Quantum Computing for Predicting Protein Folding and Interactions

Scott Longwell
*Department of Bioengineering, Stanford University**
(Dated: May 19, 2019)

Abstract to go here.

I. PERSONAL BACKGROUND

I am a Bioengineering PhD student in the Fordyce Lab, where we develop and apply experimental tools to understand how changing a protein's amino acid sequence emerges as a change in various functions (such as folding, binding, and enzymatic activity). The majority of my coursework has been in machine learning and deep learning.

II. PROBLEM BACKGROUND

Understanding the principles that govern how proteins fold and bind on another - effectively intra- and intermolecular versions of the same process - would usher in a new era in drug design and personalized medicine [1]. It is also notoriously difficult to predict the trajectory of these processes from the sequence of amino acids that comprise proteins. In the case of protein folding, the Levinthal paradox notes that for a protein of length n amino acids, there are $n-1$ phi-angles and $n-1$ psi-angles within the protein backbone, each of which has ~ 3 possible values [2, 3]. In all, this leads to an exponential number of conformations which a folding protein can sample: $3^{2(n-1)}$, ignoring a large number of sterically impossible conformations. Despite this massive number, proteins fold to low-energy minimia on incredibly short timescales (10^3 down to 10^6 seconds), suggesting that there are heuristics and principles to efficiently traverse this exponentially large conformation space.

It is often stated that the massively parallel nature of quantum computing will be the advance that renders protein folding tractable to predict [4]. There are several possible embeddings of the protein folding process [5]. A simplistic but commonly used approach is to map the amino acids of a protein sequence onto a cubic lattice[6] by placing the first two amino acids at adjacent positions and placing the remaining amino acids sequentially, each time making a move in one of the 6 possible absolute directions on the lattice. A cost function is then established that a) penalizes chain overlaps and b) accounts for interactions between adjacent amino acids. This embedding has lent itself well to quantum annealers, with Perdomo and colleagues demonstrating folding of short peptides on D-Wave systems [7, 8]. More recently, a team from ProteinQure demonstrated similar capabilities on the D-Wave system, as well as the first implementation on a gate-based universal quantum computer (19-Q Acorn at Rigetti) [9, 10].

III. PROPOSED WORK

My proposed work has a minimum requirement and possible extensions.

Minimum Requirement:

- Replicate QAOA as in Babej, et. al. 2018 [10] and document within a Jupyter notebook.

Extension Ideas:

- run on quantum computer
- consider orientation of amino acids when evaluating cost
- different lattice basis based on psi- and phi- angles (2N angles, but only 3 'moves' for each angle).

[1] K. A. Dilland J. L. MacCallum, The Protein-Folding Problem, 50 Years On, Science **338**, 1042 (2012).
[2] C. Levinthal, How to fold graciously, Mossbauer spectroscopy in biological systems **67**, 22 (1969).
[3] A. V. Melkikhand D. K. F. Meijer, On a generalized Levinthal's paradox: The role of long- and short range interactions in complex bio-molecular reactions, including protein and DNA folding, Progress in Biophysics and Molecular Biology **132**, 57 (2018).
[4] Y. Cao, J. Romero, and A. Aspuru-Guzik, Potential of quantum computing for drug discovery, IBM Journal of Research and Development **62**, 6:1 (2018), good overview.
[5] R. Babbush, A. Perdomo-Ortiz, B. O'Gorman, W. Macready, and A. Aspuru-Guzik, Construction of Energy Functions for Lattice Heteropolymer Models: A Case Study in Constraint Satisfaction Programming and Adiabatic Quantum Optimization, arXiv:1211.3422 [cond-mat, physics:math-ph, physics:physics, physics:quant-ph] , 201 (2014), arXiv:1211.3422 [cond-mat, physics:math-ph,

* longwell@stanford.edu

physics:physics, physics:quant-ph].
[6] A further simplification is to use a planar lattice.
[7] A. Perdomo, C. Truncik, I. Tubert-Brohman, G. Rose, and A. Aspuru-Guzik, Construction of model Hamiltonians for adiabatic quantum computation and its application to finding low-energy conformations of lattice protein models, Physical Review A **78**, 012320 (2008).
[8] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik, Finding low-energy conformations of lattice protein models by quantum annealing, Scientific Reports **2**, 571 (2012).
[9] M. Fingerhuth, T. Babej, and C. Ing, A quantum alternating operator ansatz with hard and soft constraints for lattice protein folding, arXiv:1810.13411 [quant-ph] (2018), proteinQure-II (Rigetti), arXiv:1810.13411 [quant-ph].
[10] T. Babej, C. Ing, and M. Fingerhuth, Coarse-grained lattice protein folding on a quantum annealer, arXiv:1811.00713 [quant-ph] (2018), proteinQure-I (D-Wave), arXiv:1811.00713 [quant-ph].