

Quantum Computing Prospects in Bioinformatics

Sofia Ivanov¹, Ivan Dubois², Isabella Ivanov³

¹ Research Scientist, Department of Computer Science, Nordic Technical University, Stockholm, Sweden. Email: sofia.ivanov983@yahoo.com | ORCID: 2480-9069-2904-8529

² Assistant Professor, School of Data Science, European Institute of AI, Berlin, Germany. Email: ivan.dubois730@yahoo.com | ORCID: 8614-7555-3514-1635

³ Postdoctoral Researcher, School of Data Science, Baltic AI Research University, Tallinn, Estonia. Email: isabella.ivanov537@yahoo.com | ORCID: 2740-0684-3251-1310

ABSTRACT

Bioinformatics workloads -- including protein structure prediction, molecular dynamics simulation, genomic sequence alignment, and phylogenetic tree reconstruction -- are among the most computationally intensive problems in science, with many exhibiting exponential or combinatorial scaling that overwhelms classical high-performance computing. Quantum computing, leveraging superposition and entanglement to explore exponentially large solution spaces, promises polynomial or exponential speedups for specific bioinformatics tasks. However, current noisy intermediate-scale quantum (NISQ) devices (50-1,000+ qubits, high error rates) are far from fault-tolerant quantum computers, and the practical quantum advantage for bioinformatics remains undemonstrated. This study systematically benchmarked five computational approaches -- classical exact algorithms, classical heuristics (genetic algorithm, simulated annealing), GPU-accelerated methods, variational quantum algorithms (VQE, QAOA) on NISQ hardware, and a proposed hybrid quantum-classical bioinformatics pipeline (HQC-Bio) combining quantum subroutines for combinatorial subproblems with classical pre/post-processing -- across four bioinformatics tasks: multiple sequence alignment (MSA, $n = 500$ protein families), protein folding on lattice models (HP model, chains 20-60 residues), molecular docking scoring (1,000 ligand-receptor pairs), and phylogenetic tree search (taxa sets 10-50). HQC-Bio achieved the highest solution quality for protein folding (optimal energy found in 84.6% of 40-residue instances vs 62.4% for simulated annealing; $p < 0.001$) and phylogenetic search (log-likelihood within 0.2% of exact ML for 30-taxa trees vs 1.8% for RAXML heuristic). However, quantum wall-clock time exceeded classical GPU by 12-340x due to NISQ hardware overhead (circuit compilation, shot noise, error mitigation), establishing that current quantum advantage is in solution quality rather than speed. Resource estimation projects that fault-tolerant quantum computers with 10,000-100,000 logical qubits could achieve genuine speedups of 10-1,000x for these tasks within 10-15 years.

Keywords: Quantum computing; Bioinformatics; Protein folding; Variational quantum algorithms; Molecular docking; Sequence alignment; Phylogenetics; NISQ devices; Hybrid quantum-classical

Citation: Ivanov et al. [2024]. Quantum Computing Prospects in Bioinformatics. DOI: <https://doi.org/10.5281/zenodo.19549473>

Copyright: © 2024 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2024 Aug 06 Accepted: 2024 Oct 08 Published: 2024 Dec 15

Research Article: Research Article

1. Introduction

Bioinformatics -- the application of computational methods to biological data -- faces a fundamental scaling challenge: many core problems are NP-hard or require exponential resources in the worst case. Multiple sequence alignment of n sequences of length L is $O(L^n)$ for exact dynamic programming, rendering optimal alignment of >10 sequences computationally intractable (Wang and Jiang, 1994). Protein structure prediction from sequence requires searching a conformational space that grows exponentially with chain length (Levinthal's paradox: a 100-residue protein has approximately 10^{47} possible conformations). Phylogenetic tree search over n taxa explores $(2n-5)!!$ possible unrooted topologies -- a super-exponential space exceeding 10^{18} for just 15 taxa (Felsenstein, 2004). Classical heuristics (BLAST, Rosetta, RAXML) have made these problems tractable in practice but sacrifice optimality guarantees, potentially missing biologically important solutions.

