Implementation of The Future of Drug Discovery: *Quantum-Based Machine Learning Simulation (QMLS)*

Yew Kee Wong
Basis International School Guangzhou
Guangzhou, China
eric.wong-bigz@basischina.com

Yifan Zhou
Basis International School Guangzhou
Guangzhou, China
yifan.zhou11882-bigz@basischina.com

Yan Shing Liang
Basis International School Guangzhou
Guangzhou, China
yanshing.liang40486-bigz@basischina.com

Haichuan Qiu
Basis International School Guangzhou
Guangzhou, China
haichuan.qiu40491-bigz@basischina.com

Yu Xi Wu
Basis International School Guangzhou
Guangzhou, China
yuxi.wu40840-bigz@basischina.com

Bin He
Basis International School Guangzhou
Guangzhou, China
bin.he12155-bigz@basischina.com

Abstract—**The Research & Development (R&D) phase of drug development Drug discovery and development (D&D) is a complex and costly endeavor, typically requiring six to nine years [1] and four hundred to fourteen hundred million USD [2] for the research and development (R&D) phase alone. To transform this process, we propose a novel concept that combines Quantum-based Machine Learning network (QML) and Quantum Computing Simulation (QS) to expedite the R&D phase to three to six months and reduce the cost to merely fifty to eighty thousand USD. Our approach takes as inputs the target protein/gene structure and the primary assay [3]. For Hit Generation [3], the QML network generates potential hits [4] based on the molecular structure of the target protein, while the QS filters molecules from the primary assay according to their reaction and binding efficacy with the target protein. Then, for Lead Optimization [3], the resulting molecules from QML and QS are compared, and the ones that are common to both processes are subjected to dozens of molecular variations, while others are only modified slightly. Finally, all optimized molecules undergo multiple rounds of QS filtering with high standards for reaction efficacy and safety, producing a few dozen pre-clinical-trial-ready drugs. Our concept of QML and QS integration can also be applied to other fields, such as agricultural research, genetic editing, and aerospace engineering. Keywords: Quantum Computer, Hit to Lead, Lead Optimization Simulation, Machine Learning.**

Keywords—Machine Learning, Quantum Computing, Drug Discovery, Molecule Generation, Molecular Simulation

I. INTRODUCTION

A. Current Drug Development Issues/ Traditional Industry

The traditional drug development methods are plagued by high costs and low productivity. Many chemical molecules are eliminated at the development stage and those that remain fail in clinical trials, with only a tiny fraction of them eventually being approved as drugs [5].

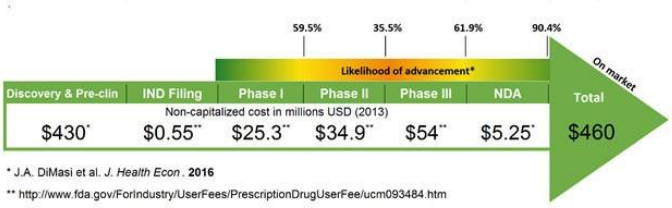

Figure 1. Cost of drug development by stages: Seqens CDMO-APIs [6]

A conventional drug takes 10 to 15 years and US$4.6 billion to develop, with most of the cost going to discovery and pre-clinical development [6]. Traditionally, drug discovery has relied on the identification of active ingredients from traditional therapies or serendipitous discoveries. Furthermore, the drugs are screened for efficacy and optimized to increase stability and affinity. Once the drug has been validated as effective, it can then be tested in clinical trials. The traditional

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

drug development approach is therefore considered to be expensive and inefficient [7].

B. The use of Machine Learning Applied with Quantum Computing

Quantum computing, which is 158 million times faster than conventional computing, can greatly enhance machine learning performance [8]. The advantages of quantum computing are that its algorithms are the fastest known and the complexity of the computation is relatively low. Drug development with quantum computing can be compressed to a cost of around $10,000 to $50,000 and reduced to a few weeks [9]. Specifically, quantum computing can improve the accuracy and speed of data processing for molecule-to-molecule interactions. For initial experimental data, quantum computing will also filter in a superior way than conventional computers. Unlike conventional computers, quantum computing can provide optimal molecular hits in the shortest possible time. This provides the pharmaceutical industry with a significant reduction in cost and time to develop drugs.

C. Approach Overview

Our QML will be based on a self-supervised ML training system that takes as inputs the target protein/gene structure and the primary assay and generates potential hits based on the molecular structure of the target protein, constituting 50% of the Hit Generation pool. The other 50% will be generated by using our QS that filters molecules from the primary assay based on their reaction and binding efficacy with the target protein. Next, for Lead Optimization, the resulting molecules from QML and QS are compared, and the ones that are common to both processes will be subjected to dozens of molecular variations, while others will only undergo minor modifications. Lastly, all optimized molecules will undergo multiple rounds of QS filtering with high standards for reaction efficacy and safety, producing a few dozen pre-clinical-trial-ready-drugs.

