Position Papers of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 69–71
DOI: 10.15439/2024F3588
ISSN 2300-5963 ACSIS, Vol. 40

A generic method of pose generation in molecular docking via quadratic unconstrained binary optimization

Pei-Kun Yang
Research Center for Applied Sciences,
Academia Sinica
Taipei, Taiwan
peikun@gate.sinica.edu.tw

Jung-Hsin Lin
Biomedical Translation Research Center, Research Center for Applied Sciences,
Academia Sinica
Taipei, Taiwan
jhlin@gate.sinica.edu.tw

***Abstract*—Docking of a ligand onto the binding pocket of its protein target, designated as the molecular docking problem, is a very important method for structure-based drug design. We have implemented a generic pose generation method for molecular docking by solving the quadratic unconstrained binary optimization (QUBO) problem with the Fujitsu digital annealer. In combination with the AutoDock 4 scoring function, the success rate for predicting the binding poses to be sufficiently close to their experimental binding poses, namely, with the root mean squared deviation (RMSD) less than 2 Å, was 84.3 %, when benchmarking against part of the PDBbind core set (242 protein-ligand complexes). To our best knowledge, this is the first implementation of molecular docking that conforms with the QUBO formalism demonstrating a performance comparable with the conventional methods.**

***Index Terms*—molecular docking, quadratic unconstrained binary optimization, QUBO, pose generation, AutoDock**

I. Introduction

MOLECULAR docking is an essential method for structure-based drug design and virtual screening of chemical libraries for finding chemical skeletons for creating novel chemical entities. Docking of a ligand onto the binding pocket of its protein target, generally consists of two parts: pose generation and binding affinity evaluation. In the first step, a myriad of ligand conformations at the protein surface (usually at the binding pocket) need to be generated, and these conformations should include the ones that are very close to the experimentally determined binding poses. Typical experimental methods are protein X-ray crystallography, nuclear magnetic resonances, and cryogenic electron microscopy. In the second step, the binding affinities of these ligand conformations at the protein binding pocket will be evaluated with a scoring function (or a free energy functional), and the poses with best binding affinities should be very close the experimental binding poses. It is considered as a successful molecular docking when the second step can be achieved, i.e., the root mean squared deviation (RMSD) of the predicted binding pose with the best score (binding affinity) from the experimental binding pose is less than, e.g., 2 Å.

Quadratic unconstrained binary optimization (QUBO), sometimes also known as unconstrained binary quadratic programming (UBQP), is a class of combinatorial optimization problems with a huge variety of applications. QUBO is known as an NP hard problem. Many classical problems from theoretical computer science, e.g., maximum cut, graph coloring and the partition problem, have been formulated into QUBO. Due to its close connection to Ising models, QUBO constitutes a major class of computational problems for adiabatic quantum computation, where it can be solved through a physical process named quantum annealing.

D-Wave and Fujitsu are two well-known companies that strive to develop computers to efficiently solve the QUBO problems with quantum annealer and quantum-inspired (or physics-inspired) annealers, respectively. Although many important applications have been embedded into the QUBO formulism, molecular docking is still not yet implemented and it is not clear whether such an implementation can indeed lead to practically useful applications.

II. Methods

Many important problems in molecular biology, including protein folding, protein-protein binding, protein-DNA (or RNA) binding, and protein-ligand binding, are problems of searching for free energy minimum, from the perspective of statistical thermodynamics. The problem of finding the minimum value in one dimension can be easily solved by, e.g., the Newton-Raphson method, etc. The difficulty of finding the global minimum exponentially escalates as the dimensionality increases. Compared with algorithms that have been developed for decades, the emerging hardware such as those developed by D-Wave and Fujitsu have the opportunity to find solutions with lower function values at high dimensions with dramatically less time than the conventional methods. The QUBO formulism generally reads:

$$F(\mathbf{X}) = \mathbf{X}^{\mathrm{T}}\mathbf{J}\mathbf{X} + \mathbf{H}\mathbf{X} \qquad (1)$$

In Equation (1), **X** has n binary variables, and **J**, **H** and **X** are $(n \times n)$, $(1 \times n)$ and $(n \times 1)$ matrices respectively. Given the element values J_{ij} and H_i of the **J** and **H** matrices, QUBO solvers can be used to find a set of solutions to X that minimize the value of F**(X)**.

