

A Comparative Analysis of Quantum Computing Variational Quantum Eigensolver Algorithm and Molecular Dynamics Simulations for Peptide Folding

Akshay Uttarkar

R V College of Engineering

Vidya Niranjana (✉ vidya.n@rvce.edu.in)

R V College of Engineering

Research Article

Keywords: Quantum Computing, Protein folding, Variational quantum eigensolver, MD simulation

Posted Date: October 27th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3460426/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Quantum Information Processing on February 6th, 2024. See the published version at <https://doi.org/10.1007/s11128-024-04261-9>.

Abstract

Quantum computing in biology is one of the most rapidly evolving field of technology. Protein folding is one of the key challenges which requires accurate and efficient algorithms with a quick computational time. Structural conformations of proteins with disordered regions need colossal amount of computational resource to map its least energy conformation state. In this regard, quantum algorithms like Variational quantum eigensolver (VQE) are applied in the current research work to predict the lowest energy value of 50 peptides of 7 amino acids each. VQE is initially used to calculate the energy values over which Variational Quantum Optimization is applied via Conditional Value at Risk (CVaR) over 100 iterations of 500000 shots each to obtain least ground state energy value. This is compared to the molecular dynamics-based simulations of 50 nanoseconds each to calculate the energy values along with the folding pattern. The results suggest efficient folding outcomes from CvaR-VQE compared to MD based simulations. With the ever-expanding quantum hardware and improving algorithms the problem of protein folding can be resolved to obtain in depth insights on the biological process and drug design.

1. Introduction

Protein folding is a complex process that involves the spontaneous collapse of a linear polypeptide chain into a three-dimensional structure. This process is essential for protein function, as the three-dimensional structure of a protein determines its biological activity. Molecular dynamics (MD) simulations are a powerful tool for studying protein folding. MD simulations can be used to simulate the motion of atoms in a protein over time, and to track how the protein's structure changes as it folds. This information can be used to understand the factors that influence protein folding, and to design new proteins with desired properties.

One of the most important applications of MD simulations in protein folding is to study the folding kinetics of proteins. Folding kinetics refers to the rate at which a protein folds, and the pathway that it takes to reach its native state. MD simulations can be used to measure the folding times of proteins, and to identify the key steps involved in the folding process [1]. Use of enhanced sampling techniques, such as replica exchange molecular dynamics and meta-dynamics, to overcome the limitations of traditional molecular dynamics simulations. They discuss the use of these techniques to study protein folding and the advantages and disadvantages of each approach [2]. MD simulations have also been used to study the thermodynamics of protein folding. The thermodynamics of protein folding refers to the energy changes that occur during folding. MD simulations can be used to calculate the free energy of different protein conformations, and to identify the most stable conformation [3][4].

MD simulations have also been used to study the effects of mutations and environmental factors on protein folding. Mutations can alter the folding kinetics and thermodynamics of proteins and can lead to misfolded proteins that are associated with diseases such as Alzheimer's and Parkinson's. MD simulations can be used to predict how mutations will affect protein folding, and to design drugs that can stabilize misfolded proteins [5]. In recent years, there have been significant advances in the accuracy and

efficiency of MD simulations. This has made it possible to simulate the folding of increasingly large and complex proteins. For example, in 2016, researchers at the University of Washington used MD simulations to simulate the folding of a protein called villin, which is one of the largest and most complex proteins known [6].

One of the main drawbacks of MD simulations is that they are computationally expensive. This is because MD simulations need to track the motion of every atom in a protein over time, which can require a lot of computing power. Quantum folding is a promising new approach that has the potential to overcome the limitations of MD simulations.

Quantum folding simulations are still in their early stages of development, but they have the potential to revolutionize the study of protein folding. Quantum folding simulations could be used to design new proteins with desired properties, and to develop new drugs and therapies for diseases that are associated with protein misfolding and regions of intrinsic disorder of proteins [7].

Variational quantum eigensolvers (VQEs) are a type of quantum algorithm that can be used to calculate the ground state energy of a quantum system. VQEs work by iteratively preparing a quantum circuit and measuring its output. The quantum circuit is parameterized by a set of variational parameters, which are adjusted at each iteration to minimize the energy of the system. Since, iteration-based minimization of energy is an optimal solution for protein folding, VQE's are one of the important approaches in quantum protein folding [8].

In the paper [9], proposes a model of Hamiltonian with $O(N^4)$ scaling and a robust optimization scheme. The algorithm is successfully applied to the folding of a 10 amino acid and a 7 amino acid peptide. The model proposes a efficient system with noise tolerant quantum algorithms.

Few theoretical papers have proposed that the k UpCCGSD ansatz has been shown to achieve excellent accuracy while offering linear scaling, making it a good trade off in cost to accuracy among ansätze proposed at the time of writing. Adaptive ansätze could be a reliable alternative subject to further studies on their expected computational cost. Layered (Hardware-Efficient ansatz) HEA are likely not suitable for large systems as they require a large number of parameters and need to span a significant proportion of the Hilbert space to guarantee a good ground state approximation can be reached. The excerpt also mentions that VQE may be resilient to barren plateaus if the right design choices are made in the ansatz, such as local encodings for the Hamiltonian, an ansatz that is problem tailored, constructed adaptively during the optimization, using specific initialization techniques, and use of a local mapping for the Hamiltonian [10–17].

Recently, a hybrid classical-quantum digitized-counter diabatic algorithm to tackle the protein folding problem on a tetrahedral lattice. They outperform state-of-the-art quantum algorithms using problem-inspired and hardware-efficient variational quantum circuits. They applied the method to proteins with up to 9 amino acids, using up to 17 qubits on quantum hardware [18, 19].

VQEs work by optimizing a cost function that contains information about the solution to the problem. The quantum part of a VQA consists of a parameterized quantum circuit (PQC), also known as a circuit ansatz, which is used to generate trial quantum states. The classical part of a VQA consists of an optimization routine that is used to find the optimal parameters for the PQC [20, 21]

The choice of PQC has a significant impact on the performance of a VQA. PQCs can be broadly divided into two categories: problem-inspired and hardware-efficient. Problem-inspired ansätze are designed to exploit the properties of the problem Hamiltonian to efficiently reach the desired state. Hardware-efficient ansätze are designed to minimize the noise introduced by deep circuits and unimplementable connections [22, 23].