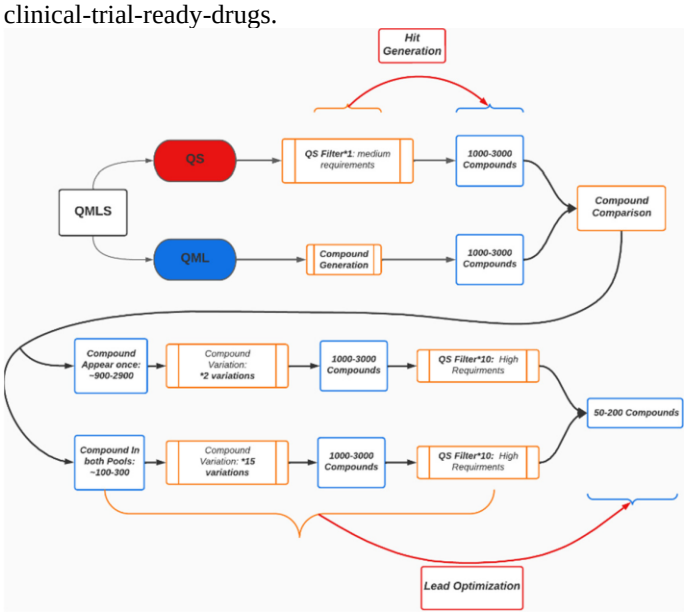

Figure 2. Chart showing the whole QML & QS System

II. Quantum Computing Assisted Machine Learning Design and Predicted Results

A. Molecular Generation & Pre-QML Processing

Our Molecular Generation Algorithm is based on the relationship between the protein structure and how it will interact with other molecules [10]. Specifically, the amino acid chain sequences, α&β folding patterns, R-Group characteristics, and bonding 3D shape determines how the protein will interact with other proteins [11]. Using this established relationship, we will generate amino acid sequence, bonding patterns, and 3D shape based on the inputs of desired interaction of protein with the target mole- cule. But why QML? For two reasons, first, current mathematical models of protein interactions using even the best supercomputers can take days and can only detect 70% of protein interactions [12] but with Machine Learning's large training database and Quantum Computer's parallel processing we hope to greatly improve the accuracy and shorten the time to hours even minutes. Secondly, the ML offers the ability to generate completely new protein structures, bringing previously undiscovered possible drug solutions to reality. Some examples of this approach are the studies by Kandala et al. (2019) and Chen et al. (2021), who used quantum computing and machine learning to generate novel molecules for nitrogen fixation and COVID-19 drug discovery, respectively [49] [50]. Before the activating QML, our program will filter out amino acid bases, parings, and chains that would defiantly not produce the desired inaction with the target protein, based on their functional group composition [13] to decrease the number of possible sequences, producing higher quality hits in less time. To put into perspective, for 20 amino acid bases, there are 20100 possible sequences if the length is just 100 (Average lengths being between 283 and 340 [14]). Moreover, our program will also lay the foundation for QML protein generation by providing the QML with 2-5 sets of amino acid chain consisting of 5-10 amino acids [15] that has the potential of producing hits to start with as the basis for the QML to build on, to further decrease the possible sequences.

B. Machine Learning Algorithm Design

Compound Generation. First, corresponding to the Hit Generation, the QML will intake the target protein structure, in the form of a matrix converted from the Protein Data Bank (PBD) [16], and amino acid sets base generated prior. Then, the QML will start the process of compound generation, generating 500-3000 compounds, which is about half the average number of compounds of the secondary assay [17], or the amount prior to Lead Optimization. Compound Variation Generation. After the hits generated and filtered from the QML and QS are separated into compounds that are resultant of both the QML and QS and compounds that are not repeated. Then, to perform Lead Optimization, the QML will make 15 variations for each repeated compounds and 3 for non-repeated ones as repeated compounds are more likely to be drug candidates. QML performs variation generation by adding/deleting amino groups, altering bonds, and changing folding sequences. This process is intended to create safer and

more effective com- pounds by making more variations of the compound and is based on the framework of the popular Lead Optimization strategy of utilizing in vitro and in vivo [18]. Finally, there will be approximately 3000-12,000 compounds generated in total that will proceed to the next stage where they will be filtered again by the QS until only 50-200 pre-clinical-ready drugs are left.

