HELMHOLTZ Information

JÜLICH Forschungszentrum

Program 1, Topic 1 : Enabling Computational & Data-Intensive Science and Engineering

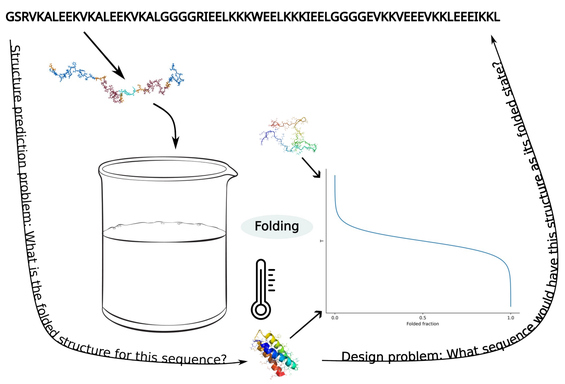
Protein Folding and Design using Quantum Annealing

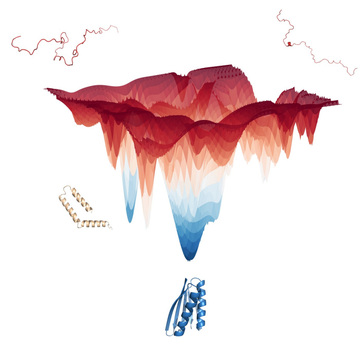
Sandipan Mohanty

Overview

- Quantum annealing achieves 100% success in protein folding and design challenges
- Our results establish a clear proof-of-concept for the application of quantum annealing in biophysics.

Protein folding, design, structure prediction





- Structure prediction: Given sequence, predict folded structure
- Folding: Unter what physical conditions does the transition to the folded state happen? What temperature? How long does it take? How does the interaction with other nearby molecules affect the transition?…
- Design: Inverse problem. Given a desired folded structure find a sequence which would fold to it

- A very rough funnel like energy landscape with the folded native state at the bottom
- Challenging sampling problem: high dimensionality, exponentially growing search space
- Tough physics problem: large number of mildly malleable entities with complex atomically anisotropic interactions

Quantum annealing

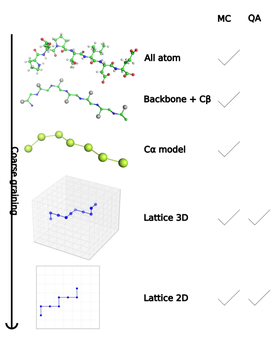
- D-Wave quantum annealer is an analogue computing device which exploits quantum mechanics to solve suitable computing problems
- D-Wave Advantage, like JUPSI, has > 5000 supercomputing qubits with a connectivity of 15

- The quantum annealer effectively generates sample bit configurations $\{x_i\}$ which minimise the QUBO energy

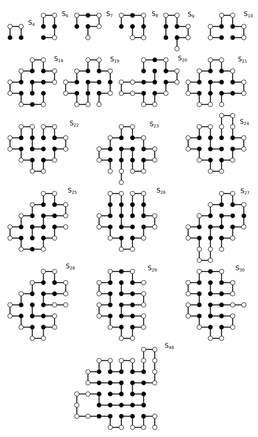
$$E = \sum_{i=1}^{N} q_i x_i + \sum_{i \neq j}^{N} q_{ij} x_i x_j$$

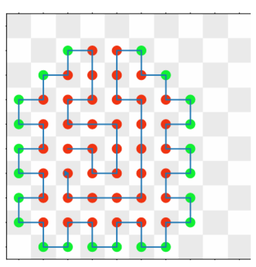
- Strategy: frame the problem to be solved as an energy minimisation problem over binary variables, and let the annealer produce solutions

Protein folding with quantum annealing



- Coarse graining reduces computational complexity sacrifising specificity in favour of general principles and conceptual frameworks
- Lattice models, with a reduced amino acid alphabet (H = Hydrophobic, P = Polar), are thoroughly studied by classical methods, sometimes with exact results
- Easily translated into optimisation problems on bit-strings.
- Our encoding maintains the same algebraic form for the energy irrespective of system size and composition.





$$E = -\sum_{|f-f'|>1} \sum_{n.n.s,s'} \sigma_s^f \sigma_{s'}^{f'} h_f h_{f'} + \lambda_1 \sum_{f=1}^{N} \Big(\sum_s \sigma_s^f - 1\Big)^2 + \frac{1}{2}\lambda_2 \sum_{f \neq f'} \sum_s \sigma_s^f \sigma_s^{f'} + \lambda_3 \sum_{f=1}^{N-1} \sum_s \sigma_s^f \sum_{\|s'-s\|>1} \sigma_{s'}^{f+1}$$

