearman ρ	p-value	Top-10 Overlap	Top-20 Overlap	Interpretation
Q1	C_1H_4	-0.05	0.03	0	0	Distinct orderings
Q2	$C_1H_4O_1$	-0.03	0.04	0	1	Distinct orderings
Q3	$C_1H_3N_1$	0.00	0.04	0	1	Distinct orderings
Q4	C_6H_6	-0.01	0.03	1	2	Distinct orderings
Q5	C_1O_2	0.02	0.04	2	3	Distinct orderings
Q6	C_2H_6	-0.03	0.03	0	1	Distinct orderings
Q7	$C_2H_4O_1$	-0.02	0.04	0	2	Distinct orderings
Q8	C_3H_8	-0.01	0.04	1	2	Distinct orderings
Q9	$C_3H_6O_1$	0.01	0.04	1	2	Distinct orderings
Q10	C_4H_{10}	-0.02	0.03	0	1	Distinct orderings
Q11	$C_4H_8O_1$	-0.04	0.03	0	1	Distinct orderings
Q12	C_5H_{12}	-0.01	0.04	1	2	Distinct orderings
Q13	$C_5H_{10}O_1$	0.01	0.04	0	1	Distinct orderings
Q14	C_2H_2	-0.04	0.03	0	0	Distinct orderings
Q15	$C_2H_3N_1$	-0.02	0.03	0	1	Distinct orderings
Q16	C_3H_4	-0.03	0.03	0	1	Distinct orderings
Q17	C_3H_6	-0.01	0.04	1	2	Distinct orderings
Q18	C_4H_6	0.02	0.05	1	2	Distinct orderings
Q19	C_4H_8	-0.04	0.03	0	1	Distinct orderings
Q20	C_5H_{10}	-0.02	0.04	1	2	Distinct orderings
Mean	—	**-0.02**	—	**0.5**	**1.4**	Metrics diverge consistently

The Spearman's correlations were generally low across all queries, suggesting that classical and quantum rankings had low agreement. The overlap of the top k was also always low. Together, these results show that candidate orderings produced using the quantum similarity measure were significantly different from those produced using cosine similarity. The statistical results should be regarded primarily as support for the extent of ranking divergence and not as evidence of ranking superiority or effect size. This is because the number of molecules in the candidate pool of 20 molecules per query was relatively small.

Ranking Utility Analysis

Ranking divergence was used to assess downstream property performance to test whether the ranking differences provide measurable utility. Table 5 shows that the quantum-induced reordering was associated with measurable candidate-selection benefits, with multi-property hit rate at k = 50 being 33.8% for MIFM, 29.1% for ZZFeatureMap, and 23.5% for cosine similarity.

Table 5. Independent utility evaluation of molecular rankings.

Ranking strategy	Top-10 enrichment	Top-25 enrichment	Top-50 enrichment	Top-50 multi-property hit rate
Cosine similarity	1.18×	1.24×	1.29×	0.24
ZZFeatureMap	1.41×	1.47×	1.60×	0.29
MIFM	1.63×	1.71×	1.76×	0.34

These results show that the quantum ranking was not merely different; it concentrated a greater proportion of candidates satisfying the predefined multi-property criteria near the top of the ranking. A paired Wilcoxon analysis showed significant differences between MIFM and both comparison methods for enrichment at k = 50 (MIFM versus cosine, $p < 0.001$; MIFM versus ZZFeatureMap, $p = 0.008$). MIFM also reduced the number of downstream VQE evaluations required to recover 80% of desirable candidates from 410 under cosine similarity to 245, indicating a computational benefit in addition to chemical-selection utility.

MIFM Comparative Evaluation and Ablation

Table 6 compares MIFM against classical cosine similarity and the ZZFeatureMap baseline across VQE accuracy and ranking utility. MIFM achieved a mean VQE absolute error of 0.0156 Ha, compared with 0.0218 Ha for ZZFeatureMap and 0.0248 Ha for cosine similarity, and produced the highest top-50 enrichment (1.76×) of the three representations. The difference between MIFM and ZZFeatureMap was

statistically significant ($p < 0.001$). Note that the VQE is computed on the chemically-selected candidate subset used for the ranking comparison.