In this work we construct a QUBO model for finding binding poses in molecular docking. In a 3-D lattice covering the binding pocket, QUBO solvers should obtain a solution X_i of 1, to indicate a ligand atom will locate as this lattice point. Compare the distribution of ligand atoms and X_i to obtain the possible binding positions of ligand. In addition to using Fujitsu DAU3 as QUBO solvers, we can also PyTorch to solve the QUBO model.

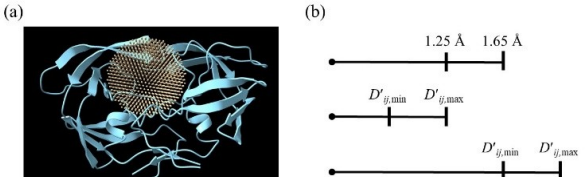

Figure 1: (a) Lattice points near to the binding pocket of the protein in the PDB ID 1a30. The protein conformation was shown in ribbon representation, colored in cyan. In the region of the protein binding pocket, the grey round points are the lattice points X_i. (b) (Up) The distance between the non-hydrogen covalent bonds are between 1.25 Å and 1.65 Å. (Middle) The maximum distance between two lattice points $D'_{ij,\max}$ is less than 1.25 Å. Lattice points i and j cannot co-exist inside an atom. (Down) The minimum distance between two lattice points $D'_{ij,\min}$ is larger than 1.65 Å. Lattice points i and j cannot form the covalent bond.

III. Results

Our major results can be seen from Figure 2 and Figure 3.

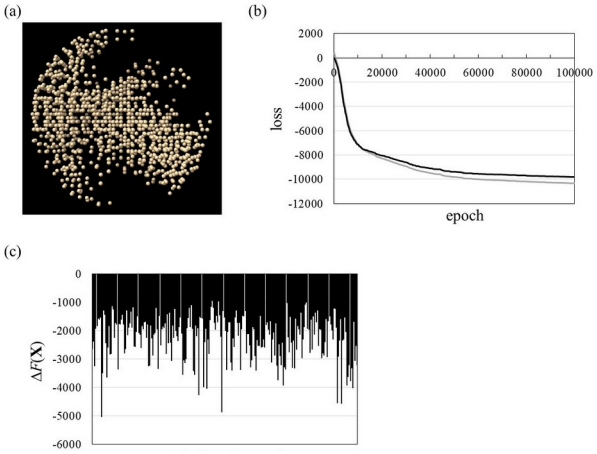

Figure 2: (a) Lattice points with $X_i = 1$, as determined by Fujitsu DAU3. Represented as small round spheres. (b) The grey line shows that $F(\mathbf{X}')$ decrease as the number of epoch increases. The black line shows the values of $F(\mathbf{X})$ in the course of epoch. (c) The difference between the minimum values determined by Fujitsu DAU3 and PyTorch.

IV. Discussion

From our results it indicates that with proper implementation and suitable parameters, it is possible to embed the molecular docking problem into the QUBO format. Currently our implementation is largely guided by our intuition and our prior understandings of the molecular docking problem, and these background and experiences greatly accelerate the progress of this work. It may be possible to employ some machine learning approaches or large language models, such as newer generation of ChatGPT, to translate the problem of interest into the QUBO formulism. However, it may take some more time to witness this to be a reality.

V. Conclusion

Molecular docking is an essential workhorse for structure-based drug design and virtual screening of chemical libraries for finding chemical skeletons for creating novel chemical entities. Emerging hardware such as those developed by D-Wave and Fujitsu have a great opportunity to find solutions with lower function values at high dimensions with dramatically less time than the conventional methods. To our best knowledge, this is the first implementation of molecular docking that conforms with the QUBO formalism demonstrating a performance comparable with the conventional methods. It can be envisioned that such an approach could evolve to become to more efficient and more accurate method for molecular docking and thereby accelerate the drug discovery process.

