

プロジェクト名(タイトル):

## 分子動力学シミュレーションを用いた生体分子の構造および反応機構の解明

利用者氏名:

杉田 有治(1)、○伊東 真吾(1)、大出 真央(1)、大滝 大樹(1、2)、新津 藍(1)、本橋 昌大(1)、八木 清(1)、若林 大貴(1)、Haeri Im(1)、Yaokun Lei(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

(2) 長崎大学 医歯薬学総合研究所 分子標的医学研究センター

Bias-exchange adaptively-biased molecular dynamics 法を用いたアミロイド  $\beta$  前駆体の2量体化プロセス解析

(担当:伊東)

1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

アミロイド  $\beta$  前駆体タンパク質 (APP) は1回膜貫通型タンパク質であり、C99 と呼ばれる 99 個のアミノ酸から  $\alpha$  ヘリックスを形成する領域が膜貫通ドメインとして存在している。C99 は2量体を形成することで安定化することが知られている[N. Miyashita, et. al., J. Am. Chem. Soc., 2009]が(図 1 A)、この  $\alpha$  ヘリックスの2量体化には、(1) Helix-Helix distance、(2) Crossing angle、(3) Crick angle of Helix A、(4) Crick angle of Helix B の4つの反応座標が重要な要素となる(図 1 B)。しかしながら、2量体化エネルギーの計算にこれら4つの反応座標すべてを取り込むことは計算コストの高さから困難であり、2つほどの反応座標を選択し計算を行うか、4次元を1つまたは2つの次元に圧縮する必要があった。このような少ない次元での自由エネルギー計算は、時として実験値と乖離した結果をもたらした。

本研究は最新の bias-exchange adaptively-biased molecular dynamics (BE-ABMD) 法を用いることにより、従来のレプリカ交換アンブレラサンプリング法などでは困難であった4つの反応座標を全て考慮した自由エネルギー計算を行うことで、安定な立体構造の同定および、2つの変異体を通じ膜貫通ドメインのどの部位が構造安定化に寄与しているかについて解明を目指した。

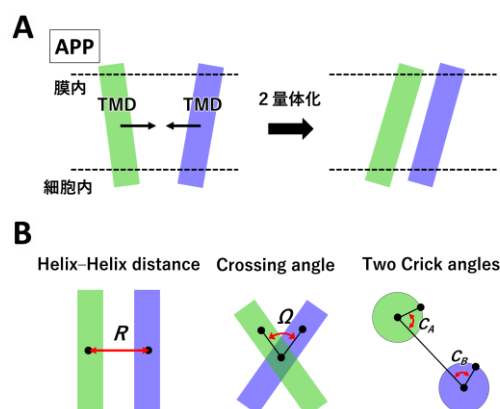


図 1. APP の 2 量体化プロセスと、キーとなる反応座標。

(A) APP-C99 における 2 量体化プロセスの概念図。

(B) 2 量体化に関する 4 つの反応座標の定義。

2. 具体的な利用内容、計算方法

APP-C99 におけるアミノ酸番号 23 番目から 55 番目までの 33 残基に対して、先ほど示した 4 つの反応座標(図 1 B) 全てを含めた自由エネルギー計算を BE-ABMD 法において 1  $\mu$ s/replica 実行することで、反応座標に沿う自由エネルギー地形を解析した。また野生型だけではなく、(1) G29LG33L 変異型および、(2) 39 番目から 52 番目の残基をすべてアラニンに置換した変異体に対しても同様の計算を行い野生型と比較を行った。

3. 結果

本計算により求められた PMF を図 2 A に示す。得られた自由エネルギー地形を解析することで、APP-C99 には Glycine-in、Glycine-side および Glycine-out の 3 つの安定な構造が存在することがわかった(図 2 B)。これは先行研究[L. Dominguez, et. al., Proc. Natl. Acad., 2016]の結果と一致する。また、2 つの変異体との比較により各安定構造の形成には、(1) Glycine zipper

