Quantum Protein Folding

Sakibul Islam Rayhan

Shahjalal University of Science and Technology, Sylhet, Bangladesh

Department of Electrical and Electronic Engineering

Date: 29 June 2023

Abstract

Proteins, as long polypeptide chains composed of 50 or more residues, are known to adopt unique structures based on their amino acid sequences. However, proteins exhibit dynamic behavior and exist as ensembles of conformations rather than rigid structures. The prediction of a protein's three-dimensional structure from its primary amino acid sequence poses a challenging problem known as protein folding. Classical algorithms offer practical solutions for sampling the conformation space of small proteins. However, the inherent complexity of the protein folding problem remains challenging even when simplified to the Hydrophobic-Polar model. On the other hand, despite the current unattainability of fault-tolerant quantum computers using state-of-the-art quantum technologies, there is promising evidence suggesting that quantum algorithms can be effectively employed in noisy quantum computers to accelerate energy optimization in frustrated systems. In this study, we employed a quantum simulator provided by IBM, utilizing the Qiskit framework, to investigate the protein folding problem with a dataset encompassing over 7 amino acids. By leveraging the power of quantum computation, we aimed to explore potential advancements in energy optimization for frustrated systems.

Introduction

The protein folding problem is of utmost importance in understanding the functions and mechanisms of proteins, as their unique structures dictate their biological properties. The complexity of the protein folding problem becomes apparent when considering the vast conformational space that proteins can explore. Even a relatively short protein chain consisting of 100 amino acids can potentially adopt an astronomical number of conformations, estimated to be the order of 10^{47}. This combinatorial explosion of possible conformations presents a daunting computational task for finding the proteins native, low-energy conformation. The Levinthal paradox further highlights the conundrum of protein folding. It states that if a protein were to explore all possible conformations sequentially, even at a rapid rate, it would take an unreasonable amount of time, far exceeding the age of the universe. This observation raises the intriguing question of how proteins manage to efficiently navigate the conformational landscape to find their native, low-energy structures without exhaustive sampling. While proteins seem to possess

inherent biases or mechanisms that guide them towards their low-energy conformations, replicating this efficiency computationally presents a significant challenge. In this work, our goal is to tackle the protein folding problem by leveraging the power of quantum computing. We aim to encode the protein's conformational search space into a qubit operator and devise quantum algorithms to explore this space efficiently. By mapping the problem onto a quantum system, we can exploit the unique properties of quantum mechanics to tackle the computational complexity associated with protein folding.

Methods and Analysis

The journey of exploring a polypeptide chain begins with the intricate task of encoding each individual bead into a qubit operator, while carefully adhering to all the physical constraints involved. This encoding process, known as Hamiltonian encoding, captures the essence of the chain's unique characteristics within the realm of quantum computation. Armed with this quantum representation, we embark on a transformative path utilizing the power of variational quantum eigen solvers (VQE). This remarkable algorithm unravels the secrets of the polypeptide's lowest energy conformation by delving into the realm of eigenvalues, seeking out the most optimal arrangement of its constituents. Yet, the true beauty lies in the flexibility offered by the parametrized nature of VQE. At each iteration, we measure the output, allowing us to glean valuable insights and data. With this wealth of information in hand, classical algorithms come to the fore, skillfully optimizing our solution, homing in on the elusive lowest energy conformation we seek.

