Protein folding using quantum computers

Rajat Kumar Vishwakarma[1]

[1]Department of Physics, University of Allahabad

October 4, 2022

Protein folding using quantum computers

Rajat Kumar Vishwakarma
University of Allahabad, Department of Physics, 211013 Prayagraj, India
Source Code
(Dated: January 5, 2022)

Abstract

Protein folding has been one of the most difficult problems for over a half-century,The random thermal motions causing conformational changes that lead energetically downhill towards the native structure, a principle captured in funnel-shaped energy landscapes.Unfolded polypeptides have a wide range of possible conformations. The search problem becomes intractable for classical computers due to the exponential growth of potential conformations with chain length. So far, there is theoretical and experimental evidence that solving such optimization problems using Quantum Computing approaches such as Quantum Annealing, VQE, and QAOA has an advantage. Although Google's DeepMind-AlphaFold has accomplished much,But we can go even further with the quantum approach.Here we show how to predict structure of protein as well as RNA folding using the Variational Quantum Eigensolver with Conditional Value at Risk (CVaR) expectation values for the solution of the problem and for finding the minimum configuration energy and our task is to identify a protein's minimal energy structure. The protein's structure is optimized to reduce energy. Also making sure that all physical constraints are met and encoding the protein folding problem into a qubit operator.

Introduction

Life can't exist without Proteins, and knowing their structures helps use large-scale experiments to mechanically understand how they work, The molecular biologist Cyrus Levinthal figured out in 1969 that if proteins were taking this plodding, check-each-possible-fold approach in our bodies, their folding process would take longer than the entire lifetime of the universe. This is called Levinthal's Paradox. About 100,000 different protein structures have been identified.This is a small fraction of the billions of known protein sequences. Accurate computational methods are needed to fill this gap and support comprehensive structural bioinformatics. Estimates the three-dimensional structure of proteins determined only by amino acids. We can determine the total number N(n) of unique lattice conformations that separate diverse protein intermediate structures for a protein with n amino acid residues. For illustration, N(4) = 4 and N(6) = 22. This

gives us a set of N(n)-items, which we refer to as the structure set and expressed by S(n) = {s1,s2,....,sN(n)}.
Let $|S_a\rangle$ denotes the state of protein structure in the a-th confirmation we have Hamiltonian in N_n-dimension Hilbert space, so that $\mathcal{H} = |sa\rangle|a = 1, 2, \cdots, Nn$,

$$\mathcal{H} = -\sum_{a,b} J_{ab}|s_a\rangle\langle s_b|, \tag{1}$$

According to their contact interaction, the n amino acids are divided into hydrophobic and hydrophilic (also known as polar) groups from a coarse-grained perspective. Since H and P are traditionally used to denote the hydrophobic and polar amino acids, respectively, a sequence of n amino acids can be labeled by q = {q1, q2, , q(n)} where q(k) with k = [1, 2, , n] denotes either H or P. As a result, there will be a total of 2^n potential sequences. Let's refer to the collection of all random sequences as the sequence set denoted by $\delta_n = \{[\nu]|\nu = 1, 2,..., 2^n\}$. We may get the total contact energy for each structure in delta for any particular sequence-[ν] specified by a **q.**

$$\varepsilon_a^{[v]} = \sum_{k<l} E_{q_k q_l}\delta|r_k^a - r_l^a|, 1(1 - \delta|k - l|, 1), \tag{2}$$

Quantum Dynamics, We consider the energy dissipation brought on by the medium in which the folding takes place in order to reach a quantum mechanical understanding. The Lindblad equation governs this situation.

$$\frac{d}{dt}\bar{p} = \frac{1}{i\hbar}[H, \bar{p}] + \mathcal{L}(\bar{p}), \tag{3}$$

Where $\mathcal{L}_{\bar{p}}$ reflects the effect of dissipation;

$$\mathcal{L}(\bar{p}) = \frac{\lambda}{2}\left(2L\bar{p}L^\dagger - \bar{p}L^\dagger L - L^\dagger L\bar{p}\right) \tag{4}$$

According to analyses of random walks with sticky vertices, L and L † are referred to as the Lindblad operators in this context. This operator L † = ∑ ab Γ$ab|sa\rangle\langle sb|$. is provided by the off diagonal part discussed before. For a two-level system that can be thought of as the two vertex graphs with a sticky vertex, actual equation.4 provides a general expression that becomes the standard one in terms of Pauli matrices.

