課題名（タイトル）：

Elucidation of biomolecular mechanisms through molecular dynamics simulation

利用者氏名：
○本橋 昌大（1）、Haeri Im（1）、Elisa Rioual（1）、水谷 淳生（1）、Tarun Maity（1）

理研における所属研究室名：
（1）理化学研究所 開拓研究本部 杉田理論分子科学研究室

Investigating the effect of glycosylation on the structural features and hydration effects of glycoproteins

(担当: Im)

１．　Background and purpose of the project, relationship of the project with other projects

Glycoproteins are comprised of protein and carbohydrate chains. The glycosylation patterns are of crucial importance for the folding, stability, and biological activity of glycoproteins. Understanding the relationship between glycosylation patterns and protein function is essential for the development of novel therapeutics and diagnostics. In addition, the structure and stability of glycoproteins are significantly affected by surrounding water molecules. Interfacial water molecules also influence the binding affinity between glycoproteins and receptors. However, the exact molecular mechanisms by which glycosylation patterns and water molecules determine the structure and binding affinity of glycoproteins are not fully understood.

In this research project, we will investigate how the glycosylation patterns impact the structure and binding affinity and how surrounding water molecules contribute to these properties. For this purpose, we will perform molecular dynamics simulations of glycoprotein using the GENESIS software and analyze the effects of glycosylation patterns on the protein structure and function. The results of this study will provide molecular insights into the relationship between glycosylation and the protein structure and function, and will provide invaluable information for developing new drugs targeting glycoproteins.

２．　Specific usage status of the system and calculation method

We performed all-atom MD simulations at 310K and 0.15M NaCl for erythropoietin complex(EPO/EPOR). MD simulations were performed for the EPO complexes.

We employed the CHARMM36m force field for the EPO/EPOR and the TIP3P model for water. Starting from the initial structures of the EPO complex with three glycans(N24N38N83) using CHARMM-GUI, we used GENESIS to perform three independent MD simulations for 1.5 microseconds for each structure.

３．　Result

The glycan at the N24 position maintained the most structurally stable conformation throughout the simulation and exhibited the longest interaction duration with the receptor EPOR. In particular, the terminal sialic acid of the glycan, carrying a negative charge, formed electrostatic interactions with specific positively charged regions on the protein surface, further strengthening glycan-protein interactions (Fig. A). These interactions were maintained throughout the simulation, indicating that the N24 glycan plays a crucial role in enhancing the binding stability with EPOR.

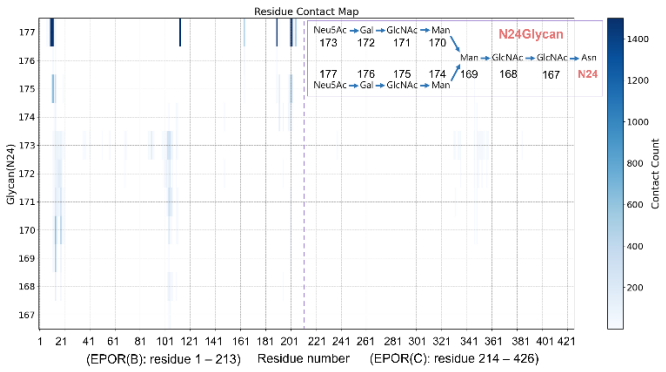

Figure A. Number of contacts between glycan(N24) and EPOR.

Additionally, the N24 glycan exhibited longer interactions with surrounding water molecules than other glycans in the same structure, leading to an increased water residence time (Fig. B).

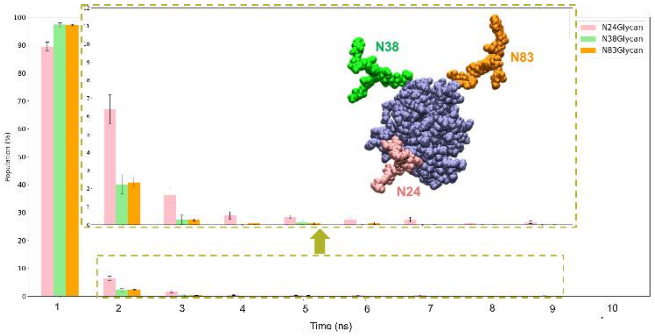

Figure B. Water residence time around glycans

Furthermore, the presence of glycans not only affected the hydration around the glycan itself but also prolonged the water residence time around both EPO and EPOR (Fig. C).

