RQuanTech
Boosting the Qubit Revolution

QUANTUM ALGORITHMS FOR STRUCTURAL MOLECULAR BIOLOGY

RQuanTech has developed an innovative algorithm for a potential use in the structural molecular biology applications.

INTRODUCTION

For billions of years, the process of evolution has optimized the sequence of amino acids that make up naturally occurring proteins to suit the needs of the organisms that make them. So we ask: can we use computation to design non naturally-occurring proteins that suit our biomedical and industrial needs?
This question is a combinatorial optimization problem, because the output of a protein design computation is a sequence of amino acids. Due to the vast diversity of naturally occurring proteins, it is possible and very useful to begin a protein design computation with a naturally occurring protein and then to modify it to achieve the desired function. These can be formalized as problems in weighted chains. We can represent each amino acid by a distinct character and a protein is represented by the amino acid character string that makes it up. The chain representing the protein is divided into sub-chains, the mass of each substring is determined, and the list of masses is compared to a database of proteins. One of the challenges for this technique is dealing with very large strings of characters, which may have several possible substrings. The number of substrings selected is critical to good results. Also is critical the scoring function for comparing these letters, to find either the best global alignment or the best local alignment between the two sequences. Unfortunately, it is not easy to find the parameters of a scoring function that best captures the similarity between amino acid types. This has led to the development of many types of scores in the form of substitution matrices in the hope of producing biologically meaningful sequence alignments. In addition, when the similarity between the two proteins to be compared is low, the quality of the corresponding sequence alignment is usually lacking. Therefore, sequence alignment techniques are usually poor methods for classifying proteins into folds or detecting homology, both essential tasks in the hope of solving the protein structure prediction problem. There have been many methods developed to circumvent these problems.
One option to improve upon these methods involves considering multiple amino acids at once. This idea has led to the concept of “alignment-free” methods which have been developed over the past three decades. the sequence alignment methods as well as the recent string kernel methods depend critically on a scoring, or substitution matrix. Those substitution matrices basically encode amino acids as arrays of numerical values, where those values are derived from statistical analyses of reference alignments (the PAM and BLOSUM matrices), or from the physical and chemical properties of amino acids. While those matrices have been used in the context of fold recognition problems, they have not been optimized for such a task. There have been attempts to perform such an optimization; none, however, have yet surpassed the well accepted BLOSUM62 matrix.

One of the big potential steps that could be achieved in the field, would be implementation of Quantum algorithms that allow to jump over the Non-Polynomial (NP) problems presented in the above biological computational methods.

EXAMPLE APPLICATION

[1]In our proposal (footnote), we present how **RQuanTech** quantum algorithms can accelerate the estimation of Knowledge-Based Statistical Force Fields and the cost functions to derive the final protein configurations. We show how the amplitude estimation algorithm could reduce drastically the number of steps required to estimate structures with high confidence.

For the modelling of protein structure problem, the main challenge is the exponential number of potential configurations to sample. The searching for the optimal configuration is often performed via Monte Carlo evaluation.

Quantum computing promises algorithmic speedups for a variety of tasks, such as factoring or optimization.

The Quantum Fourier Algorithm jointly with Shor's algorithm implemented by **RQuanTech** could be extended and generalized to function optimization, amplitude amplification and estimation, integration, quantum walk-based methods for element distinctness, and Markov chain algorithms, for example. In particular, the amplitude estimation algorithm can provide close to quadratic speedups for estimating expectation values, and thus provides a speedup to a problem for which Monte Carlo methods are used classically.

RQuanTech presents a new perspective of how to use quantum computing for the modelling of protein structure problem. We combine well-known quantum techniques, such as amplitude estimation and the quantum algorithm for Monte Carlo with the CABS coarse-grained protein model.

MONTE CARLO METHOD PAPER IMPLEMENTATION

As we can see in the below figure the paper proposes the implementation of Monte Carlo method as shown:

[1] Based on the paper: Modelling of Disordered Protein Structures Using Monte Carlo Simulations and Knowledge-Based Statistical Force Fields by: Maciej Pawel Ciemny, Aleksandra Elzbieta Badaczewska-Dawid , Monika Pikuzinska , Andrzej Kolinski and Sebastian Kmiecik ,(Int. J. Mol. Sci. 2019, 20, 606)

