Quantum Information Processing (2024) 23:48
https://doi.org/10.1007/s11128-024-04261-9

A comparative insight into peptide folding with quantum CVaR-VQE algorithm, MD simulations and structural alphabet analysis

Akshay Uttarkar[1] · **Vidya Niranjan**[1]

Received: 18 October 2023 / Accepted: 5 January 2024 / Published online: 6 February 2024

Abstract
Quantum computing in biology is one of the most rapidly evolving fields 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 seven 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 500,000 shots each to obtain least ground-state energy value. This is compared to the molecular dynamics-based simulations of 50 ns 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 and HMM-SA. 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.

Keywords Quantum computing · Protein folding · Variational quantum eigensolver · MD simulation

1 Introduction

Protein folding is a complex process that involves the spontaneous folding 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

✉ Vidya Niranjan
vidya.n@rvce.edu.in

[1] Department of Biotechnology, R V College of Engineering, (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Bangalore 560059, India

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 novel approach that has the potential to overcome the limitations of MD simulations.

Quantum folding simulations are still in their initial 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 disorderedness 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, VQEs are one of the important approaches in quantum protein folding [8].

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 ten

amino acid and a seven amino acid peptide. The model proposes an 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 ansatz proposed at the time of writing. Adaptive ansatz could be a reliable alternative subject to further studies on their expected computational cost. Layered (hardware-efficient ansatz) HEA is likely not suitable for large systems as they require a large number of parameters and need to span a considerable 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 is observed. 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 nine 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 ansatz is designed to exploit the properties of the problem Hamiltonian to efficiently reach the desired state. Hardware-efficient ansatz is designed to minimize the noise introduced by deep circuits and unimplementable connections [22, 23].

VQE offers advantages in peptide folding, including polynomial time modelling, resilience to noise, efficient folding outcomes and resource efficiency. However, it has limitations such as limited precision, scaling issues with increasing molecule size and incomplete exploration of the solution space. VQE has shown promise in addressing the geometry calculations involved in peptide folding, but these limitations need to be addressed to further improve its performance in this domain.

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, Hidden Markov Model-Structural Alphabet (HMM-SA) and VQE. The comparison provides an insight 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 seven 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 seven amino acid peptides. It is 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 five fixed bits and the seven variable bits, which will be represented by qubits. In the current work, the configuration was set to “0011q1qqqqqq” 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, 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 is important to note that beads separated by fewer than five bonds cannot establish a $1-$ NN relationship. The equation for the calculation is provided in Eq. 1.

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

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

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 four 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)

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

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 $\overline{a}$, 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 $\overline{a}$) 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).

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

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.

We have utilized a denser encoding of the polymer configuration that requires only two qubits per turn $t_i = q_{2i}$ reducing the total number of qubits to 22. This variant generated five-local (instead of three-local) terms in the qubit Hamiltonian while keeping the total number of Pauli strings within an affordable range for small instances. We also integrated the side chains with the corresponding bead along the primary sequence and neglected interactions with $\mathrm{l} > 1$ to further reduce the number of qubits.

The qubit Hamiltonian that characterizes the energy of a particular fold determined by the encoded turns and the fixed sequence of beads is defined in the following step. Physical interactions (attractive or repulsive in nature) are applied when two beads occupy neighbouring sites or are at a distance $\mathrm{l} > 1$, where l is the length of the shortest lattice path connecting them. Penalty terms are applied when physical reducing the total number of qubits to 22.

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 $f_a^{\text{ex}}\left(i^{(1)}\right)$ 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 seven residues with most disorderedness, D, H, Q, K, S, E and P, were selected. The number of possible permutations ${}^7\text{P}_6$ 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 (seven for setup and two 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 500,000 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 pi 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, cut-off 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 cut-off 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

The quantum algorithm for protein folding utilized a circuit depth of $m = 2$ in the variational quantum eigensolver (VQE) algorithm. This depth parameter is crucial for understanding the complexity and expressiveness of the quantum circuit used in the algorithm. Additionally, the optimization method employed in this algorithm was the Conditional Value-at-Risk (CVaR) VQE, a modified version of the VQE algorithm. The CVaR-VQE algorithm provides a drastic speed-up to the optimization of diagonal Hamiltonians.

