

Quantum computing opportunities in synthetic biology

Gopal Peddinti, D.Sc. (Tech),
Senior Scientist in Computational Biology,
Bioanalytics and biological data science
team, VTT

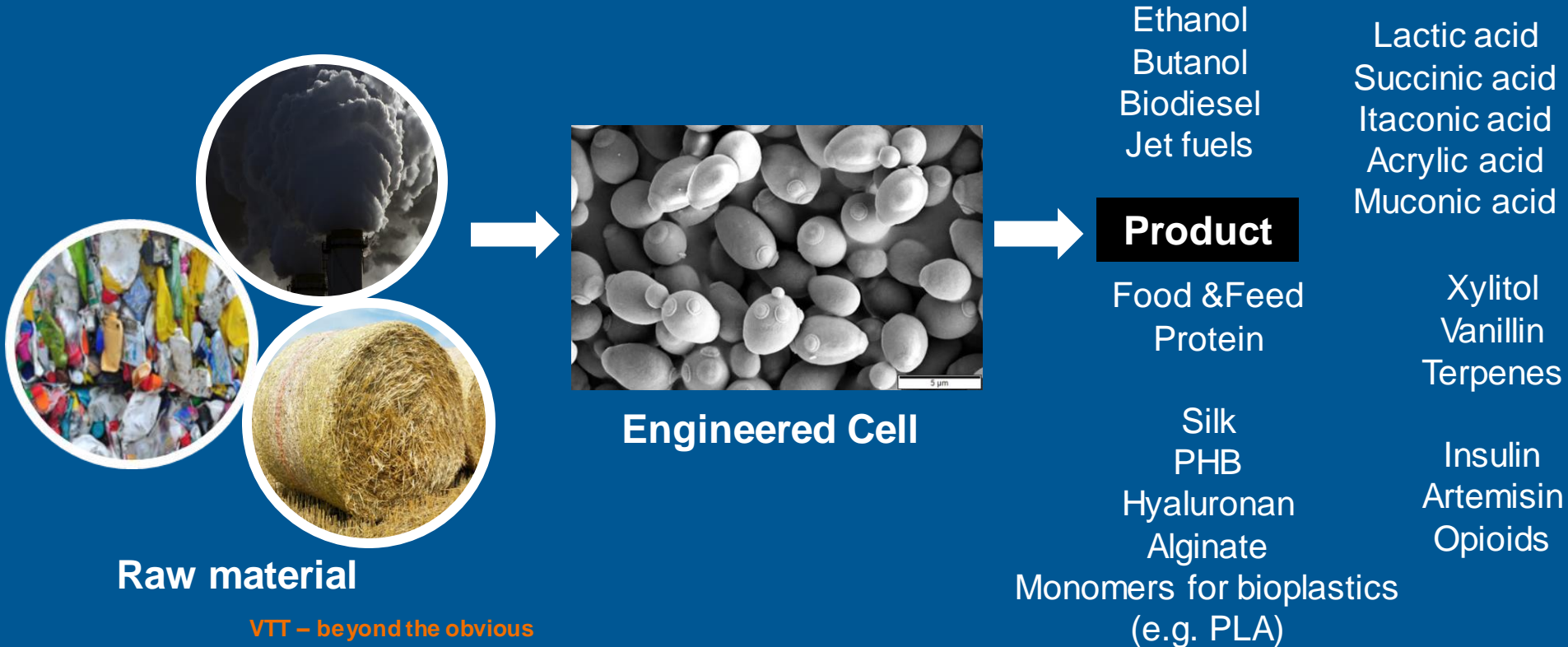
19/12/2023 VTT – beyond the obvious

Contents

- Introduction to synthetic biology
- Computational methods in synthetic biology
- Quantum computing opportunities in synthetic biology
 - Selected computational synthetic biology problems
- Conclusion

