Apply quantum algorithms to protein structure prediction

Qcourse570

QWorld

April 2022

What is a Protein?

- **Protein**: macromolecules consisting of **amino acids**
- **Peptide**: a linear chain of amino acid residues

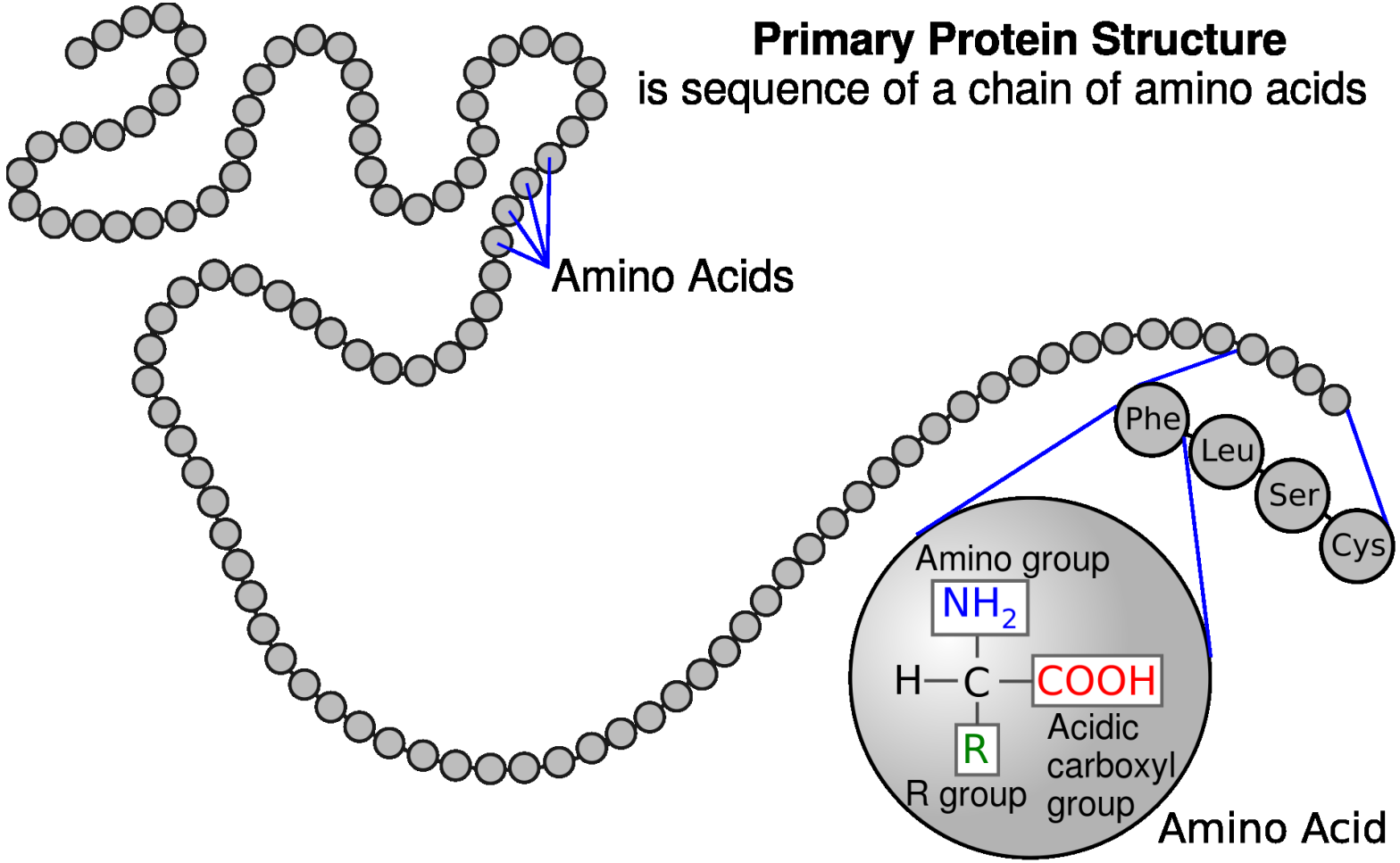
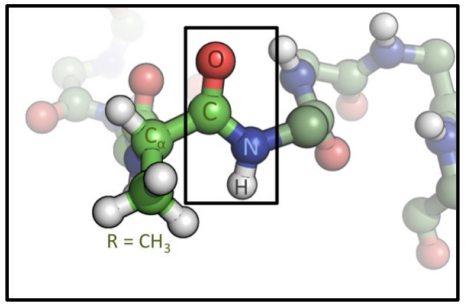