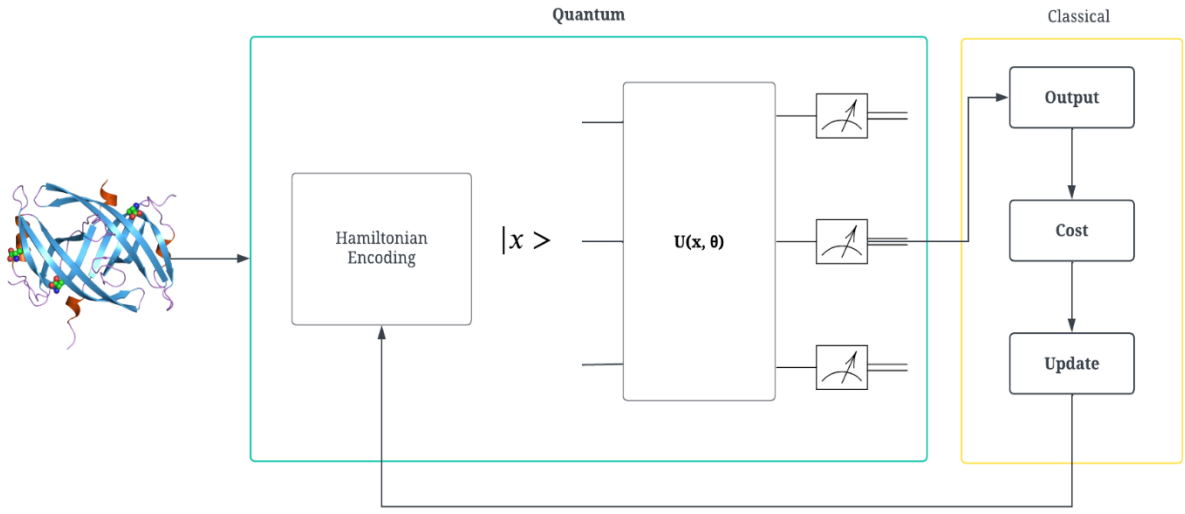

Figure-1: Protein Folding Enhanced by Quantum and Classical Algorithms

Hamiltonian Encoding:

In our quest to decipher the intricate world of protein folding, we delve into the realm of quantum computing, harnessing the power of two distinct qubit types - configuration qubits and interaction qubits. The configuration qubits eloquently capture the essence of each bead, elegantly encoding its unique position and graceful turns within the polypeptide chain. With each configuration qubit, we unlock the secrets of the bead's spatial arrangement, unveiling the delicate choreography of its structure. But our journey does not stop there. The interaction qubits add a mesmerizing dimension to the quantum landscape. These qubits intricately depict the interplay and connections between neighboring beads, allowing us to explore the subtle forces that shape the protein's folding pathway.

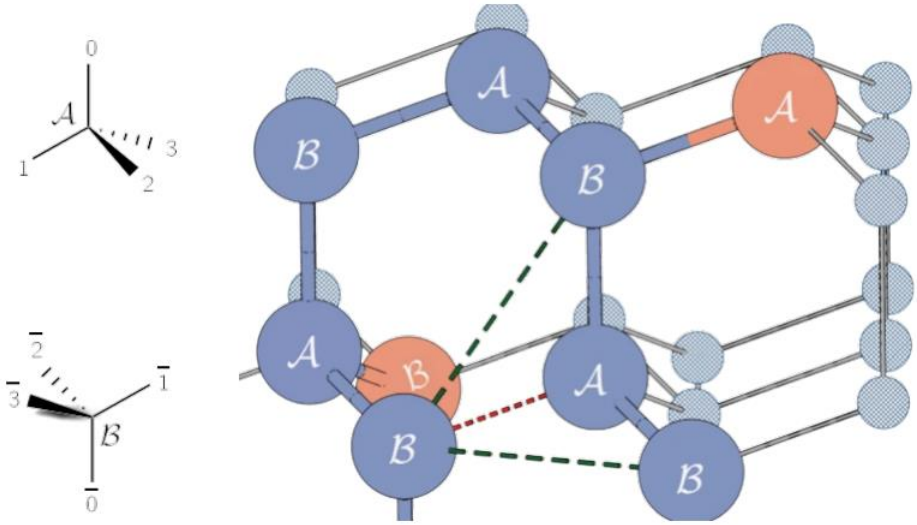

Figure-2: Labelling of co-ordinate system at the sub lattice A and B

For configuration qubit if t_i is the "turn" which defines the position of the beads i+1 relative to the previous position i. At A site t_i can only grow in the direction of t_i ε {0,1,2,3} at B sites it will be alternate to A which is t_i ε $\{\bar{0}, \bar{1}, \bar{2}, \bar{3}\}$. So, if one qubit per axis then,

