

平成28年度医療研究開発推進事業費補助金  
(創薬等ライフサイエンス研究支援基盤事業) 補助事業成果報告書

## I. 基本情報

事業名：創薬等ライフサイエンス研究支援基盤事業（創薬等支援技術基盤プラットフォーム事業）  
Platform Project for Supporting Drug Discovery and Life Science Research  
(Platform for Drug discovery, Informatics, and Structural life science)

補助事業課題名：(日本語) タンパク質の複合体構造・相互作用に関する総合的な予測・解析の実施と  
高度化（超分子モデリングと重要アミノ酸残基の推定）  
(英語) Prediction and analysis of protein interactions and protein complex  
structures (Supra-molecule modeling and inference of important  
residues)

補助事業担当者 (日本語) 量子科学技術研究開発機構 量子ビーム科学研究部門  
グループリーダー 河野秀俊

所属 役職 氏名： (英語) Hidetoshi Kono, Group Leader, Quantum Beam Science Directorate,  
National Institutes for Quantum and Radiological Science and  
Technology

実施期間： 平成 28 年 4 月 1 日 ～ 平成 29 年 3 月 31 日

## II. 成果の概要（総括研究報告）

### 原子モデル構築技術の高度化

電子顕微鏡画像や電子顕微鏡立体構造（3DEM）、それに SAXS（X 線小角散乱）や SANS（中性子線小角散乱）などの実験データをもとにして、生体分子の原子モデルの構築を行うための様々な計算機手法を開発した。これらの計算機手法には、大きく分けて 2 種類ある。一つは、X 線結晶構造などを初期構造にして、多数の変形構造を構築する計算機手法。そして、もう一つは、多数の変形構造の中から、実験データを再現するような構造を選ぶ計算機手法である。前者に関しては、様々な異なる種類の分子に対して適した手法を開発したり、ユーザーの負担を減らすため、自動化を進めたりした。後者に関しては、並列化と適切な最適化アルゴリズムを用いることによ

り、計算の高速化を行った。開発した計算方法の一部は、Scientific Reports (Matsumoto et al. 2017) に発表した。

### 分子構造変形技術を用いた支援

独自に開発した分子構造変形技術と、電子顕微鏡像やSAXS、SANSなどの低分解能実験データへの原子モデル構築あてはめ技術を用いて、分子のさまざまな状態での原子モデルの構築支援を行った。支援開始時に、X線結晶構造解析によって分子の全体構造が分かっている例はなく、既知の部分構造情報や、アミノ酸配列情報、DNA塩基配列情報を統合的に利用して、低分解能実験データを再現するような分子の全体構造の構築を行った。結果、世界に先駆けて、巨大カドヘリン分子の立体構造モデル(Tsukasaki et al., Proc. Natl. Acad. USA, 2014)、およびヌクレオソーム2量体の立体構造モデル(Kato et al. Science, 2017) を提唱することができた。

### タンパク質-DNA相互作用の解析と分子パッキング設計の計算技術の高度化

タンパク質とDNAの複合体形成において、DNA自体のもつ物性が重要であることが分かってきた。そこで、多様なDNA配列に対してMD計算を実施し、その構造及びダイナミクスデータを機械学習することで、任意のDNA配列にして、構造及びダイナミクスを推定する方法を開発した(Andrabi et al. Scientific Reports, 2017)。また、アミノ酸や塩基の変異による分子の動構造の影響を解析できる方法 (Linear Discriminant Analysis 法)を開発した (Sakuraba and Kono, J. Chem. Phys. 2016)。この方法を用いることで、分子の構造揺らぎの変化が、変異によるものか分子本来の揺らぎなのかを見極めることができるようになった。