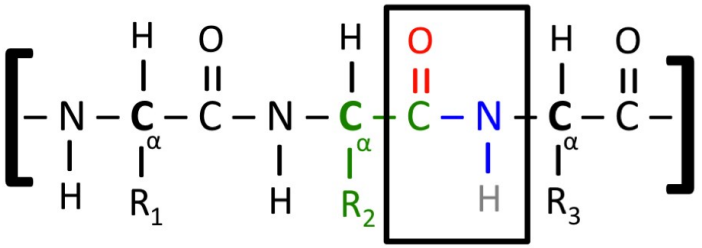
Photo from Wikipedia

Amino Acids

- Amino acids consist of:
  - amino (NH_3^+)
  - carboxylate (COOH)
  - R group (C, H)

- 20 amino acids
- https://www.rcsb.org/3d-view/1GWD



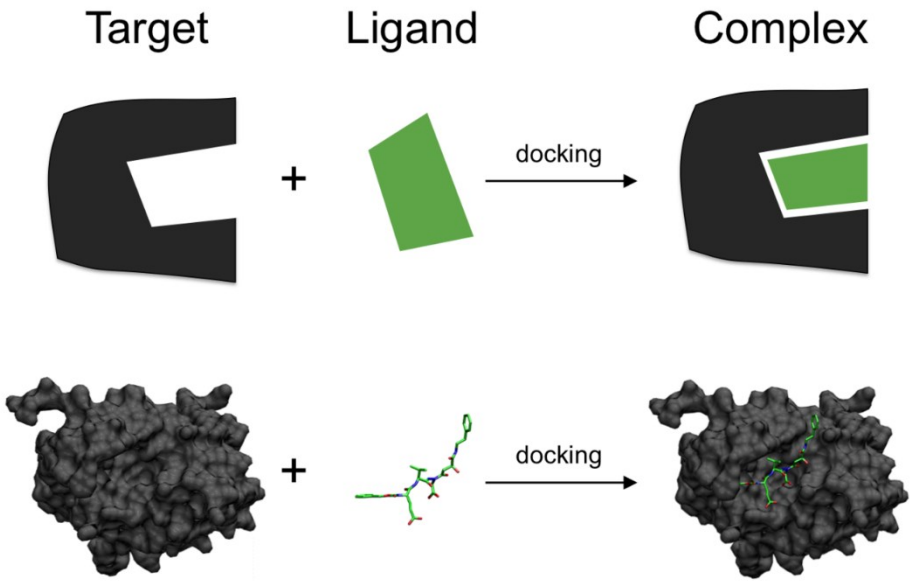


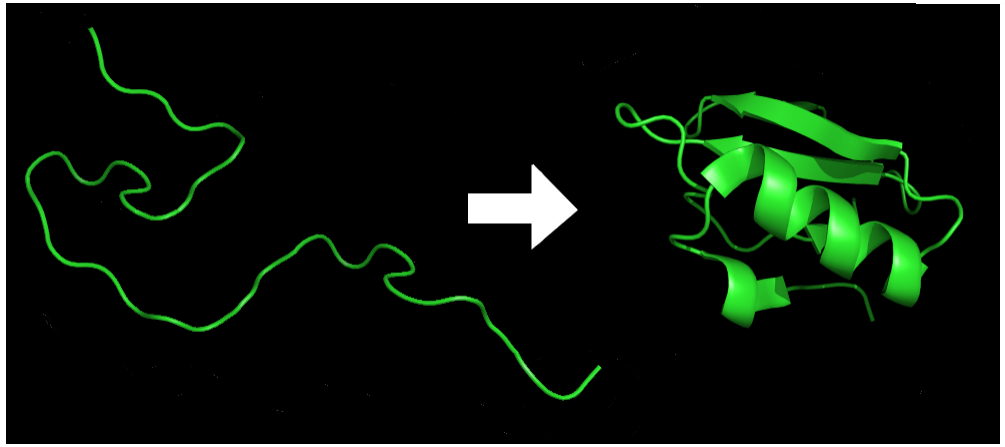
#	Amini acid	3-letter symbols	1-letter symbols
1	Alanine	Ala	A
2	Arginine	Arg	R
3	Asparagine	Asn	N
4	Aspartate	Asp	D
5	Cysteine	Cys	C
6	Glutamine	Gln	Q
7	Glutamate	Glu	E
8	Glycine	Gly	G
9	Histidine	His	H
10	Isoleucine	Ile	I
11	Leucine	Leu	L
12	Lysine	Lys	K
13	Methionine	Met	M
14	Phenylalanine	Phe	F
15	Proline	Pro	P
16	Serine	Ser	S
17	Threonine	Thr	T
18	Tryptophan	Trp	W
19	Tyrosine	Tyr	Y
20	Valine	Val	V