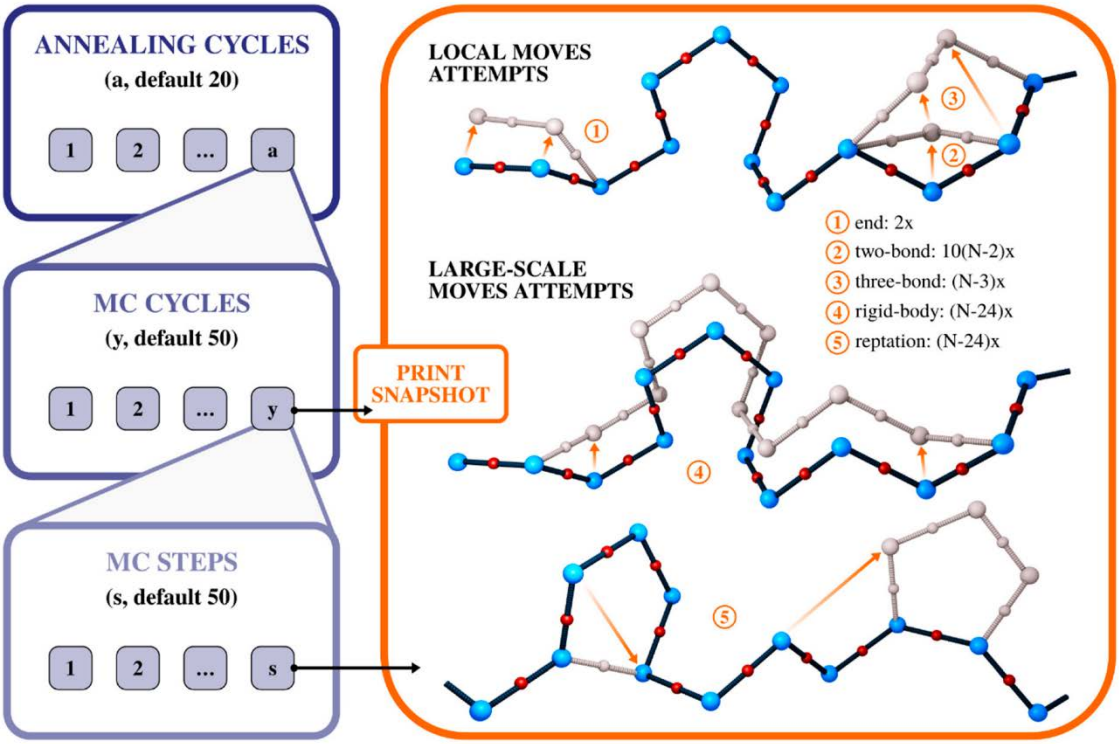


Figure 3

Sampling scheme of the CABS model. Blue panels show implementation details of Monte Carlo (MC) iterations (loops). The orange panel shows all motions that may be performed in a single MC step. The simulation is organized as a set of nested loops, in which the s number of MC steps is organized into the y number of cycles, and these in a annealing cycles (number of a, y or s cycles can be controlled by the user in CABS-flex and CABS-dock standalone packages [72]). In the orange panel, numbers 1 to 5 denote the available moves, presented together with the number of attempts to perform a move in each of the MC steps. The resulting trajectory is comprised of simulation snapshots saved at the end of each MC cycle.

RQUANTECH MONTECARLO IMPLEMENTATION

RQuanTech Monte Carlo implementation proceeds in the following way.

Assume that the optimal structure be P and P' be the approximation obtained from *k* samples. Assume that the random variable of the cost function f (S_T) is bounded in variance,

i.e. $V [f (S_T)] < \lambda^2$. Then the probability that the estimation P' is ε away from the optimal is determined by Chebyshev's inequality

$Prob [|P'-P| \geq \varepsilon] \leq \lambda^2 / k \varepsilon^2$

For a constant success probability, we thus require

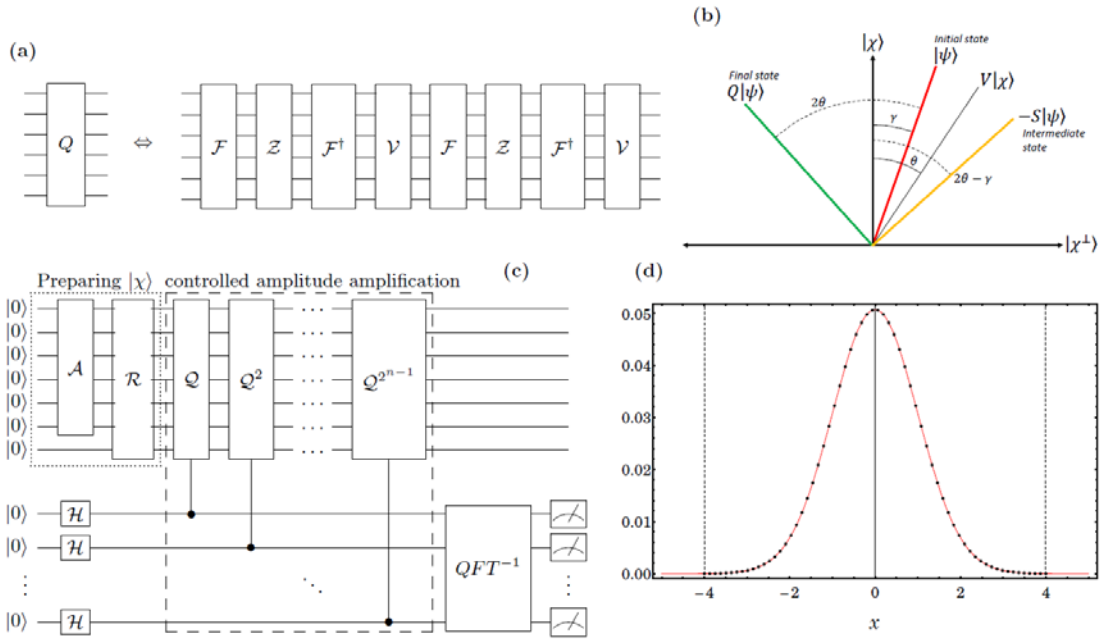
$k = O(\lambda^2 / \varepsilon^2)$

samples to estimate to additive error ε. The task of the quantum algorithm will be to improve the ε dependence from ε^2 to ε , hence providing a quadratic speedup for a given error.

Before we present the quantum algorithm, we show how to encode the structure values into a quantum algorithm and how to obtain the same ε dependency as the classical algorithm. We then show the quadratic speedup by using the fundamental quantum algorithm of amplitude estimation.

QUANTUM ALGORITHM FOR MONTE CARLO

Using amplitude estimation for the quantum Monte Carlo. FIG. (Algorithm scheme)



(a) The n + 1 qubit phase estimation unitary is written in terms of $F := R(A \otimes I_2)$, and the simple rotation Unitaries, $Z := |I_2^{n+1}\rangle - 2\,|0_{n+1}\rangle\langle 0_{n+1}|$ and $V := |I_2^{n+1}\rangle - 2\, I_2^n \otimes |1\rangle\langle 1|$.

(b) A visualization of the action of $Q := US$, with $S = VUV$ and $U = FZF^\dagger$, on an arbitrary state

$|\psi\rangle$ in the span of χ and $V(\chi)$. First, the action of $-S$ on $|\psi\rangle$ is to reflect along $V(\chi)$, resulting in the intermediate $-S|\psi\rangle$. Then, $-U$ acts on $-S|\psi\rangle$ by reflecting along $|\chi\rangle$ The resultant state $Q|\chi\rangle$ has been rotated anticlockwise by an angle 2θ in the hyperplane of $|\chi\rangle$ and $|\chi^\dagger\rangle$.

(c) The phase estimation circuit. Here, *A* encodes the randomness by preparing a superposition in |x>, while *R* encodes the random variable into the |1> state of an ancilla qubit. The output after both steps is the multiqubit state |χ>. Amplitude estimation then proceeds by invoking phase estimation to encode the rotation angle θ in a register of quantum bits that are measured to obtain the estimate ^ θ.

(d) The superposition prepared by *A* (or equivalently *G*) is a discretization of the variation of the protein structure. In this case, *R* encodes the cost function.

After the Inverse Quantum Fourier Transform, we'll obtain the best estimation for the structure of the protein.

CONCLUSION

We have described a quantum algorithm for the modelling of protein structure. We have assumed that the key features of CABS (its representation, force field and the scale of the movements used for the MC) are known and the corresponding quantum states can be prepared efficiently.

In addition, we assume efficient computability of the cost function. Under these assumptions, we exhibit a quadratic speedup in the number of samples required to estimate the structure of the protein up to a given error: if the desired accuracy is ε, then classical methods show a $1 / \varepsilon^2$ dependency in the number of samples, while **RQuanTech** quantum algorithm shows a $1 / \varepsilon$ dependency.

FURTHER CONSIDERATIONS

RQuanTech Quantum algorithm promise a significant speedup for computations related to Combinatorial algorithms used frequently in the Structural Biology Molecular field. In principle, current calculations could be reduced to much shorter time scales that would allow a much more capacity of structures to be solved.

We are looking at teaming up with the best in the Computer-Aided Protein Engineering field to demonstrate (pilot) in real settings the relevance of methodology.