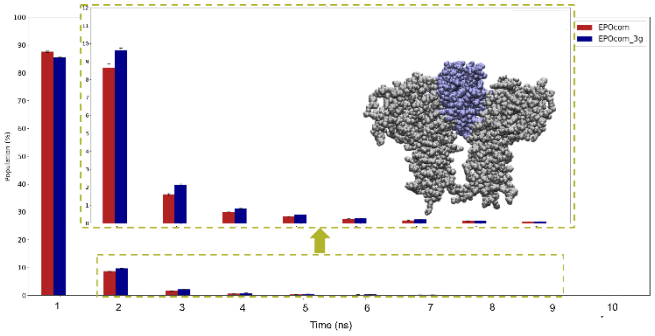

Figure C. Water residence time around EPO/EPOR except for glycans

4．　Conclusion

Glycosylation in the EPO/EPOR complex significantly influenced the structural stability of the protein and its receptor binding. The study found that the N24 glycan increased the water residence time more than glycans at other positions and also affected the hydration layer around EPO and EPOR. These results suggest that glycans may regulate binding affinity and enhance receptor interaction.

MDH – CS enzymes complex: insight of its structure by simulation

(担当: Rioual)

1．　Background and purpose of the project, relationship of the project with other projects

The Krebs cycle is a central metabolic pathway, responsible for generating key intermediates that play vital roles in various biosynthetic processes [1]. Within this cycle, two enzymes—Malate Dehydrogenase (MDH) and Citrate Synthase (CS)—catalyze consecutive reactions, forming an important regulatory node. Recent literature indicates that these enzymes may form a metabolon under dilute conditions, a spatiotemporal arrangement that promotes the efficient transfer of intermediate substrates, as supported by experimental data [2,3]. Despite this, structural information regarding the MDH-CS complex is notably absent from the Protein Data Bank (PDB), limiting our understanding of its precise architecture.

Furthermore, an increasing body of research on cellular-like environments suggests that experiments conducted in dilute conditions may not fully capture the behavior of proteins in their native, crowded settings [4-7]. This study aims to address this gap by simulating MDH and CS in more complex, crowded environments to offer a more accurate representation of their spatial and functional characteristics. In order to investigate how these enzymes behave in such conditions, we will utilize coarse-grained (CG) simulations. However, before constructing the CG model, it is critical to first explore the dynamics of MDH and CS under dilute conditions. To this end, we have conducted all-atom

simulations across 31 distinct systems, gathering valuable insights to inform the development of our CG approach. To mitigate potential bias from force field selection, we also replicated these simulations using three different force fields.

Sources:

1. Arnold PK, Finley LWS. (2023) Regulation and function of the mammalian tricarboxylic acid cycle. *J Biol Chem.* ;299(2):102838. doi: 10.1016/j.jbc.2022.102838
2. Bulutoglu, B. (2016) Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. *ACS Chem. Bio.* doi: 10.1021/acschembio.6b00523
3. Elcock, A. H. and McCammon, J. A. (1996) Evidence for Electrostatic Channeling in a Fusion Protein of Malate Dehydrogenase and Citrate Synthase. *ACS.* doi: 10.1021/bi9614747
4. Minton, A. P.; The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *Journal Of Biological Chemistry. (2001).* 276(14): 10577–10580.
5. Minton, A. P.; How can biochemical reactions within cells differ from those in test tubes?. *J. Cell Sci. (2006).* 119(14): 2863–2869.
6. Zhou, H.X., Rivas, G., and Minton, A.P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu Rev Biophys. (2008).* 37:375-97.
7. Kasahara, K., *et al.* Reduced efficacy of a Src kinase inhibitor in crowded protein solution. *Nat Commun. (2021).* 12, 4099.

2. Specific usage status of the system and calculation method

In our investigation of the dynamics of Malate Dehydrogenase (MDH) and Citrate Synthase (CS), we conducted all-atom simulations using GENESIS 2.1. Simulations were performed with three different force field combinations: CHARMM36m with the TIP3P water model, AMBER14 with TIP4PD, and AMBER99 with TIP4PEW. The initial structures of MDH and CS were derived from monomeric predictions made by AlphaFold 2, as the structure of Bacillus subtilis MDH is not available in the Protein Data Bank (PDB).