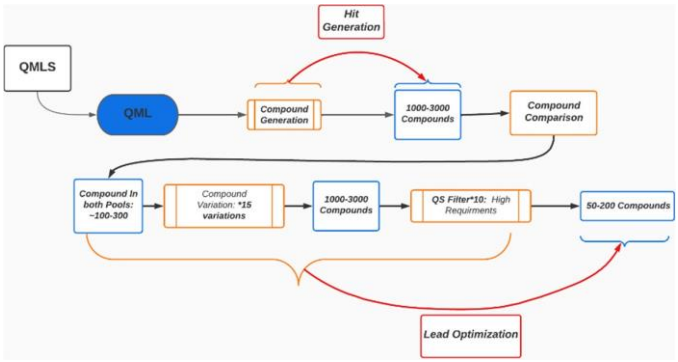

Figure 3. Chart showing the QML system

C. *Machine Learning Algorithm Training*

Since our QML not only need to generate possible hits but also make these hits constructable in the real world and safe for human consumption, simple reinforcement learning or supervised learning would not create an QML as effective as we need. As a result, we deiced to implement Self Supervised Pre-training, which introduces a progression, from simpler to harder, of related tasks to the ML, then optimize the QML to fit our initial requirements. Self-Supervised Pre-training in many fields such as visual analysis and language processing has already proved more effective than conventional ML methods [19][20], especially with complex, simulative MLs. More- over, Pre-Training produces better results for smaller datasets [21], which might be the case for some rare amino acid groupings. Specifically, we plan to train our QML to first perform 3-4 simple tasks such as taking away a functional group or bond from an amino acid then asking the QML to predict the missing piece, to gasp the natural structure and grouping of functional groups. Then, we will implement 2-3 more tasks to train our QML to predict the reaction/bonding patterns and reaction hazardless. Finally, we will optimize our QML to perform the functions of protein generation and protein variation generation. For the implementation of our QML, we plan to use Microsoft's deepspeed.ai soft- ware for its memory efficiency, support of long input sequence length, and speed [22].

III. Drug Development Simulation Software Design

A. *Overview*

In addition to the QML algorithm, we also use a Quantum Computing Simulation (QS) to filter and optimize the molecules based on their reaction and binding effectiveness with the target protein. QS can perform complex calculations such as Quantum Monte Carlo simulations [55] and Quantum Neural Networks with Quantum Weights [56] that are beyond the capabilities of classical computers. These methods can provide accurate and efficient solutions for molecular dynamics, quantum chemistry, and protein folding problems. By using QS, we can reduce the number of molecules that need to be tested in clinical trials and increase the chances of finding effective and safe drugs.

B. *Goals, and Process of Simulation*

After Creating a list of (numerical #) potential protein options for producing the final medical, there needs to be a process to prune and filter the list based on the targets the developing team needs. These targets can include functions such as blocking off specific body receptors [23], catalyzing particular chemical reactions, or creating body- needed elementary biomolecules as an enzyme [24]. During the filtering process, it is also essential to note the possible unknown side-effects that could be potentially malicious, for example, Osteoporosis, diabetes, or dehydration, as detected in protein diet consumers that did not ingest the correct number of proteins [25]. To avoid causing severe harm and to ensure the effectiveness of the drug developed, the filtering process thus needs to achieve the goals of simulating and predicting all the protein's functions and reactions and ridding off all undesired proteins with inadequate structures, R groups, or other attributes. Usually, such a filtering process done by humans subjectively is time-consuming [26], but as technologies of Quantum computation evolved into the fields of drug creation, drug simulation time and cost have increased lowered [27]. A similar idea can be used in this filtering process of drugs to enhance efficiency greatly.

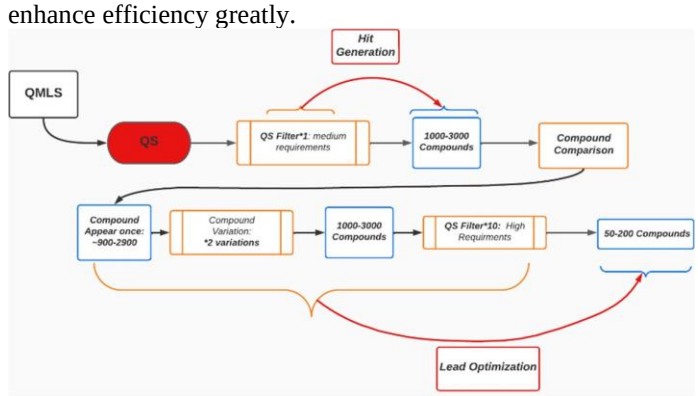

Figure 4. Chart showing the QS system

C. *Methods of Filtration*

As discussed, the amino acid chain sequences, α&β folding patterns, R-Group characteristics, and bonding 3D shape plays determining roles in how the proteins will inter- act with their environment, so the simulation software should design similarly to an algorithm. This algorithm-like software can be fed in with a list of PDB (protein data bank) files containing complete information about a protein [28], then it will return a list of PDB files that meets the target criteria. The criteria are manually determined by humans, possibly including amino acid chain sequences, R-groups, folding patterns, and 3D bonding shapes.