また、結晶構造の分解能向上やリガンド結合部位が結晶内分子パッキングによって潰れていない構造の取得を目的に、結晶内分子パッキングの改変する方法の開発に挑んだ。タンパク質の主鎖構造に適合するアミノ酸配列セットを計算するプログラムをベースにして、さまざまな相対配置をとったタンパク質 (結晶内分子パッキングに対応) に適合するアミノ酸配列セットを計算できるようにした。また、結晶系に応じた分子配置の対称性を考慮することで、計算の手数を大幅に削減した。分子の相対的な回転について、膨大な数の組合せ数からできるだけ均等に相対的な配置をサンプリングできるようにアルゴリズムを改良し、均一なサンプリングを実現した。同時に、マスターワーカールゴリズムを導入し、プログラムの並列化、計算終了条件の改良を行い、計算の高速化 (CPUに比例した高速化を実現) を達成した。

### タンパク質-DNA相互作用解析と分子パッキング改変の支援

DNA結合タンパク質TALは塩基配列特異的にDNAに結合するが、どの残基が特異性を出すのか結合実験のみからはわからない。そこで、TALタンパク質とDNAの複合体のモデルを構築、分子動力学計算、構造バイオインフォマティクス解析を行い、特異的結合を担っているアミノ酸残基の特定を支援した (Tujii et al. BBRC, 2013)。同様な方法で、肝がん特有にみられるDNA結合タンパク質の変異の影響の解析を支援した。また、高度化によって開発した、DNA配列に対する動特性を予測するプログラム DynaSeq を用いて、1312種類の転写因子の結合サイトの解析を支援した。結果、転写因子の結合サイトのの上流、下流の200bpにも有意な情報量が存在し、それらの領域を考慮することで、予測精度は向上することが分かった。(Andrabi et al. Scientific Reports, in press, 2017)。

結晶内分子パッキングの改変の支援では、 実際のタンパク質に適用し、10個の変異体の提案

を行い、結晶構造解析を行った。変異を入れたタンパク質の結晶構造解析に成功したが、分解能は変異前と変わらなかった。変異サイトのパッキング改善が見られたものの、周辺のパッキングが変化したことによって、結果的に分解能が同程度になったと推定された。

### **Advancement of atomic model construction technology**

We have developed many computational methods for constructing atomic models of biomolecules based on experimental data such as electron microscope images, electron microscope three-dimensional structure (3DEM), SAXS (small angle X-ray scattering) and SANS (small angle neutron scattering). There are roughly two types in these computational methods. One is a method to construct numerous deformed structures from the initial structure such as X-ray crystal structure. The other is a method to select a structure that reproduces experimental data from among many deformed structures. Regarding the former, we have developed appropriate methods for different kinds of molecules. Also, in order to reduce the burden of users, we have proceeded with automation. With regard to the latter, we have improved the computational speed by using parallelization and an appropriate optimization algorithm. A part of the computational methods was reported in Scientific Reports (Matsumoto et al. 2017).

### **Support for molecular modeling using molecular structure deformation technology**

We have supported the construction of atomic models of molecules in various functional states using our original computational techniques for building deformed atomic models and for fitting the atomic models to low resolution experimental data, such as electron microscope images, SAXS, and SANS. In any case, the whole structure of the molecule was not known by the X-ray crystal structure analysis at the start of the support. Thus, by comprehensively using the known partial structure information, the amino acid sequence information, and the DNA base sequence information, we constructed the whole structures of the molecules that well reproduced the low resolution experimental data. As a result, we were able to propose the three dimensional structural models of the giant cadherin molecules (Tsukasaki et al. Proc. Natl. Acad. USA, 2014) and of the nucleosome dimer (Kato et al. Science, 2017).

### **Development of protein-DNA interaction method and molecular packing improvement method**

The knowledge on the importance of physico-chemical properties has been accumulated, however, it is not well utilized in protein-DNA interaction analysis. We carried out molecular dynamics simulations on various DNA sequences and derived structural and dynamical parameters using a machine learning technique. Using the parameters, we developed a method (DynaSeq) for characterizing DNA sequence-dependent conformation and dynamics (Andrabi et al., Scientific Reports, 2017). In addition, we developed a computational method (Linear Discriminant Analysis method) which can detect differences between two trajectories from MD simulations carried out under different conditions (Sakuraba and Kono, J. Chem. Phys. 2016). This method allows us to evaluate the effect caused by simulation conditions such as mutation or with/without a ligand.