A total of 20 simulation boxes were generated, each simulated for 100 ns. To compare our results with existing literature, we also included simulations of CS and MDH from different species, such as *Sus scrofa* and *Bacillus anthracis*. Specifically, we simulated chains from the following PDB entries: 1CTS (CS from *Sus scrofa*), 2C6X (CS from *Bacillus subtilis*), 1MLD (MDH from *Sus scrofa*), and 3TL2 (MDH from *Bacillus anthracis*). This approach allowed us to examine whether the dynamics observed for Bacillus subtilis MDH and CS align with those in other species and to compare our results with experimental data for different species.

3. Result

The structure of Malate Dehydrogenase (MDH) remained stable throughout the 100-ns simulation. The average Cα-RMSD from the initial structure was consistently below 2 Å. In contrast, Citrate Synthase (CS) exhibited greater dynamics, with an average Cα-RMSD of 6 Å. This increased flexibility can be attributed to the dynamics of its C-terminal segment. Upon careful examination of the available dimeric structure of CS in the PDB, we observed that this segment interacts with another monomer to form a dimer, leading to the hypothesis that this dimeric conformation stabilizes the C-terminal segment.

Additionally, we identified hinge regions adjacent to the binding sites of both CS and MDH, which are likely essential for their functional activity:

- CS: [0–50], [100–110], [170–250], [270–300], [350–]

- MDH: [0–10], [70–100], [200–270]

The binding sites and critical amino acids were located through sequence alignment with previous literature and comparison to sequences from different species. The presence of these hinges is critical for constructing an appropriate coarse-grained (CG) model of these two enzymes. Notably, the dynamics observed in the simulations of both the PDB and predicted structures were similar, reinforcing our hypothesis that this dynamic behavior is crucial for the enzymatic functions of CS and MDH.

４．　Conclusion

In this study, we have explored the dynamics of Malate Dehydrogenase (MDH) and Citrate Synthase (CS) through all-atom simulations, shedding light on their structural stability and flexibility. Our results show that MDH remains stable, while CS exhibits significant dynamics, particularly in its C-terminal segment. This increased flexibility appears to be linked to its dimerization, which stabilizes the C-terminal region. Furthermore, we identified key hinge regions adjacent to the binding sites of both enzymes, which are likely crucial for their functional activities. These insights, derived from simulations of both predicted and PDB structures, are essential for the development of a coarse-grained model.

By conducting sequence alignment with previous literature and different species, we successfully pinpointed the binding sites and important amino acids, providing a more comprehensive understanding of the enzymes' functional mechanisms. Our findings set the stage for further investigation into the interactions between MDH and CS, as well as the binding of their substrates. Ultimately, this work aims to contribute to the construction of an accurate CG model to study the formation of the metabolon in crowded cellular environments, offering a more nuanced understanding of the spatial and functional properties of these enzymes in complex conditions.

５．　Schedule and prospect for the future

We plan to study the interactions between Malate Dehydrogenase (MDH) and Citrate Synthase (CS), as well as the binding of their respective substrates. By incorporating all the structural and dynamic information gathered from the simulations, we aim to gain a deeper understanding of MDH and CS. This will enable us to construct an accurate coarse-grained (CG) model, which is essential for studying the formation of the metabolon in crowded conditions.

MELD を利用した粗視化モデルでのタンパク質 DNA
（担当：水谷）

１．　本課題の研究の背景、目的、関係する課題との関係

転写因子をはじめとする DNA 結合タンパク質と DNA との相互作用は、遺伝子発現を含む生体内の多くのプロセスに関与している。この相互作用を理解するためには、タンパク質による DNA 認識メカニズムを解明することが重要である。タンパク質の構造予測については、AlphaFold2 などの手法により大きな進展があった(K. Tunyasuvunakool, et. al, *Nature*, 2021)。しかし、タンパク質-DNA 複合体の予測には依然として多くの課題が残されている。

DNA 結合部位の予測やターゲット配列の同定など、実験から得られる情報は豊富であり、これらを活用することで予測精度の向上が期待できる。このような実験情報を利用した複合体の予測手法の一つに、Alberto Perez らが開発した MELD（Modeling Employing Limited Data）がある(J. L. MacCallum, et. al, *PNAS*, 2015)。MELD は、分子動力学シミュレーションとベイズ推定を組み合わせることで、シミュレーションを通じて入力された実験データに適合する構造を得る手法である。彼らはすでに、隠溶媒系の全原子モデルを用いた手法により、タンパク質-DNA 複合体の予測を行っている(R. Esmaeeli, et. al., *NAS*, 2023)。しかし、全原子モデルは計算コストが高く、また一部のタンパク質については正確な構造を得ることが難しいという