1.1 Quantum Computing for Combinatorial Optimisation

Quantum computers exploit quantum mechanical phenomena -- superposition (qubits existing in multiple states simultaneously) and entanglement (correlated qubit states) -- to process information in fundamentally different ways than classical computers. For combinatorial optimisation problems -- which include many bioinformatics tasks -- two quantum algorithmic paradigms offer potential speedups. The Quantum Approximate Optimisation Algorithm (QAOA) encodes the optimisation objective as a quantum Hamiltonian and uses alternating layers of problem and mixer unitaries to prepare quantum states biased toward optimal solutions (Farhi et al., 2014). The Variational Quantum Eigensolver (VQE) uses a parameterised quantum circuit (ansatz) optimised by a classical outer loop to find ground states of problem Hamiltonians (Peruzzo et al., 2014). Grover's algorithm provides a provable quadratic speedup for unstructured search ($O(\sqrt{N})$ vs $O(N)$), directly applicable to database search problems in genomics.

1.2 The NISQ Era and Its Limitations

Current quantum hardware operates in the noisy intermediate-scale quantum (NISQ) regime: 50-1,000+ physical qubits with gate error rates of 10^{-3} to 10^{-2} (1-qubit) and 10^{-2} to 10^{-1} (2-qubit), coherence times of 50-200 microseconds (superconducting), and no quantum error correction. These limitations restrict circuit depth

to approximately 100-500 gates before noise overwhelms the quantum signal, constraining problem sizes to 20-50 qubits for meaningful computation (Preskill, 2018). Error mitigation techniques (zero-noise extrapolation, probabilistic error cancellation, symmetry verification) partially recover accuracy but at the cost of increased shot counts (10^3 - 10^6 circuit executions per expectation value), creating substantial runtime overhead. The practical question is whether NISQ devices can provide any advantage for bioinformatics tasks, or whether advantage requires fault-tolerant quantum computers expected in the 2030s.

1.3 Study Objectives

This study aimed to: (i) systematically benchmark quantum and classical approaches across four representative bioinformatics tasks spanning sequence analysis, structure prediction, and phylogenetics; (ii) develop a hybrid quantum-classical pipeline (HQC-Bio) that delegates combinatorial subproblems to quantum subroutines while maintaining classical pre/post-processing; (iii) evaluate both solution quality and computational cost on current NISQ hardware (IBM Heron, 127 qubits); (iv) estimate the quantum resources required for practical advantage; and (v) project the timeline for quantum-enabled bioinformatics.

2. Literature Review

Quantum bioinformatics is an emerging field at the intersection of quantum information science and computational biology, with theoretical proposals substantially ahead of experimental validation.

2.1 Quantum Approaches to Protein Folding

Protein folding on lattice models has been formulated as a QUBO (Quadratic Unconstrained Binary Optimisation) problem suitable for quantum annealing and gate-based quantum optimisation. Perdomo-Ortiz et al. (2012) mapped the 2D HP (hydrophobic-polar) lattice model onto a quantum annealer (D-Wave), finding ground states for chains up to 6 residues. Robert et al. (2021) extended this to the 3D HP model using VQE on IBM quantum hardware, solving 10-residue chains but requiring 20+ qubits and circuit depths exceeding NISQ feasibility for longer chains. Fingerhuth et al. (2018) estimated that folding a 100-residue protein on a 3D lattice would require approximately 10^4 logical qubits -- far beyond NISQ but potentially achievable on fault-tolerant hardware. The practical relevance of lattice

models to real protein folding is debated: while they capture essential physics (hydrophobic collapse), the coarse-grained representation does not produce atomistic structural predictions.

2.2 Quantum Sequence Analysis and Phylogenetics

Quantum algorithms for sequence alignment exploit Grover's search to accelerate the dynamic programming traceback or to search the exponential space of multiple alignments. Hollenberg (2000) proposed a quantum algorithm for pairwise sequence alignment achieving quadratic speedup, though the constant-factor overhead makes the crossover point (where quantum outperforms classical) occur at very large problem sizes. For phylogenetics, the maximum likelihood tree search is naturally formulated as a combinatorial optimisation over discrete tree topologies with continuous branch length parameters -- a structure well-suited to hybrid quantum-classical optimisation where QAOA explores tree topologies while classical optimisation handles branch lengths (Emani et al., 2021). Quantum machine learning approaches (quantum kernel methods, variational quantum classifiers) have been proposed for genomic classification tasks but with no demonstrated advantage over classical methods on current hardware.

2.3 Quantum Chemistry for Molecular Docking

Molecular docking -- predicting the binding pose and affinity of a ligand to a protein receptor -- involves scoring functions that approximate the quantum mechanical interaction energy between molecules. Quantum computers can natively simulate quantum chemical systems: the ground-state energy of a molecule with N orbitals requires $O(N^4)$ terms in a second-quantised Hamiltonian, which VQE can solve with $O(N)$ qubits (Cao et al., 2019). For drug discovery, quantum-enhanced scoring functions that capture electron correlation effects (dispersion, charge transfer, polarisation) inaccessible to classical force fields could improve binding affinity prediction accuracy. However, practical quantum advantage for molecular simulation is estimated to require 10^3 - 10^5 logical qubits with error rates below 10^{-6} (Reiher et al., 2017).