We also tried to develop a method which improves molecular packing to achieve a higher resolution analysis and alters molecular packing in a crystal so that a ligand binding region can be exposed to the solvent. Based on an algorithm we developed for protein design (Kono and Saven, J. Mol. Biol. 2001), we calculated packing energies over changing relative orientations and relative distances between molecules in a crystal unit. By incorporating the master-worker algorithm, the calculation now can be carried out in parallel and efficiently.

### Support for protein-DNA interactions analysis and for molecular packing design

A DNA-binding protein, TAL binds to DNA specifically, but it was not known which residues are responsible for the binding. Using molecular modeling and molecular dynamics simulations, we supported to detect key residues (Tujii et al. BBRC, 2013). In a similar way, we also supported to examine the effect of amino acid mutations of a DNA-binding protein which are often observed in hepatoma.

We supported analysis of binding sites of 1312 transcription factors using DynaSeq we developed. The data showed that the physico-chemical properties of DNA sequences are significantly different not only at the binding sites but also 200 bp upstream or downstream of the binding sites, indicating that the binding site prediction can be improved by considering a longer DNA sequence than currently used (Andrabi et al, Scientific Reports, in press, 2017).

In support for crystal packing improvement, we proposed 10 mutants each of which has one mutation. Among them, we prepared several mutants to crystallize them. We successfully solved a crystal structure for one of them at the same resolution as the wild type. The packing was improved at the mutation site, but packings for the surrounding residues were perturbed. We concluded that this was why the resolution remained at the same level of the wild type.

### III. 成果の外部への発表

(1) 学会誌・雑誌等における論文一覧（国内誌 0 件、国際誌 6 件）

1. Tsujii, S., Futaki, S., Imanishi, M., Creating a TALE protein with unbiased 5'-T binding, *Biochemical and Biophysical Research Communications*, 2013, **441**, 262-265.
2. Tsukasaki Y., Miyazaki N., Matsumoto A., Nagae S., Yonemura S., Tanoue T., Iwasaki K., and Takeichi M. Giant cadherins Fat and Dachous self-fold to organize properly spaced intercellular junctions, *Proc. Natl. Acad. Sci. U.S.A.*, 2014, **111**, 16011-16016.
3. Sakuraba, S.; Kono, H. Spotting the difference in molecular dynamics simulations of biomolecules. *J. Chem. Phys.* 2016, 145, 074116.
4. Matsumoto, A., Miyazaki, N., Takagi, J. & Iwasaki, K. 2D hybrid analysis: Approach for building three-dimensional atomic model by electron microscopy image matching, *Scientific Reports*, 2017, **7**, 377.
5. Kato, D.; Osakabe, A.; Mizukami, Y.; Adachi, F.; Arimura, Y.; Saikusa, K.; Akashi, S.; Nishimura, Y.; Park, S.-Y.; Matsumoto, A.; Kono, H.; Inoue, R.; Sugiyama, M.; Kurumizaka, H. Crystal structure of the overlapping dinucleosome composed of hexasome and octasome. *Science* 2017, **356**, 205-208.
6. Andrabi, M.; Hutchins, A. P.; Miranda-Saavedra, D.; Kono, H.; Nussinov, R.; Mizuguchi, K.; Ahamd, S. Genome-wide transcription factor activities are explained by intrinsic conformational dynamics of binding-sites and distal flanking-regions. *Scientific Reports*, Accepted.