と呼ばれる G29-G33 間の相互作用、(2) G33-L34 および M35-G37 の相互作用、そして (3) G38 から A42 付近の疎水性相互作用が重要な役割を果たすことを突き止めた。

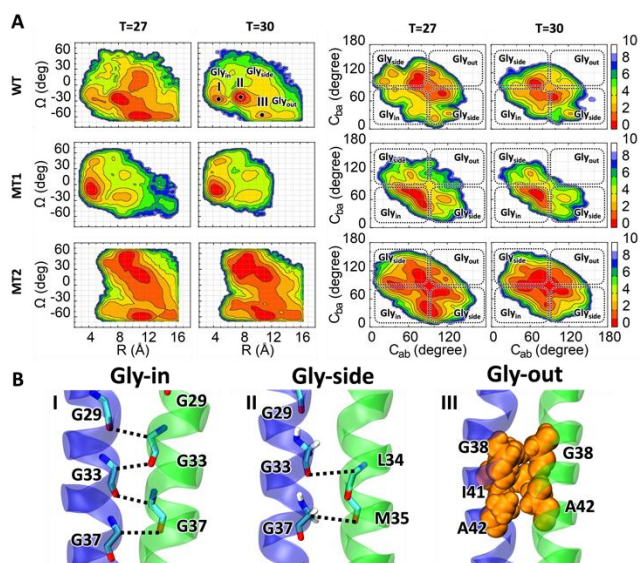


図2. 得られた (A) 自由エネルギー地形および、(B) 3つの安定な立体構造。

#### 4. まとめ

本研究において BE-ABMD 法を用いることにより、初めて 2 量体化のキーとなる 4 つの反応座標全てを取り扱った自由エネルギー計算を行なった。また、本計算により、APP-C99 の 2 量体には 3 つの安定構造があることを明らかにするとともに、それらの形成においてどの残基間相互作用が重要であるかを突き止めることに成功した。

#### 5. 今後の計画・展望

本研究により、BE-ABMD 法が膜貫通タンパク質の多量体化メカニズムの解析に有効な手法であることがわかった。本手法を用いることで、APP-C99 以外の膜貫通タンパク質における多量体化メカニズムの解明に挑戦していきたい。

### 量子化学計算による高輝度な蛍光性分子の探索

(担当：八木)

#### 1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

蛍光性分子は、励起光を吸収し蛍光を発光する分子の総称で、塗料や衣服などに幅広く活用されている身近な物質である。近年、蛍光顕微鏡による細胞イメー

ジングが盛んになり、生命科学研究においても、蛍光性分子は必要不可欠となっている。そのため、新しい光化学的性質を持つ蛍光性分子の発見は様々な化学分野に大きなインパクトをもたらす。しかし、現在、新規蛍光性分子の探索は極めて困難である。

本課題では、量子化学計算により、新しい蛍光性分子を探索するワークフローを確立する。実験と量子化学計算で得られるデータを活用し、与えられた候補分子の光化学的性質を素早く評価する機械学習モデルを構築する。このモデルを用いて、網羅的に生成された分子の中から有力候補分子を絞り込む。バイアスなく候補分子を探索することで、従来の常識では考えられなかった新発見につなげる。

#### 2. 具体的な利用内容、計算方法

本年度は機械学習のベースとなるデータを取得するため、既知の蛍光性分子に対する量子化学計算を実施した。蛍光性分子の鋳型として、

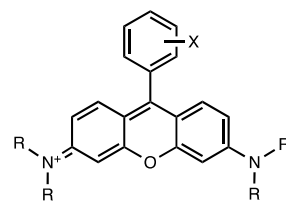


図3. ローダミンの分子式。

3)。置換基やヘテロ原子の

異なる様々なローダミン類、153 個を文献から探した。これらの分子に対し、最安定構造、HOMO-LUMO エネルギー準位、励起エネルギーなどを量子化学計算により求めた。計算レベルには、密度汎関数 (DFT) 法の B3LYP 汎関数と基底関数 cc-pVDZ を用いた。励起エネルギーは時間依存密度汎関数(TD-DFT)法を用いた。プログラムには Gaussian16 を用いた。