Introduction to synthetic biology

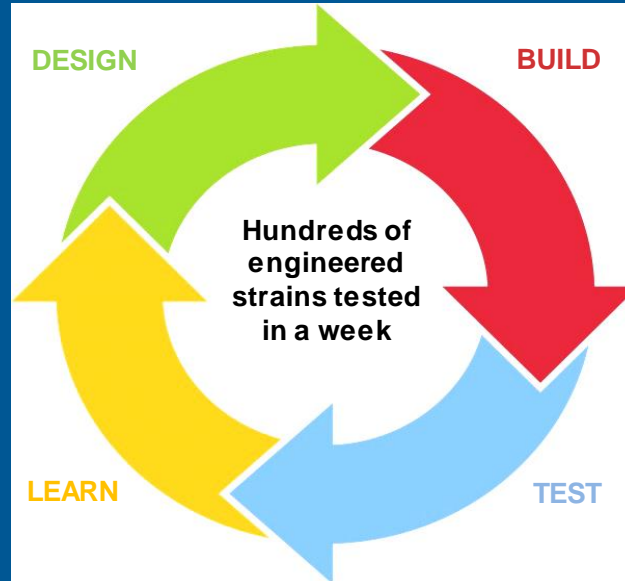
Synthetic biology (or engineering biology) drives sustainable breakthroughs in Bioeconomy



The Design-Build-Test-Learn cycle of Synthetic Biology

Computer Aided DESIGN
based on computational
models of biological
systems

**LEARN: learning from
the TEST data**



**BUILD: implementing the
biological DESIGN in
the wet lab**

**TEST: verifying the BUILD
(data collection)**

Computational methods in synthetic biology

The Design-Build-Test-Learn cycle of Synthetic Biology

Computational biology methods used (non-exhaustive list)

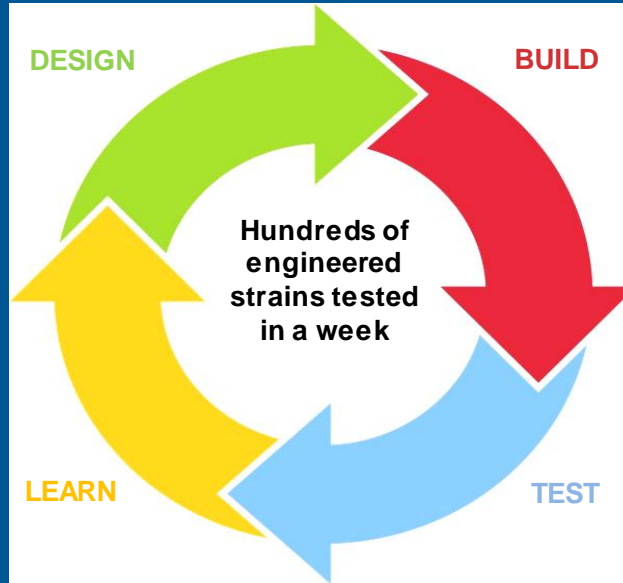
DESIGN: Computational (re)design

- AI-based *de novo* enzyme design
- Enzyme discovery (sequence data mining and AI)
- Host optimization guided by genome-scale metabolic models
- Reaction and pathway feasibility evaluation (biothermodynamics)
- Design of new-to-nature [biosynthetic] pathways

LEARN Analysis and decisions

- Machine learning & AI
- Molecular dynamics of biomolecular systems
- Docking & other molecular simulations (e.g., QM/MM)
- Microbial community modelling
- Lipid modelling

VTT – beyond the obvious



BUILD: Building of production strains

- (Design of biological constructs)
- (Molecular cloning, genome editing, ...)
- (Robotics)

TEST: Cultivation and measurement

- Next-generation sequencing data
 - *De novo* genome/transcriptome assembly
 - Read mapping/alignments
- Metabolomics, Proteomics data
- Fermentation data
- Raman spectroscopy data
- Enzyme activity verification
- (Robotics)