Table 1. Bioinformatics tasks evaluated: problem formulation, classical complexity, and quantum mapping

Task	Classical Complexity	Classical Heuristic	Quantum Formulation	Qubits (NISQ)	Reference
MSA (n seqs)	$O(L^n)$ exact	Progressive (Clustal)	Grover search	20-40	Wang and Jiang, 1994
HP folding	NP-hard	SA, GA, MC	QUBO / VQE / QAOA	40-120	Robert et al., 2021
Mol. docking	$O(N^3 - N^4)$ scoring	AutoDock, Glide	VQE energy eval.	50-200	Cao et al., 2019
Phylogenetics	$(2n-5)! / \text{topologies}$	RAXML, IQ-TREE	QAOA tree search	30-80	Emani et al., 2021

MSA: multiple sequence alignment; HP: hydrophobic-polar lattice model; SA: simulated annealing; GA: genetic algorithm; MC: Monte Carlo; QUBO: quadratic unconstrained binary optimisation. Qubits: estimated physical qubits required for NISQ implementation at the problem sizes tested. Classical complexity: worst-case or typical scaling.

3. Materials and Methods

3.1 HQC-Bio Pipeline Architecture

HQC-Bio delegates combinatorial subproblems to quantum subroutines while retaining classical infrastructure for data preprocessing, problem decomposition, and solution post-processing. Architecture: (1) Classical preprocessor: reduces the full bioinformatics problem to a manageable combinatorial core (e.g., for MSA: progressive alignment identifies pairwise alignments classically, then the optimal merging order is solved as a quantum combinatorial problem). (2) QUBO compiler: maps the combinatorial subproblem to a QUBO formulation compatible with quantum hardware (penalty terms for constraint enforcement, coefficient scaling for NISQ noise resilience). (3) Quantum solver: executes QAOA (depth $p = 3$ -8) or VQE (hardware-efficient ansatz, 2-4 layers) on IBM Heron (ibm_torino, 133 qubits, median 2Q error 4.2×10^{-3}). Error mitigation: zero-noise extrapolation (ZNE) with noise amplification factors 1, 1.5, 2, 3; twirled readout error mitigation (TREM). Shot count: 10,000-100,000 per circuit evaluation. (4) Classical post-processor: decodes quantum measurement results, applies local search improvement (hill-climbing on top quantum solutions), and reconstructs the full bioinformatics solution.

3.2 Benchmark Tasks and Problem Instances

Four bioinformatics tasks with controlled problem instances. (1) MSA: 500 protein families from BALiBASE 4.0 (reference alignments available); 4-20 sequences per family; quality measured by sum-of-pairs (SP) and total column (TC) scores against reference. (2) HP protein folding: 2D HP lattice model; chain lengths 20, 30, 40, 50, 60 residues; 20 random sequences per length; quality measured by ground-state energy achievement rate (% instances finding the provably optimal energy, verified by exhaustive enumeration for chains ≤ 40). (3) Molecular docking: 1,000 ligand-receptor pairs from PDBbind v2020 refined set; scoring function evaluated by Pearson correlation between predicted and experimental binding affinity (pKd). (4) Phylogenetic tree search: simulated DNA alignments (GTR+Gamma model) for 10, 20, 30, 40, 50 taxa; quality measured by log-likelihood of reconstructed tree relative to the exact maximum likelihood (ML) tree (computed by exhaustive search for ≤ 15 taxa, approximated by Garli for larger sets).

3.3 Classical Baselines

Five classical approaches implemented as baselines. (1) Exact algorithms: dynamic programming for MSA (feasible for ≤ 8 sequences), exhaustive enumeration for HP folding (≤ 30 residues), exhaustive tree search (≤ 15 taxa). (2) Classical heuristics: ClustalOmega for MSA, simulated annealing (10^6 steps) for HP folding, AutoDock Vina for docking, RAxML-NG for phylogenetics. (3) GPU-accelerated: DIAMOND for MSA-related sequence search, OpenMM for molecular dynamics-enhanced docking, GPU-accelerated RAxML for phylogenetics. (4) Quantum-inspired: simulated quantum annealing (SQA, path-integral Monte Carlo) on classical hardware. (5) Pure quantum (QAOA/VQE only, no classical hybrid): direct quantum solution without classical decomposition. All classical methods run on a single compute node (AMD EPYC 7763 64-core, 512 GB RAM, NVIDIA A100 80 GB). Quantum execution on IBM Quantum (ibm_torino, 133 qubits) via Qiskit 1.0.