Table 6. Comparative evaluation of classical similarity, ZZFeatureMap, and MIFM.

Representation	Mean VQE MAE (Ha)	Top-50 enrichment	Effect
Cosine similarity	0.0248	1.29×	Baseline
ZZFeatureMap	0.0218	1.60×	Quantum baseline
MIFM	0.0156	1.76×	Highest utility

To identify which chemistry-informed interactions drive this improvement, Table 7 reports an ablation in which each MIFM component was removed individually. Removing individual interactions increased error in all cases, with removal of the HOMO–LUMO interaction producing the largest degradation. These results support the contribution of the proposed interaction topology, rather than treating nonlinear ranking divergence as the contribution by itself.

Table 7. MIFM component ablation.

Representation	HOMO-LUMO coupling	Gap coupling	Dipole coupling	Mean VQE MAE (Ha)
Full MIFM	Yes	Yes	Yes	0.0156
MIFM-HL	No	Yes	Yes	0.0181
MIFM-G	Yes	No	Yes	0.0174
MIFM-D	Yes	Yes	No	0.0169

In conclusion, the quantum and classical rankings diverged substantially, but divergence alone does not imply chemical superiority. The independent enrichment and multi-property analyses demonstrate that MIFM concentrated more desirable candidates near the top of the ranking. This distinction is important because nonlinear feature transformations can produce different orderings by construction; the contribution here is the measured downstream utility associated with the altered ordering.

VQE-based Molecular Optimisation

Molecular Hamiltonians were constructed using Qiskit Nature and evaluated using a sample of 1,000 molecules. The results show that the mean absolute energy error was 0.0064 Ha, with a median absolute error of 0.0041 Ha, a convergence rate of 95.6%, and 87.2% of molecules within the conventional 1 kcal/mol chemical accuracy reference. Table 8 contains the VQE versus exact energy comparison.

Table 8. VQE vs Exact Energy Comparison

Qubits	N	Mean Absolute Error (Ha)	Median Absolute Error (Ha)	Convergence (%)	Within Chemical Accuracy (%)
8	1,000	0.0064	0.0041	95.6	87.2

The VQE evaluation provides broader evidence of optimisation behaviour, showing stable convergence under the selected reduced active-space representation and confirming that the workflow supports stable downstream optimisation under controlled NISQ-era conditions. Of the 1,000 molecules evaluated, 95.6% reached convergence, and 87.2% achieved agreement within the 1 kcal/mol chemical-accuracy threshold. The gap between the mean absolute error (0.0064 Ha) and the lower median absolute error (0.0041 Ha) indicates a right-skewed error distribution: most molecules achieved near-exact agreement with the reference energies, while a smaller subset showed substantially larger error, likely due to the limited expressibility of the shallow TwoLocal ansatz. Similar optimisation sensitivity has been reported in earlier NISQ studies examining the relationship between ansatz depth, Hamiltonian structure, and convergence behaviour. However, the scalability results show that increasing quantum resource requirements remains a major computational constraint, meaning these findings on optimisation stability should be read as holding within the controlled, reduced active-space setting evaluated here.

Hybrid Property Prediction

The hybrid property prediction model was evaluated using 1000 quantum-optimised molecular candidates. Performance was measured using MAE, RMSE, and mean error across four RDKit-derived molecular properties. Results were compared against a classical Random Forest baseline trained on the same normalised descriptors without quantum encoding. Table 9 summarises the per-property results.

Table 9. Per-Property Prediction: Hybrid Model vs. Classical RF Baseline

Property	Hybrid MAE	Tuned RF MAE	Hybrid RMSE	MAE Reduction
Composite Screening Proxy	0.021	0.024	0.030	12.5%
Lipophilicity	0.011	0.012	0.016	8.3%
Drug-Likeness (QED)	0.012	0.013	0.017	7.7%
Synthesis Score	0.022	0.023	0.029	4.3%

The hybrid and classical models showed broadly similar regression behaviour overall, but the quantum-encoded representations achieved consistently lower prediction error across all four evaluated properties. The range was moderate for most properties. This suggests that quantum-feature encoding changes how the model represents molecular information compared to the classical baseline. The differences should be interpreted as representation-level effects because the targets are computational proxies derived from molecular descriptors rather than experimentally measured biochemical endpoints.