In the current work, we are comparing the folding of 50 peptides consisting of amino acids which are in abundance in the disordered regions of proteins. The methods under comparison are MD simulation and VQE. The comparison provides a insights on pros and cons of both the methods and provides an unbiased rationale for the future of protein folding specially for regions with disorderedness.

2. Materials

2.1. Configuration of Qubits and Hamiltonians

The 50 proteins sequences selected are of 7 amino acids in length. The study assumes the coarse grained (CG) model of proteins. Amino acids, symbolized as "beads," can move across the lattice and engage in interactions with one another. The connections between amino acids can assume one of four orientations, corresponding to the corners of a tetrahedron. These four orientations can be encoded using a pair of qubits.

2.1.1. Qubit Configuration

A protein comprising N beads, where N equals 7 in this context, can undergo $N-1$ rotations. Consequently, a total of 12 bits are essential to describe the bonds within the 7 amino acid peptides. It's worth noting that, without limiting generality, the first two rotations can be designated as 01 and 00. Additionally, one of the bits in the third rotation is predetermined due to symmetry considerations. The mapping known as "turn2qubit" illustrates these 12 bits, encompassing the values of the 5 fixed bits and the 7 variable bits, which will be represented by qubits. In the current work, the configuration was set to "0011q1qqqqq" For more in-depth information, please refer to the comprehensive encoding scheme outlined [9].

2.1.2 Interaction Qubits

The techniques outlined in references [9] and [18] are capable of accounting for interactions among any number of nearest neighbours (NN). However, in this specific instance, we are only examining interaction terms involving 1-NN, which are feasible for certain beads based on the lattice's configuration. It's important to note that beads separated by fewer than 5 bonds cannot establish a 1-NN relationship. The equation for the calculation is provided in Eq. 1.

$$H^{(1)} = \sum_{i=1}^{N-4} \sum_{\substack{j \geq i+5 \\ j-i=1}}^N h_{i,j}^{(1)}$$

1

For pairwise interactions between various amino acids, a look-up table of contact energy values is utilised.

2.1.3 Hamiltonian function

The energy for any protein fold conceivable is determined by the exact Hamiltonian function. Only 1-NN interactions are considered, and irrational configurations result in energy penalties.

For the sparser encoding, we need to impose that one and only of the four qubits that define a turn is equal to one. On a quantum circuit, this can be achieved easily by using a valid initialization of the qubits together with gate operations that conserved the number of 1's in each set of 4 qubits that encodes a turn.

When this is not possible, a penalty function with a large positive λ can be used to impose this constraint.

Several constraints are needed to prevent the growths of the chain towards unphysical geometries (e.g., to prevent that the chain (main or side) at side i folds back into itself). To this end, we compute function $T(i,j)$ defined as in Eq. 2

$$T(i, j) = \sum_{a=\{0,1,2,3\}} f_a(i) f_a(j)$$

(2)

for each pair of beads i and j . $T(i,j)$ returns a 1 if and only if the turns t_i and t_j are along the same axis (a and a , respectively). Note that $T(i,j)$ is composed of 2-local terms for the sparse encoding. Firstly, we need to eliminate sequences where the same axis (a and a) is chosen twice in a row. since this will give rise to a chain folding back into itself. To this end we apply a penalty the following penalty term in Eq. 3.

$$H_{gc} = \sum_{i=3}^{N-1} \lambda_{back} T(i, i+1)$$

(3)

with large positive λ "back". Note that we can easily control the number of 2-local terms appearing in the sum (linear number of two local terms). In the general case, for a degree of branching $s > 1$, similar terms need to be added to prevent the overlap of two consecutive bonds within the side chains.

Natural polymers have a well-defined chirality that must be imposed in our model. In proteins, the position of the side chains at the insertion point with the main chain determines the chirality of each residue. The

position of the first side chain bead i^1 on the main bead i is imposed by the choice of the (main chain) turns $t_{(i-1)}$ and t_i .

To enforce the correct chirality, we add a constraint H_{cr} to the Hamiltonian. The required (expected) chirality at i^1 is encoded in the function $f_a^{ex}(i^{(1)})$ which is a function of $f_a(i-1)$ and $f_a(i-1)$.

2.2 Selection of sequences

For disorder prediction, separating disordered from ordered proteins is crucial. Identifying biases in the amino acid composition is one of the first stages in identifying a characteristic that separates IDPs from non-IDPs. The amino acids W, C, F, I, Y, V, L, and N are hydrophobic and uncharged, whereas the hydrophilic, charged amino acids A, R, G, Q, S, P, E, and K have been identified as amino acids that promote order [24]. The remaining amino acids, H, M, T, and D, can be present in both structured and unstructured regions, making them confusing. In a more recent investigation, amino acids were rated according to how likely they were to generate disordered areas. W, F, Y, I, M, L, V, N, C, T, A, G, R, D, H, Q, K, S, E, P list of residues mentioned order promoting to disorder promoting [25].

The 7 residues with most disorder D, H, Q, K, S, E, P was selected. Number of possible permutations 7P_6 is 5040. Out of this random 50 sequences were selected for analysis.

2.3 Protein Folding and Variational Quantum Eigensolvers

To fully model a fold, nine qubits are needed (7 for setup and 2 for interaction). To determine the fold with the least amount of energy, use the exact Hamiltonian function directly with each possible bit string (each representing a potential fold). This value can afterwards be compared to the outcome of the quantum-based optimisation. For tiny proteins, this exhaustive search is feasible, but it is impractical for larger proteins because of the exponential growth in the number of possible fold combinations. The number shots were set to 500000 in the local quantum simulator for accurate results.

We convert the optimisation problem over the set of bit strings into a different optimisation problem, this time one over the set of angles between π and $-\pi$, using a quantum circuit. The angles between the protein beads and these angles are unrelated. The quantum circuit uses the set of angles provided by the goal function to generate a variety of bit strings with differing probability. Invoke the exact Hamiltonian function on the bit strings that the quantum circuit returned and set the goal function's value to the weighted average of the lesser energies. The number of iterations for each peptide was set to 100.