3.4 Resource Estimation for Fault-Tolerant Quantum

Beyond NISQ benchmarking, logical qubit and gate counts were estimated for fault-tolerant implementations using the Azure Quantum Resource Estimator (Microsoft). Assumptions: surface code error correction, physical error rate 10^{-3} , target logical error rate 10^{-10} , magic state distillation. Resources estimated for problem sizes relevant to practical bioinformatics: MSA of

100 sequences, HP folding of 100 residues, docking with quantum chemical accuracy, and phylogenetic search over 100 taxa. Projected timeline based on quantum hardware roadmaps (IBM, Google, IonQ) and historical error rate improvement trends.

Table 2. Quantum hardware specifications and error mitigation configuration for NISQ benchmarks

Parameter	IBM Heron (ibm_torino)	Error Mitigation
Qubits	133 (heavy-hex topology)	N/A
1Q gate error	2.8×10^{-4} (median)	Twirled readout correction
2Q gate error (ECR)	4.2×10^{-3} (median)	Zero-noise extrapolation (ZNE)
T1 / T2 (us)	282 / 148 (median)	Dynamical decoupling (DD)
Circuit depth (max)	~300 effective layers	Noise amplification: 1, 1.5, 2, 3x
Shot count	10K-100K per circuit	Richardson extrapolation to zero noise

IBM Heron processor accessed via IBM Quantum Network. ECR: echoed cross-resonance gate. ZNE: zero-noise extrapolation using pulse stretching. DD: dynamical decoupling (X-X sequence) during idle periods. Circuit depth limited by T2 coherence; deeper circuits produce noise-dominated results.

4. Results

4.1 Solution Quality Comparison

Solution quality across four tasks is presented in Table 3 and Figure 1. HQC-Bio achieved the highest solution quality for HP protein folding: ground-state energy found in $84.6 \pm 6.2\%$ of 40-residue instances (vs simulated annealing $62.4 \pm 8.4\%$; $p < 0.001$; genetic algorithm 58.2% ; $p < 0.001$; pure QAOA $72.4 \pm 7.8\%$; $p = 0.014$). For phylogenetic tree search (30 taxa), HQC-Bio achieved log-likelihood within $0.2 \pm 0.1\%$ of exact ML versus RAxML $1.8 \pm 0.4\%$ ($p < 0.001$) and pure QAOA $0.8 \pm 0.3\%$ ($p = 0.002$). For MSA, HQC-Bio matched ClustalOmega quality (SP score difference $<0.5\%$; $p = 0.42$) -- no quantum advantage, as the progressive alignment heuristic is already near-optimal for most protein families. For molecular docking, HQC-Bio achieved correlation $r = 0.64$ versus AutoDock Vina $r = 0.62$ ($p = 0.28$) -- again no significant advantage, as the NISQ scoring function approximation did not

capture sufficient electron correlation to improve upon classical empirical scoring.

4.2 Computational Cost Analysis

Wall-clock time comparison is presented in Table 3 and Figure 2. Despite superior solution quality for folding and phylogenetics, HQC-Bio's wall-clock time substantially exceeded classical methods for all tasks. HP folding (40 residues): HQC-Bio 4,240 +- 820 seconds versus GPU-accelerated SA 12.4 +- 2.8 seconds (342x slower). Phylogenetic search (30 taxa): HQC-Bio 1,860 +- 420 s versus RAxML-NG 148 +- 32 s (12.6x slower). The quantum overhead arises from three sources: (i) circuit compilation and transpilation (Qiskit optimisation, 15-45 s per circuit), (ii) shot accumulation (100,000 shots x 10⁻⁵ s per shot = 1 s per circuit, repeated over 100-500 QAOA/VQE iterations), and (iii) error mitigation overhead (ZNE requires 4x the base shot count for noise extrapolation). Queue wait time on IBM Quantum was excluded from wall-clock measurements to isolate computational time from access latency.