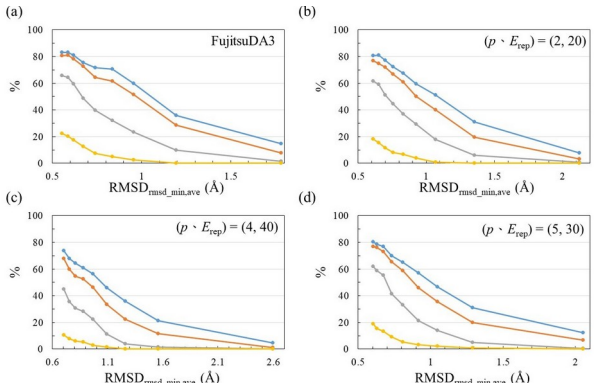

Figure 3: Resulting using (a) FujitsuDAU3, (b) PyTorch (p, E_{rep}) = (2, 20), (c) (4, 40) and (d) (5, 30) for solving QUBO. The fraction of RMSD $_{bfe\_min}$ < 2.0 Å (blue line), < 1.5 Å (red line)、< 1.0 Å (grey line) and < 0.6 Å (yellow line)。

Acknowledgment

JHL would like to acknowledge the financial support of National Science of Technology Council (NSTC), under the grants NSTC 111-2112-M-001-053-MY3 and NSTC 112-2119-M-006-004. JHL would like to also knowledge the financial support and Research Center for Applied Sciences (RCAS), Biomedical Translational Research Center (BioTReC), both are research institutes of Academia Sinica (AS).

References

[1] Maia, E. H. B.; Assis, L. C.; De Oliveira, T. A.; Da Silva, A. M.; Taranto, A. G., Structure-based virtual screening: from classical to artificial intelligence. *Frontiers in chemistry* **2020,** *8*, 343.

[2] Varela-Rial, A.; Majewski, M.; De Fabritiis, G., Structure based virtual screening: Fast and slow. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2022,** *12* (2), e1544.

[3] Burley, S. K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chao, H.; Chen, L.; Craig, P. A.; Crichlow, G. V.; Dalenberg, K.; Duarte, J. M., RCSB Protein Data Bank (RCSB. org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Research* **2023,** *51* (D1), D488-D508.

[4] Tingle, B.; Tang, K.; Castanon, J.; Gutierrez, J.; Khurelbaatar, M.; Dandarchuluun, C.; Moroz, Y.; Irwin, J., ZINC-22-A free multi-billion-scale database of tangible compounds for ligand discovery. **2022**.

[5] Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L., Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of chemical information and modeling* **2012,** *52* (11), 2864-2875.

[6] Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge structural database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016,** *72* (2), 171-179R.

[7] Fu, H.; Chen, H.; Blazhynska, M.; Goulard Coderc de Lacam, E.; Szczepaniak, F.; Pavlova, A.; Shao, X.; Gumbart, J. C.; Dehez, F.; Roux, B., Accurate determination of protein: ligand standard binding free energies from molecular dynamics simulations. *Nature Protocols* **2022,** *17* (4), 1114-1141..

[8] Heinzelmann, G.; Gilson, M. K., Automation of absolute protein-ligand binding free energy calculations for docking refinement and compound evaluation. *Scientific reports* **2021,** *11* (1), 1-18.

[9] Wang, E.; Fu, W.; Jiang, D.; Sun, H.; Wang, J.; Zhang, X.; Weng, G.; Liu, H.; Tao, P.; Hou, T., VAD-MM/GBSA: A Variable Atomic Dielectric MM/GBSA Model for Improved Accuracy in Protein–Ligand Binding Free Energy Calculations. *Journal of Chemical Information and Modeling* **2021**.

[10] Dittrich, J.; Schmidt, D.; Pfleger, C.; Gohlke, H., Converging a knowledge-based scoring function: DrugScore2018. *Journal of chemical information and modeling* **2018,** *59* (1), 509-521.