$$t_i = q_{4i-3}q_{4i-2}q_{4i-1}q_{4i} \quad (1)$$

Total number of qubits required to encode a conformation q_{cf} corresponds to,

$$N_{cf} = 4(N-3) \quad (2)$$

While total number of strings as the number of monomers increases will be,

$$N_t = a(N-b)^4 \quad (3)$$

And for dense coding it will be twice the number of its initial part.

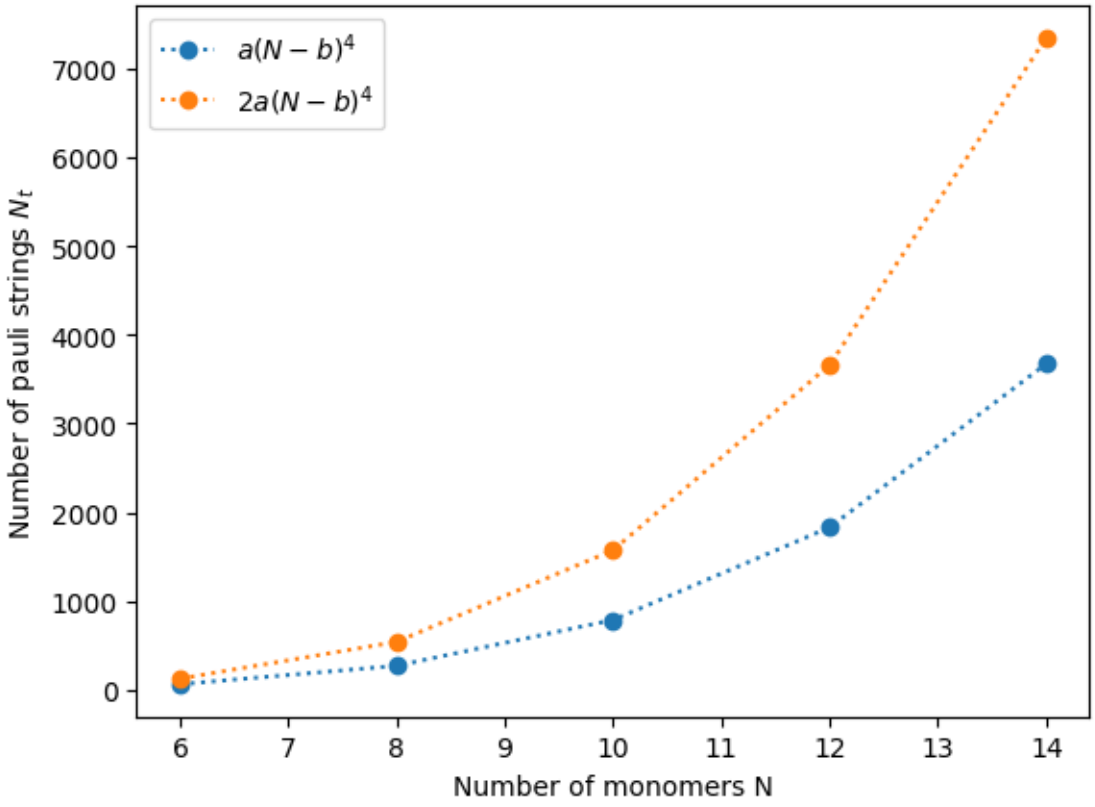

Figure-3: Number of Pauli strings with respect to the number of monomers for the sparse and dense encoding models

For interaction qubit which is q_{in}, composed of $q_{i,j}^{l}$ for each l^{th} nearest neighbor interaction on the lattice. In this work we are using three types of interaction available by the IBM Qiskit. Which is Miyazawa Jernigan Interaction, Mixed Interaction and Random Interaction.