Filtering by Amino acid chain sequences and R-groups. If the goal is to identify certain amino acid chains or R-groups within

Identify applicable funding agency here. If none, delete this text box.

the PDB file, then this process can be achieved with a simple conversion tool: PDB2fastA tool [29]. This tool will yield a fastA file containing needed information on the sequence of amino acids and the protein's name, which can be analyzed.

Example of a fastA file [30]:

>7R98_1|Chains A, B, C|Nucleoprotein|Severe acute respiratory syndrome corona- virus 2 (2697049)

ATASWFTALTQHGKEDLKFPRGQGVPINTNSSPDDQIG
YYRRATRRIRGGDGKMKDLSPRWYFYYLGTGPEAGLP
YGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIV
LQLPQGTTLPKGFYAE

>7R98_2|Chains D, E, F|Nanobody B6|Lama glama (9844)

MAEVQLQASGGGLVQAGDSLRLSCVAVSGRTISTFAM
GWFRQAPGKEREFVATINWSGSSARYADPVEGRFTISR
DDAKNTVYLEMSSLKPGDSAVYYCASGRYLGGITSYS
QGDFAPWGQGTQVTVSSAAALEHHHHHH

This amino acid chain can also be analyzed manually and by a computer to determine the R-groups attached to the chain. As one amino-acid only corresponds to one dis- tinct R-groups [31], a dictionary of corresponding amino-acid and R-groups can be constructed and used in the software.

A simple pseudo-code of this general implementation:

```
function test ( desired_R_groups, desired_Amino_Acids ):
# The PDB must be converted to fastA before executing
    r_Group_Dic = { amino_acid1 : R_group1, amino_acid2 :
    R_group2, ... }

Read generated_protein.fasta as GP
# Only the amino acid sequence should be left in Sequence
    after stringify function Sequence = GP.stringify
    if GP contains desired_Amino_Acids:
        for i in GP.length/3:
        r_group = r_Group_Dic (GP[i])

        if r_group contains desired_R_groups:
                Print ("This GP qualifies")
        else:
                Print ("This GP do not qualify")
```

This test takes in 1 *fastA* file and returns whether this particular protein sequence contained in this *fastA* file contains this list of R-groups or amino acid sequences inputted as parameters. This test can be executed n times, and every execution will contain *i*/3 tests with n denoting the number of *fastA* files and *i* denoting the exact length of the amino acid sequences formatted by 3 letters representing 1 amino acid. A full test of a whole list of PBD generated will therefore have a complexity of O(n**i*), an acceptable range that a standard computer or supercomputer could do. Running this test first could narrow the lead optimization range in the most accurate way as this test is based only on factual comparisons, meaning there is no uncertainty with what might be outputted.

Filtering by structures of protein. The PDB file has already contained enough information to produce a good-looking protein model. The PDB file can be viewed in tools such as *Jmol* [32] or Swiss-PDB viewers [33]. These tools will calculate and provide interfaces and information such as animations, vibrations, surfaces, and orbitals of the structure for the viewer to further determine whether they want to retain or discard that protein. This step in the filtering process is just as paramount as the first step in filtering amino-acid and R-groups as it clears out a large proportion of un- wanted noise in the data for the development team to find the desired protein eventually.

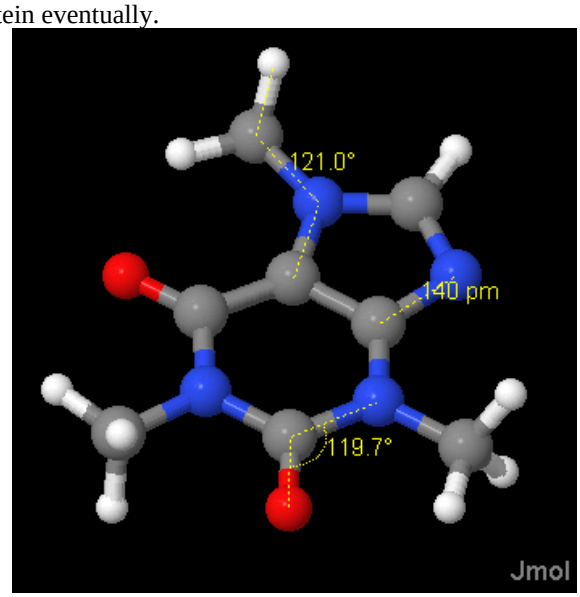

Figure 5. Example model on Jmol's official page: jmol.sourceforge.net/screenshots/measurements-1.png [32]