Specifically, after the energy is computed for each observed fold, the associated probabilities are sorted by energy. The objective function returns an expectation energy computed from the tail end of the probability distribution, cutoff by an alpha parameter. This expectation energy is a conditional value at risk (CVaR). An alpha value of 0.05 was used experimentally, but for a noise-free simulation ProteinVQEObjective uses a smaller cutoff value of 0.025.

2.4 Creation of Circuit Ansatz

The configuration is represented by qubits 1 through 7, while the interaction is represented by the remaining three qubits (including a helper qubit that is not measured). An effective quantum structure architecture for VQE is available at [26].

2.5 Iterative Simulations of CVaR-VQE

Over the configuration and interaction qubits, there are two levels of R_x , R_y and R_z rotation gates. The variational circuit uses these rotation degrees as learnt parameters. To identify the set of R_x , R_y and R_z rotation angles that yield the lowest expected energy, use the *surrogateopt* (Global Optimisation Toolbox) algorithm with the objective function. The rotation angles are determined via a genetic method in the original work [9]. Although *surrogateopt* converges to values that are comparable, its internal use of intervals causes its convergence behaviour to be noticeably different.

2.6 MD Simulation

The protein sequences were imported into 3D builder of Maestro, Schrodinger. The sequence was entered an extended structure was built. The structures predicted, were subjected to protein preparation [27] using EpiK [28, 29] and OLPS3e [30] force field was added. The peptides were not minimized. An orthorhombic system was built for simulation with TIP3P [31] as solvent and no additional ions were added to naturalize the system and maintain its native nature.

Desmond [32] was used to run MD simulations for 50 ns each. A detailed steps involved in minimization and equilibration is available in [33, 34]. Using the trajectory files the protein structures at an interval of every 10ns were extracted to capture the folding pattern. A total of 6 structure for each sequence (0,10,20,30,40,50th frame) was subjected to minimization and evaluation of LJ interaction energy.

3. Results

3.1 Molecular dynamics-based simulations to achieve protein folding event.

The MD simulation was carried out for 50 sequences for 50 nanoseconds to verify and analyse the simulation for every 10 nanoseconds as mentioned in the methods section 2.6.

The total interaction energy was calculated for each of the frame (6 frames) and compared the time frame at which the least energy was achieved. From the analysis it was observed that not all peptides fold to achieve the energy $< \Delta G = 0$ which is stable and thermodynamically feasible.

Over the period of simulation, a mean energy values for 50 sequences at native energy fold (0th frame) is found to be 9.586 kcal/mol, at 10 nanoseconds it is 8.484 kcal/mol, at 20 nanoseconds it is 6.663 kcal/mol, at 30 nanoseconds it is 6.154 kcal/mol and 40 nanoseconds it is at 3.781 kcal/mol and at 50 nanoseconds it is at 2.684 kcal/mol. The results suggest that over the simulation period, the protein folding occurs better with higher simulation period as suggested by the mean value and it is aligned to the

conventional understanding of protein folding theories with MD simulation. Energy values at each time frame along with mean and outliers are shown in Fig. 1A.

Nevertheless, it doesn't always refer to the time at which the least energy was achieved only after the end of simulation which is 50ns in our case.

In the observations made in the current research work that minimum energy ground state was achieved for 26 peptides (52%) at 50 ns, 6 peptides (12%) at 40 ns, 9 peptides (18%) at 30 ns, 8 peptides (16%) at 20 ns and 1 peptide (2%) at 10 ns. The number of minimum energy events occurred is shown in Fig. 1B.

A spline fitting curve in which the data is split into multiple polynomial pieces each of the degree K . The polynomial pieces are joined together at their endpoints in such a way that the function is continuous to order $K - 1$ at each join point. It was plotted to obtain the best simulation time frame to obtain ground state energy which is found to be the range of 40 to 50 nanoseconds at an approximation of -4 kcal/mol which also in total corresponds to 64% of peptides under consideration. The spline fit curve is plotted in Fig. 1C.

Figure 1: A) Box and whisker showing the least ground state energy values for all the peptides at specific time frame across the simulation time frame. The average value is shown with a black line within the box. B) A bar plot with count of minimum energy event attained by the peptide at specific simulation time. C) The spline fitting curve with energy in kcal/mol versus time in nanoseconds.

Sequence 16 with a peptide sequence, "DPSHQKE" has shown a classical mode of peptide folding. The energy values at native state is 10.65 kcal/mol, at 10ns it is 1.23 kcal/mol, at 20ns it is 4.23 kcal/mol, at 30ns it observed to be 6.46 kcal/mol, similarly at 40ns at it is 2.62 kcal/mol and finally at 50ns it is -1.11 kcal/mol reaching a energy value with $< \Delta G = 0$.

The peptide gradually folds to achieve ground state energy over the period of simulation for 50ns and achieved an RMSD change of 4Å at the end of simulation period. Figure 2A provides a comprehensive view of the peptide structures its folding pattern with energy values in kcal/mol overlaid on the root mean square deviation (RMSD) plot versus time.

On the contrary, sequence 19 with a sequence "KSQHPED" has shown the folding pattern wherein it reached the lowest energy value of -0.98 kcal/mol at time frame of 10ns from 6.32 kcal/mol from the native state. At 20ns the energy value was found to be 0.04 kcal/mol, at 30ns it is at 5.78 kcal/mol, at 40ns it is observed to be 2.52 kcal/mol and finally 1.06 kcal/mol at 50ns at the end of simulation period. In Fig. 2B, a peptide folding over the period of simulation is provided overlaid on RMSD plot along with energy values in kcal/mol.

As previously mentioned, 52% of the peptides reached the lowest ground state energy at the end of simulation period. In other cases, the lowest ground state energy is reached during intermediate phases of simulation. This has always been a potential drawback of MD simulation wherein the time frame of achieving lowest energy state is not always time based and setting a threshold value is not feasible.

This clearly suggests that while simulation is quintessentially important for achieving for the ground state energy and mere prediction of 3D structure is not enough for the prediction of ground state energy.