#### 3. 結果

153 個のローダミン分子に対する量子化学計算を終えた。一例として、ID10 の結果を図4に示す。ID10 では、アミンの置換基がメチル基となり、ヘテロ原子が Si になっている。さらに、Si にはメチル基とフェニル基がついている。分子を文字化した SMILES でリストを与えると、Gaussian のインプットを生成し、実行する自動化スクリプトを作成した。現在、Gaussian のアウトプットから重要な物性値を抜き出すスクリプトを作成している。

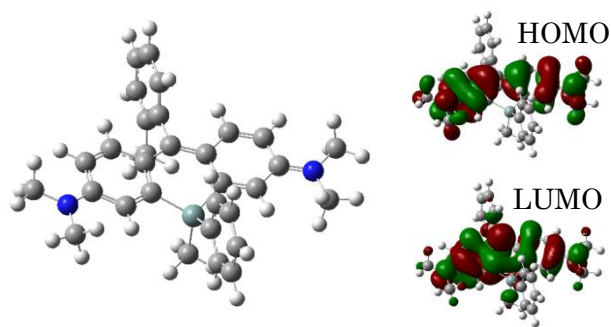


図4. ID10 の構造と HOMO, LUMO。

#### 4. まとめ

蛍光性分子であるローダミン類に対し、既知の分子 153 個を文献から集め、その構造と物性値を量子化学計算により求めた。ワークフローを自動化するスクリプトを作成した。

#### 5. 今後の計画・展望

ローダミン類の光化学的性質に影響を与える物性値を抜き出し、それらを説明変数、蛍光強度を目的変数とする機械学習モデルを構築する。

### Development of accurate and general Machine Learning force field for enzymatic reaction

(担当 : Lei)

1. Background and purpose of the project, relationship of the project with other projects  
Machine learning (ML) method emerges as an efficient and accurate surrogate model of high-level electronic structure theory. However, those ML models have long been restricted to closed chemical systems without being exposed to external potential applied by environment media or artificially. In previous work, we propose to include the high order terms of Taylor expansion of electrostatic operator and use equivariant model which can generate high-order tensors covariant with rotation as base model so that we can use multipole expansion equation to derive useful representation accounting for polarization and intermolecular interaction. However, there are two places to be further improved. Firstly, the truncation of Taylor expansion will result in inferior performance because interaction

between MM atoms near the QM region and QM atoms include higher order terms. Secondly, the high order terms of Taylor expansion from atoms will decrease very quickly so that the contribution of MM atoms at large distance can be overlooked. Considering these two problems, we partition MM atoms into two groups based on a chosen cutoff distance: one group within the cutoff is termed short-range MM atoms (SRMM), while the other one is referred as long-range MM atoms (LRMM). We use message passing/long-range descriptor to capture the interaction between SRMM/LRMM and QM atoms. Moreover, to prevent excessive computation of high-order term of long-range descriptor from MM atoms at large distances, we introduce a range-separated function. This function forces high order terms to decay to 0 at specified distance.

#### 2. Specific usage status of the system and calculation method

We performed 10 picoseconds Replica-Exchange Umbrella Sampling (REUS) simulation with DFTB/MM potential to sample the configurations making up training dataset/test dataset. Then we perform single point energy calculation for all sampled configurations with DFT method. At last, we carried out replica exchange umbrella sampling to calculate the free energy using DFT/MM. We apply this procedure to the following system: 6'-deoxychalcone in water/enzymatic environment, methyl chloride and Fluoride ion in acetonitrile, toluene, water and acetone, 5-nitrobenzisoxazole catalyzed by Kemp eliminase (R1-7/10H variant), a proton transfer reaction of dihydroxyacetone phosphate (DHAP) catalyzed by triosephosphate isomerase (TIM).