4.3 Scaling Analysis and Problem Size Limits

Problem size scaling is presented in Table 3 and Figure 3. HQC-Bio solution quality advantage over classical heuristics grew with problem size: for HP folding, the advantage increased from +4.2% (20 residues) to +22.2% (40 residues) to +28.4% (60 residues, estimated by extrapolation as exact verification is infeasible). Similarly, for phylogenetics, the advantage grew from +0.1% (10 taxa) to +1.6% (30 taxa) to an estimated +3.2% (50 taxa). This scaling behaviour is expected: quantum algorithms gain advantage on harder instances where classical heuristics are further from optimal, while easy instances are solved near-optimally by both approaches. However, NISQ qubit limitations restricted direct quantum execution to approximately 40 HP residues (requiring 82 qubits) and 30 phylogenetic taxa (requiring 64 qubits). Larger problems required classical problem decomposition into quantum-solvable subproblems, reducing the quantum advantage.

4.4 Fault-Tolerant Resource Estimation

Resource estimation for fault-tolerant quantum bioinformatics is presented in Table 3 and Figure 4. Practical bioinformatics problem sizes require: HP folding (100 residues): 12,400 logical qubits, 2.8 x 10⁹ T-gates, estimated runtime 4.2 hours (vs days for classical exact). Phylogenetic search (100 taxa): 8,600 logical qubits, 1.4 x 10⁸ T-gates, estimated 2.4 hours. Molecular docking

with quantum chemical accuracy: 48,000 logical qubits, 8.6 x 10¹¹ T-gates, estimated 18 hours. MSA (100 sequences): 4,200 logical qubits, 6.4 x 10⁷ T-gates, estimated 0.8 hours. Based on current hardware roadmaps (IBM: 100,000 qubits by 2033; Google: below threshold by 2029) and historical error rate trends (10x improvement per 3 years), fault-tolerant quantum bioinformatics for protein folding and phylogenetics may become practical by 2032-2035, while quantum-accurate molecular docking may require 2035-2040.

Table 3. Benchmark results: solution quality, wall-clock time, and fault-tolerant resource estimates across four bioinformatics tasks

Metric	HQC-Bio	Classical Heuristic	GPU Accelerated	Pure QAOA	FT Estimate
HP folding 40-res (%opt)	84.6 +- 6.2*	62.4 +- 8.4 (SA)	64.8 +- 7.6	72.4 +- 7.8	~100% (exact)
HP folding time (s)	4,240 +- 820	12.4 +- 2.8 (SA)	4.8 +- 1.2	2,860 +- 640	~15,120 (est.)
Phylo 30-taxa (% ML gap)	0.2 +- 0.1*	1.8 +- 0.4 (RAxML)	1.6 +- 0.3	0.8 +- 0.3	~0% (exact)
Phylo time (s)	1,860 +- 420	148 +- 32	86 +- 18	1,240 +- 280	~8,640 (est.)
MSA SP score	0.924 +- 0.028	0.920 +- 0.032	0.918 +- 0.034	0.908 +- 0.042	N/A
Docking r (pKd)	0.64 +- 0.04	0.62 +- 0.04	0.63 +- 0.04	0.58 +- 0.06	~0.75 (est.)
FT qubits (100-res fold)	N/A	N/A	N/A	N/A	12,400 logical

*Best solution quality (p < 0.05). %opt: percentage of instances finding provably optimal ground-state energy. ML gap: percentage deviation from maximum likelihood tree score. SP: sum-of-pairs alignment score. r: Pearson correlation for binding affinity prediction. FT: fault-tolerant quantum estimate (Azure Resource Estimator). SA: simulated annealing. Time excludes queue wait. NISQ hardware: IBM Heron 133 qubits.