Photo from Wikipedia

Quantum algorithms for protein problems

- Protein Folding (Quantum Annealing / QAOA / VQE)
- Protein Ligand Docking (VQE)
- Protein Design (Grover)

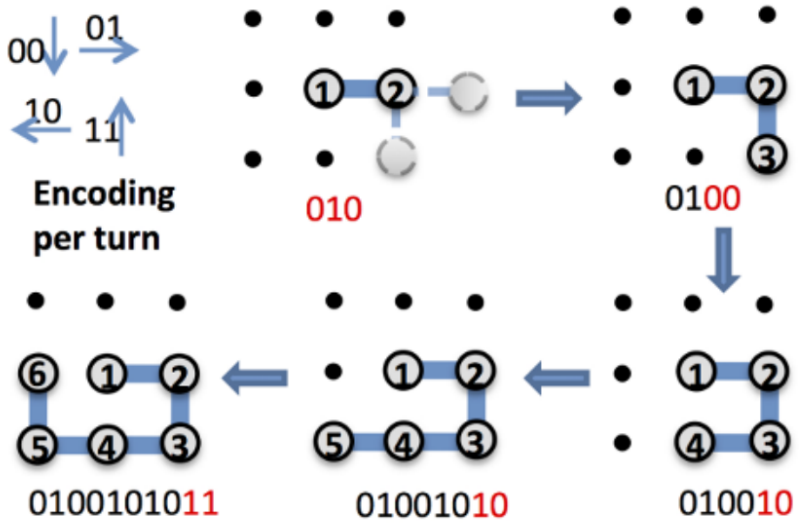


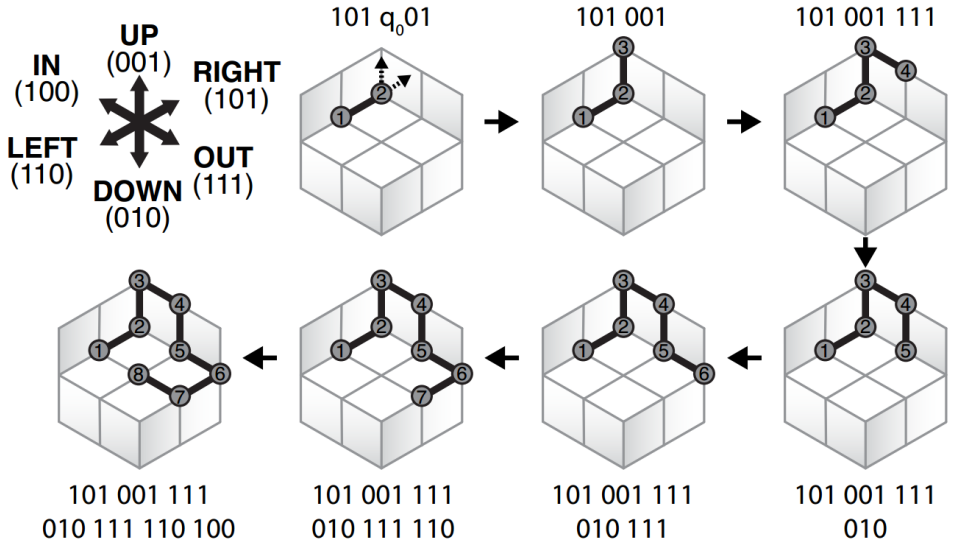


Photos from Wikipedia

Protein Folding Problem

- Definition: The protein folding problem is finding the lowest energy conformations of a protein given its primary sequence of amino acids.
- large-scale problem
- **Lattice Models**
  - self-avoiding walk (SAW)
- Coarse-grained representation
  - Points: amino acids





Photos from [8-9]

Previous works

- There are several lattice models, but only two of them are used in quantum computing [2].
  - Hydrophobic-polar model (HP) [1]
  - Miyazawa-Jernigan model [3]