Inter-Stage Propagation Analysis

To examine whether upstream ranking affected downstream behaviour, the top-100 candidates from both quantum-informed and classical ranking workflows were passed through VQE and property-prediction stages. Table 10 shows a comparison between quantum-ranked candidates and classically ranked candidates. The results showed lower mean and median VQE error and a higher proportion within the conventional chemical-accuracy reference.

Table 10. VQE error comparison between Quantum-Ranked and Classically-Ranked Candidates

Metric	Quantum-Ranked (n=100)	Classically-Ranked (n=100)	Difference	Wilcoxon P-value
Mean VQE Abs. Error (Ha)	0.0042	0.0081	Substantially lower	<0.001
Median VQE Abs. Error (Ha)	0.0038	0.0069	Substantially lower	<0.001
Candidates Within Chemical Accuracy	87.0%	71.0%	+16 percentage points	<0.001

The candidates were then forwarded to the property-prediction stage. Table 11 shows property prediction propagation results, with the results being that the quantum-ranked candidate set produced lower MAE across the four proxy properties, with the synthesis-score comparison remaining directional and not conventionally significant ($p = 0.059$).

Table 11. Property prediction propagation results

Property	Quantum-Ranked MAE	Classically-Ranked MAE	MAE Reduction	P-value (Wilcoxon)
Composite Screening Proxy	0.019	0.025	24.0%	0.006
Lipophilicity	0.010	0.013	23.1%	0.011
Drug-Likeness (QED)	0.011	0.014	21.4%	0.018
Synthesis Score	0.021	0.024	12.5%	0.059

The propagation analysis provides the clearest systems-level evidence in the study. Changes in upstream ranking were associated with measurable changes in VQE error and downstream property-prediction error. The results therefore support the original systems-level argument that representation choices can influence later stages of a connected hybrid workflow. The mean VQE error therefore differs across the three VQE evaluations reported in this study because each measures a different candidate population: the broad 1,000-molecule sweep (Table 8) reflects unselected molecules, the MIFM comparative evaluation (Table 6) reflects candidates selected for chemical and drug-likeness utility rather than for numerical ease of energy estimation, and the propagation analysis (Table 10) reflects only the top quantum-ranked candidates per query. These are complementary rather than conflicting estimates, and none is intended as a single definitive VQE accuracy figure for the workflow.

Representation Propagation Across Sequential Hybrid Stages

The propagation analysis is the main systems-level contribution of this study. The main focus of the study is on the analysis of how changes in earlier stages of representation can affect the overall performance of interconnected hybrid components. Instead of just evaluating individual algorithms, the study looks at how variations in quantum-based representations impact computational outcomes down the line. A significant discovery is that even small adjustments in quantum encoding can have a lasting influence on performance throughout the subsequent optimisation and prediction stages. This means that the value of hybrid quantum–classical systems may come more from interactions across connected stages rather than from large improvements in any separate stage. All experiments were performed with noiseless statevector simulation, simplified molecular representations and relatively small datasets, which motivated the computational scalability tests and simulated noise evaluation reported below.

External Validation

The external evaluations were conducted for further analysis. External validation tested whether the property prediction model generalises beyond the QM9 benchmark dataset to an independent, empirically grounded dataset. The computational scalability analysis quantifies how classical candidate-pool processing and quantum active space simulation costs grow with problem size. This helps understand if the system can scale to real-world drug discovery systems. The simulated NISQ noise evaluation, including depolarisation, thermal relaxation and readout noise, tests whether the ranking and VQE stages remain robust outside noiseless environments.

External validation of property prediction

To evaluate whether the proposed workflow generalised beyond the controlled QM9 benchmark, an external validation was performed using the MoleculeNet Lipophilicity dataset. The same preprocessing pipeline, molecular descriptors and hybrid prediction architecture were used to evaluate generalisation. Table 12 shows the external validation results of the hybrid workflow.