The Folding Time

Now that we have the idea of the mean first passage time, we can develop a definition for the protein folding time. The first passage probability vanishes, $Fa,b(\tau 0\ ) = 0$, which really happens for the aforementioned quantum walk, throughout the mean first-passage time from a starting state $|s_a\rangle$ to a destination state $|s_b\rangle$, which is given by

$$\int_{t=0}^{\tau_0} tF_{a,b}(t)dt / \int_{t=0}^{\tau_0} F_{a,b}(t)dt \tag{5}$$

$$P_{a,b}(t) \ = \ \int_0^t F_{a,b}\left(t'\right) P_{b,b}\left(t - t'\right) dt' \tag{6}$$

Here $P_{a,b}(t)$ denotes the probability of state |b> at time period t starting from the state |a> at initial time t=0.
Protein folding, which gives proteins their native structure, takes place when the starting state is chosen to be $|s1\rangle$ and the target states are those that are the most compact. For any positive n, they are for instance, $|s(n)\rangle$. The following is the formula for calculating the folding time:

$$T_{\tau_{fd}} = \frac{\int_0^{\tau_0} tF_{1,c}(t)dt}{\int_0^{\tau_0} F_{1,c}(t)dt} \tag{7}$$

Lattice conformational model

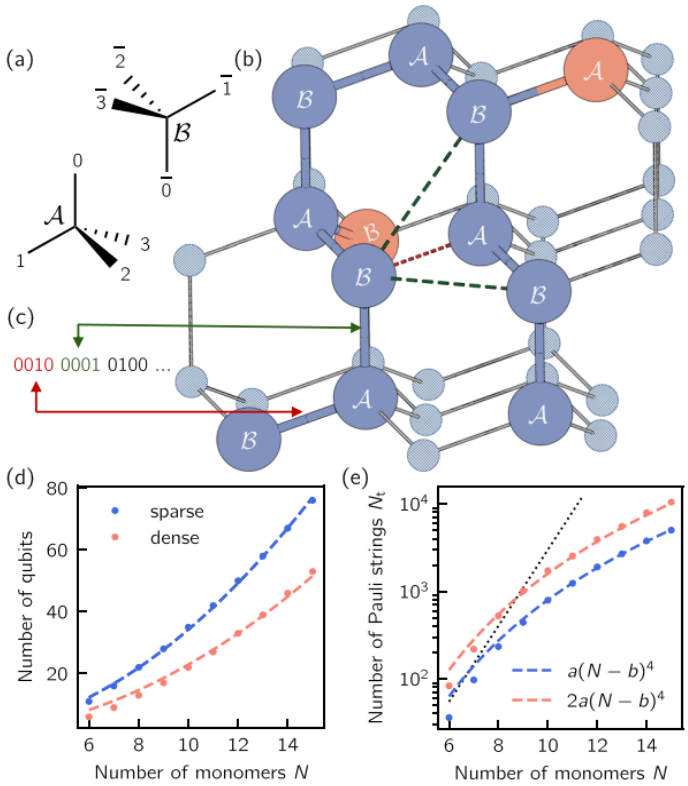
We call the backbone bead of index i ∈ {1, . . . , N} in the polymer's primary sequence and denote the beads constituting its side chain by i (1), i(2), ..., i(s). Discussions about the locality of the Hamiltonian in relation to the sparser encoding are held in the absence of additional information.

Encoding(sparse and dense)

Encoding with Sparser The polymer sequence is generated by specifying the series of lattice turns beginning with bead i = 1. Each turn is encoded on n qubits q4i3q4i2q4i1q4i, each of which represents a tetrahedral lattice direction. Only one of the n qubits will have a value of one (the others will be set to zero). We encode side chain turns in $i\ as\ q^{(k-1)}{}_k$, where the upper index in parenthesis labels n qubits describing the turn of side chain bead i(k), and the lower indices of the form i(k) label the bead k along the side chain at main chain position i.
This encoding requires **4(N − 3)** qubits to totally define a conformation.

Encoding Denser The turn is represented by n qubits. As a result, we will encode the turns in a given side chain using . Without losing generality, we choose qubit state function Ψ as the first nth turns. If the main chain's bead does not bear a side chain, another qubit can be saved without breaking any symmetry. The side chains can be encoded using the same rules as the sparser encoding. The general form of a string of bits defining the entire conformation will have the side chain qubits in parentheses. To completely define a conformation, this encoding requires **2(N - 3)** qubits.