The qubit Hamiltonian describes the energy of a given fold defined by the sequence of beads and the encoded turns. The different contributions to the polymer Hamiltonian are,

$$H(q) = H_{gc}(q_{cf}) + H_{ch}(q_{in}) + H_{in}(q) \quad (4)$$

Where H_{gc} defines the energy of the geometrical constraint, H_{ch} which defines the chirality and H_{in} which defines the interaction energy terms.

Variational Quantum Eigensolver (VQE):

The aim of using variational quantum eigensolver is to find the lowest eigenvalue of a given matrix. If $|\psi\rangle$ is a wavefunction then the time independent Schrodinger equation will be,

$$\hat{H}|\psi\rangle = E\,|\psi\rangle \quad (5)$$

Where $\hat{H}$ is the Hamiltonian and E yields a specific energy. Our goal is to find the smallest eigenvalue which is why we can say the expectation value of the Hamiltonian is bounded by a certain energy which is $\langle\psi|\hat{H}|\,\psi\rangle \geq E_0$. Where we want to optimize the value of E_0 which is unknown by minimizing the expected value of Hamiltonian.

$$Min_\theta \,\langle\psi(\theta)|\hat{H}|\,\psi(\theta)\rangle \quad (6)$$

Now we must map this Hamiltonian to a quantum computer. To address this problem let's see what the Hamiltonian is,

$$\hat{H} = \hat{T} + \hat{V} \quad (7)$$

Hamiltonian describes the energy of the system which is kinetic ($\hat{T}$) and potential energy ($\hat{V}$) for the system. Where $\hat{V}$ depends on the position of the system. $\hat{V}$ may refer to the energy that is between the position of the polypeptide chain also considering all the constraint regarding the factor. We can think of three different parts for VQE.

Pauli Encoding:

For a fermionic system, it is often most convenient to qubitize: that is to write the many-body Hamiltonian of the system using second quantization, and then use a mapping to write the creation-annihilation operators in terms of Pauli operators. Once the Hamiltonian $\hat{H}$ is written in terms of Pauli operators and irrelevant states are discarded, it would consist of linear combination of Pauli strings $\hat{P}_i$ consisting of tensor product of Pauli operators, such that.

$$\hat{H} = \sum_i \alpha_i \hat{P}_i \quad (8)$$

Where α_i are the numerical coefficients. Based on the coefficient, the number of Pauli strings can be reduced in order to optimize the problem.

Ansatz and Initial Trial Function:

The choice of ansatz state depends on the system of interest. In gate-based quantum computing, the ansatz is given by a parametrized quantum circuit, whose parameters can be updated after each run. The ansatz must be adaptable enough to not miss the desired state.

Measurement:

The expectation value of a given state $|\psi(\theta_1, \theta_2, \ldots., \theta_n)\rangle$ with parameters $\{\theta_i\}_{i=1}^{n}$, has an expectation value of the energy or cost function given by,

$$E(\theta_1, \ldots, \theta_n) = \langle \widehat{H} \rangle = \sum_i \alpha_i \left\langle \psi(\theta_1, \theta_2, \ldots., \theta_n) \middle| \widehat{P}_i \middle| \psi(\theta_1, \theta_2, \ldots., \theta_n) \right\rangle \quad 9$$

In order to obtain the expectation value of the energy, one can measure the expectation value of each Pauli string. This step corresponds to measuring each qubit in the axis provided by the Pauli string.

Output and Optimization:

Using classical algorithms in a digital computer, the parameters of the ansatz can be optimized. For this minimization, it is necessary to find the minima of a multivariable function. Classical optimizers using Constrained Optimization by Linear Approximation (COBYLA) optimizer can be used for this purpose. By running the circuit many times and constantly updating the parameters to find the global minima of the expectation value of the desired observable, one can approach the ground state of the given system and store it in a quantum processor as a series of quantum gate instructions.