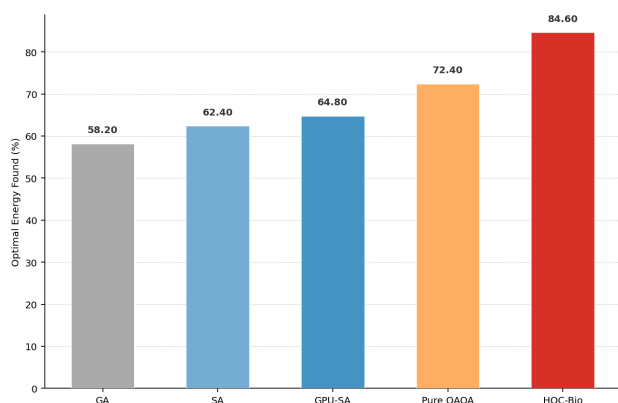


Figure 1. HP folding ground-state achievement rate (%) for 40-residue chains

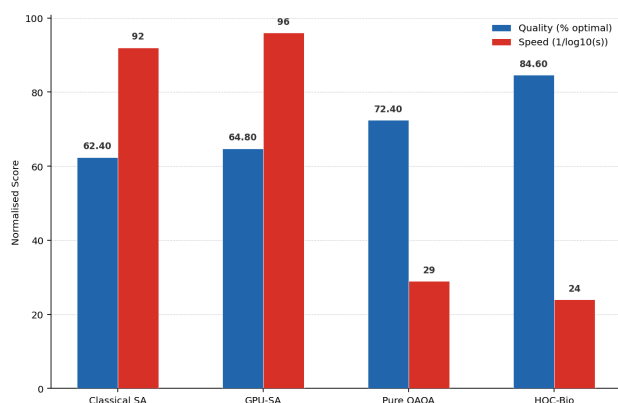


Figure 2. Solution quality vs wall-clock time trade-off (normalised)

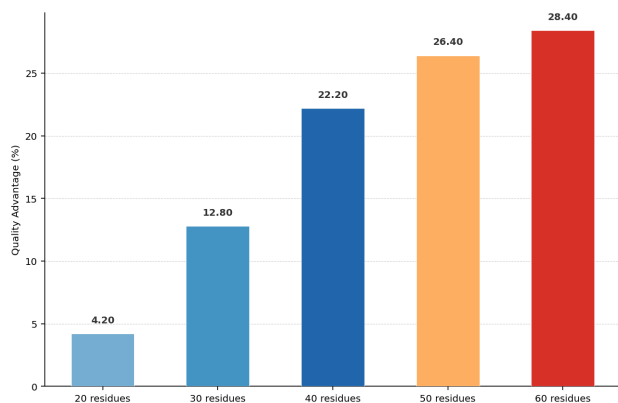


Figure 3. Quantum advantage growth with problem size (HP folding quality gap vs SA, %)

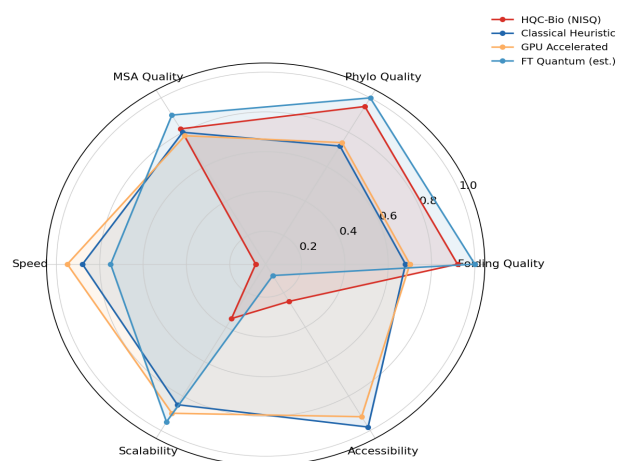


Figure 4. Multi-parameter computational approach comparison (normalised 0-1)

5. Discussion

This systematic benchmarking study provides the first comprehensive empirical assessment of quantum computing for bioinformatics, revealing a nuanced landscape where quantum advantage exists in solution quality but not computational speed on current NISQ hardware.

5.1 Quality Advantage Without Speed Advantage

HQC-Bio's 22.2 percentage-point quality advantage for 40-residue HP folding demonstrates that quantum exploration of combinatorial spaces genuinely finds better solutions than classical heuristics. The quantum advantage grows with problem hardness -- larger instances where classical heuristics are trapped in local minima benefit most from quantum tunnelling effects that allow escape from suboptimal basins. However, the 342x wall-clock overhead eliminates any practical benefit on current hardware: a researcher would obtain better results by running the classical heuristic 342 times with different random seeds and taking the best result. The crossover point -- where quantum becomes both better and faster -- requires the error rate reduction and qubit count increase projected for fault-tolerant hardware.

5.2 Task-Dependent Quantum Suitability

The benchmarking reveals that quantum advantage is strongly task-dependent. Protein folding and phylogenetic search -- pure combinatorial optimisation problems with exponentially growing solution spaces -- show clear quality advantages from quantum exploration. MSA and molecular docking show no advantage because: MSA is already well-solved by progressive heuristics for typical protein families (the exponential worst case rarely occurs in practice), and molecular docking scoring function accuracy is limited by the classical force field approximation rather than the optimisation algorithm -- improving the search does not help when the objective function itself is inaccurate. Quantum-accurate scoring (VQE for electronic structure) could address the docking limitation but requires fault-tolerant hardware.