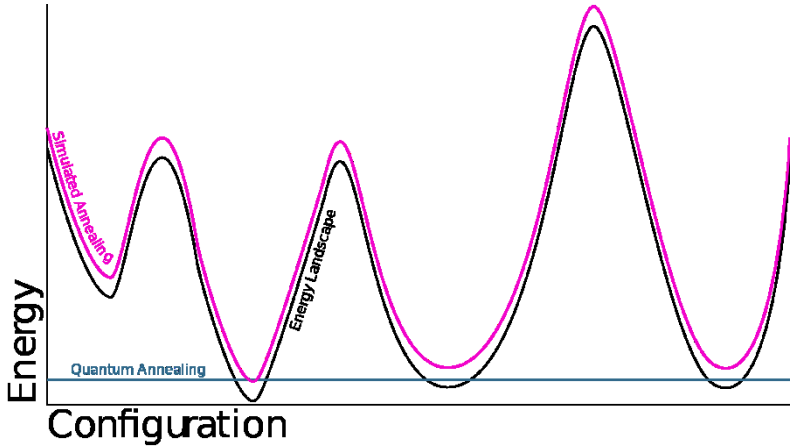
Table 3. Contact energies in RT units; e_{ij} for upper half and diagonal and e'_{ij} for lower half

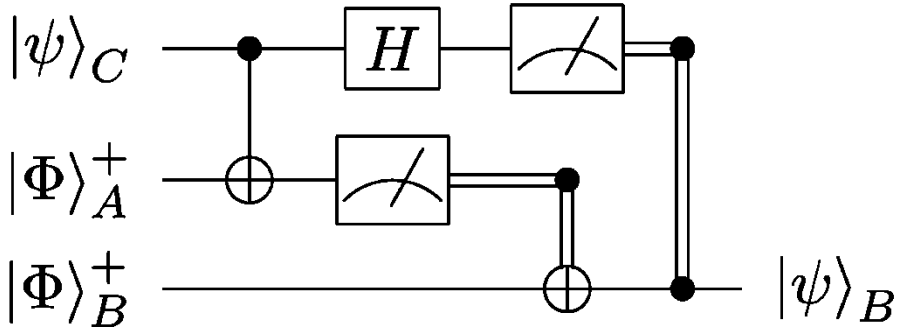
		Cys	Met	Phe	Ile	Leu	Val	Trp	Tyr	Ala	Gly	Thr	Ser	Asn	Gln	Asp	Glu	His	Arg	Lys	Pro	
	Cys	−5.44	−4.99	−5.80	−5.50	−5.83	−4.96	−4.95	−4.16	−3.57	−3.16	−3.11	−2.86	−2.59	−2.85	−2.41	−2.27	−3.60	−2.57	−1.95	−3.07	Cys
	Met	0.46	−5.46	−6.56	−6.02	−6.41	−5.32	−5.55	−4.91	−3.94	−3.39	−3.51	−3.03	−2.95	−3.30	−2.57	−2.89	−3.98	−3.12	−2.48	−3.45	Met
	Phe	0.54	−0.20	−7.26	−6.84	−7.28	−6.29	−6.16	−5.66	−4.81	−4.13	−4.28	−4.02	−3.75	−4.10	−3.48	−3.56	−4.77	−3.98	−3.36	−4.25	Phe
	Ile	0.49	−0.01	0.06	−6.54	−7.04	−6.05	−5.78	−5.25	−4.58	−3.78	−4.03	−3.52	−3.24	−3.67	−3.17	−3.27	−4.14	−3.63	−3.01	−3.76	Ile
	Leu	0.57	0.01	0.03	−0.08	−7.37	−6.48	−6.14	−5.67	−4.91	−4.16	−4.34	−3.92	−3.74	−4.04	−3.40	−3.59	−4.54	−4.03	−3.37	−4.20	Leu
	Val	0.52	0.18	0.10	−0.01	−0.04	−5.52	−5.18	−4.62	−4.04	−3.38	−3.46	−3.05	−2.83	−3.07	−2.48	−2.67	−3.58	−3.07	−2.49	−3.32	Val
	Trp	0.30	−0.29	0.00	0.02	0.08	0.11	−5.06	−4.66	−3.82	−3.42	−3.22	−2.99	−3.07	−3.11	−2.84	−2.99	−3.98	−3.41	−2.69	−3.73	Trp
	Tyr	0.64	−0.10	0.05	0.11	0.10	0.23	−0.04	−4.17	−3.36	−3.01	−3.01	−2.78	−2.76	−2.97	−2.76	−2.79	−3.52	−3.16	−2.60	−3.19	Tyr
	Ala	0.51	0.15	0.17	0.05	0.13	0.08	0.07	0.09	−2.72	−2.31	−2.32	−2.01	−1.84	−1.89	−1.70	−1.51	−2.41	−1.83	−1.31	−2.03	Ala
	Gly	0.68	0.46	0.62	0.62	0.65	0.51	0.24	0.20	0.18	−2.24	−2.08	−1.82	−1.74	−1.66	−1.59	−1.22	−2.15	−1.72	−1.15	−1.87	Gly
	Thr	0.67	0.28	0.41	0.30	0.40	0.36	0.37	0.13	0.10	0.10	−2.12	−1.96	−1.88	−1.90	−1.80	−1.74	−2.42	−1.90	−1.31	−1.90	Thr
	Ser	0.69	0.53	0.44	0.59	0.60	0.55	0.38	0.14	0.18	0.14	−0.06	−1.67	−1.58	−1.49	−1.63	−1.48	−2.11	−1.62	−1.05	−1.57	Ser