Result and Discussion

In our quest to unravel the mysteries of protein folding, we embarked on a remarkable journey with an eight-bead polypeptide chain, gracefully composed of the sequence "APRLRFRY". Adding to its allure, we adorned this chain with two exquisite side chain beads, amplifying the complexity of our exploration. The interactions that govern the folding of this captivating protein were brought to life through the enchanting interplay of mixed interactions. We carefully orchestrated the delicate dance between the beads, allowing them to weave a tapestry of intricate connections. Yet, our scientific pursuit did not stop there. We ventured into the realm of different interaction models, uncovering the nuanced influences that shaped the protein's folding pathway. Each interaction model whispered its own secrets, enriching our understanding of the profound interplay between structure and function. As we expanded our investigation to encompass larger polypeptide chains, we became acutely aware of the interplay between complexity and time. The graceful dance of discovery unfolded at its own pace, with the time required to unveil the lowest eigenvalue increasing as the number of beads in the protein grew. The intricate nature of protein folding demanded patience and computational prowess as we ventured deeper into the conformational landscape.

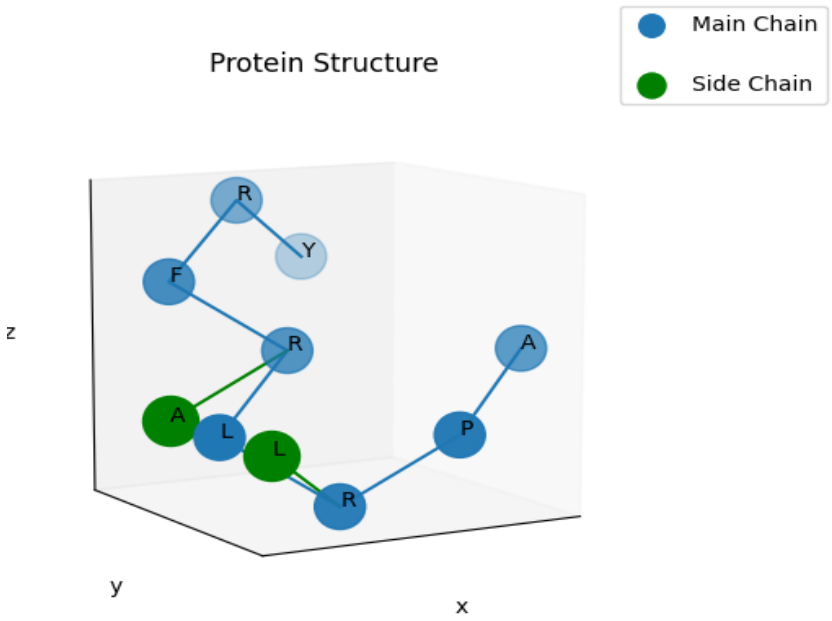

Fig: 3D protein structure

VQE, with its quantum prowess, became a conduit for this exploration, providing us with a treasure trove of conformations per energy. Each measured energy level offered a glimpse into the protein's equilibrium state, where it found solace in the depths of its lowest energy conformations. The convergence of quantum computation and energetic analysis became a gateway to unlocking the secrets hidden within the protein's structural enigma.