From the simulations, 36 were achieved ground state energy $< \Delta G = 0$ within 50 ns, whereas 14 peptides were $> \Delta G = 0$ even after completion of 50ns.

Ground state energy values for each peptide at different time frames along with the least energy value and statistical data (Mean and Standard deviation). All the values in kcal/mol. The complete data is provided in Supplementary File 1.

3.2 Protein folding using VQE

Protein folding using VQE which is a quantum algorithm using resource efficient method using Amazon Bracket local simulator and SV1 state vector simulator. The 50 peptides were subjected to VQE based protein folding. The ground state energy for each of the peptide was recorded along with the best outcome qubit configuration outcome.

The quantum ansatz circuit was designed with enhanced number of conditional rotational gates compared to previous studies [9]. The detailed circuit is provided in Fig. 3A. The increased number of gates increases the robustness of the circuit with dual conditional rotational gates. The first seven qubits represent the for seven amino acids and two are the interaction qubits. The qubit configuration and dense coding was done corresponding to guidelines from [9].

The ground state energy for the each of the 49 peptides (98%) was found to be with $< \Delta G = 0$ and 1 peptide (2%) with $> \Delta G = 0$. This in comparison with MD simulation is better in terms of folding.

The comparison between the ground state energy values was subjected to statistical analysis (two-tailed t-test) at significant level of 95% ($\alpha = 0.05$). The conclusion suggests that there is significant change in values with a R-squared value of 0.8032 suggesting an effective fitness of data into the model. A violin plot Fig. 3B, figure provides crucial insights into the mean value and kernel density. The count of the energy values is provided as a histogram provided in Fig. 3C. The complete list of peptides along with the energy values for both MD simulations and VQE is provided in Supplementary File 2.

The mean value of energy from MD simulation is 2.302 kcal/mol and that of VQE is -5.885 kcal/mol. With higher number of shots (500000) we achieved minimum outputs for every peptide. Out of 50 peptides under consideration 33 peptides (66%) of the folds had less than 5 outputs with minimum being 1 output and the remainder of 17 peptides (34%) had greater than 5 outputs with the maximum being at 13. This provides conclusive evidence on the optimization of output folds for each of the peptide.

In comparison to the peptides folded in MD simulation, sequence 16 with residues "DPSHQKE" provided 4 outputs with 4 qubit configurations. An ground state energy was achieved at -2.97 kcal/mol in comparison to -1.11 kcal/mol with 44% probability. Figure 4A provides a plot on the energy values calculated at every iteration and stability achieved after 42 iterations.

The probability of peptide folding across the 4 outputs is plotted in Fig. 4B along with qubit configuration in x-axis. A bead-based representation of folded peptide for the best output probability is provided in Fig. 4C.

Similarly, sequence 19 with residues “KSQHPED” provided 8 outputs with 8 qubit configurations. The best output has a probability of 29% over other outcomes.

The energy value attained is -6.05 kcal/mol compared to -0.98 kcal/mol from MD simulations at the time frame of 10ns. Figure 4D provides a plot on the energy values calculated at every iteration and stability achieved after 63 iterations.

The probability of peptide folding across the 8 outputs is plotted in Fig. 4E along with qubit configuration in x-axis. A bead-based representation of folded peptide for the best output probability is provided in Fig. 4F. The plots for all the 50 peptides for the minimization of energy values across 100 iterations, the output plot with probability and peptide fold with residues highlighted in provided in supplementary file 3, 4 and 5 respectively.

A comparative analysis of peptide folding using VQE and MD simulations-based methods is discussed in the next section.

4. Discussion

In the current research work, we have used and an optimized circuit with additional conditional rotation gates, a quantum algorithm for solving the Protein Folding (PF) problem on a regular tetrahedral lattice. Our model Hamiltonian represents a sequence, with beads symbolizing amino acids and the option to include side chains as extra beads connected to the main chain. Interactions between amino acids are based on contacts between neighbouring beads on the lattice, extended to I-NN ($l > 1$) contacts along lattice edges for medium to long-range interactions. We've also incorporated penalty terms to prevent bead overlaps. This Hamiltonian effectively models complex coarse-grained interactions like Lennard-Jones and Coulombic forces, demonstrating the accurate reproduction of secondary structure elements through simple adjustments to the interaction potential map.

In the research paper we advocate the accuracy and effectiveness of VQE in providing an optimal ground energy state value compared to conventional MD simulations to identify fold with a minimum simulation time of 50ns to each of the peptides.

This work in our opinion the first of its kind comparison (till submission of manuscript) on the MD simulation based folding and Quantum algorithm.

In the methodology and parameters of the previously reported studies are provided in Table 1. The data was available as supplementary file with [9].

Table 1

Comparison of existing models for 3D protein native structure prediction using quantum algorithms

	Perdomo et al. [35]	Robert, A. et al [9]	Perdomo et al. [36]	Babbush et al. [37]	Babej et al. [38, 39]	Current project
Model	Hydrophobic-Polar (HP)	Coarse-Grained	Coarse-Grained	Coarse-Grained	Coarse-Grained	Coarse-Grained
Lattice	all	Tetrahedral	all	all	all	Tetrahedral
Types	2	∞	∞	∞	∞	∞
Interactions	Nearest	l^{th} Nearest	Nearest	Nearest	Nearest	l^{th} Nearest
Locality	$\log_2 N$	$l + 2$	N	4	N	$l + 2$
Qubits	$N \log(N)$	$N^2 \exp(l)$	$N^2 \log(N)$	$N^3 \log(N)$	N	$N^2 \exp(l)$
Scaling	N^8	N^4	$\exp(N)$	$N^{12} \log(N)$	$\exp(N)$	N^4
Experiment	No	IBM QPU	D-Wave	No	Rigetti QPU	Local Simulator/ Amazon Bracket QPU

Similar to other hybrid algorithms created for future quantum processors, the proposed CVaR-VQE algorithm is a hybrid quantum-classical algorithm. The genetic algorithm creates a population of parameters (i.e., an offspring) during each iteration of the VQE algorithm. This population records the ancestors and incorporates some randomness (by mutation rates) into a highly corrugated potential energy surface-driven process. This does not imply that other optimizers would perform worse in this application than the genetic algorithm, though. More details into utilising CVaR-VQE with the COBYLA classical optimizer is available at Barkoutsos, P et al [40].