#### 3. Result

From Table 1-2, we can see that the performance of new model (MLMM2) is much better than original one (MLMM1) especially for the prediction accuracy of MM force.

Table 1 Mean Absolute Error of model of TIM system.

Model	Energy (kcal/mol)	Force of MM	Force of QM
		Atoms (kcal/mol/Å)	atoms (kcal/mol/Å)
MLMM2	0.121/0.163	0.000486/0.00049	0.159/0.184
MLMM1	0.190/0.358	0.001892/0.00137	0.123/0.212

Train Error/Test Error

Table 2 Mean Absolute Error of model of SN2 reaction in water.

Model	Energy (kcal/mol)	Force of MM	Force of QM
		Atoms (kcal/mol/Å)	atoms (kcal/mol/Å)
MLMM2	0.111/0.272	0.00092/0.00124	0.078/0.208
MLMM1	0.091/0.554	0.00328/0.00405	0.288/0.499

Train Error/Test Error

Furthermore, we examine the transferability of model by evaluating MAE of energy and force on SN2 dataset with different solvent from training dataset. The model is trained only using SN2 dataset (water solution). The results shown in Table 3-4 shows that new model has better transferability.

Table 3 Mean Absolute Error of model of SN2 reaction in toluene.

Model	Energy (kcal/mol)	Force of MM	Force of QM
		Atoms (kcal/mol/Å)	atoms (kcal/mol/Å)
MLMM2	0.5574	0.00285	1.814
MLMM1	1.026	0.00388	1.872

Train Error/Test Error

Table 4 Mean Absolute Error of model of SN2 reaction in acetone.

Model	Energy (kcal/mol)	Force of MM	Force of QM
		Atoms (kcal/mol/Å)	atoms (kcal/mol/Å)
MLMM2	0.3454	0.00193	0.864
MLMM1	0.450	0.00328	0.777

Train Error/Test Error

#### 4. Conclusion

In this work, we adopt a multiscale strategy to treat the interaction between SRMM/LRMM and QM region. The results show that with this multiscale strategy the model's accuracy and transferability has been much improved.

#### 5. Schedule and prospect for the future

Our model is transferrable among various solution especially when solvent has similar polarity with the one used in training dataset. Therefore, it's expected that with extensive sampling along the polarity of solvents our model can achieve good prediction accuracy in reaction in various environment medias. It is the hope that our strategy, along with the efficient sampling method, will enable the creation of next-generation QM/MM potentials for a wide spectrum of free-energy applications, ranging from drug discovery to enzyme design. Therefore, we are inspecting whether our model trained with dataset in simple solvents can perform well in heterogenous system (enzymatic reaction).

#### Investigation of structural properties of erythropoietin based on glycosylation patterns using molecular dynamics simulations.

(担当: Im)

1. Background and purpose of the project, relationship of the project with other projects  
Glycoproteins comprise protein and carbohydrate chains that are crucial for the folding, stability, and biological activity of proteins. Recently, the glycosylation patterns of erythropoietin (EPO), one of the glycoproteins, have been found to be significant for the activity of EPO. However, the exact molecular mechanisms by which glycosylation patterns determine the structural properties of EPO are not fully understood. In this study, we investigate how the glycosylation patterns of EPO impact the structural characteristics and how surrounding water molecules contribute to these properties. We perform all-atom molecular dynamics

(MD) simulations of the EPO structures for different glycosylation patterns and analyze the effects of glycosylation patterns on structural properties and free energy. The results of this study will provide molecular insights into the relationship between glycosylation patterns and glycoprotein structure and function.