(2) 学会・シンポジウム等における口頭・ポスター発表

ポスター

1. A New Method for Evaluating the Specificity of Indirect Readout in Protein - DNA Recognition. Yamasaki S, Terada T, Kono H, Shimizu K, & Sarai A *European Conference on Computational Biology 2012* (2012/09) 国外.
2. Atomic model building from low resolution electron microscope images of integrin. Matsumoto A, Iwasaki K, & Takagi J, *Biophysical Society 57th Annual Meeting*. (2013/02) 国外.
3. Superimposition of a crystal structure over EM images. Iwasaki K, Matsumoto A, & Takagi J *Keystone Symposia "Structural Analysis of Supramolecular Assemblies by Hybrid Methods"* (2013/03) 国外.
4. A new approach to build 3D atomic model from single electron microscope image. Matsumoto A, Takagi J, & Iwasaki K 第51回日本生物物理学会年会 (2013/10) 国内.
5. 2D hybrid analysis:A new approach to build 3D atomic model from 2D EM image. Matsumoto A, Takagi J, & Iwasaki K 第52回日本生物物理学会年会 (2014/09) 国内.
6. Designing a new artificial transcription factor based on engrailed homeodomain. Sunami T & Kono H 第52回日本生物物理学会年会 (2014/09) 国内.
7. 2Dハイブリッド法による2D電顕画像からの3D原子モデル構築. 松本淳, 高木淳一, 岩崎憲治 日本顕微鏡学会第71回学術講演会 (2015/05) 国内.
8. Decomposition of averaged EM images by 2D hybrid analysis. Matsumoto A, Takagi J, & Iwasaki K. 第54回日本生物物理学会年会 (2016/11) 国内.
9. 生体分子MDの間違い探し. 櫻庭俊、河野秀俊 第28回分子シミュレーション討論会 (2014/11) 国内.
10. エングレイルドホメオドメインを用いた新たな転写因子の設計、角南智子、河野秀俊、第15回日本蛋白質科学会年会 (2015/06) 国内.
11. エングレイルドホメオドメインを用いた新たな転写因子の設計. 角南智子、河野秀俊、第38回日本分子生物学会年会/第88回日本生化学会大会合同大会(BMB 2015) (2015/12) 国内.
12. エングレイルドホメオドメインを用いた新規転写因子設計. 角南智子、河野秀俊、第16回日本蛋白質科学会年会 (2016/06) 国内.

口頭

13. 二次元 EM イメージへの原子構造のフィッティング. 岩崎憲治, 高木淳, 松本淳, *The 69th Annual Meeting of Japanese Society of Microscopy*. (2013/05) 国内.
14. 立体構造から探るタンパク質-DNA 認識機構. 河野秀俊 2013 年年会生命医薬情報学連合大会 (2013/10) 国内.
15. Understanding Protein-DNA Recognition by Structural Bioinformatics and Molecular Dynamics Simulation. Kono H., *Second BMIRC International Symposium on Advances in Bioinformatics and Medical Engineering* (2014/01) 国内.
16. Atomic model building from low resolution electron microscopy images. Matsumoto A, Takagi J, & Iwasaki K, *The 14th Annual Meeting of the Protein Science Society of Japan* (2014/06) 国内

17. 2D hybrid analysis: A new approach to build 3D atomic model from 2D EM image, Matsumoto, A. 大阪大学蛋白質研究所セミナー (2015/3) 国内
18. 2D hybrid analysis; An Approach to build 3D atomic model from 2D EM image. Matsumoto A 第53回日本生物物理学会年会 (2015/09) 国内
19. 転写因子のアクティビティは結合サイト及びその周辺領域の DNA 特性の関係で可変、河野秀俊、第16回日本蛋白質科学会年会 (2016/06) 国内.
20. 細胞内超分子構造から、細胞外蛋白質の構造まで. 岩崎憲治, 宮崎直幸, 松本淳 第16回日本蛋白質科学会年会 (2016/06) 国内.

(3) 「国民との科学・技術対話社会」に対する取り組み

1. 松本淳, 石田恒, 河野秀俊 (2014/12) 電子顕微鏡像など低分解能構造データに適合する原子モデルの構築. PDIS (創薬等支援技術基盤プラットフォーム) ワークショップ、実験とバイオインフォマティクスの協奏による構造生命科学, お台場、東京, 国内
2. 河野秀俊 (2016/09/06) 構造多形をもつ生体高分子の全原子モデルの構築. JSBi 関西地域部会 第21回バイオメディカル研究会, 大阪、国内

(4) 特許出願

なし