	Asn	0.97	0.62	0.72	0.87	0.79	0.77	0.30	0.17	0.36	0.22	0.02	0.10	−1.68	−1.71	−1.68	−1.51	−2.08	−1.64	−1.21	−1.53	Asn
	Gln	0.64	0.20	0.30	0.37	0.42	0.46	0.19	−0.12	0.24	0.24	−0.08	0.11	−0.10	−1.54	−1.46	−1.42	−1.98	−1.80	−1.29	−1.73	Gln
	Asp	0.91	0.77	0.75	0.71	0.89	0.89	0.30	−0.07	0.26	0.13	−0.14	−0.19	−0.24	−0.09	−1.21	−1.02	−2.32	−2.29	−1.68	−1.33	Asp
	Glu	0.91	0.30	0.52	0.46	0.55	0.55	0.00	−0.25	0.30	0.36	−0.22	−0.19	−0.21	−0.19	0.05	−0.91	−2.15	−2.27	−1.80	−1.26	Glu
	His	0.65	0.28	0.39	0.66	0.67	0.70	0.08	0.09	0.47	0.50	0.16	0.26	0.29	0.31	−0.19	−0.16	−3.05	−2.16	−1.35	−2.25	His
	Arg	0.93	0.38	0.42	0.41	0.43	0.47	−0.11	−0.30	0.30	0.18	−0.07	−0.01	−0.02	−0.26	−0.91	−1.04	0.14	−1.55	−0.59	−1.70	Arg
	Lys	0.83	0.31	0.33	0.32	0.37	0.33	−0.10	−0.46	0.11	0.03	−0.19	−0.15	−0.30	−0.46	−1.01	−1.28	0.23	0.24	−0.12	−0.97	Lys
	Pro	0.53	0.16	0.25	0.39	0.35	0.31	−0.33	−0.23	0.20	0.13	0.04	0.14	0.18	−0.08	0.14	0.07	0.15	−0.05	−0.04	−1.75	Pro
$e_{rr} - 2.55$	e_{rr}	−3.57	−3.92	−4.76	−4.42	−4.81	−3.89	−3.81	−3.41	−2.57	−2.19	−2.29	−1.98	−1.92	−2.00	−1.84	−1.79	−2.56	−2.11	−1.52	−2.09	
$e_r - 3.60$	e_t	−4.29	−4.73	−5.57	−5.29	−5.71	−4.72	−4.41	−3.87	−3.17	−2.53	−2.63	−2.27	−2.14	−2.35	−2.02	−2.07	−2.94	−2.43	−1.82	−2.53	
$f_r - 3.60$	f_t	−5.58	−6.14	−7.39	−7.09	−7.88	−6.15	−5.34	−4.60	−3.24	−2.22	−2.48	−1.92	−1.74	−1.93	−1.54	−1.49	−2.91	−2.07	−1.17	−1.97	
N_{tr}/N_t	2.096	2.723	2.722	2.780	2.811	2.893	2.728	2.537	2.493	2.143	1.840	1.973	1.771	1.699	1.720	1.598	1.508	2.075	1.787	1.343	1.629	
q_t 7.162	6.281	6.646	6.137	5.870	6.042	6.087	6.155	5.793	6.037	6.334	6.284	6.486	6.582	6.574	6.469	6.487	6.235	6.241	6.318	6.569	5.858	