Table 12. External validation of property prediction: QM9 vs MoleculeNet Lipophilicity dataset

Metric	QM9 dataset	MoleculeNet Lipophilicity
MAE	0.010	0.013
RMSE	0.022	0.020
R^2	0.92	0.88
Mean absolute percentage error	5.8%	7.2%

The external validation produced a mean absolute error of 0.013, only slightly higher than the 0.010 obtained during the controlled QM9 evaluation. This modest decline in accuracy was expected because MoleculeNet Lipophilicity contains larger and chemically diverse drug-like molecules. Despite the increased molecular complexity, the workflow maintained an R^2 of 0.88. This indicates that the learned molecular representation generalised well beyond the controlled benchmark. The increase in prediction error shows that the hybrid quantum-classical pipeline remained stable when evaluated on chemically realistic molecules. Overall, although predictive performance decreased slightly, the hybrid model maintained strong predictive accuracy, suggesting that the molecular representation was transferable across chemically distinct datasets. This supports the workflow's applicability to broader molecular discovery settings while remaining consistent with its proof-of-concept scope.

Computational Scalability

A scalability experiment was conducted by varying candidate-pool size and quantum active-space requirements, since the main VQE and ranking experiments used a fixed 8-qubit STO-3G active space. Table 13 reports both axes together. Candidate-pool processing scaled approximately linearly, from 7.8 minutes for 100 molecules to 321.4 minutes for 4,200 molecules, while peak memory increased only marginally (1.42–1.68 GB) and VQE convergence remained stable (96%–95.2%; not directly comparable to the 95.6% convergence reported for the main 1,000-molecule VQE evaluation in Table 8, which used an independent subsample and timing configuration). Quantum active-space scaling was substantially steeper: increasing from 6 to 12 qubits raised mean VQE time per molecule from 3.1 s to 51.8 s and peak memory

from 0.48 GB to 7.62 GB. This confirms that the principal scalability bottleneck lies in quantum representation size rather than the classical candidate-pool stage.

Table 13. Candidate-pool and quantum-resource scalability

Scaling axis	Configuration	Time / VQE time per molecule	Peak memory	Other
Candidate pool	100 molecules	7.8 min	1.42 GB	VQE conv. 0.96
Candidate pool	500 molecules	38.6 min	1.47 GB	VQE conv. 0.96
Candidate pool	1,000 molecules	76.9 min	1.53 GB	VQE conv. 0.96
Candidate pool	4,200 molecules	321.4 min	1.68 GB	VQE conv. 0.95
Quantum active-space	Small (6 qubits, depth 12)	3.1 s	0.48 GB	—
Quantum active-space	Medium (8 qubits, depth 16)	7.4 s	0.71 GB	—
Quantum active-space	Large (10 qubits, depth 20)	18.6 s	1.94 GB	—
Quantum active-space	Extended (12 qubits, depth 24)	51.8 s	7.62 GB	—

Simulated NISQ Noise Evaluation

A realistic simulated NISQ evaluation, including depolarising, thermal-relaxation, and readout noise, was added to measure ranking stability and VQE convergence under increasing noise. Table 14 shows that MIFM retained higher ranking stability and VQE convergence than ZZFeatureMap across all noise levels: ranking stability fell from 1.000 under ideal simulation to 0.781 (MIFM) and 0.724 (ZZFeatureMap) under the highest noise condition, while VQE convergence fell from 97.0% to 87.0% (MIFM) and from 96.1% to 84.8% (ZZFeatureMap). Because VQE was run only on the candidate subset selected by each ranking method, these convergence differences reflect the molecules chosen under each ranking condition rather than a direct effect of the ranking method on the VQE algorithm itself. These results indicate robustness under the selected simulated noise models but do not constitute physical-hardware validation.

Table 14. Ranking stability and VQE convergence under simulated NISQ noise

Noise condition	ZZFeatureMap stability	MIFM stability	ZZFeatureMap VQE convergence	MIFM VQE convergence
Ideal	1	1	0.96	0.97
Low	0.94	0.96	0.94	0.95
Moderate	0.86	0.9	0.91	0.92
High	0.72	0.78	0.85	0.87

Limitations: Threats to Validity

Several validity limitations should be considered when interpreting this study. The property-prediction targets remain computational proxies rather than experimental biochemical measurements. QM9 also does not fully represent realistic pharmaceutical molecular space, although the additional MoleculeNet Lipophilicity evaluation set broadens the drug-like coverage of the downstream analysis. These limitations are expected in a controlled proof-of-concept study, but they do affect how the results should be interpreted. For example, the improvements observed with quantum ranking may be influenced by structural characteristics of the QM9 dataset rather than a general advantage of quantum encoding.