For the process of MD simulation, the simulation period plays a very crucial role in achieving the minimum energy state. In the current work, with 50 peptides of 7 residues each, 28% of the peptides were not able to fold and achieve $< \Delta G = 0$ within the simulation time. This leads to performing larger simulation runs and hence setting a gold standard threshold value for all the proteins/peptides is not a feasible task. For each peptide/protein the simulation time differs, and it becomes a more tedious task if the proteins have regions of high disorder.

In MD based simulation for peptide folding, the best fold with least ground state energy can reach at any time frame over the total period of simulation and the to achieve this user need to cluster the peptides

based on RMSD and analyse the energy values. This becomes a hectic task when we handle large protein datasets with proteins of high disorderness.

With the advent of quantum computing and VQE based algorithms, we can calculate the number of possible folding outcomes for each peptide in short span of time. The ground state energy values are calculated along with the folding pattern. Even though the technology, hardware and the algorithms are still in the initial stages. They provide a promising prospect for the future of protein folding and the biochemistry involved within.

5. Conclusion

In the study, the amino acids part of disordered regions on proteins are selected at random as peptides for folding prediction. This is because the disordered regions pose a bigger challenge to protein folding problem compared to other domain of protein secondary structure.

Our results clearly show that VQE based algorithms can effectively fold the peptides compared to MD simulation.

Nevertheless, the current VQE model too has certain biases with respect to defining the qubit configuration and interaction qubits. The current qubits in QPU are limited which also prevents in incorporating large peptide or protein sequences in the purview of QPU.

With more advancements in algorithms like Quantum Approximate Optimization Algorithm (QAOA) [41] and its integration with VQE more accurate folding ground state energies can be predicted in near future.

Declarations

Supplementary Materials

Supplementary File 1: Ground state energy values for each peptide at different time frames along with the least energy value and statistical data (Mean and Standard deviation). All the values in kcal/mol.

Supplementary File 2: The complete list of peptides sequences along with energy values from MD simulations and VQE in kcal/mol.

Supplementary File 3: The energy values for the individual peptide with minimization of energy values over the iterations in provided. Please use the bookmarks to navigate for specific sequence.

Supplementary File 4: The output qubit configurations plot for the individual peptide with probability score on y-axis and qubit configuration on x-axis. Please use the bookmarks to navigate for specific sequence.

Supplementary File 5: The 3D representation of the peptide fold occurred for the each of the peptide with respect to its best output configuration is provided. Please use the bookmarks to navigate for specific

sequence.

Author Contributions

V.N. was in for ideation and conceptualization. A.U. contributed to the computational analysis and drafting the manuscript. Both the authors have read and agreed to the published version of the manuscript.

Funding

This research work was funded by Ministry of Electronics and Information Technology under the Meity QCaL Cohort-II which provides the access to Amazon Bracket.

Acknowledgments

The authors thank Dr. Anala M R from the Department of Computer Science and Engineering, the R V College of Engineering, and Bangalore for providing GPU (NVIDIA A100) computational support. The authors thank the support of Dr. Venugopal K, from the Department of Mathematics. A heartfelt thanks to Suman A for the unconditional support. A warm heartfelt thanks to the staff and administration at the R V College of Engineering for their support.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Shea J-E, Brooks lii CL. FROM FOLDING THEORIES TO FOLDING PROTEINS: A Review and Assessment of Simulation Studies of Protein Folding and Unfolding. *Annual Review of Physical Chemistry*. 2001;52(1):499-535.
2. Scheraga HA, Khalili M, Liwo A. Protein-Folding Dynamics: Overview of Molecular Simulation Techniques. *Annual Review of Physical Chemistry*. 2007;58(1):57-83.
3. Freddolino PL, Liu F, Gruebele M, Schulten K. Ten-Microsecond Molecular Dynamics Simulation of a Fast-Folding WW Domain. *Biophys J*. 2008;94(10):L75-L7.
4. Daidone I, Amadei A, Roccatano D, Nola AD. Molecular Dynamics Simulation of Protein Folding by Essential Dynamics Sampling: Folding Landscape of Horse Heart Cytochrome c. *Biophys J*. 2003;85(5):2865-71.
5. Beck D. Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods*. 2004;34(1):112-20.
6. Sonavane UB, Ramadugu SK, Joshi RR. Study of Early Events in the Protein Folding of Villin Headpiece using Molecular Dynamics Simulation. *Journal of Biomolecular Structure and Dynamics*. 2008;26(2):203-14.