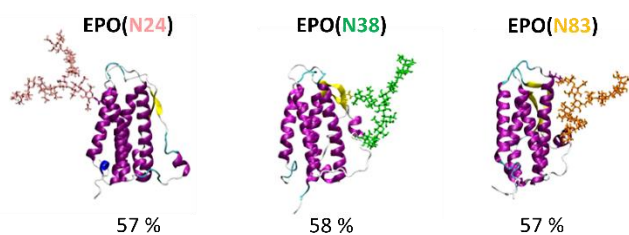
## 2. Specific usage status of the system and calculation method

We performed all-atom MD simulations at 310K and 0.15M NaCl for erythropoietin (EPO) glycoforms. MD simulations were performed for three EPO glycoforms (EPO(N24), EPO(N38), and EPO(N83)).

We employed the CHARMM36m force field for EPO glycoforms and the TIP3P model for water. We used GENESIS to perform three independent MD simulations for 525 ns for each structure, starting from the initial structures using CHARMM-GUI.

## 3. Result

The  $\alpha$ -helical structure in EPO glycoforms remains consistent despite differing glycosylation configurations (Fig. 5). This consistency in  $\alpha$ -helicity underscores fundamental structural stability in EPO glycoforms, irrespective of the glycan arrangements.



**Fig. 5.  $\alpha$ -helicity of EPO glycoforms**

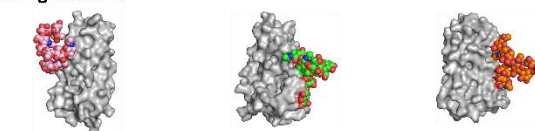
The average solvent accessible surface area (SASA) values of each of the EPO glycoforms are  $13360 (\pm 121) \text{ \AA}^2$  (N24),  $13414 (\pm 75) \text{ \AA}^2$  (N38), and  $13113 (\pm 79) \text{ \AA}^2$  (N83), respectively. In contrast to the stability observed in  $\alpha$ -helicity, the degree of surface exposure in EPO glycoforms varies in response to the glycosylation pattern. Glycoforms with different glycosylation patterns exhibit a diverse range of

solvent exposure, suggesting that glycosylation directly influences the interaction between EPO glycoforms and their surrounding solvent environment.

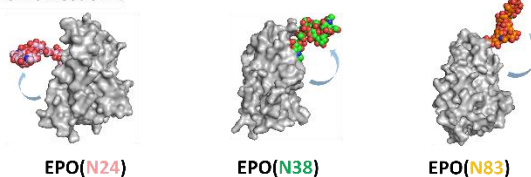
Likewise, the solvation free energy (SFE) of EPO glycoforms also depends on their glycosylation patterns (EPO(N24):  $1544.1 (\pm 33.2) \text{ kcal/mol}$ , EPO(N38):  $1507.8 (\pm 17.1) \text{ kcal/mol}$ , and EPO(N83):  $1552.3 (\pm 37.1) \text{ kcal/mol}$ ). We observed that different glycosylation arrangements result in varying levels of solvation free energy, implying that the position of glycans plays a critical role in determining the hydrophobicity of these glycoforms.

The hydrophobicity of EPO glycoforms is intricately linked to the proximity of glycans to the EPO surface. Glycoforms with glycans closer to the surface exhibit increased hydrophobicity, whereas those with more distantly positioned glycans, encompassing a wider exposed area, display enhanced hydrophilicity (Fig. 6). This correlation between glycan positioning and hydrophobicity offers valuable insights into the functions of EPO glycoforms in various biological contexts.

### Structures of highest SFE:



### Structures of lowest SFE:



**Fig. 6. Structural properties of EPO glycoforms by the tendency of hydrophobicity.**

## 4. Conclusion

Our study of EPO glycoforms using molecular dynamics simulations has revealed that while glycosylation patterns do not affect the  $\alpha$ -helicity, they significantly influence surface exposure and solvation free energy. This indicates a link between glycan positioning and hydrophobic or hydrophilic properties of EPO structures. The findings

demonstrate that glycosylation patterns are crucial in determining the structural stability and hydrophobicity of EPO glycoforms through interactions between glycan and EPO structures.