Quantum computing opportunities in synthetic biology

Emani, P. S., et al., (2021). <https://doi.org/10.1038/s41592-020-01004-3>

Outeiral, C., et al., (2021). <https://doi.org/10.1002/WCMS.1481>

Marx, V. (2021). <https://doi.org/10.1038/s41592-021-01199-z>

Sarkar, A., et al., (2021). <https://doi.org/10.1371/JOURNAL.PONE.0249850>

Reiher, M., et al., (2017). <https://doi.org/10.1073/PNAS.1619152114>

The Design-Build-Test-Learn cycle of Synthetic Biology

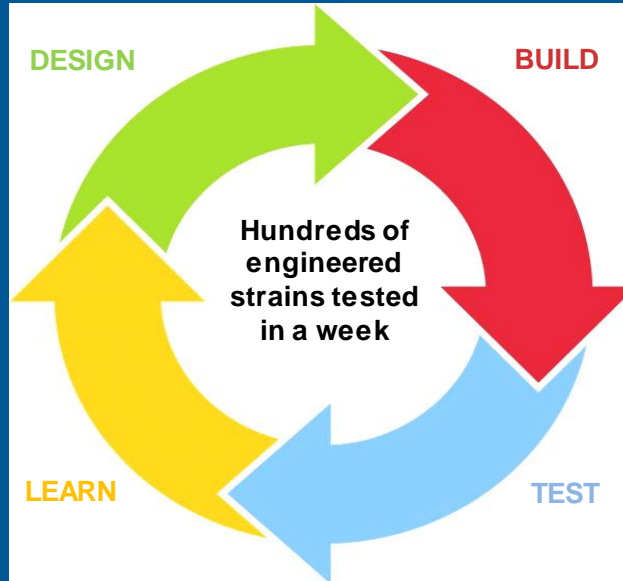
Quantum computing opportunities!

DESIGN: Computational (re)design

- AI-based *de novo* enzyme design
- Enzyme discovery (sequence data mining and AI)
- Host optimization guided by genome-scale metabolic models
- Reaction and pathway feasibility evaluation (biothermodynamics)
- Design of new-to-nature [biosynthetic] pathways

LEARN Analysis and decisions

- Machine learning & AI
- Molecular dynamics of biomolecular systems
- Docking & other molecular simulations (e.g., QM/MM)
- Microbial community modelling
- Lipid modelling



BUILD: Building of production strains

- (Design of biological constructs)
- (Molecular cloning, genome editing, ...)
- (Robotics)

TEST: Cultivation and measurement

- Next-generation sequencing data
 - De novo genome/transcriptome assembly
 - Read mapping/alignments
- Metabolomics, Proteomics data
- Fermentation data
- Raman spectroscopy data
- Enzyme activity verification
- (Robotics)

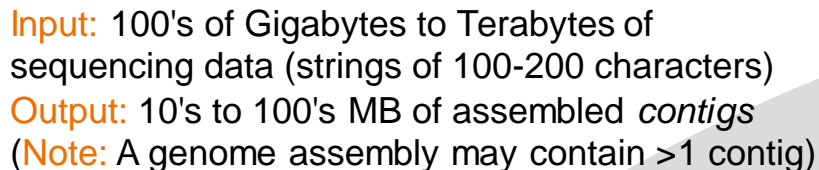
In theory, quantum computing is more widely applicable in synthetic biology

Because, at the core, almost every computational biology procedure involves ...

| Task | Application | Advantage | Remark |
|---|--|------------------------|---------------------------------|
| Searching (typically with dynamic programming / HMM) | Sequence / structure alignments, read mapping | Quadratic | QRAM is a challenge, QHMM, QUBO |
| Optimization (mixed integer linear programming, linear/quadratic optimization, genetic algorithms, etc.,) | Host optimization, and many more | Quadratic or heuristic | QUBO etc., |
| Machine learning (supervised or unsupervised) | Structure prediction, Function design, genotype-phenotype mapping, and many more | Quadratic or heuristic | Linear models, QVAE, QML |
| Simulations of physical movements of electrons, atoms and molecules | Molecular dynamics, reaction simulations (QM/MM) | Exponential | Inherently quantum |

Quantum computing opportunities in synthetic biology: selected applications

- De novo genome assembly
- Sequence alignment
- Metabolic optimization
- Protein function design
- Enzymatic reaction simulations



Meta-transcriptomics – microbial community

- The genomic/transcriptomic content of (non-model) organisms
 - What biological functions does an organism have?
 - What is the best species for specific industrial application?
- Differences between species
 - Verification of engineered strains
 - Why do differences occur (mutations)
 - Species identification
- Who (which microbes) are in my sample?
 - How do I cultivate them?
 - Valuable enzymes from species that we can't yet cultivate
- ...