7. Pal S, Bhattacharya M, Lee S-S, Chakraborty C. Quantum Computing in the Next-Generation Computational Biology Landscape: From Protein Folding to Molecular Dynamics. *Molecular Biotechnology*. 2023.
8. GhÉlis C, Yon J. Introduction to Considerations of Protein Folding Deduced from Characteristics of Folded Proteins. *Protein Folding*: Elsevier; 1982. p. 35-6.
9. Robert A, Barkoutsos PK, Woerner S, Tavernelli I. Resource-efficient quantum algorithm for protein folding. *npj Quantum Information*. 2021;7(1).
10. Vogt N, Zanker S, Reiner J-M, Marthaler M, Eckl T, Marusczyk A. Preparing ground states with a broken symmetry with variational quantum algorithms. *Quantum Science and Technology*. 2021;6(3):035003.
11. Choquette A, Di Paolo A, Barkoutsos PK, S n chal D, Tavernelli I, Blais A. Quantum-optimal-control-inspired ansatz for variational quantum algorithms. *Physical Review Research*. 2021;3(2).
12. Cerezo M, Verdon G, Huang H-Y, Cincio L, Coles PJ. Challenges and opportunities in quantum machine learning. *Nature Computational Science*. 2022;2(9):567-76.
13. Tilly J, Chen H, Cao S, Picozzi D, Setia K, Li Y, et al. The Variational Quantum Eigensolver: A review of methods and best practices. *Physics Reports*. 2022;986:1-128.
14. Cerezo M, Sone A, Volkoff T, Cincio L, Coles PJ. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat Commun*. 2021;12(1).
15. Uvarov AV, Biamonte JD. On barren plateaus and cost function locality in variational quantum algorithms. *Journal of Physics A: Mathematical and Theoretical*. 2021;54(24):245301.
16. Lee J, Huggins WJ, Head-Gordon M, Whaley KB. Generalized Unitary Coupled Cluster Wave functions for Quantum Computation. *Journal of Chemical Theory and Computation*. 2018;15(1):311-24.
17. Holmes Z, Sharma K, Cerezo M, Coles PJ. Connecting Ansatz Expressibility to Gradient Magnitudes and Barren Plateaus. *PRX Quantum*. 2022;3(1).
18. Chandarana P, Hegade NN, Montalban I, Solano E, Chen X. Digitized Counterdiabatic Quantum Algorithm for Protein Folding. *Physical Review Applied*. 2023;20(1).
19. Bharti K, Cervera-Lierta A, Kyaw TH, Haug T, Alperin-Lea S, Anand A, et al. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics*. 2022;94(1).
20. Wecker D, Hastings MB, Troyer M. Progress towards practical quantum variational algorithms. *Physical Review A*. 2015;92(4).
21. Wiersema R, Zhou C, de Sereville Y, Carrasquilla JF, Kim YB, Yuen H. Exploring Entanglement and Optimization within the Hamiltonian Variational Ansatz. *PRX Quantum*. 2020;1(2).
22. Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow JM, et al. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature*. 2017;549(7671):242-6.
23. Farhi E, Goldstone J, Gutmann S, Zhou L. The Quantum Approximate Optimization Algorithm and the Sherrington-Kirkpatrick Model at Infinite Size. *Quantum*. 2022;6:759.

24. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*. 2001;19(1):26-59.
25. Campen A, Williams R, Brown C, Meng J, Uversky V, Dunker A. TOP-IDP-Scale: A New Amino Acid Scale Measuring Propensity for Intrinsic Disorder. *Protein & Peptide Letters*. 2008;15(9):956-63.
26. Du Y, Huang T, You S, Hsieh M-H, Tao D. Quantum circuit architecture search for variational quantum algorithms. *npj Quantum Information*. 2022;8(1).
27. Madhavi Sastry G, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*. 2013;27(3):221-34.
28. Greenwood JR, Calkins D, Sullivan AP, Shelley JC. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *Journal of Computer-Aided Molecular Design*. 2010;24(6-7):591-604.
29. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. *Journal of Computer-Aided Molecular Design*. 2007;21(12):681-91.
30. Roos K, Wu C, Damm W, Reboul M, Stevenson JM, Lu C, et al. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *Journal of Chemical Theory and Computation*. 2019;15(3):1863-74.
31. Mark P, Nilsson L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A*. 2001;105(43):9954-60.
32. Bowers KJ, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE, Chow E, et al. Molecular dynamics—Scalable algorithms for molecular dynamics simulations on commodity clusters. *Proceedings of the 2006 ACM/IEEE conference on Supercomputing - SC '06*: ACM Press; 2006.
33. Uttarkar A, Niranjana V. Brefeldin A variant via combinatorial screening acts as an effective antagonist inducing structural modification in EPAC2. *Molecular Simulation*. 2022;48(17):1592-603.
34. Niranjana V, Uttarkar A, Ramakrishnan A, Muralidharan A, Shashidhara A, Acharya A, et al. De Novo Design of Anti-COVID Drugs Using Machine Learning-Based Equivariant Diffusion Model Targeting the Spike Protein. *Current Issues in Molecular Biology*. 2023;45(5):4261-84.
35. Perdomo A, Truncik C, Tubert-Brohman I, Rose G, Aspuru-Guzik A. Construction of model Hamiltonians for adiabatic quantum computation and its application to finding low-energy conformations of lattice protein models. *Physical Review A*. 2008;78(1).
36. Perdomo-Ortiz A, Dickson N, Drew-Brook M, Rose G, Aspuru-Guzik A. Finding low-energy conformations of lattice protein models by quantum annealing. *Sci Rep*. 2012;2(1).
37. Babbush R, Perdomo-Ortiz A, O'Gorman B, Mcready W, Aspuru-Guzik A. Construction of Energy Functions for Lattice Heteropolymer Models: Efficient Encodings for Constraint Satisfaction Programming and Quantum Annealing. *Advances in Chemical Physics*: Wiley; 2014. p. 201-44.
38. Fingerhuth M, Babej T, Wittek P. Adiabatic quantum computation. *Definitions: Qeios*; 2019.

39. Fingerhuth M, Babej T, Wittek P. Open source software in quantum computing. PLoS One. 2018;13(12):e0208561.
40. Barkoutsos PK, Nannicini G, Robert A, Tavernelli I, Woerner S. Improving Variational Quantum Optimization using CVaR. Quantum. 2020;4:256.
41. Boulebnane S, Lucas X, Meyder A, Adaszewski S, Montanaro A. Peptide conformational sampling using the Quantum Approximate Optimization Algorithm. npj Quantum Information. 2023;9(1)