5. Schedule and prospect for the future

We will perform the same study on systems with more diverse glycosylation patterns of EPO glycoforms. Furthermore, we plan to analyze the contribution of each residue of EPO through thermodynamic analysis.

2023 年度 利用研究成果リスト

【雑誌に受理された論文】

1. S. Ito, K. Yagi, and Y. Sugita, “Allosteric regulation of  $\beta$ -reaction stage I in tryptophan synthase upon the  $\alpha$ -ligand binding” J. Chem. Phys. 158 (2023) 115101.

【口頭発表】

1. 八木清, 「QM/MM 法の開発と酵素反応・光化学反応への応用」, 第 23 回日本蛋白質科学会年会, 2023 年 7 月 6 日, 名古屋.
2. K. Yagi, “Anharmonic Vibrational Calculations of Biomolecules and Polymers Based on Local Coordinates”, TACC2023, Sep. 4-9, 2023 (Sapporo, Japan)
3. 八木清, “Development of non-adiabatic QM/MM molecular dynamics method and applications to light-driven proteins”, 第 61 回日本生物物理学会年会, 2023 年 11 月 14-16 日, 名古屋.

【ポスター発表】

1. S. Ito, K. Yagi, and Y. Sugita, “Allosteric regulation of  $\beta$ -reaction stage I in tryptophan synthase” The 61st Annual Meeting of the Biophysical Society of Japan, November 2023 (Nagoya, Japan)
2. S. Ito, K. Yagi, and Y. Sugita, “Allosteric regulation of  $\beta$ -reaction stage I in tryptophan synthase” CCP2023-34th IUPAP Conference on Computational Physics, August 2023 (Kobe, Japan)
3. S. Ito, K. Yagi, and Y. Sugita, “Allosteric regulation of  $\beta$ -reaction stage I in tryptophan synthase” The 23rd Annual Meeting of the Protein Science Society of Japan, July 2023 (Nagoya, Japan)
4. K. Yagi and Y. Sugita, “Development of high-performance QM/MM-MD methods in GENESIS”, Biomolecules and Nanostructures 8, June 14-18, 2023 (Krakow, Poland)
5. K. Yagi and Y. Sugita, “Development of the hybrid QM/MM method in GENESIS and applications to enzyme reactions”, CCP2023, Aug. 4-8, 2023 (Kobe, Japan)
6. 八木清, 伊東真吾, 「大規模 QM/MM 分子動力学計算に基づく酵素反応の自由エネルギー地形探索」, 第 10 回 HPCI 成果報告会, 2023 年 10 月 25, 26 日, 東京.
7. Y. K. Lei, K. Yagi, and Y. Sugita, “Learning QM/MM Potential by Equivariant Multiscale Model” The 61st Annual Meeting of the Biophysical Society of Japan, November 2023 (Nagoya, Japan)
8. Y. K. Lei, K. Yagi, and Y. Sugita, “Learning QM/MM Potential by Equivariant Multiscale Model” CCP2023-34th IUPAP Conference on Computational Physics, August 2023 (Kobe, Japan)
9. Haeri Im, Song-Ho Chong, and Yuji Sugita, "Investigating the Impact of Glycosylation Patterns on the Structural Characteristics of Erythropoietin Using Molecular Dynamics Simulations", CCP2023-34th IUPAP Conference on Computational Physics, August 2023 (Kobe, Japan)
10. Haeri Im, Song-Ho Chong, and Yuji Sugita, "Exploring the Structural Characteristics of Erythropoietin Based on Glycosylation Patterns Using Molecular Dynamics Simulations", The 61st Annual Meeting of the Biophysical Society of Japan, November 2023 (Nagoya, Japan)