De novo sequence assembly: overlap-layout-consensus graph approach

Genome to be determined

TTCCGGAGAGGGAGCCTGAGAAATGGCTACCACATCCACGGAGAGG

READS

GCCTGAGAAATGGCTACCACATC

CCACATCCACGGAGAGG

TTCCGGAGAGGGAGCCTGAG

TTCCGGAGAGGGAGCCTGAGAAATGGCTACCACATCCACGGAGAGG

Hamiltonian path finding (TSP)

Overlap graph

Reflections

- QC does not always exponentially speed you up :-)
 - For TSP, we get quadratic speed-up at the best.
- A typical NGS data set has millions of reads: overlap graph with millions of nodes, through which to find Hamiltonian paths
 - Number of logical qubits in a reference implementation $\sim 2^N$ ($N = n(\text{reads})$) [*]
 - Quantum computers will not likely scale up in our lifetime for this naïve QUBO implementation of the OLC algorithm!
- Major simplifications/approximations are needed
 - Alternative assembly approaches need to be studied in the context of QC
- Related trends in the domain:
 - Sequencing isolates vs microbial communities (sequence assembly gets harder and harder)
 - Increased length of sequencing reads (sequence assembly becomes easier and easier)

[*] Sarkar et al., 2021, DOI: 10.1371/JOURNAL.PONE.0249850

Quantum computing opportunities in synthetic biology: selected applications

- De novo genome assembly
- **Sequence alignment**
- Metabolic optimization
- Protein function design
- Enzymatic reaction simulations

Sequence alignment is like Google search for biological sequences!

- Used for studying the sequence similarity arising from structural, functional or evolutionary relationships.
 - Especially to find hits similar to a query sequence from large sequence databases

| | | |
|------------|---|-----|
| Hsa_ZNF226 | KGFTLSSNLQA--HQRVHTGKPYKCNE--CGKSFRRNSHYQVHLVVH----- | 525 |
| Mac_ZNF226 | KGFTLSSNLQA--HQRVHTGKPYKCSE--CGKSFRRNSHYQVHLVVH----- | 489 |
| Odo_ZNF226 | KGFTLSSNLQA--HQRVHTGKPYKCDE--CGKSFRRNSHYQVHLVVH----- | 528 |
| Mir_ZNF226 | KGFTLSSNLQA--HQRVHTGKPYKCDE--CGKSFRRNSHYQVHLVVH----- | 528 |
| Cam_ZNF226 | KGFTLSSNLQA--HQRVHTGKPYKCGE--CGKSFRRNSHYQVHLVVH----- | 524 |
| Fuk_ZNF226 | KGFTLSSNLQA--HVRVHTGKPYTCEE--CGKSFRRNSHYQVHLVVH----- | 493 |
| Man_ZNF226 | KGFTLSSNLQA--HQRVHTGKPYRCDE--CGKSFRRNSHYQVHLVVH----- | 546 |
| Cap_ZNF226 | KAFSLWSKLN--HILVHTGKPHCESV--CKKAFAQRSSLKIHMLVH----- | 284 |
| Nem_ZNF226 | KCFTELETLKT--HLMHSGEKPYKCDE--CGKCFQSGNLKRHLMIH----- | 309 |
| Rhi_ZNF226 | KNYTQWGHQR--HMSHTGEFPYKCNVPECEMSFFRSYELKEHII SHNSTTSNEIILYN | 347 |
| Ech_ZNF226 | RGFRQASHLLS--HERVHSGEKPFKCEE--CGKFFSSRSNLQTHHRIH----- | 218 |
| Ara_ZNF226 | KEFPQKRNLNV--HYRHTNEKPYSCDV--CQKDFSWKGDVKVHYRIH----- | 281 |
| Mli_ZNF226 | KCFSQQISLRT--HRRVHTGKPFKDE--CDKFAQLNVLLAHRRGH----- | 303 |
| Dro_ZNF226 | KAFKNNSHLQE--HLRTHQEARPFKCSH--CSKSFKLRSILQKHLLTH----- | 243 |
| The_ZNF226 | KRFSSSSNLSIHKHKRTHGKPYMCDV--CTKRFSFYSHLYTHHRIH----- | 248 |
| Sal_ZNF226 | FKTAYKRTLTT--HKRIHTGKPYKCDQ--CSFKTAHKSTLARHKRTH----- | 363 |
| Ber_ZNF226 | RAFRNSSNLRT--HERLHTDERPYRCRY--CDRAFSGSGNLHAHERVH----- | 188 |
| Sen_ZNF226 | QAFCHSNLTA--HMRHTGEQPYVCQT--CG----- | 139 |
| Den_ZNF226 | KSFTHASDLKI--HQRIHTGKPYQCVE--CGKSFTQTSNLKIHQRIH----- | 493 |
| Oct_ZNF226 | KSFSHNGHLVT--HNRHTGKPYQCDI--CGKSFSHNGHLVTHNRIH----- | 437 |