Table from [3]

Quantum Computing

- Richard Fineman: "Nature isn't classical, and if you want to make a simulation of nature, you'd better make it quantum mechanical“
- Quantum computing models:
  - Adiabatic Quantum Computing (Quantum Annealing) [6-9]
  - Digital Quantum Computing (Gate-based model) [10-11]





Photos from Wikipedia

Gate-based model

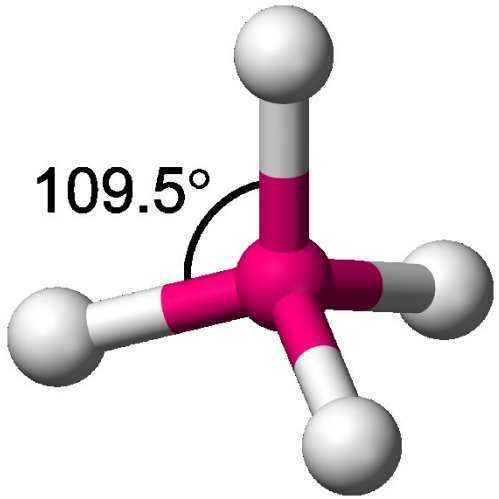
- Variational Quantum Algorithms (VQA):
  - Quantum Approximate Optimization Algorithm (QAOA) [10]
  - Variational Quantum Eigensolver (VQE) [11]

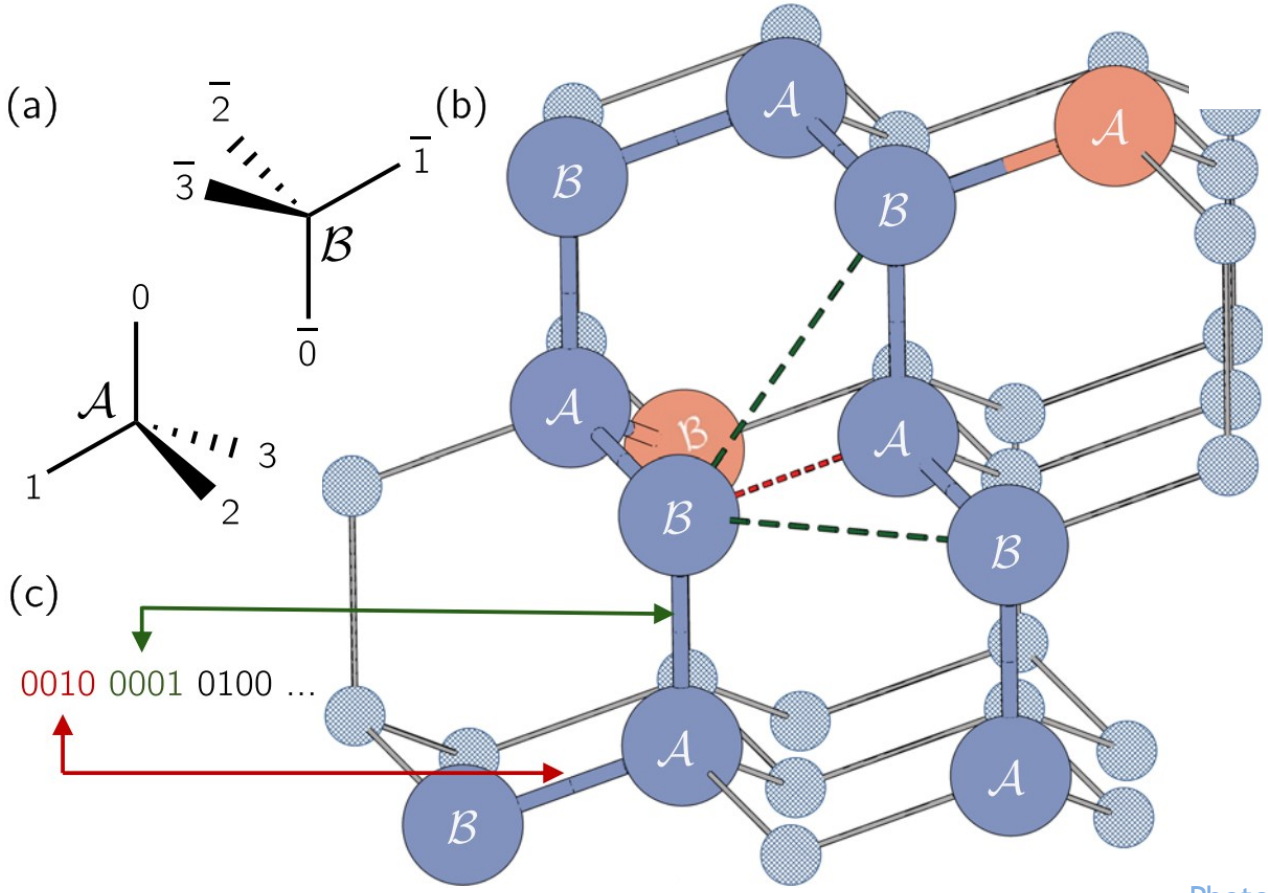
- Ground State Problem (finding eigenvalue):
  - Quantum Phase Estimation (QPE)
  - Variational Quantum Eigensolver (VQE)