Figures

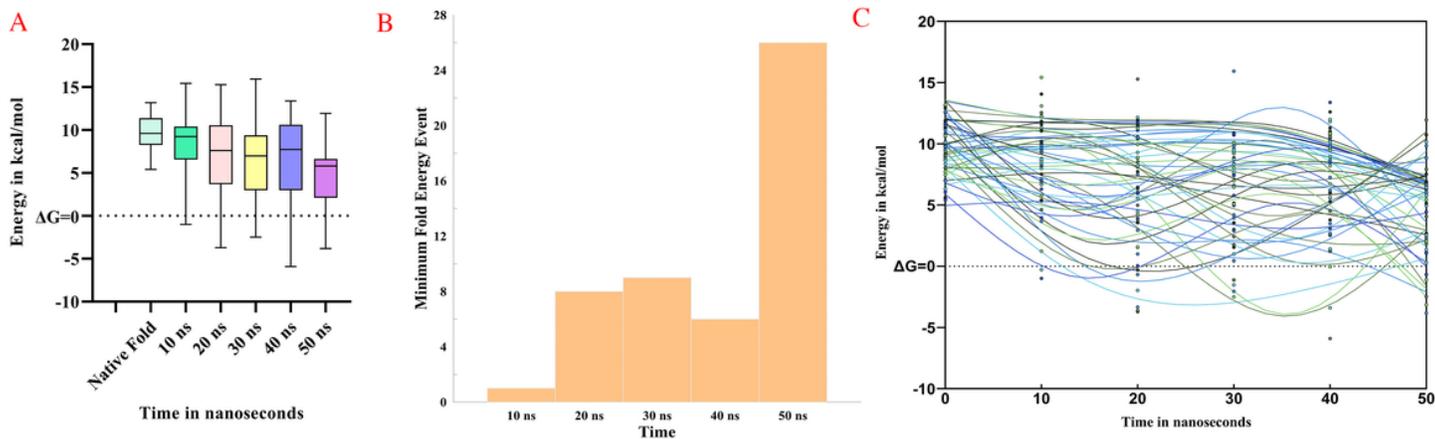


Figure 1

A) Box and whisker showing the least ground state energy values for all the peptides at specific time frame across the simulation time frame. The average value is shown with a black line within the box. B) A bar plot with count of minimum energy event attained by the peptide at specific simulation time. C) The spline fitting curve with energy in kcal/mol versus time in nanoseconds.

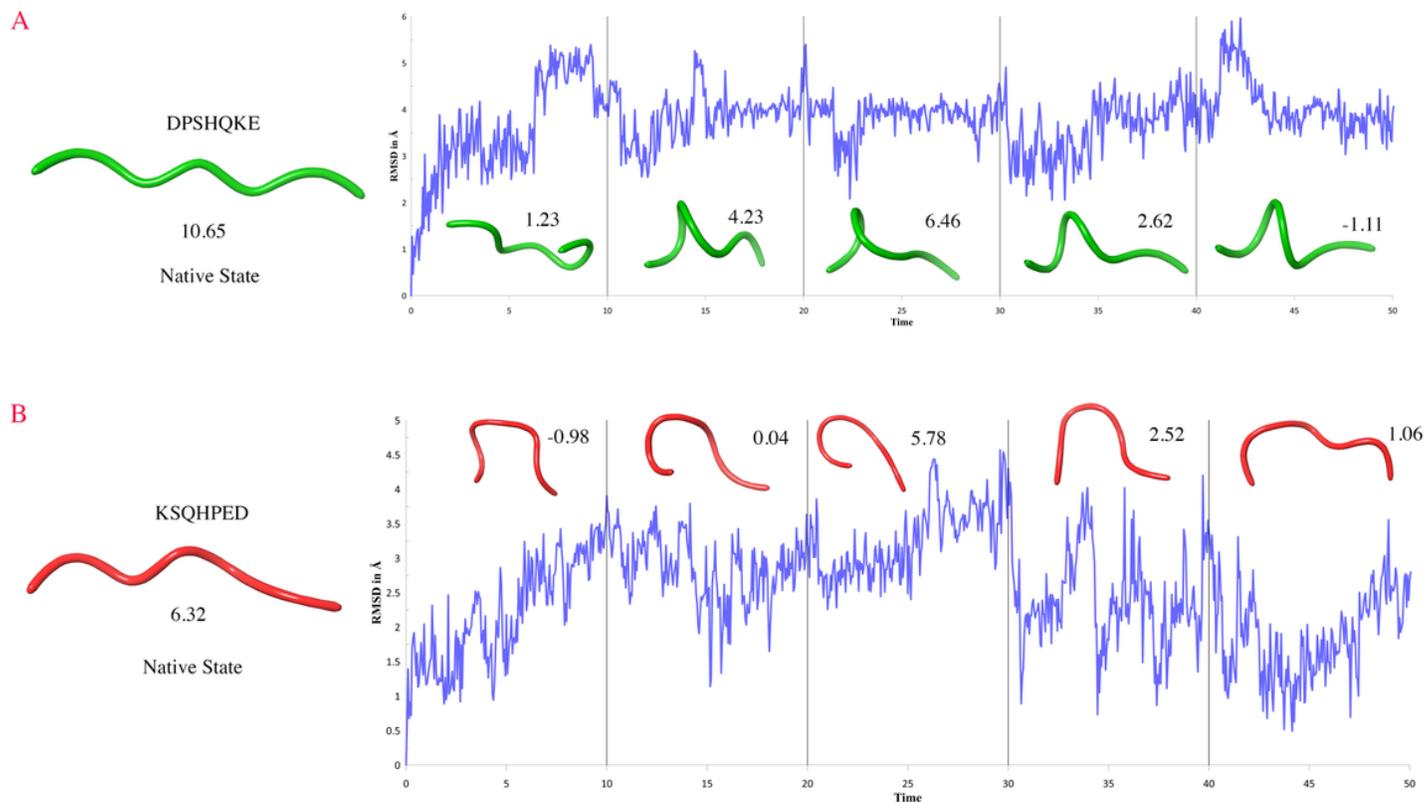


Figure 2

A) The evolution of peptide folding (green) over the period of simulation time for the peptide 16 sequence “DPSHQKE” along with energy values in kcal/mol. The folding is overlaid on the RMSD plot with deviation in angstrom versus time in nanoseconds.

B) The evolution of peptide folding (red) over the period of simulation time for the peptide 19 sequence “KSQHPED” along with energy values in kcal/mol. The folding is overlaid on the RMSD plot with deviation in angstrom versus time in nanoseconds.