* * * * *

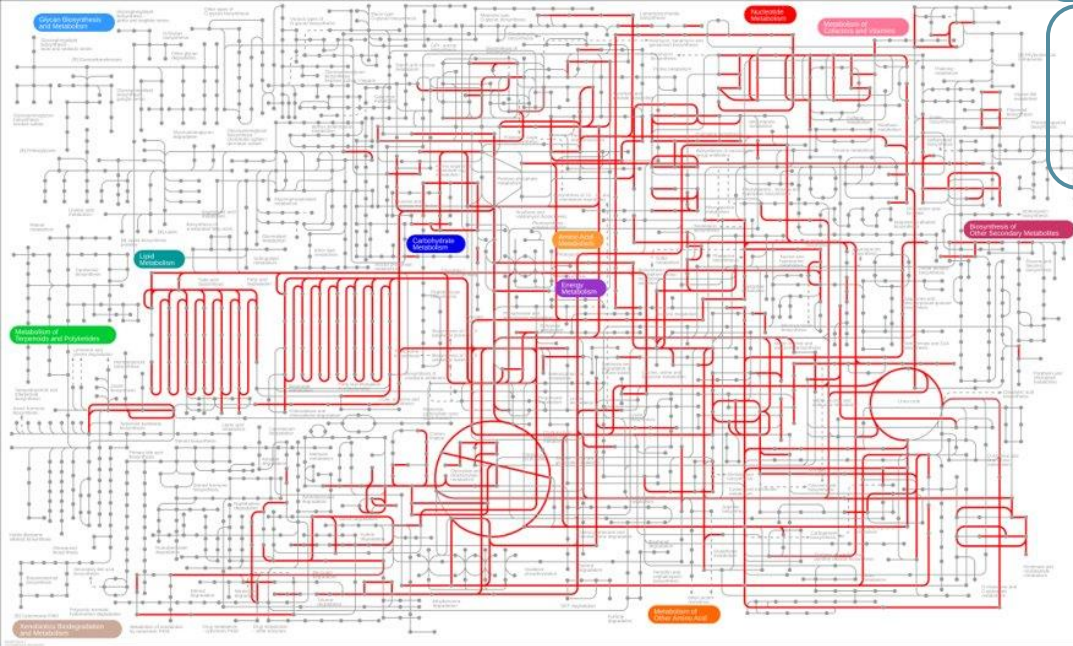
Reflections

- Related approaches are also used for aligning sequencing reads to reference genomes or transcriptomes for studying mutations or gene expression etc.,
- Sequence alignment concepts have also become directly applicable for searching large structure databases such as the AlphaFold DB.
- These algorithms inherently involve search and optimization
 - Hence naturally benefit from Grover's search and other approaches => Quadratic time improvement.
- **Quantum RAM might be the biggest challenge** for the applicability of quantum algorithms because of the sizes of the sequences databases and reference genome sequences