課題が明らかになった。そこで私たちは、アミノ酸 1 分子を 1 粒子で表現する粗視化モデルと、Cheng Tan らが開発した PWMcos を MELD と組み合わせることで、より高精度かつ計算コストの低い予測を目指した（C. Tan, S. Takada, *JCTC*, 2018）。PWMcos は、多数の DNA 配列に対するタンパク質の特異性を定量化した Position-Weight Matrix（PWM）を活用し、タンパク質-DNA 間の微妙な特異性を表現できる手法である。私たちは、このモデルと MELD を組み合わせることで、タンパク質-DNA 複合体の高精度な予測が可能になると考えている。

２．　具体的な利用内容、計算方法

本課題では研究の前段階としてタンパク質モデル及び PWMcos のパラメータの最適化を行った。本研究ではタンパク質モデルとして AICG2+モデル(W. Li, et. al., PNAS, 2014)を用いている。しかし、このモデルでは図１a で示されているような二量体のシミュレーションを行う際、単量体間の相互作用のパラメータを適切に調整しないと図 1b に示されているように構造が解離する可能性がある。そこで単量体間の相互作用のパラメータを 1.0 から 2.0 まで 0.2 刻みでスケーリングした複数のパラメータセットを作成し、それぞれについて 1.0×10^8 ステップのシミュレーションを実施した。また、PWMcos モデルをて適用するためには２つのパラメータλ及びε’をタンパク質 DNA 間の解離定数 K_d を用いて最適化する必要がある。この最適化のために先の最適化で得られたタンパク質モデルを用いて両方のパラメータを変化させながら 1.0×10^8 ステップのシミュレーションを行い、それぞれの解離定数を計算した。

３．　結果

二量体についてサンプリングを行った結果、1.4 以上ではシミュレーション中で二量体の安定が観察された(図 1c は 1.4 倍にスケーリングした結果)。また、PWMcos のパラメータについてもサンプリングの結果、実験値に合うように最適化された。

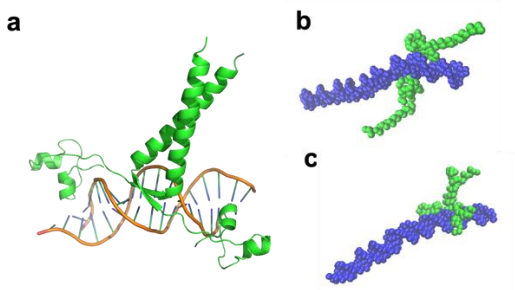

図１ タンパク質PUT3-DNA複合体の天然構造及びシミュレーションのスナップショット (a) PUT3-DNA複合体の構造(PDB ID: 1ZME)(b,c)パラメータ最適化前後でのシミュレーションのスナップショット(緑:タンパク質、青: DNA)

４．　まとめ

構造予測の前段階としてタンパク質、DNA 複合体の系のパラメータの最適化を行った。それぞれについてパラメータを変えてサンプリングを行った結果、実験結果を再現するようなパラメータが得られた。

５．　今後の計画・展望

本計算により、タンパク質モデルと PWMcos の適切なパラメータを得ることができた。しかし、これらを MELD とともに利用するためには PWMcos のポテンシャル関数の改良が必要である。今後は、PWMcos の改良を進めるとともに、MELD との併用による構造予測を実施する予定である。

F_1-ATPase の 40° サブステップ回転のメカニズム解明（担当：本橋）

６．　利用がなかった場合の理由

本年度は令和 6 年度 HPCI システム利用研究課題（ID: hp240047）の資源を利用したため HOKUSAI は利用しなかった。

2024 年度　利用研究成果リスト

【ポスター発表】

1. Haeri Im, Song-Ho Chong, Isseki Yu, and Yuji Sugita, “Impact of glycosylation on the structural features and hydration effects of glycoproteins”, IUPAB2024, 2024.06.24~28, ICC Kyoto, Kyoto, Japan

2. Elisa Rioual, Yangyang Zhang, Erika Fukuhara, Paola Laurino, Yuji Sugita, “MDH – CS enzymes complex: insight of its structure by simulation”, ACS Spring 2025, March 2025, San Diego, USA