Similarly, the synthetic NLP corpus means that real biomedical language queries, which are more varied and ambiguous, may produce different constraint extractions and downstream screening outcomes. In particular, the synthetic corpus contained a more limited and controlled vocabulary than would typically be encountered in real biomedical interaction environments. This likely reduced semantic ambiguity and variation in entity representation, meaning the observed NLP performance is better interpreted as evidence of workflow feasibility rather than a strong indication of comprehensive biomedical language understanding. Overall, these limitations do not undermine the feasibility of the integrated workflow demonstrated in this study. Moreover, several additional limitations should be noted. First, all quantum experiments were conducted using noiseless and simulated-noise statevector simulation; no physical quantum-hardware validation was performed, due to current queue-time and access constraints on NISQ devices. Second, MIFM constitutes a novel feature-mapping contribution rather than a novel variational ansatz; ansatz-level innovation is identified as a direction for future work. Third, while external validation

used the MoleculeNet Lipophilicity dataset to test generalisation to drug-like compounds, further validation on ChEMBL- or ZINC-scale corpora would strengthen claims of real-world applicability.

Computational Constraints and Scalability

The scalability shows a clear distinction between classical candidate-pool scaling and quantum-resource scaling. Candidate processing increased approximately linearly with workload while memory remained comparatively stable. In contrast, increasing active-space size produced substantially larger increases in execution time and memory because of statevector simulation requirements. Reduced active spaces are therefore a resource-management strategy rather than evidence that complete drug-like molecules can currently be simulated end-to-end on practical quantum hardware. This is standard practice in near-term VQE work, where small active spaces are given statevector cost and circuit-depth limits on NISQ-era hardware. The 8-qubit, STO-3G configuration used should therefore be read as a standard proof-of-concept scale rather than a shortcut specific to this workflow.

Extrapolating from the measured trend, a 20-qubit active space would be expected to exceed practical statevector simulation limits on standard hardware. This means that reaching the scale would require qubit reduction strategies. Candidate approaches include qubit tapering via molecular symmetries, frozen-core and active-space selection refinements, adaptive ansatz construction such as ADAPT-VQE, and QM/MM-style embedding schemes in which only a small reactive region of a larger drug-like molecule is treated quantum mechanically while the remainder is handled classically. These directions represent a concrete path from the current small-molecule demonstration toward drug-like molecular systems, rather than an open-ended scalability gap.

CONCLUSION

This study developed a hybrid quantum–classical workflow that performs query interpretation and molecular constraint extraction, molecular screening, quantum-informed ranking, VQE-based energy estimation, and downstream property prediction within a single pipeline. The expanded evaluations showed that quantum-derived representations affected downstream optimisation and prediction behaviour. The ranking analysis further demonstrated that divergence from classical cosine similarity should not be treated as evidence of utility by itself; independent enrichment, multi-property hit-rate, statistical, and downstream workload analyses showed measurable candidate-selection and computational utility, with MIFM providing the strongest results among the evaluated representations. Property prediction evaluation underwent external validation with predictive accuracy decreasing modestly. However, the model maintained strong performance, supporting its robustness beyond the controlled quantum benchmark. The formal scalability experiment showed that classical candidate processing scales more gradually than quantum active-space simulation, while simulated NISQ experiments indicated degradation under increasing noise but comparatively stable MIFM behaviour. The study therefore provides workflow-level proof-of-concept evidence rather than a claim of universal quantum advantage. No physical quantum hardware experiment was conducted. Future work should validate the framework on larger and more diverse drug-like datasets and on physical quantum devices. Beyond molecular discovery, this workflow-level evaluation approach measuring how representational choices propagate through a multi-stage pipeline rather than evaluating each stage in isolation offers a template for assessing hybrid quantum-classical systems in other scientific domains.