Fig. 1



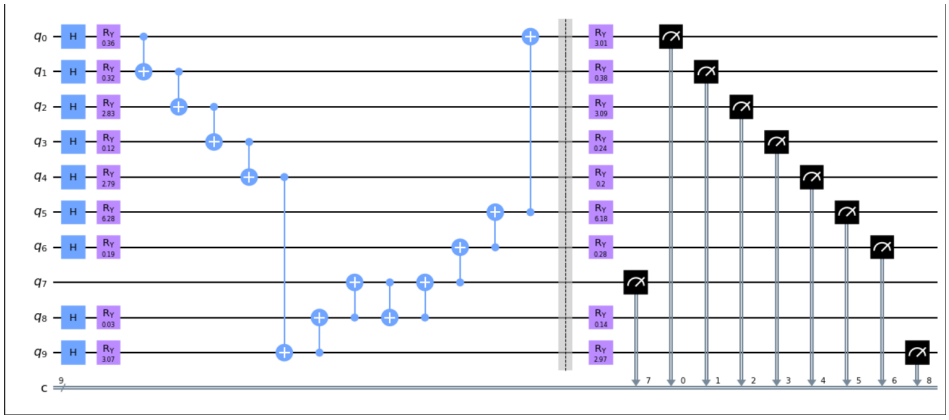
Tetrahedral lattice polymer model

Interaction qubits

To describe the interactions, we introduce a new qubit register q_{in}, which is made up of $q^{l}_{i,j}$ for each nearest neighbor (l-NN) interaction on the lattice (see red and green dashed lines for i= 1 and i = 2 in Fig. 1, b). The application of these registers will be discussed in conjunction with the definition of interaction energy terms. The number of qubits comprising the interaction register, , is entirely determined by the polymer skeleton (i.e., including the side chains), regardless of the color of the beads, and scales as $\mathcal{O}(\mathcal{N}^2)$. It is worth noting that two 1-NN beads occupy different sub-lattices (A or B). For i > 1, however, all beads in both sub-lattices have the potential to interact.The pairwise interaction energies i;j between the beads at distance l can be randomly

chosen given a basic sequence or they can be modified from pre-existing models, such as the one proposed by Miyazawa and Jernigan (MJ) for 1-NN interactions, to produce an unique fold.

Fig.2



Results of an experiment on the folding of a neuropeptide of seven amino acids For the creation of the protein configurations, a parametrized quantum circuit is used. The ideal fold is reconstructed using the best possible set of qubit gate Rotations.

The Hamiltonian

The second process is to create the qubit Hamiltonian that characterizes the energy of a specific fold determined by a fixed bead sequence and encoded turns. Physical interactions (attractive or repulsive in nature) are applied when two beads occupy neighboring sites or are at a distance l > 1, where l is the length of the shortest lattice path connecting them. Penalty terms are applied when physical constraints are violated (for example, when beads occupy the same position on the lattice or overlap). Therefore, for (with q = {q(cf), (qin)}) the different contributions to the polymer Hamiltonian are,"

$$H(q) = H_{gc}(q_{cf}) + H_{ch}(q_{cf}) + H_{in}(q) \qquad (8)$$

The definitions of the chirality constraint (Hch), which enforces the correct stereochemistry of the side chains if present, and the geometrical constraint (Hgc), which governs the expansion of the primary sequence without splitting or say bifurcation.
Also the number of terms (or Pauli strings) in the n-qubit Hamiltonian H is what we use to determine the algorithm's scaling (q) given by,

$$H(q) = \sum_{\gamma}^{N_t} h_{\gamma} \bigotimes_{i=1}^{n} q_i^{\gamma_i} \qquad (9)$$

Where; h_γ is real coeff. $q_i = (1 - \alpha_i^z)/2$, α_i^z is the Z-Pauli matrix, $\gamma(i) \in \{0, 1\}$, and N_t is the total number of terms.

The variational quantum algorithm

For the 128 shots (iter=1024 shots) simulation with an "fulll" entangling scheme on Qiskit, the noisy simulations were carried out with α= 1% (resp. α= 0.1%). Every circuit was built with a VQE depth of m = 2. Given a run of n qubits, the suggested method, dubbed CVaR-VQE, is a hybrid quantum-classical algorithm that functions similarly to all hybrid schemes proposed for upcoming quantum processors. At every single VQE iteration, the genetic algorithm creates a "population" of parameters (i.e., an offspring), which preserves the memory of previous generations and introduces a degree of stochasticity (through mutation rates) in a process where analytic gradients are very ineffective at directing the minimization process.