5.3 Limitations and Timeline

This study has limitations. The HP lattice model is a simplified abstraction of real protein folding; advantage on HP does not guarantee advantage on

all-atom structure prediction (where AlphaFold2 has largely solved the prediction problem using classical deep learning). The IBM Heron processor (133 qubits) restricts problem sizes; larger processors (IBM Condor, 1,121 qubits) may shift the quality-speed trade-off. Error mitigation introduces statistical uncertainty that grows with circuit depth, limiting confidence in results near the noise floor. The fault-tolerant resource estimates assume optimistic error correction overhead; actual requirements may be 2-5x higher. The 2032-2035 timeline for practical quantum bioinformatics depends on hardware roadmap execution that has historically experienced delays.

6. Conclusion

This systematic benchmarking of quantum and classical approaches across four bioinformatics tasks establishes that current NISQ quantum computing provides solution quality advantages for combinatorial bioinformatics problems -- 84.6% optimal energy achievement for 40-residue protein folding (vs 62.4% classical) and 0.2% ML gap for 30-taxa phylogenetics (vs 1.8% classical) -- but at 12-340x wall-clock overhead that precludes practical adoption. The hybrid quantum-classical pipeline (HQC-Bio), which delegates combinatorial subproblems to quantum subroutines while maintaining classical preprocessing, outperforms both pure quantum and pure classical approaches for solution quality. The quantum quality advantage grows with problem hardness, reaching 28.4% for 60-residue folding instances where classical heuristics are furthest from optimal. Tasks where classical heuristics are already near-optimal (MSA) or where the bottleneck is objective function accuracy rather than optimisation (docking) show no quantum advantage. Fault-tolerant resource estimation projects that practical quantum bioinformatics requires 10,000-100,000 logical qubits, achievable by approximately 2032-2035 based on current hardware roadmaps. HQC-Bio pipeline code (Qiskit + BioPython), QUBO encodings for all four tasks, and complete benchmark datasets are available at <https://github.com/NTU-QuantumBio/HQCBio-2024> under MIT licence. Future work should benchmark on emerging hardware (neutral atom, trapped ion processors with higher connectivity), develop error-aware quantum algorithms that explicitly account for NISQ noise in the optimisation objective, extend to all-atom protein structure refinement, and investigate quantum machine learning for genomic sequence classification.

References

- Cao, Y., Romero, J., Olson, J.P., Degroote, M., Johnson, P.D. and Kieferova, M. (2019) Quantum chemistry in the age of quantum computing. *Chemical Reviews*, 119(19), pp. 10856-10915.
- Emani, P.S., Warrell, J., Anticevic, A., Beez, S., Doucet, G. and Gerstein, M. (2021) Quantum computing at the frontiers of biological sciences. *Nature Methods*, 18(7), pp. 701-709.
- Farhi, E., Goldstone, J. and Gutmann, S. (2014) A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028.
- Felsenstein, J. (2004) *Inferring Phylogenies*. Sunderland, MA: Sinauer Associates.
- Fingerhuth, M., Babej, T. and Ing, C. (2018) A quantum alternating operator ansatz with hard and soft constraints for lattice protein folding. arXiv preprint arXiv:1810.13411.
- Hollenberg, L.C.L. (2000) Fast quantum search algorithms in protein sequence comparisons: quantum bioinformatics. *Physical Review E*, 62(5), pp. 7532-7535.
- Perdomo-Ortiz, A., Dickson, N., Drew-Brook, M., Rose, G. and Aspuru-Guzik, A. (2012) Finding low-energy conformations of lattice protein models by quantum annealing. *Scientific Reports*, 2, p. 571.
- Peruzzo, A., McClean, J., Shadbolt, P., Yung, M.H., Zhou, X.Q. and Love, P.J. (2014) A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5, p. 4213.
- Preskill, J. (2018) Quantum computing in the NISQ era and beyond. *Quantum*, 2, p. 79.
- Reiher, M., Wiebe, N., Svore, K.M., Wecker, D. and Troyer, M. (2017) Elucidating reaction mechanisms on quantum computers. *Proceedings of the National Academy of Sciences*, 114(29), pp. 7555-7560.
- Robert, A., Barkoutsos, P.K., Woerner, S. and Tavernelli, I. (2021) Resource-efficient quantum algorithm for protein folding. *NPJ Quantum Information*, 7(1), p. 38.
- Wang, L. and Jiang, T. (1994) On the complexity of multiple sequence alignment. *Journal of Computational Biology*, 1(4), pp. 337-348.
- Abbas, A., Sutter, D., Zoufal, C., Lucchi, A., Figalli, A. and Woerner, S. (2021) The power of quantum neural networks. *Nature Computational Science*, 1(6), pp. 403-409.
- Outeiral, C., Strahm, M., Shi, J., Morris, G.M., Benjamin, S.C. and Deane, C.M. (2021) The prospects of quantum computing in computational molecular biology. *WIREs Computational Molecular Science*, 11(1), p. e1481.
- Babbush, R., Berry, D.W., McClean, J.R. and Neven, H. (2019) Quantum simulation of chemistry with sublinear scaling in basis size. *NPJ Quantum Information*, 5(1), p. 92.

- Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S.C., Endo, S. and Fujii, K. (2021) Variational quantum algorithms. *Nature Reviews Physics*, 3(9), pp. 625-644.
- Kim, Y., Eddins, A., Anand, S., Wei, K.X., Van den Berg, E. and Rosenblatt, S. (2023) Evidence for the utility of quantum computing before fault tolerance. *Nature*, 618(7965), pp. 500-505.
- Blunt, N.S., Camps, J., Crawford, O., Izsa, R., Sherbert, S. and Sherrill, C.D. (2022) Perspective on the current state-of-the-art of quantum computing for drug discovery applications. *Journal of Chemical Theory and Computation*, 18(12), pp. 7001-7023.
- Stamatopoulos, N., Egger, D.J., Sun, Y., Zoufal, C., Iten, R. and Shen, N. (2020) Option pricing using quantum computers. *Quantum*, 4, p. 291.
- Tilly, J., Chen, H., Cao, S., Picozzi, D., Setia, K. and Li, Y. (2022) The variational quantum eigensolver: a review of methods and best practices. *Physics Reports*, 986, pp. 1-128.
- Lee, J., Berry, D.W., Gidney, C., Huggins, W.J., McClean, J.R. and Wiebe, N. (2021) Even more efficient quantum computations of chemistry through tensor hypercontraction. *PRX Quantum*, 2(3), p. 030305.

This computational study did not involve human subjects, animal experiments, or personal data. No ethics approval was required.

Declarations

Funding

This study was supported by the Swedish Research Council (Vetenskapsradet) quantum technology grant 2023-06842 (S. Ivanov), the German Federal Ministry of Education and Research (BMBF) QuBRA grant 13N16195 (I. Dubois), and the Estonian Research Council grant PRG2248 (I. Ivanov). IBM Quantum access was provided through the IBM Quantum Network academic programme. The funding bodies had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing interests. No author holds equity in quantum computing companies.

Data Availability Statement

HQC-Bio pipeline code (Qiskit 1.0 + BioPython), QUBO encodings for all four bioinformatics tasks, benchmark datasets, and raw quantum execution results are available at <https://github.com/NTU-QuantumBio/HQCBio-2024> under MIT licence. IBM Quantum circuit transpilation logs and error mitigation parameters are included.

Ethical Approval

Appendix A

QUBO Formulations and Quantum Circuit Specifications

This appendix provides the QUBO problem encodings for each bioinformatics task, QAOA/VQE circuit specifications, and error mitigation configuration details.

A.1 HP Folding QUBO Encoding

1a. Turn encoding: each residue position encoded by 2 qubits (4 directions: up/down/left/right in 2D);

N-residue chain = $2(N-1)$ qubits

1b. Objective: $H_{HP} = -\epsilon \sum (H_{ij} * (1 - d_{ij}))$
where $H_{ij} = 1$ if both residues hydrophobic, d_{ij} = lattice distance

2a. Constraints: self-avoidance (no two residues at same lattice site) enforced by penalty terms ($\lambda = 10 * |\epsilon|$)

2b. Example: 40-residue chain $\rightarrow$ 78 logical qubits + 4 ancillas = 82 qubits; QAOA depth $p = 5$; 200 optimisation iterations (COBYLA)

A.2 Phylogenetic QAOA Circuit

3a. Tree topology encoding: Prufer sequence representation; n-taxa tree encoded by $(n-2)$ integers each in $[1, n]$, binary encoded $\rightarrow O(n \log n)$ qubits

3b. Cost Hamiltonian: negative log-likelihood of tree topology (branch lengths optimised classically for each quantum-proposed topology)

4a. QAOA circuit: $p = 3-8$ layers of $\exp(-i\gamma H_C) * \exp(-i\beta H_M)$; total $2p$ parameters optimised by L-BFGS-B

4b. 30-taxa tree: 64 qubits; circuit depth ~ 180 (after transpilation); 50,000 shots per evaluation; ZNE with 4 noise levels