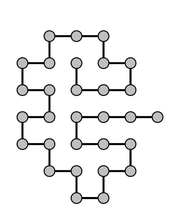
Id	Sequence	E_{min}	QA success
S4	HPPH	-1	100%
S6	HPPHPH	-2	100%
S7	PHPPHPH	-2	100%
S8	HPHPHPPH	-3	100%
S9	HHPPHPPHP	-3	100%
S10	HPPHPPHPPH	-4	100%
S14	HHHPPPHPPHPPPH	-5	100%
S18	HHHPPHPPHPHPPHPHPH	-9	100%
S19	PHPHPHPPHPHPPHPPHHH	-9	100%
S20	HPHPHPPHPHPPHPPPPHHH	-9	100%
S21	PHHPPHPHPPHPHPPHPHHH	-10	100%
S22	HPPHPPHPHPPHPHPPHPHHH	-11	100%
S23	PPHHHHPPHPPHPHPPHPHPPHP	-10	100%
S24	HPPPPHPPHPHPPHPHPPHPPHHH	-11	100%
S25	PHPHPHPHPPHPHPHPPHPPHHHHH	-13	100%
S26	HHHHPPHHPPHPHPPHPHPPHHPPHH	-14	100%
S27	PHPHPHPHPPHPHPHPPHPPPPHHHHH	-13	100%
S28	PPHHHPPHPPHPHPHPPHPHPPHPPHHH	-13	100%
S29	PHPHPHPPHHPPHPHPPHPPHHHHPPHHH	-15	100%
S30	PPHHHHPPHPPHPHPPHHPPHPHPHPPHHH	-15	100%
S48	PPHPPHHPPHHPPPPPHHHHHHHHHHPPPPPPHHPPHHPPHPPHHHHH	-23	100%
S64	HHHHHHHHHHHHPHPHPPHHPPHHPPHPPHHPPHHPPHPPHHPPHHPPHPH PHPHHHHHHHHHHHH	-42	100%

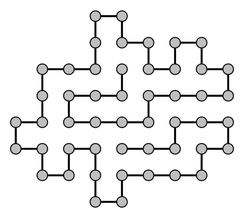
Application to protein design

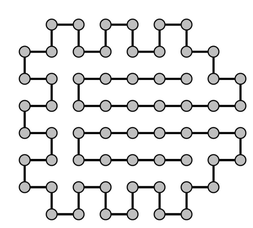
- What sequence of H and P amino acids will spontaneously fold into the given structures?
- Vary the identities of each bead in the model between H and P, and find sequences which have a low HP modle energy
- Screen resulting sequences based on their ability to fold to the structure
- Sequence optimisation QUBO energy function is then given as:

$$E(\{s\}) = -\sum_{1 \le i < j \le N} w_{ij} s_i s_j + \lambda \Big(\sum_{i=1}^{N} s_i - N_H\Big)^2$$

- 12 small structures with chain lengths from 8 to 20, and 3 larger targets of length 30, 50 and 64 shown here to the left were used to verify our approach
- In every instance, the D-Wave hybrid annealer produces multiple candidate sequences for different values of the hydrophobic content
- For the small sequences, a direct comparison with exhaustive enumerations was used to test the validity of the solutions.
- The three larger structures, were tested using the folding method also using the annealer as described to the left in this poster. Multiple solutions were found in each case, each having the target structure as a structure with the lowest energy
- One of the candidates for the 64 bead structure from our folding study: HHHHHP HHHHHH PHPHPP HHPPHH PPHPPH PPPPHP PHPPHH PPHHPP HPHPHH HHHHPH HHHH
- This sequence is an improvement over the sequence used in our folding study, since it has the target structure as the *unique* ground state!
- The simpler Hamiltonian of the sequence optimisation problem results in much smaller resource requirements (a few seconds) for each design challenge







3 year outlook

- Lattice protein folding in 3D with QA (done) and comparison with classical MC (pending)
- 20 letter amino acid alphabet
- Aggregation, droplet formation
- Reproduce results using gate based quantum computers
- Reduce qubit expense and noise to tackle bigger problems

Collaborations

LUND UNIVERSITY

CHALMERS

References

- "Folding lattice proteins with quantum annealing", Anders Irbäck, Lucas Knuthson, Sandipan Mohanty, and Carsten Peterson, Phys. Rev. Research 4, 043013 (2022)
- "Using quantum annealing to design lattice proteins", Anders Irbäck, Lucas Knuthson, Sandipan Mohanty, and Carsten Perterson, Phys. Rev. Research 6, 013162 (2024)