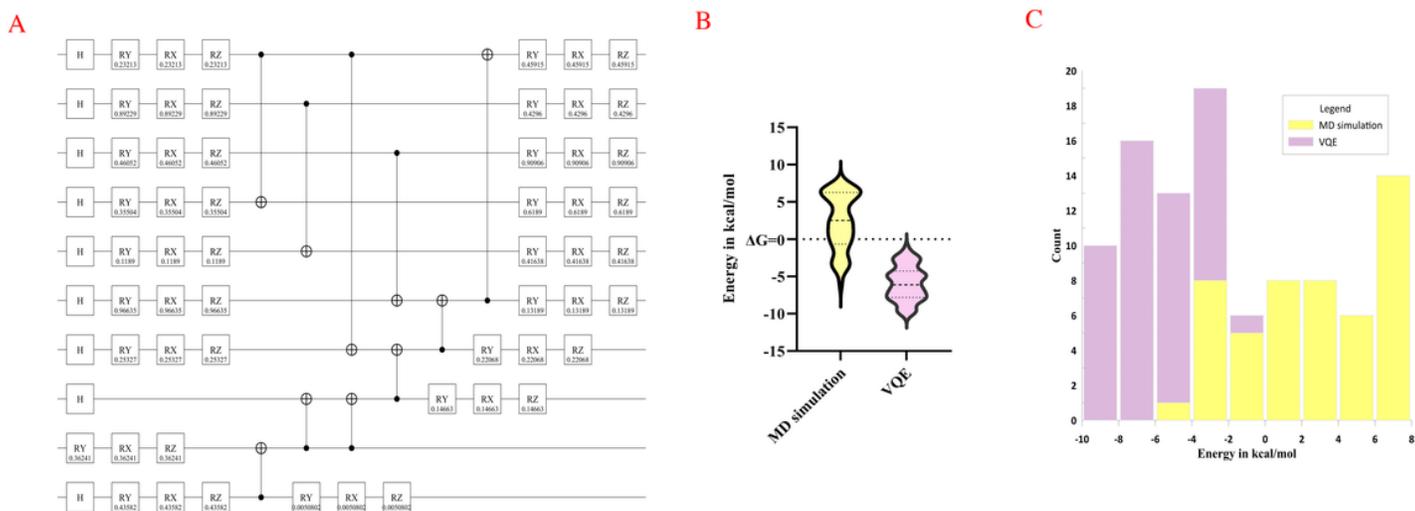


Figure 3

A) A detailed pictorial representation of the quantum ansatz circuit with Hadamard, conditional rotational gates and CNOT gates. B) A violin plot comparing the energy values in kcal/mol for protein folding achieved via MD based simulation (yellow) and VQE (purple), with black dashed line showing average and thin line on either side with quartile range. C) A comparative plot of count of number of minimum energy events with energy range on X-axis in kcal/mol occurred between MD based simulation (yellow) and VQE (purple).

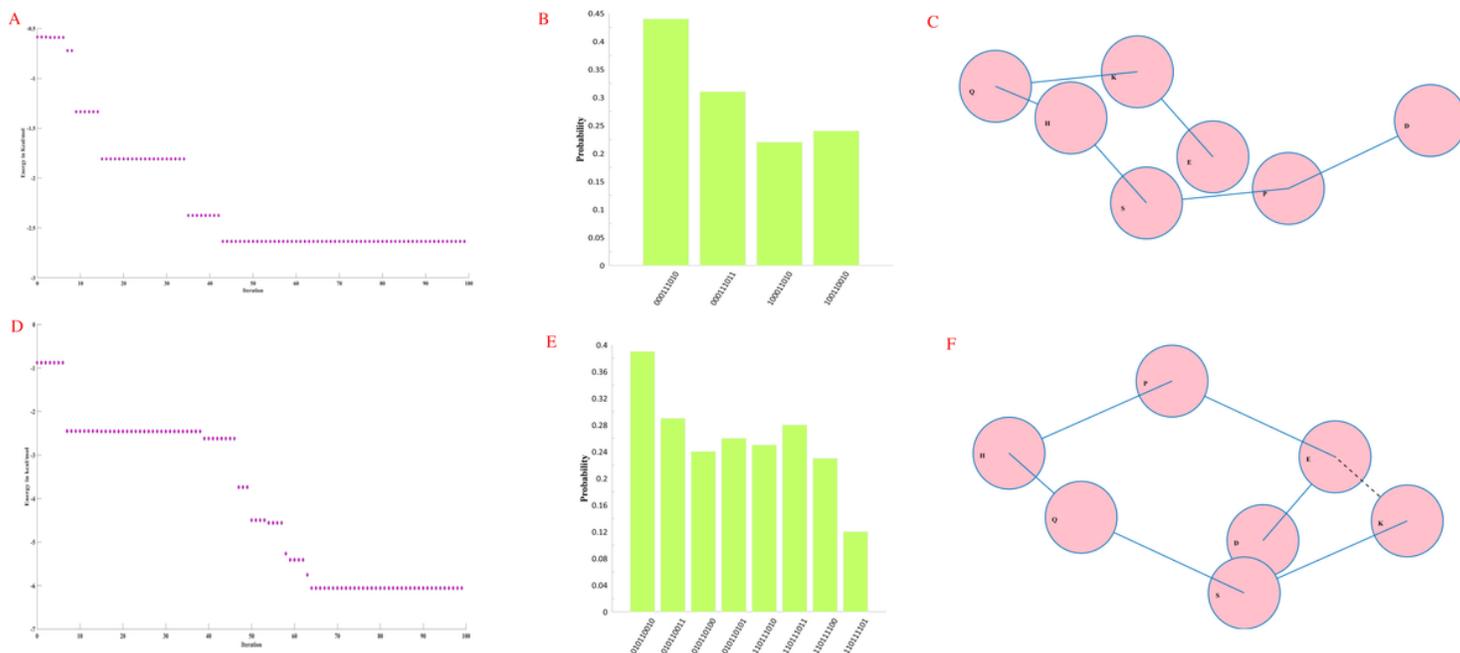


Figure 4

A) Energy optimization via VQE based simulation for peptide number 16 over number iterations are plot with energy in kcal/mol. B) The number of output configurations obtained along with the probability values in x-axis. C) The folding pattern for peptide for the best output is shown with amino acids as beads and single letter code mentioned against each of them. D) Energy optimization via VQE based simulation for peptide number 19 over number iterations are plot with energy in kcal/mol. E) The number of output configurations obtained along with the probability values in x-axis. F) The folding pattern for peptide for the best output is shown with amino acids as beads and single letter code mentioned against each of them.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFile1.pdf](#)
- [SupplementaryFile2.pdf](#)

- [SupplementaryFile3.pdf](#)
- [SupplementaryFile4.pdf](#)
- [SupplementaryFile5.pdf](#)