[11] Bao, J.; He, X.; Zhang, J. Z., Development of a New Scoring Function for Virtual Screening: APBScore. *Journal of Chemical Information and Modeling* **2020,** *60* (12), 6355-6365.

[12] Cavasotto, C. N.; Aucar, M. G., High-throughput docking using quantum mechanical scoring. *Frontiers in chemistry* **2020,** *8*, 246.

[13] Kadukova, M.; Machado, K. d. S.; Chacón, P.; Grudinin, S., KORP-PL: a coarse-grained knowledge-based scoring function for protein–ligand interactions. *Bioinformatics* **2021,** *37* (7), 943-950.

[14] Yang, C.; Zhang, Y., Lin_F9: A Linear Empirical Scoring Function for Protein–Ligand Docking. *Journal of Chemical Information and Modeling* **2021**.

[15] Schneider, C.; Buchanan, A.; Taddese, B.; Deane, C. M., DLAB: deep learning methods for structure-based virtual screening of antibodies. *Bioinformatics* **2022,** *38* (2), 377-383.

[16] Li, H.; Sze, K. H.; Lu, G.; Ballester, P. J., Machine-learning scoring functions for structure-based virtual screening. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2021,** *11* (1), e1478.

[17] Ricci-Lopez, J.; Aguila, S. A.; Gilson, M. K.; Brizuela, C. A., Improving structure-based virtual screening with ensemble docking and machine learning. *Journal of Chemical Information and Modeling* **2021,** *61* (11), 5362-5376.

[18] McNutt, A. T.; Francoeur, P.; Aggarwal, R.; Masuda, T.; Meli, R.; Ragoza, M.; Sunseri, J.; Koes, D. R., GNINA 1.0: molecular docking with deep learning. *Journal of cheminformatics* **2021,** *13* (1), 1-20.

[19] Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S., AutoDock Vina 1.2. 0: New docking methods, expanded force field, and python bindings. *Journal of Chemical Information and Modeling* **2021,** *61* (8), 3891-3898.

[20] Allen, W. J.; Balius, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C., DOCK 6: Impact of new features and current docking performance. *Journal of computational chemistry* **2015,** *36* (15), 1132-1156.

[21] Ypma, T. J., Historical development of the Newton–Raphson method. *SIAM review* **1995,** *37* (4), 531-551.

[22] Willsch, D.; Willsch, M.; Gonzalez Calaza, C. D.; Jin, F.; De Raedt, H.; Svensson, M.; Michielsen, K., Benchmarking Advantage and D-Wave 2000Q quantum annealers with exact cover problems. *Quantum Information Processing* **2022,** *21* (4), 141.

[23] Şeker, O.; Tanoumand, N.; Bodur, M., Digital annealer for quadratic unconstrained binary optimization: a comparative performance analysis. *Applied Soft Computing* **2022,** *127*, 109367.

[24] Woods, B. D.; Kochenberger, G.; Punnen, A. P., QUBO Software. In *The Quadratic Unconstrained Binary Optimization Problem*, Springer: 2022; pp 301-311

[25] Zaman, M.; Tanahashi, K.; Tanaka, S., PyQUBO: Python library for mapping combinatorial optimization problems to QUBO form. *IEEE Transactions on Computers* **2021,** *71* (4), 838-850.

[26] Kochenberger, G.; Hao, J.-K.; Glover, F.; Lewis, M.; Lü, Z.; Wang, H.; Wang, Y., The unconstrained binary quadratic programming problem: a survey. *Journal of combinatorial optimization* **2014,** *28* (1), 58-81.

[27] Tavares, G., New algorithms for Quadratic Unconstrained Binary Optimization (QUBO) with applications in engineering and social sciences. Rutgers The State University of New Jersey-New Brunswick: 2008.

[28] Koh, Y. W.; Nishimori, H., Quantum and classical annealing in a continuous space with multiple local minima. *Physical Review A* **2022,** *105* (6), 062435.

[29] Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S., The PDBbind database: methodologies and updates. *Journal of medicinal chemistry* **2005,** *48* (12), 4111-4119.

[30] Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L., Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **2019,** *32*.