Given a random variable X with quantile function $F - 1$, it is defined the value at risk, denoted by VaR, as

$$\mathrm{VaR}[X;\, p] = F^{-1}(p), \text{ for all } p \in (0, 1) \tag{4}$$

For conditional value at risk

$$\mathrm{CVaR}X;\, p = EX - F^{-1}(p) \mid X > F^{-1}(p) = mF^{-1}(p) \tag{5}$$

for all $p \in (0,1)$, where m denotes the mean residual life function of X. It is the additional expected loss to VaR given that the loss exceeds its VaR [27].

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 Rx, Ry and Rz rotation angles that yield the lowest expected energy, use the *surrogateopt* (Global Optimisation Toolbox) [28] algorithm with the objective function [29]. 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 and an extended structure was built. The structures predicted were subjected to protein preparation [30] using EpiK [31, 32] and OLPS3e [33] force field was added. The peptides were not minimized. An orthorhombic system was built for simulation with TIP3P [34] as solvent and no additional ions were added to naturalize the system and maintain its native nature.

Desmond [35] was used to run MD simulations for 50 ns each. A detailed steps involved in minimization and equilibration is available in [36, 37]. Using the trajectory files, the protein structures at an interval of every 10ns were extracted to capture the

folding pattern. A total of six structures for each sequence (0, 10, 20, 30, 40, 50th frame) were subjected to minimization and evaluation of LJ interaction energy.

2.7 Folding of peptides via HMM-based structural alphabet

A de novo method, PEP-FOLD [38], is used to infer peptide conformations from amino acid sequences. This approach combines the projected series of structured letters (SA) letters superseded by HMM algorithm [39] to a greedy algorithm and a coarse-grained force field, based on the structural alphabet SA letters to characterize the conformations of four successive residues.

A forward backtrack (FBT) generator is used, and Monte Carlo steps set to 30,000. A total of 100 models were built for each peptide.

3 Results

3.1 Molecular dynamics-based simulations and SA method to achieve protein folding event

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

The total interaction energy was calculated for each of the frame (six 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 value for 50 sequences at native energy fold (0th frame) is found to be 27.39 kcal/mol, at 10 ns it is 5.69 kcal/mol, at 20 ns it is 1.14 kcal/mol, at 30 ns it is − 0.36 kcal/mol, at 40 ns it is at − 1.74 kcal/mol, and at 50 ns it is at − 0.17 kcal/mol. The results suggest that over the simulation period, the protein folding occurs better with higher simulation period till 40 ns and decreases gradually at 50 ns as suggested by the mean value and it is aligned to the conventional understanding of protein folding theories with MD simulation. The energy values at each time frame along with mean and outliers are shown in Fig. 1A. Nevertheless, it

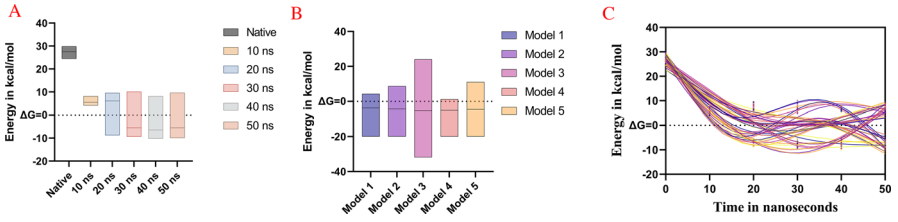

Fig. 1 **A** Box plot showing the least ground-state energy values for all the peptides at specific time frame across the simulation time frame. The average value is shown with a black line within the box. **B** Box plot showing the least ground-state energy values for all models built using SA. The average value is shown with a black line within the box. **C** The spline fitting curve with energy in kcal/mol versus time in nanoseconds

does not always refer the at the least energy was achieved only after end of simulation which is 50 ns in our case.

In the observations made in the current research work that minimum energy ground state was achieved for 3 peptides (6%) at 10 ns, 10 peptides (20%) at 20 ns, 14 peptides (28%) at 30 ns, 19 peptides (38%) at 40 ns and 2 peptide (4%) at 50 ns.

The best energy values for any peptide can occur at any time frame across 50 ns for comparison with other classical and quantum algorithms.

The surface algorithm (SA) was used to predict the folds to peptides. For each peptide, five models will be generated and shortlisted based on sOPEP (Optimized Potential for Efficient structure Prediction) scores.

The ground-state energy for each model is calculated and the values are plotted in Fig. 1B. The mean energy values for each model are below $< \Delta G = 0$. This method has provided better energy values compared to MD simulations. To best model with least energy values is selected for comparison with quantum algorithms.

For MD simulations, 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 25–40 ns 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.

Sequence 1 with a peptide sequence, "DSPKHQE," has shown a classical mode of peptide folding. The energy values at native state are 28.22 kcal/mol, at 10 ns it is 4.12 kcal/mol, at 20 ns it is 9.51 kcal/mol, at 30 ns it observed to be -7.25 kcal/mol, similarly at 40 ns at it is 6.21 kcal/mol, and finally at 50 ns it is -6.44 kcal/mol reaching an energy value with $\Delta G > 0$.

The peptide gradually folds to achieve ground-state energy over the period of simulation for 30 ns becomes unstable and regains the stable nature at 50 ns. They achieved a RMSD change of 2.25 Å at the end of simulation period. The RMSD is very stable after 35 ns nanoseconds and Fig. 2A provides a comprehensive view of the peptide structures its folding pattern with energy values in kcal/mol overlayed on the root mean square deviation (RMSD) plot versus time.

For peptide folds available for sequence 1 "DSPKHQE" using SA algorithm gives five models. These five models are shown along with energy values in kcal/mol. The surface representation with set transparency is observed. The colour set is white to red based on its secondary structure.

As previously mentioned, 38% of the peptides reached the lowest ground-state energy at 40 ns of simulation period. In other cases, the lowest ground-state energy is reached during intermediate phases of simulation or at 50 ns. 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 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.

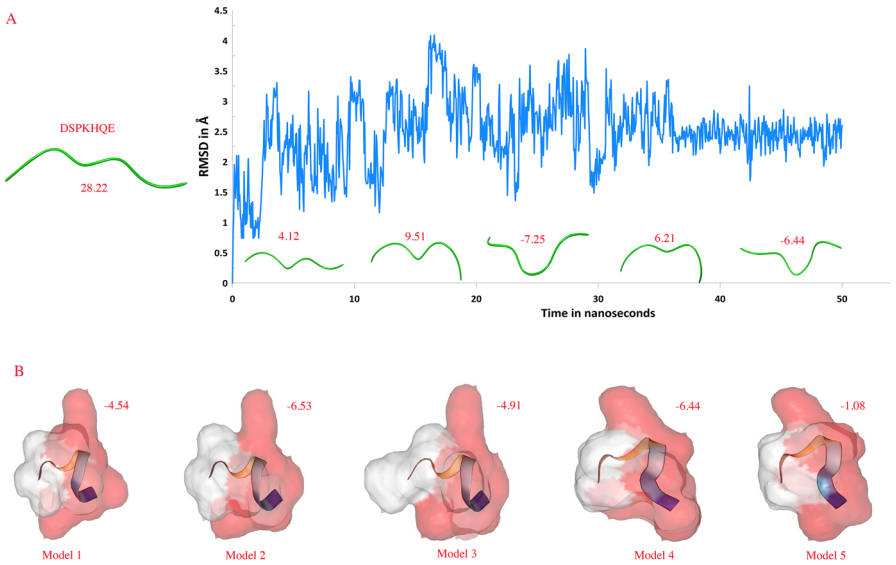

Fig. 2 **A** The evolution of peptide folding (green) over the period of simulation time for the peptide 1 sequence "DSPKHQE" along with energy values in kcal/mol is observed. The folding is overlayed on the RMSD plot with deviation in angstrom versus time in nanoseconds. **B** The peptide folds for "DSPKHQE" across all the five models are shown along with the energy in kcal/mol. The peptides are shown in ribbon form with transparent surface with colour coding set to secondary structure from white to red

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 are observed. 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 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 50 peptides was found to be with $< \Delta G = 0$. This in comparison with mean values for each peptide with MD simulation and SA is higher in terms of folding, whereas the range of values and prediction is very efficient.

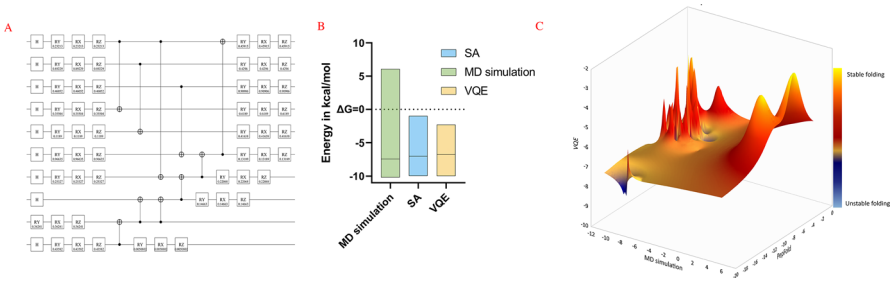

Fig. 3 **A** A detailed pictorial representation of the quantum ansatz circuit with Hadamard, conditional rotational gates and CNOT gates. **B** A box plot comparing the energy values in kcal/mol for protein folding achieved via MD-based simulation (green), VQE (brown) and SA (blue), with black line showing average value. **C** A 3D plot with energy values from MD simulation, VQE and SA provides a better comparison of stable folding of peptides, with colour scale set to bright yellow for stable and blue for unstable

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. In a box plot Fig. 3B, the figure provides crucial insights into the mean value. A 3D plot showcasing the folding energies of the peptides across three methods is plotted. It can be observed that VQE and SA have shown good stable folding (yellow) of peptides, whereas a plateau (red) and drop towards unstable (blue) is for MD simulations. The complete list of peptides along with the energy values for SA, MD simulations and VQE is provided in Supplementary File 2.

The mean value of energy from MD simulation is − 6.67 kcal/mol and that of VQE is − 7.37 kcal/mol.

With higher number of shots (500,000), we achieved minimum outputs for every peptide. Out of 50 peptides under consideration, 33 peptides (66%) of the folds had less than five outputs with minimum being one output and the remainder of 17 peptides (34%) had greater than five outputs with the maximum being at 13. This provides conclusive evidence on the optimization of output folds for each of the peptide.

In comparison with the peptides folded in MD simulation, sequence 1 with residues "DSPKHQE" provided three outputs with three qubit configurations. A ground-state energy was achieved at − 5.01 kcal/mol in comparison with − 1.11 kcal/mol with 32% probability. Figure 4A provides a plot on the energy values calculated at every iteration and stability achieved after 45 iterations.

The probability of peptide folding across the three 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.

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, SA and MD simulations-based methods is discussed in the next section.

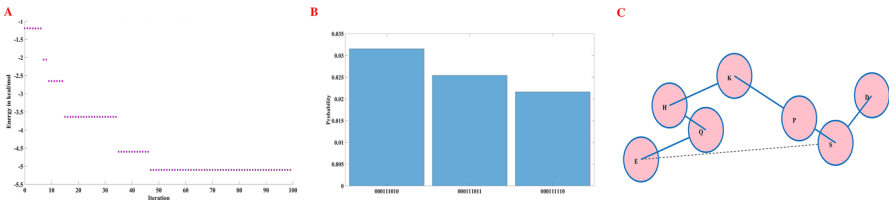

Fig. 4 **A** Energy optimization via VQE-based simulation for peptide 1 over number iterations is plotted with energy in kcal/mol. **B** The number of output configurations obtained along with the probability values in x-axis. **C** The folding pattern for peptide for the best output is shown with amino acids as beads and single letter code mentioned against each of them

4 Discussion

The protein folding problem, a fundamental challenge in structural biology, involves predicting a protein's three-dimensional structure from its amino acid sequence. Classically, this problem is computationally complex, often requiring extensive simulations and search strategies, which may be infeasible for large proteins due to the exponentially growing number of possible conformations. However, variational quantum eigensolver (VQE)-based algorithms have emerged as a promising solution. VQE leverages quantum computing to efficiently explore the vast conformational space of proteins, offering a more scalable approach compared to classical methods.

Classical methods for protein folding, such as molecular dynamics simulations and Monte Carlo methods, face challenges in efficiently exploring the conformational space, especially for large proteins. The computational complexity of these methods limits their applicability to larger and more complex protein structures.

On the other hand, VQE-based algorithms, which utilize quantum circuits to encode and optimize the protein folding problem, have shown promise in providing a more efficient and scalable approach. These algorithms can leverage the inherent parallelism of quantum computing to explore a large number of conformations simultaneously, potentially overcoming the limitations of classical methods in handling the combinatorial complexity of the protein folding problem solutions, potentially revolutionizing the field of protein structure prediction.

Classical methods, such as molecular dynamics simulations, often require extensive simulations and search strategies, which can be time-consuming and may not be feasible for large proteins. In contrast, VQE-based algorithms can leverage the inherent parallelism of quantum computing to explore a large number of conformations simultaneously, potentially overcoming the limitations of classical methods in handling the combinatorial complexity of the protein folding problem.

Overall, VQE-based algorithms offer a promising avenue for more efficient and scalable solutions to the protein folding problem, compared to classical methods. These algorithms can leverage the inherent parallelism of quantum computing to explore a large number of conformations simultaneously, potentially overcoming the limitations of classical methods in handling the combinatorial complexity of the protein folding problem.

In the current research work, we have used 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 $l-$ NN ($l > 1$) contacts along lattice edges for medium- to long-range interactions. We have 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 50 ns 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, structured alphabet and Quantum algorithm.

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

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 utilizing CVaR-VQE with the COBYLA classical optimizer are available at Barkoutsos et al. [45].

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 seven residues each, all the peptides reached a ground-state energy fold within 50 ns. But this is not always true and there is a need 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 disorderedness.

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.

Table 1 Comparison of existing models for 3D protein native structure prediction using quantum algorithms

	Perdomo et al. [40]	Robert et al. [9]	Perdomo et al. [41]	Babbush et al. [42]	Babej et al. [43, 44]	Current work
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	lth Nearest	Nearest	Nearest	Nearest	Lth Nearest
Locality	$\log_2 N$	$1+2$	N	4	N	$1+2$
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

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. Another disadvantage is that the barren plateau of VQE is always questioned for its accuracy in larger systems.

With more advancements in algorithms like quantum approximate optimization algorithm (QAOA) [46] and its integration with VQE more accurate folding ground-state energies can be predicted in near future.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11128-024-04261-9.

Acknowledgements The authors thank Department of Computer Science and Engineering, the R V College of Engineering, 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 her unconditional support. A warm heartfelt thanks to the staff and administration at the R V College of Engineering for their support.

Author contributions VN was in for ideation and conceptualization. AU 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 Braket.

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

1. Shea, J.-E., Brooks Iii, C.L.: FROM folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu. Rev. Phys. Chem. **52**(1), 499–535 (2001)
2. Scheraga, H.A., Khalili, M., Liwo, A.: Protein-folding dynamics: overview of molecular simulation techniques. Annu. Rev. Phys. Chem. **58**(1), 57–83 (2007)
3. Freddolino, P.L., Liu, F., Gruebele, M., Schulten, K.: Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. Biophys. J. **94**(10), L75–L77 (2008)
4. Daidone, I., Amadei, A., Roccatano, D., Nola, A.D.: Molecular dynamics simulation of protein folding by essential dynamics sampling: folding landscape of horse heart cytochrome c. Biophys. J. **85**(5), 2865–2871 (2003)
5. Beck, D.: Methods for molecular dynamics simulations of protein folding/unfolding in solution. Methods **34**(1), 112–120 (2004)

6. Sonavane, U.B., Ramadugu, S.K., Joshi, R.R.: Study of early events in the protein folding of Villin headpiece using molecular dynamics simulation. J. Biomol. Struct. Dyn. **26**(2), 203–214 (2008)
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. Mol. Biotechnol. (2023). https://doi.org/10.1007/s12033-023-00765-4
8. GhÉLis, C., Yon, J.: Introduction to considerations of protein folding deduced from characteristics of folded proteins. In: Ghélis, C. (ed.) Protein folding, pp. 35–6. Elsevier, Amsterdam (1982)
9. Robert, A., Barkoutsos, P.K., Woerner, S., Tavernelli, I.: Resource-efficient quantum algorithm for protein folding. npj Quantum Inform. **7**(1), 38 (2021)
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 Sci. Technol. **6**(3), 035003 (2021)
11. Choquette, A., Di Paolo, A., Barkoutsos, P.K., Sénéchal, D., Tavernelli, I., Blais, A.: Quantum-optimal-control-inspired ansatz for variational quantum algorithms. Phys. Rev. Res. **3**(2), 023092 (2021)
12. Cerezo, M., Verdon, G., Huang, H.-Y., Cincio, L., Coles, P.J.: Challenges and opportunities in quantum machine learning. Nat. Comput. Sci. **2**(9), 567–576 (2022)
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. Phys. Rep. **986**, 1–128 (2022)
14. Cerezo, M., Sone, A., Volkoff, T., Cincio, L., Coles, P.J.: Cost function dependent barren plateaus in shallow parametrized quantum circuits. Nat. Commun. **12**(1), 1791 (2021)
15. Uvarov, A.V., Biamonte, J.D.: On barren plateaus and cost function locality in variational quantum algorithms. J. Phys. A Math. Theor. **54**(24), 245301 (2021)
16. Lee, J., Huggins, W.J., Head-Gordon, M., Whaley, K.B.: Generalized unitary coupled cluster wave functions for quantum computation. J. Chem. Theory Comput. **15**(1), 311–324 (2018)
17. Holmes, Z., Sharma, K., Cerezo, M., Coles, P.J.: Connecting ansatz expressibility to gradient magnitudes and barren plateaus. PRX Quantum. **3**(1), 010313 (2022)
18. Chandarana, P., Hegade, N.N., Montalban, I., Solano, E., Chen, X.: Digitized counterdiabatic quantum algorithm for protein folding. Phys. Rev. Appl. **20**(1), 014024 (2023)
19. Bharti, K., Cervera-Lierta, A., Kyaw, T.H., Haug, T., Alperin-Lea, S., Anand, A., et al.: Noisy intermediate-scale quantum algorithms. Rev. Modern Phys. **94**(1), 015004 (2022)
20. Wecker, D., Hastings, M.B., Troyer, M.: Progress towards practical quantum variational algorithms. Phys. Rev. A. **92**(4), 042303 (2015)
21. Wiersema, R., Zhou, C., de Sereville, Y., Carrasquilla, J.F., Kim, Y.B., Yuen, H.: Exploring entanglement and optimization within the Hamiltonian variational ansatz. PRX Quantum. **1**(2), 020319 (2020)
22. Kandala, A., Mezzacapo, A., Temme, K., Takita, M., Brink, M., Chow, J.M., et al.: Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. Nature **549**(7671), 242–246 (2017)
23. Farhi, E., Goldstone, J., Gutmann, S., Zhou, L.: The quantum approximate optimization algorithm and the sherrington-kirkpatrick model at infinite size. Quantum **6**, 759 (2022)
24. Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., et al.: Intrinsically disordered protein. J. Mol. Graph. Model. **19**(1), 26–59 (2001)
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 Pept. Lett. **15**(9), 956–963 (2008)
26. Du, Y., Huang, T., You, S., Hsieh, M.-H., Tao, D.: Quantum circuit architecture search for variational quantum algorithms. npj Quantum Inform. **8**(1), 62 (2022)
27. Belzunce, F., Riquelme, C.M., Mulero, J.: An Introduction to Stochastic Orders. Academic Press, London (2015). https://doi.org/10.1016/B978-0-12-803768-3.00001-6
28. López, C.P.: Optimization techniques via the optimization toolbox. In: Lopez, C. (ed.) MATLAB optimization techniques. Apress, Berkeley, CA (2014). https://doi.org/10.1007/978-1-4842-0292-0_6
29. Gutmann, H.-M.: A radial basis function method for global optimization. J. Global Optim. **19**, 201–227 (2001)
30. MadhaviSastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., Sherman, W.: Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J. Comput. Aided Mol. Des. **27**(3), 221–234 (2013)

31. Greenwood, J.R., Calkins, D., Sullivan, A.P., Shelley, J.C.: Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. J. Comput. Aided Mol. Des. **24**(6–7), 591–604 (2010)
32. Shelley, J.C., Cholleti, A., Frye, L.L., Greenwood, J.R., Timlin, M.R., Uchimaya, M.: Epik: a software program for pK a prediction and protonation state generation for drug-like molecules. J. Comput. Aided Mol. Des. **21**(12), 681–691 (2007)
33. Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J.M., Lu, C., et al.: OPLS3e: Extending force field coverage for drug-like small molecules. J. Chem. Theory Comput. **15**(3), 1863–1874 (2019)
34. Mark, P., Nilsson, L.: Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J. Phys. Chem. A **105**(43), 9954–9960 (2001)
35. Bowers, K.J., Sacerdoti, F.D., Salmon, J.K., Shan, Y., Shaw, D.E., Chow, E., et al.: Molecular dynamics—Scalable algorithms for molecular dynamics simulations on commodity clusters. In: Proceedings of the 2006 ACM/IEEE conference on Supercomputing—SC '06, ACM Press (2006)
36. Uttarkar, A., Niranjan, V.: Brefeldin A variant via combinatorial screening acts as an effective antagonist inducing structural modification in EPAC2. Mol. Simul. **48**(17), 1592–1603 (2022)
37. Niranjan, 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. Curr. Issues Mol. Biol. **45**(5), 4261–4284 (2023)
38. Rey, J., Murail, S., de Vries, S., Derreumaux, P., Tuffery, P.: PEP-FOLD4: a pH-dependent force field for peptide structure prediction in aqueous solution. Nucleic Acids Res. **51**(W1), W432–W437 (2023). https://doi.org/10.1093/nar/gkad376
39. Camproux, A.C., et al.: A hidden markov model derived structural alphabet for proteins. J. Mol. Biol. **339**(3), 591–605 (2004). https://doi.org/10.1016/j.jmb.2004.04.005
40. 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. Phys. Rev. A. **78**(1), 012320 (2008)
41. 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. **2**(1), 1–7 (2012)
42. Babbush, R., Perdomo-Ortiz, A., O'Gorman, B., Macready, W., Aspuru-Guzik, A.: Construction of Energy Functions for Lattice Heteropolymer Models: Efficient Encodings for Constraint Satisfaction Programming and Quantum Annealing. In: Prigogine, I., Rice, S.A. (eds.) Advances in chemical physics, pp. 201–44. Wiley, London (2014)
43. Fingerhuth, M., Babej, T., Wittek, P.: Adiabatic quantum computation. Definitions: Qeios, (2019)
44. Fingerhuth, M., Babej, T., Wittek, P.: Open-source software in quantum computing. PLoS ONE **13**(12), e0208561 (2018)
45. Barkoutsos, P.K., Nannicini, G., Robert, A., Tavernelli, I., Woerner, S.: Improving variational quantum optimization using CVaR. Quantum **4**, 256 (2020)
46. Boulebnane, S., Lucas, X., Meyder, A., Adaszewski, S., Montanaro, A.: Peptide conformational sampling using the quantum approximate optimization algorithm. npj Quantum Inform. **9**(1), 70 (2023)