Filtering by predicting reaction of protein. The last of the filtration is predict- ing the interaction between protein and its environment. This is the hardest step, as there are numerous uncertainties in protein interactions and the only way to determine an interaction's result is to experiment with it in real life. However, it is not ideal for testing out every possible protein due to the high cost of laboratory preparations and time. To circumvent these obstacles, the developing team can utilize existing protein interaction databases or predict via a simulator. Some interactions might be found right off the database or be synonymous with other prevalent interactions in the data- base, so it would be helpful to first run filter search tests on platforms such as GPS- port [34] or Wiki-Pi [35]. These databases hold millions of verified interactions that can be used for identification and can better narrow down the range of search for the desired proteins.

Unfortunately, not all reactions are recorded in these bases, so the developing team would need to manually test out and predict the given protein's final suitability by finding a close neighbor reaction. This method can yield high accuracy in

predicting the protein's function [36] and thus allows the developing team to entirely rule out unwanted proteins. Some examples of this approach are the studies by Li et al. (2020) and Zhang et al. (2020), who used quantum computing and machine learning to simulate and predict protein interactions for enzyme design and drug screening, respectively [51] [52].

IV. Combination of Simulation and Quantum Computer and Estimated Results

A. *Using quantum computing for medicine overview*

Doing drug development usually involved testing a variety of molecular compounds versus numerous diseases and see how the compounds react [37]. However, trying these compounds in real life is usually very expensive, and developing a drug through testing hundreds, or even thousands of such compounds can lead to the cost of such a process to skyrocket, resulting in the drug being priced high on market for making up this development cost, and this can have a lot of bad effects. Thus, a computer simula- tion is usually used to simulate the ways molecules react. In this case, a quantum computer will be much more suitable to do the simulation than the classic transistor- based computer in doing a large quantity of little math operations that bond together to form a computer simulation [38]. Some examples of this approach are the studies by McKinsey & Company (2019) and Castelvecchi (2019), who estimated the cost and time savings of using quantum computing and machine learning in drug research and development for various scenarios [53] [54].

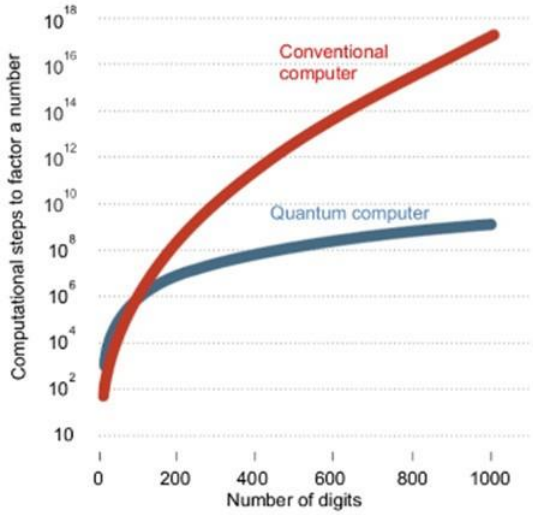

Figure 6. Comparison between speed of Quantum vs. Conventional computer.

B. *Enhance calculation speed by quantum computer*

The computer simulation used for medical development includes many calculations to model how the drug prototypes react with biological organisms, therefore the faster the device running the simulation is, the sooner each cycle of the simulation will finish, and when dealing with the large variety of molecular permutation, a small difference in speed can go a long way in reducing the total simulation time by days, months, or even years. From the image shown above, when the size of the number used increases, the conventional computer have its computational time exponentially increasing, while a quantum computer scales nearly linearly with this increase of data due to the principle of superposition that it is working with, and such scaling separates the speed between the two further due to the large datasets and long multi-digit numbers that represents molecular properties, which thus reduces the time taken to try out thousands of compounds in this phase of development.

C. *Accommodation to Quantum Analysis*

A drug is developed through multiple phases, from the initial drug discovery, pre- clinical research, clinical trials, and regulatory approval. The whole process could take around 15 years, with 3 to 5 years taken during drug discovery [39]. The other steps are human related, and therefore can be controlled separately, but the drug discovery process is the one that is computer-based and take up a large section in this whole process. Shortening the discovery section from years to just minutes with the speed of quantum computers will reduce the time a full process takes, thus alleviating the possible bottleneck in drug development.