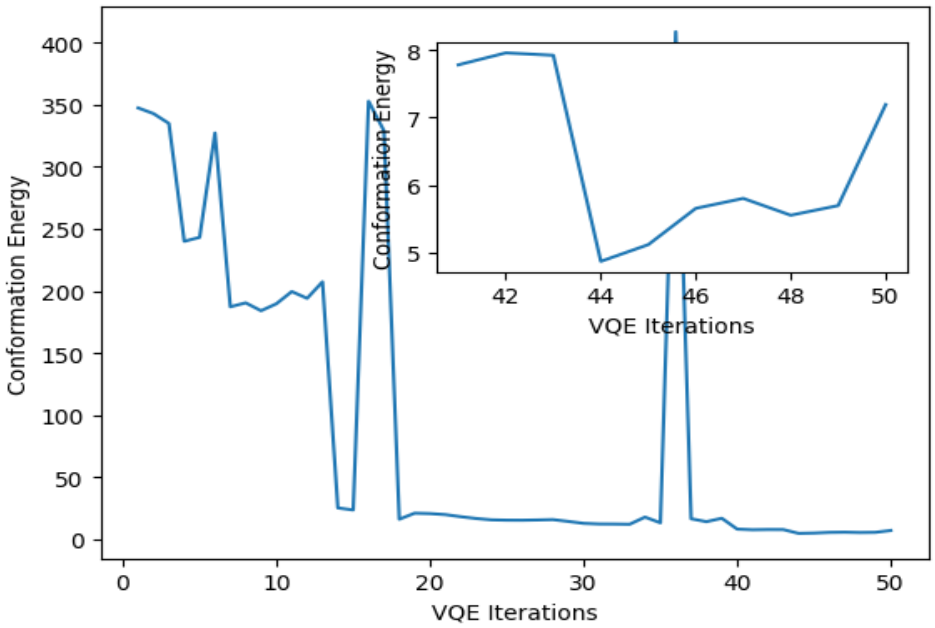

Fig: Conformation energy per iteration using VQE

By using VQE the minimum eigenvalue we get is,

Eigenvalue	4.867014277800888
Bitstring	1000001001001101001
Value	-1.3570000000017899+0j
Probability	7.6681920831e-06
State	266857

Fig: Result of the sampling eigen solver

The quest for solutions to larger polypeptide chains proved time-consuming, straining classical computational resources. Limited availability of vital resources further intensified our pursuit. However, within the tapestry of these challenges lie profound opportunities. Custom interaction models offer tailored insights, while Quantum Approximate Optimization Algorithm (QAOA) beckons with its transformative power. By refining the Variational Quantum Eigen solver (VQE) ansatz, we unlock efficiency, reducing the qubit count. Together, we venture towards a future where the beauty of protein folding unfolds seamlessly, transcending computational boundaries.

Acknowledgements

This work is done under QuantumAI. An Indian quantum-based platform. Also, the source code is given below.

……..

References

1. Robert, A., Barkoutsos, P.K., Woerner, S. *et al.* Resource-efficient quantum algorithm for protein folding. *npj Quantum Inf* **7**, 38 (2021). https://doi.org/10.1038/s41534-021-00368-4
2. Sanzo Miyazawa, Robert L. Jernigan, Residue – Residue Potentials with a Favorable Contact Pair Term and an Unfavorable High Packing Density Term, for Simulation and Threading, Journal of Molecular Biology, Volume 256, Issue 3, 1996, Pages 623-644,ISSN 0022-2836, https://doi.org/10.1006/jmbi.1996.0114.
3. (n.d.). *Lavinthal Paradox*. Wikipedia. https://en.wikipedia.org/wiki/Levinthal's_paradox
4. Panagiotis Kl. Barkoutsos, Giacomo Nannicini, Anton Robert, Ivano Tavernelli, and Stefan Woerner. Improving the Variational Quantum Optimizer using CvaR. Quantum 4, 256(2020). https://doi.org/10.22331/q-2020-04020-256
5. Roshni KG (2021) Protein folding, misfolding, and coping mechanism of cells–A short discussion. Open J Cell Protein Sci 4(1): 001-004. DOI: 10.17352/ojcps.000003
6. (n.d). VQE. PENNYLANE. https://pennylane.ai/qml/demos/tutorial_vqe

7. Hasan Mustafa, Sai Nandan Morapakula, Prateek Jain, Srinjoy Ganguly. Variational Quantum Algorithms for Chemical Simulation and Drug Discovery. 15 November 2022. https://doi.org/10.48550/arXiv.2211.07854
8. Vishwakarma, R. K. (2022, October 2). Protein folding using quantum computers. https://doi.org/10.31219/osf.io/8byzf