- Noisy Intermediate-Scale Quantum (NISQ) era

Qiskit Nature

- Tetrahedral lattice model
  - angle(s): ≈ 109.5°





Photos from Wikipedia and [11]

VQE algorithm

- VQE algorithm can solve the Schrödinger equation:
- $H|\psi(\vec{\theta})\rangle = E|\psi(\vec{\theta})\rangle$

  where:

  - E is energy / a number
  - E_0 is the lowest energy / eigenvalue of H / a number
  - $\psi(\vec{\theta})$ is the eigenstate of H / a vector
  - H is the Hamiltonian / an operator / a matrix
  - $\vec{\theta}$ is a vector of independent parameters and $\vec{\theta} = \{(\theta_1, \ldots, \theta_n)\}^T$.

  The lowest answer for this equation is the ground state (E_0).

  $E_0 \leq \langle\psi(\vec{\theta})|H|\psi(\vec{\theta})\rangle$

VQE algorithm

- VQE algorithm consists of two important parts:
  - A quantum circuit (ansatz)
  - An optimization method (classical computers)

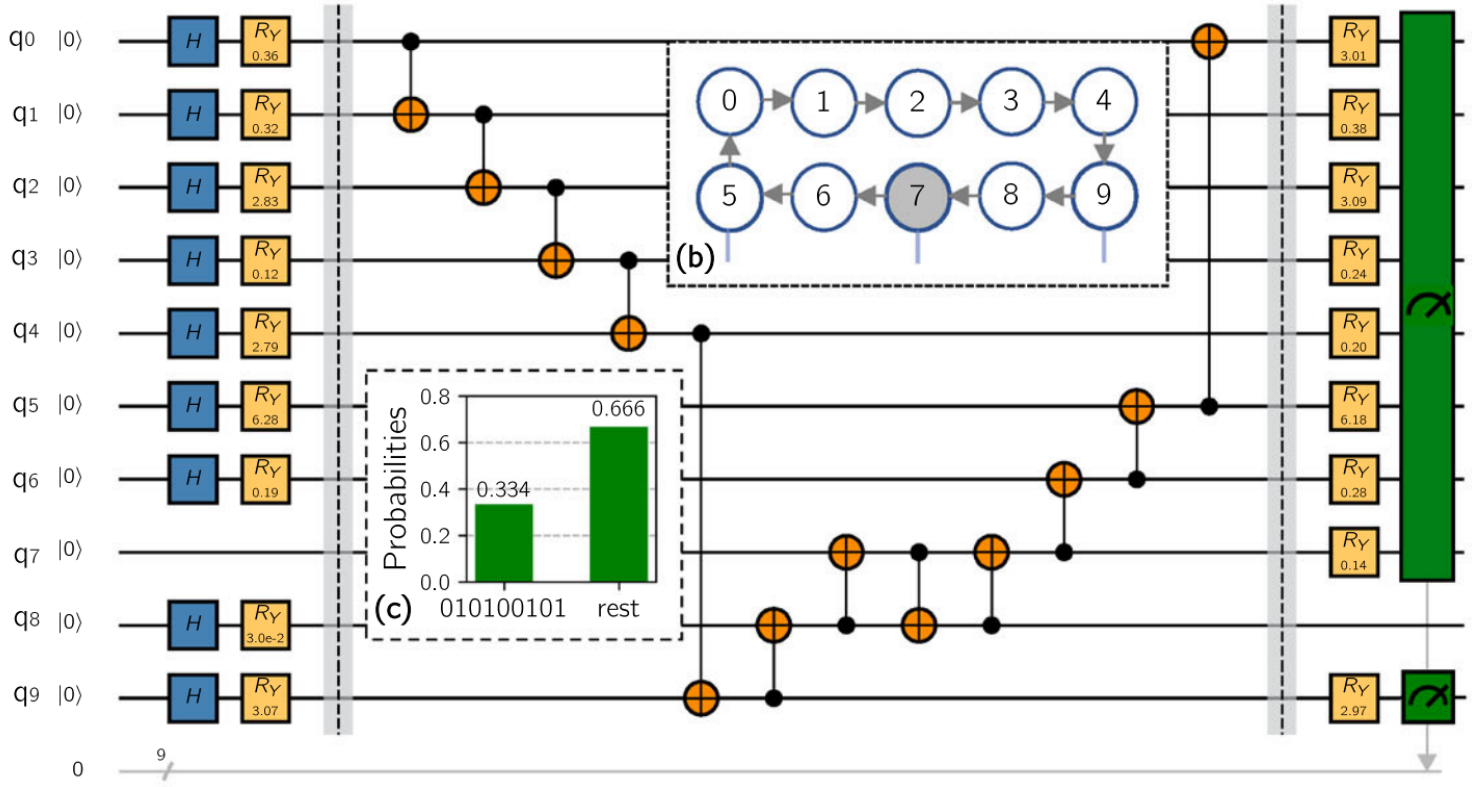


Photo from [11]

Hamiltonian

- They defined two kinds of qubits:

1. q_{cf}: *Configuration qubits that are used for configurations and the relative position of the amino acids*
2. q_{in}: *Interaction qubits that encode interactions between the amino acids*

- The Hamiltonian consists of some parts:

$H = H_{gc}(q_{cf}) + H_{ch}(q_{cf}) + H_{in}(q_{cf}, q_{in})$ *where:*

- H_{gc} *is the geometrical constraint term*
- H_{ch} *is the chirality constraint*
- H_{in} *is the interaction energy terms of the system.*

References

- [1] Dill KA (1985). "Theory for the folding and stability of globular proteins". Biochemistry. 24 (6): 1501–9.
- [2] Outeiral, C., Strahm, M., Shi, J., Morris, G. M., Benjamin, S. C., & Deane, C. M. (2020). "The prospects of quantum computing in computational molecular biology". Wiley Interdisciplinary Reviews: Computational Molecular Science, 11(1), e1481.
- [3] Miyazawa, S., & Jernigan, R. L. (1985). "Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation". Macromolecules, 18(3), 534-552.
- [4] Berger B, Leighton T (1998). "Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete". Journal of Computational Biology. 5 (1): 27–40.
- [5] Feynman, R. (1982). "Simulating physics with computers". Int. J. Theor. Phys. 21, 467–488.
- [6] Perdomo, A., Truncik, C., Tubert-Brohman, I., Rose, G., & Aspuru-Guzik, A. (2008). "Construction of model Hamiltonians for adiabatic quantum computation and its application to finding low-energy conformations of lattice protein models". Physical Review A, 78(1), 012320.
- [7] Babbush, R., Perdomo-Ortiz, A., O'Gorman, B., Macready, W., & Aspuru-Guzik, A. (2012). "Construction of energy functions for lattice heteropolymer models: a case study in constraint satisfaction programming and adiabatic quantum optimization."
- [8] Perdomo-Ortiz, A., Dickson, N., Drew-Brook, M., Rose, G., & Aspuru-Guzik, A. (2012). "Finding low-energy conformations of lattice protein models by quantum annealing". Scientific reports, 2(1), 1-7.
- [9] Babej, T., & Fingerhuth, M. (2018). "Coarse-grained lattice protein folding on a quantum annealer".
- [10] Fingerhuth, M., & Babej, T. (2018). "A quantum alternating operator ansatz with hard and soft constraints for lattice protein folding".
- [11] Robert, A., Barkoutsos, P. K., Woerner, S., & Tavernelli, I. (2021). Resource-efficient quantum algorithm for protein folding. npj Quantum Information, 7(1), 1-5.
- [12] Bharti, K., Cervera-Lierta, A., Kyaw, T. H., Haug, T., Alperin-Lea, S., Anand, A., ... & Aspuru-Guzik, A. (2022). "Noisy intermediate-scale quantum algorithms". Reviews of Modern Physics, 94(1), 015004.
- [13] Aleksandrowicz, G., and Alexander T., Amy, M., ... & Mantas. (2021). "Qiskit: An Open-source Framework for Quantum Computing".

Thanks For Attention