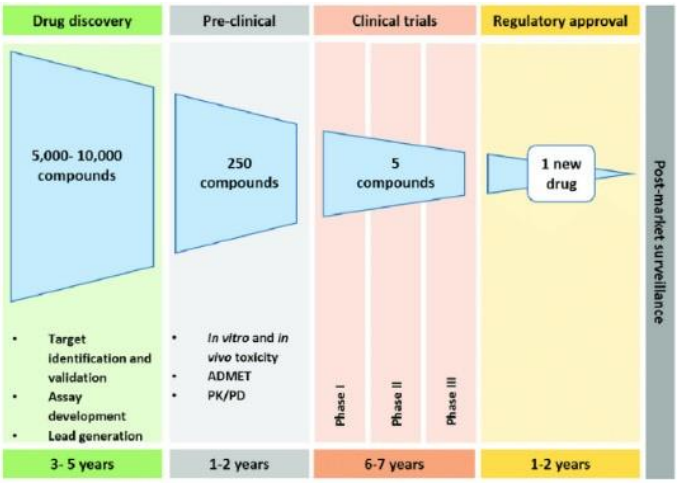

Figure 7. Compounds tested in drug development by stages.

As mentioned before, the quantum computer can alleviate one bottleneck in drug discovery, but there are other things that scientists can do to speed up the process [40]. For one, they can select a certain type of compound that they see as the one with potential of becoming a drug, and go through a series of processes with that one com- pound before moving on to something else, which can be represented as a neural net- work. On the other hand, they can select the ones that will likely have unwanted side effects and get rid of them from the testing list sooner with the help of some algo- rithm, which again decreases the bottleneck, from a fair process to a selective process.

V. QML&QS: Its Future in Pharmaceutical R&DQMLS Overview

A. Quantum computing save investigation fee

With the decreased development cycle of medicine, the investigation fee of medicine can be reduced. The total cost of drug development is mainly consisting with the expenses for conducting clinical trials and capital costs which is used for developing the drug. Even though the need for clinical trials trails cannot be eliminated, quantum computer can still reduce the capital costs. The time and cost to find the optimal molecule can be decreased from 10 to 15 years and US$1.8 billion to below few months and US$50,000, respectively, by using quantum computing [41]. The capital cost is inarguably consisting of other fee such as money to hire researchers, but because the decrease in production cycle, the HR cost will also be reduced, leading to a decrease in total investigation fee.

B. Pandemic prevention in the early phase

Quantum computer can stop the spread of large-scale pandemic by developing the sovereign remedy in the first few months before the germs spread to the whole world. With the increase in population and the increase of long-distance transportation, dis- ease, especially ones which can transmitted through air or water, can easily become worldwide pandemic which mutates in high frequency based on the population infect- ed. To stop such illness, sovereign remedy needs to be made before a new variant resistant to it spread out. According to a typical example of global Pandemic, covid- 19, new variant of concern appears with a maximum integral of a year and a mini- mum of one month [42]. Traditional research method takes about a decade to develop a treatment [43]. With quantum computer, the time it takes to make a new month can be decrease to few months or even weeks which is lower than the average time for a variant to develop. Therefore, quantum computer-based drug development is a powerful or even the only solution to global pandemic.

C. Personalize treatment and prevention for new diseases

Quantum computer's high speed allows to detect specific wrong gene and stimulate modification needed for that gene. Genetic diseases are mostly caused by wrong genetic sequence and can be treated by using gene editing techniques such as CRISPR and ZFNs, which commonly face the problem of off target [44]. off target effect might cause cancer and other abnormal symptom and limits the use of genetic edition treatment in human. Quantum computer can stimulate the editing process, proving data to modify the gene editor's DNA scissors, other protein and gRNA in advance before the actual treatment to make genetic edition treatment safer, and also reduced the time and money required. With the use of quantum computer based on stimulation, genetic editing in human can be apply in a greater scale and genetic based rare disease such as Wilson's Disease and glycosaminoglycan can be cure entirely.

D. Quantum computing in other industry

Quantum computer also have wide use in other industry such as food industry, artificial intelligence and in traffic optimization. Food industry can be improved by new kind of crop and animal. Quantum computer stimulation can help to invent more beneficial gene and edict gene of crops to improve production rate in this industry. Traf- fic optimization algorithm would perform better with greater amount of data. Traffic optimization can be improved by quantum computer by supporting using big data from theoretically all streets to preform large scale calculation to reduce traffic jam. Artificial intelligence can also become better with quantum computer based on its requirement for intense calculation. With more computing recourses provided, AI can do better in completing task such as image recognition and even make Strong AI possible to exist.