Fig. 3, **Convergence of VQE energy per iterations**

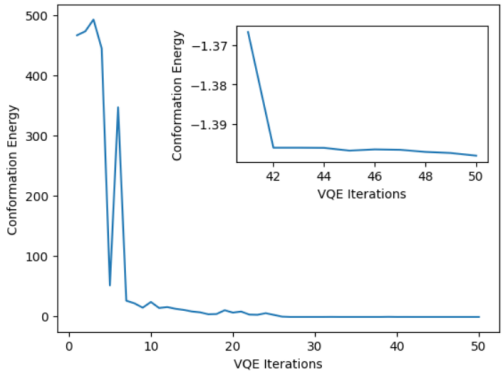
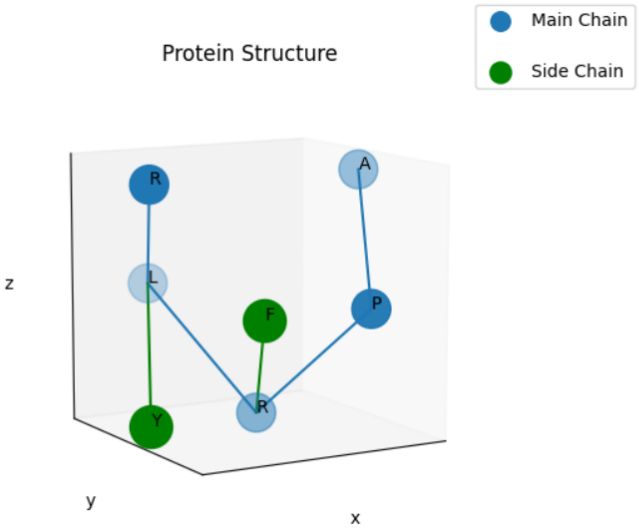


Fig. 4 **Stable 7 beads protein structure**



DATA AVAILABILITY

The data not directly reported in the main text are available from the corresponding author upon reasonable request.

CODE AVAILABILITY

All source codes and materials are available here: https://github.com/rajat709/Quantum-Protein-Folding

ACKNOWLEDGEMENTS

Thanks to A. Robert and QETCI for helpful guides on the Project. This work was done on(QETCI, open research project) supported by the AU(Quantum-Physics), prayagraj Uttar pradesh 211013, source code and data availability statements, provided in a public GitHub profile.

REFERENCES

1.Resource-efficient quantum algorithm for protein folding(2021).
2.Zwanzig, R., Szabo, A. & Bagchi, B. Levinthal's paradox. Proc. Natl. Acad. Sci. U.S.A 89, 20–22 (1992).
3.Onuchic, J. N. & Wolynes, P. G. Theory of protein folding. Curr. Opin. Struct. Biol. 14, 70–75 (2004).
4.Piana, S., Klepeis, J. L. & Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Curr. Opin. Struct. Biol. 24, 98–105 (2014).
5.Duan, Y. Pathways to a Protein Folding Intermediate Observed in a 1- Microsecond Simulation in Aqueous Solution. Science 282, 740–744 (1998).
6.Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How Fast-Folding Proteins Fold. Science 334, 517–520 (2011).
7.Levitt, M. & Warshel, A. Computer simulation of protein folding. Nature 253, 694 (1975).
8. Hinds, D. A. & Levitt, M. A lattice model for protein structure prediction at low resolution. Proc. Natl. Acad. Sci. U.S.A 89, 2536–2540 (1992).
9. Perdomo Ortiz, 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 (2008).
10. 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 (2012).
11.Cai, J., Macready, W. G. & Roy, A. A practical heuristic for finding graph minors. Preprint at: arXiv:1406.2741 [quant-ph] http://arxiv.org/abs/1406.2741 (2014).
12. Miyazawa, S. & Jernigan, R.L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term,for simulation and threading. J. Mol. Biol. 256, 623–644 (1996).
13. Peruzzo,A.et al. A variational eigenvalue solver on a photonic quantum processor. Nat. Commun. 5(2014).
14.McClean, J. R., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. New J. Phys. 18, 023023 (2016).
15.Barkoutsos, P. K., Nannicini, G., Robert, A., Tavernelli, I. & Woerner, S. Improving variational quantum optimization using CVaR. Quantum 4, 256 (2020).
16.Storn, R. & Price, K. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. J. Global Optim. 11, 341–359 (1997).

17.Hadfield, S. et al. From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. Algorithms 12, 34 (2019).
18.Das, S., Mullick, S. S. & Suganthan, P. Recent advances in differential evolution – an updated survey. Swarm Evol. Comput. 27, 1–30 (2016).
19.Kempe, J., Kitaev, A. & Regev, O. The complexity of the local hamiltonian problem. SIAM J. Comput. 35, 30 (2006).
20.Babbush, R., Perdomo-Ortiz, A., O'Gorman, B., Macready, W. & Aspuru-Guzik, A. Construction of energy functions for lattice heteropolymer models: a case study in constraint satisfaction programming and adiabatic quantum optimization. arXiv preprint arXiv:1211.3422 (2012).
21.Barkoutsos, P., Nannicini, G., Robert, A., Tavernelli, I. & Woerner, S. Improving variational quantum optimization using CVaR. Quantum 4, 256–271 (2020).