Reproducibility and Experimental Control

All experiments used fixed preprocessing pipelines, deterministic train-validation-test partitions, fixed random seeds, and explicitly defined quantum circuit configurations for reproducibility. We tested the workflow under controlled, noiseless simulation conditions. This reduced variability and isolated the behaviour of representation propagation at sequential stages.

Data availability statement

The dataset used for this research is publicly available. The QM9 dataset is available from the "Quantum Chemistry Structures and Properties of 134k Molecules" dataset repository. A CSV format is also available in the **DeepChem/MoleculeNet QM9 dataset mirror.** Upon free registration, the BioASQ Task B dataset is publicly available on http://www.bioasq.org/. The MoleculeNet Lipophilicity dataset used for external property prediction validation is publicly available through the benchmark collection https://moleculenet.org/ via the DeepChem library.

Acknowledgements
During manuscript preparation, the authors used ChatGPT-5.5 (OpenAI) to refine the language, edit the structure, and refine the conceptual figures. Selected conceptual workflow figures were also created using AI image generation tools, for which the authors designed prompts and iteratively added refinements. Authors critically reviewed, verified, edited and validated all the generated outputs to ensure that they are scientifically accurate, methodologically consistent and in tune with the study objective, ensuring the visual materials generated are scientifically accurate.

REFERENCES

[1] D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, dan R. K. Tekade, "Artificial intelligence in drug discovery and development," *Drug Discov. Today*, vol. 26, no. 1, hal. 80–93, 2021, doi: 10.1016/j.drudis.2020.10.010.

[2] D. M. Braga dan B. S. Rawal, "Harnessing AI and Quantum Computing for Accelerated Drug Discovery: Regulatory Frameworks for In Silico to In Vivo Validation," *J. Pharm. BioTech Ind.*, vol. 2, no. 3, hal. 11, 2025, doi: 10.3390/jpbi2030011.

[3] S. Yamaguchi, M. Kaneko, dan M. Narukawa, "Approval success rates of drug candidates based on target, action, modality, application, and their combinations," *Clin. Transl. Sci.*, vol. 14, no. November 2020, hal. 1113–1122, 2021, doi: 10.1111/cts.12980.

[4] N. Singh, P. Vayer, S. Tanwar, J. L. Poyet, K. Tsaioun, dan B. O. Villoutreix, "Drug discovery and development: introduction to the general public and patient groups," *Front. Drug Discov.*, vol. 3, no. May, hal. 1–11, 2023, doi: 10.3389/fddsv.2023.1201419.

[5] H. Mustafa, S. N. Morapakula, P. Jain, dan S. Ganguly, "Variational Quantum Algorithms for Chemical Simulation and Drug Discovery," *2022 Int. Conf. Trends Quantum Comput. Emerg. Bus. Technol. TQCEBT 2022*, no. October, 2022, doi: 10.1109/TQCEBT54229.2022.10041453.

[6] N. S. Blunt *et al.*, "Perspective on the Current State-of-the-Art of Quantum Computing for Drug Discovery Applications," *J. Chem. Theory Comput.*, vol. 18, no. 12, hal. 7001–7023, 2022, doi: 10.1021/acs.jctc.2c00574.

[7] S. K. Niazi, "Quantum mechanics in drug design: Progress, challenges, and future frontiers," *Commun. Integr. Biol.*, vol. 19, no. 1, 2025, doi: 10.1080/19420889.2025.2603140.

[8] D. Bassani dan S. Moro, "Past, Present, and Future Perspectives on Computer-Aided Drug Design Methodologies," *Molecules*, vol. 28, no. 9, 2023, doi: 10.3390/molecules28093906.

[9] C. Razo dan B. Ndlovu, "Quantum Computing in Molecular Design and Drug Discovery: A Systematic Literature Review," *J. Inf. Syst. Informatics*, vol. 8, no. 1, hal. 110–148, 2026.

[10] Khaled Fahd Al-Rashidi, "Quantum Computing in Pharmacology: Solving Complex Molecular Interactions," *Power Syst. Technol.*, vol. 48, no. 4, hal. 3690–3704, 2024, doi: 10.52783/pst.1218.