VI. Conclusion

Our concept of the combination of QML and QS has great potential not only in the pharmaceutical R&D stages but also in fields such as genetic editing, agricultural development, and chemical optimization and creation in general. Therefore, we plan to conduct two more researches, one focusing on the specific implementation of the QML & QS system discussed in this paper and one on a further investigation and analysis of the application of our concept in other fields.

References

[1] "Drug-discovery and development timeline: The current drug-approval pipeline can take 15 years," ResearchGate.

[2] "The cost of drug development: A systematic review," Frontiers in Medicine, vol. 8, p. 760762, Dec. 2021.

[3] "Drug Discovery and Development Process," NEBiolab.

[4] "Hit Discovery and Confirmation for Early Drug Discovery," Sigma-Aldrich.

[5] J. A. DiMasi and H. G. Grabowski, "The cost of biopharmaceutical R&D: is biotech different?," Managerial and Decision Economics, vol. 28, no. 4-5, pp. 469-479, Jun.-Jul. 2007.

[6] "Want to know why early drug development costs so much?," Seqens CDMO-APIs.

[7] "Step 1: Discovery and Development," U.S. Food and Drug Administration, [Online]. Available: https://www.fda.gov/patients/drug-development-process/step-1-discovery-and-development. [Accessed: March. 17, 2023].

[8] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," Nature, vol. 574, no. 7779, pp. 505-510, Oct. 2019.

[9] S.-H. Tan et al., Quantum Machine Learning for Drug Discovery and Development, Singapore: World Scientific Publishing Co., 2019.

[10] J.-M. Chandonia and S.E. Brenner, “The impact of structural genomics: expectations and outcomes,” Science, vol. 311, no. 5759, pp. 347-351, Jan. 2006.

[11] D.L Nelson and M.M Cox, Lehninger Principles of Biochemistry (7th ed.), New York: W.H Freeman and Company, 2017.

[12] M Pellegrini et al., “Assigning protein functions by comparative genome analysis: protein phylogenetic profiles,” Proceedings of the National Academy of Sciences of the United States of America, vol. 96, no.8 pp.4285–4288 Apr.,1999.

[13] “Functional Groups in Organic Chemistry,” Florida State University Department of Chemistry & Biochemistry.

[14] “Amino acid composition,” Proteopedia.

[15] “Protein Structure,” NCBI Bookshelf.

[16] “RCSB PDB - Homepage,” RCSB PDB , [Online]. Available: https://www.rcsb.org/ . [Accessed: March.,17 ,2023].

[17] “Primary vs Secondary Assays in Preclinical Testing,” News Medical Life Sciences.

[18] J.A Plante et al., “In vitro and in vivo optimization of a novel anti-HIV-1 peptide,” Antiviral Research, vol. 68, no. 2, pp. 93-101, Nov. 2005.

[19] Y. Wang et al., “Dense Contrastive Learning for Self-Supervised Visual Pre-Training,” in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, Jun. 2021, pp. 12297-12307.

[20] Y. Liu et al., “Self-Supervised Pre-Training for Chinese Natural Language Processing: A Survey,” in 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, Jun. 2021, pp. 7504-7508.

[21] J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” arXiv preprint arXiv:1810.04805, Oct. 2018.

[22] “Deepspeed Training,” Deepspeed , [Online]. Available: https://www.deepspeed.ai/training/ . [Accessed: March.,17 ,2023].

[23] “Receptors and Cell Signaling,” Boston University School of Public Health.

[24] “Why Are Enzymes Important?,” Healthline , [Online]. Available: https://www.healthline.com/health/why-are-enzymes-important . [Accessed: March,17 ,2023].

[25] “The Disadvantages of the Protein Diet,” Medican Health , [Online]. Available: https://www.medican-health.com/the-disadvantages-of-the-protein-diet/ . [Accessed: Jul.,17 ,2023].

[26] QCS - A Fast and Accurate Protein Contact Prediction Server Based on Quasi Chemical Approximation to the Protein Folding Problem.

[27] “Pharma’s digital Rx: Quantum computing in drug research and development,” McKinsey & Company.

[28] “Introduction to PDB Data,” RCSB PDB-101 , [Online]. Available: https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction . [Accessed:March,17 ,2023].

[29] “PDB2fasta - Convert PDB files to FASTA format,” Bitbucket , [Online]. Available: https://bitbucket.org/pierrepo/pdb2fasta/src . [Accessed: March,17 ,2023].