Quantum computing opportunities in synthetic biology: selected applications

- De novo genome assembly
- Sequence alignment
- **Metabolic optimization**
- Protein function design
- Enzymatic reaction simulations

Genome-scale metabolic models and optimization of host metabolism



Mass balance constraints
 $Sv = 0$

Directionality constraints from reaction thermodynamics
 $LB < v < UB$

+

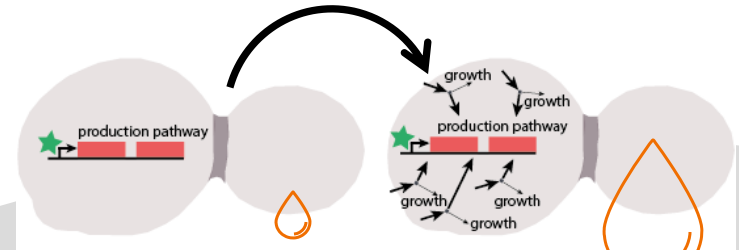
Biological optimality principles

$\max c^T v$



Metabolic phenotype
(optimization: LP/QP)

Optimizing host metabolism for production: coupling growth (biomass formation) with production of the target compound



Current phenotype

Optimized phenotype

Reflections

- Metabolic optimization is typically achieved via Bi-level optimization (MILP), genetic algorithms
 - Can be formulated as QUBO (Quadratic Unconstrained Binary Optimization) suitable for QAOA
- Metabolic optimization is an exponential scale problem
 - 2^N possible configurations (N = reactions)
 - $N \sim 3000$ to 10000 for typical genome-scale metabolic models
 - $N \sim 500000$ to **a few millions** for extended metabolic space containing all elementary chemical transformations
 - $N \sim 10$ - 100 for modeling selected subsystems (e.g., energy metabolism) or for simplified (lumped-reaction) models

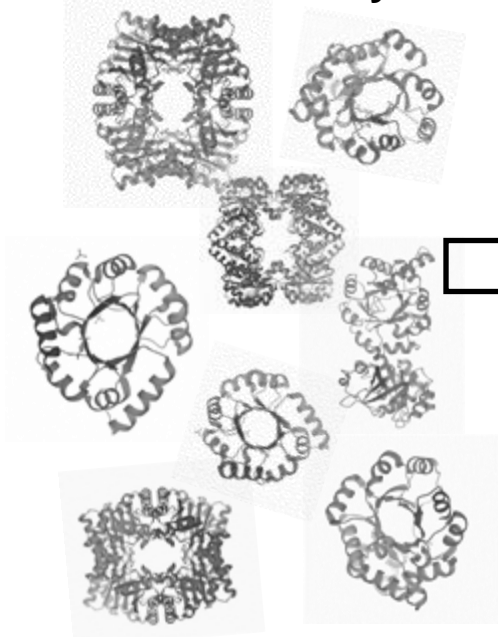
Quantum computing opportunities in synthetic biology: selected applications

- De novo genome assembly
- Sequence alignment
- Metabolic optimization
- Protein function design
- Enzymatic reaction simulations

Design of enzymes for new-to-nature functions

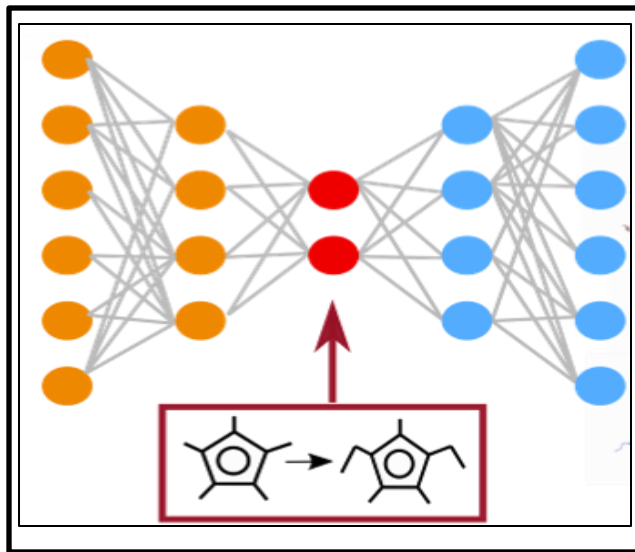
Acknowledgement: iBEX-2021 Chem2Bio team

Training data: a large set of
TIM Barrel enzymes

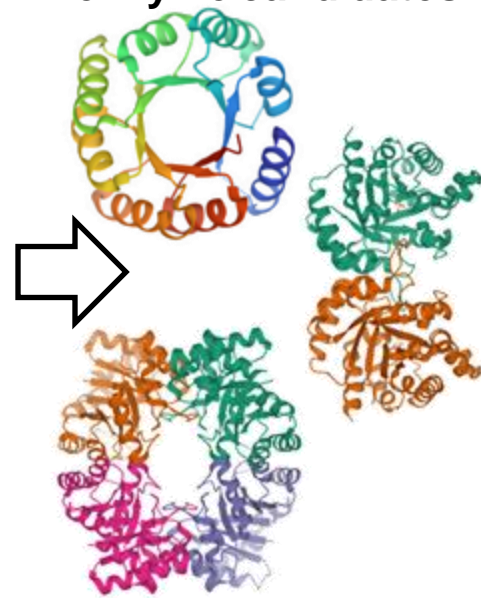


Application case: TIM Barrel enzymes

Generative AI model



Output: new-to-nature
enzyme candidates



Test case: Fructose biphosphate aldolase

Reflections

- As computational biologists, we use ML approaches in our daily life
 - VAE, cVAE, CNN, HMM, Multivariate models, ...
- Can quantum computing be used for better training of the ML models?
 - more efficient training of the hyper parameters, or
 - reducing the number of hyper parameters needed using quantum neurons
 - The GAN work by Kao et al., 2023 seems to demonstrate that using only 4 qubits, they can build better performing GANs (not more efficient)
- Can quantum computing allow us to build better performing models even if they do not provide advantage time complexity-wise?
 - QVAE, QHMM, Quantum Least Squares (QLS), QGAN?, ...

Quantum computing opportunities in synthetic biology: selected applications

- De novo genome assembly
- Sequence alignment
- Metabolic optimization
- Protein function design
- Enzymatic reaction simulations

Reflections

Quantum computing to understand and predict enzyme functions

- We are interested to predict and understand enzyme catalyzed chemical reactions from atomistic details (physical movements of electrons and atoms)
- We can simulate conformational fluctuations with enzymes using classical molecular dynamics (MD) simulations, but this method does not allow simulations of chemical reactions
- **This is potentially the most unique area for quantum computing**
 - **QC potentially gives exponential time complexity improvement**
 - **QC potentially allows us to simulate this process, which can't be done with classical computers.**
- How far are we from QM/MM simulations where classical MD steps are performed with CPU/GPU computing and QM steps with quantum computer?

Conclusion

In theory, quantum computing is widely applicable in computational synthetic biology.

BUT!

- The scale of the problems (data) poses a big challenge
 - Biological sequence data analysis might need qRAM developments
- Quantum literature is full of toy examples
 - Challenge: how do you get funding for trying toy-scale problems!?
- Not all quantum algorithms make you exponentially faster
 - Are the quadratic, heuristic improvement cases really worth trying?
- Can we envisage better models with quantum computing, even though not faster? (i.e., can we identify such problems with proofs, not empirically?)

Regardless of these challenges, our quest will continue!

Acknowledgements

- Bioanalytics and *Biological Data Science* team
 - Dorothee Barth
 - Peter Blomberg
 - Sandra Castillo
 - Samuli Ollila
 - Tuula Tenkanen
 - Lorena Martinez-España
- Merja Penttilä
- Paula Jouhten (Aalto)
- IBEX 2021 team
 - Sandra Castillo, Paula Jouhten, Harry Boer, Martina Blomster, Anna Borisova

bey⁰nd

the obvious

Gopal Peddinti
gopal.peddinti@vtt.fi
+358 40 142 7562

@VTTFinland

www.vtt.fi