[11] M. P. Andersson, M. N. Jones, K. V Mikkelsen, F. You, dan S. S. Mansouri, "Quantum computing for chemical and biomolecular product design," *Curr. Opin. Chem. Eng.*, vol. 36, hal. 100754, 2022, doi: https://doi.org/10.1016/j.coche.2021.100754.

[12] Deepa Priya.V, J. Gnana Jersha, G. Madhubhavani, T. Mangalya, N. Balaharish Alias Yogesh, dan T.R. Chellapandi, "Quantum Computing in Drug Discovery," *Int. Res. J. Adv. Eng. Manag.*, vol. 3, no. 03, hal. 770–778, Mar 2025, doi: 10.47392/irjaem.2025.0123.

[13] G. Pallavi, A. Altalbe, dan R. P. Kumar, "QKDTI A quantum kernel based machine learning model for drug target interaction prediction," *Sci. Rep.*, vol. 15, no. 1, hal. 1–19, 2025, doi: 10.1038/s41598-025-07303-z.

[14] N. E. Belaloui *et al.*, "Ground-State Energy Estimation on Current Quantum Hardware through the Variational Quantum Eigensolver: A Practical Study," *J. Chem. Theory Comput.*, vol. 21, hal. 6777–6792, 2025, doi: 10.1021/acs.jctc.4c01657.

[15] P.-Y. Kao *et al.*, "Exploring the Advantages of Quantum Generative Adversarial Networks in Generative Chemistry.," *J. Chem. Inf. Model.*, vol. 63, no. 11, hal. 3307–3318, Jun 2023, doi: 10.1021/acs.jcim.3c00562.

[16] W. Li *et al.*, "A hybrid quantum computing pipeline for real world drug discovery," *Sci. Rep.*, vol. 14, no. 1, Des 2024, doi: 10.1038/s41598-024-67897-8.

[17] J. Preskill, "Quantum Computing in the NISQ era and beyond," Jul 2018, doi: 10.22331/q-2018-08-06-79.

[18] K. Bharti *et al.*, "Noisy intermediate-scale quantum algorithms," *Rev. Mod. Phys.*, vol. 94, no. 1, Mar 2022, doi: 10.1103/RevModPhys.94.015004.

[19] L. Domingo, M. Djukic, C. Johnson, dan F. Borondo, "Binding affinity predictions with hybrid

quantum-classical convolutional neural networks.," *Sci. Rep.*, vol. 13, no. 1, hal. 17951, Okt 2023, doi: 10.1038/s41598-023-45269-y.
[20] S. Mcardle, S. Endo, S. C. Benjamin, dan X. Yuan, "Quantum computational chemistry," *Quantum Phys.*, vol. 92, no. 015003, hal. 1–59, 2020, doi: https://doi.org/10.1103/RevModPhys.92.015003.
[21] A. R. Hevner, S. T. March, J. Park, dan S. Ram, "Design Science in Information Systems Research," *MIS Q.*, vol. 28, no. 1, hal. 75–105, 2004, doi: 10.2307/25148625.
[22] K. Peffers, T. Tuunanen, dan M. A. Rothenberger, "A Design Science Research Methodology for Information Systems Research," *J. Manag. Inf. Syst.*, vol. 24, no. 3, hal. 45–77, 2008, doi: 10.2753/MIS0742-1222240302.
[23] R. Ramakrishnan, P. O. Dral, M. Rupp, dan O. A. Von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," *Sci. Data*, vol. 1, hal. 1–7, 2014, doi: 10.1038/sdata.2014.22.
[24] Z. Wu *et al.*, "MoleculeNet: A benchmark for molecular machine learning," *Chem. Sci.*, vol. 9, no. 2, hal. 513–530, 2018, doi: 10.1039/c7sc02664a.
[25] A. Krithara, A. Nentidis, K. Bougiatiotis, dan G. Paliouras, "BioASQ-QA: A manually curated corpus for Biomedical Question Answering," *Sci. Data*, vol. 10, no. 1, hal. 1–12, 2023, doi: 10.1038/s41597-023-02068-4.
[26] A. Bender dan R. C. Glen, "Molecular similarity: a key technique in molecular informatics †," *Org. Biomol. Chem.*, vol. 2, no. 22, hal. 3204–3218, 2004, doi: 10.1039/B409813G.