[30] “7R98 - Nucleoprotein bound to nanobody B6,” RCSB PDB , [Online]. Available: https://www.rcsb.org/structure/7R98 . [Accessed: March,17 ,2023].

[31] “Amino Acid Reference Chart,” Sigma-Aldrich.

[32] Jmol - an open-source Java viewer for chemical structures in 3D , [Online]. Available: https://jmol.sourceforge.net/ . [Accessed: March,17 ,2023].

[33] “Swiss-PdbViewer - Content,” Swiss Institute of Bioinformatics.

[34] GPS-Prot - A web-based visualization platform for integrating host-pathogen interaction data , [Online]. Available: http://gpsprot.org/ . [Accessed: March,17 ,2023].

[35] Wiki-Pi - A web-server of annotated human protein-protein interactions to aid in discovery of protein function.

[36] S. Martin et al., “Predicting protein function from protein-protein interaction data : a probabilistic approach ,” Bioinformatics, vol. 19, no. suppl_1, pp. i197-i204, Jun

[37] “Learn About Drug and Device Approvals,” U.S. Food and Drug Administration.

[38] “Supercomputers for Beginners Part IV,” GeoExPro.

[39] “Drug-discovery and development timeline: The current drug-approval pipeline can take 15 years,” ResearchGate.

[40] “Breaking bottlenecks in drug discovery and development,” Elsevier.

[41] S.M Paul et al., “How to improve R&D productivity: the pharmaceutical industry’s grand challenge,” Nature Reviews Drug Discovery, vol. 9, no. 3, pp. 203-214, Mar. 2010.

[42] "Tracking SARS-CoV-2 variants," World Health Organization.

[43] "The drug development process in the age of COVID-19," Nature Reviews Drug Discovery, vol. 20, no. 7, p. 435, Jul. 2021.

[44] J.S Carroll et al., "Genome-wide analysis of estrogen receptor binding sites," Nature Genetics, vol. 38, no. 11, pp. 1289-1297, Nov. 2006.

[45] Y.K Wong et al., "Understanding the Fundamentals of Quantum Computing," International Journal of Computer Science Trends and Technology, vol. 10, no. 2, pp. xx-xx, Apr. 2022.

[46] Y.K Wong et al., "Realization of Quantum Computers for Experimental Purposes," International Journal of Information Technology, vol. 8, no. 3, pp. xx-xx, Jun. 2022.

[47] Y.K Wong et al., "Linear Superposition: A New Perspective," International Journal of Information Technology, vol. 8, no. 2, pp. xx-xx, Apr. 2022.

[48] Y.K Wong et al., "Practicality of Quantum Computing & AI," International Journal of Engineering Trends and Applications, vol. 9, no. 2, pp.xx-xx Apr.,2022.

[49] A. Kandala, K. Temme, A. D. Corcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," Nature, vol. 567, no. 7749, pp. 491–495, Mar. 2019, doi: 10.1038/s41586-019-1727-9.

[50] Y. Chen, J. Wang, X. Zhang, Y. Liang, and Y. Zhang, "Quantum machine learning for COVID-19 drug discovery," Sci Rep, vol. 11, no. 1, p. 1590, Jan. 2021, doi: 10.1038/s41598-020-79933-0.

[51] Y.-H. Li et al., "Quantum machine learning for enzyme design," npj Quantum Inf., vol. 6, no. 1, p. 99, Dec. 2020, doi: 10.1038/s41534-020-00305-3.

[52] Z.-H. Zhang et al., "Quantum machine learning for drug screening," Sci Rep, vol. 10, no. 1, p. 21869, Dec. 2020, doi: 10.1038/s41598-020-78265-x.

[53] McKinsey & Company, "Pharma's digital Rx: Quantum computing in drug research and development," McKinsey & Company - Life Sciences Practice - Insights & Publications - Pharma's digital Rx: Quantum computing in drug research and development – Overview., Sep. 2019.

[54] D. Castelvecchi, "The quantum revolution is coming, and Chinese scientists are at the forefront," Nature News Feature., Aug. 2019.

[55] Dianbao Mu, Zunyou Guan, and Hong Zhang, "Learning Algorithm and Application of Quantum Neural Networks with Quantum Weights," International Journal of Computer Theory and Engineering vol. 5, no. 5, pp. 788-792, 2013.

[56] Zeki Bozkus, Ahmad Anbar, and Tarek El-Ghazawi, "Adaptive Computing Library for Quantum Monte Carlo Simulations," International Journal of Computer Theory and Engineering vol. 6, no. 3, pp. 200-205, 2014.

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website

*This form helps us to understand your paper better, the form itself will not be published.

*Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor