（様式 10）

【16am0101047j0005】

平成 29 年　5 月 26 日

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）実験データを取り入れたフレキシブル・ドッキングによる
タンパク質複合体解析パイプラインの構築、支援と高度化
（英　語）Flexible docking pipeline using experimental restraints

補助事業担当者　（日本語）大阪大学　微生物病研究所　教授　Daron Standley
所属 役職 氏名：（英　語）Daron Standley, Professor,
Research Institute for Microbial Diseases, Osaka University

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）

【和文】

創薬等ライフサイエンス研究支援基盤事業において、「支援」と「高度化」を調和的に行い、創薬に直接貢献することを目指してきた。我々の最初の提案内容と支援の必要性に基づき、いくつかの領域における高度化に注力した。(1) 実験から得られた制約条件を考慮したタンパク質のドッキング、(2) タンパク質-核酸相互作用の予測、(3) 原子レベルの解像度の B 細胞受容体のモデリング、(4) 1-3 と、他の領域のバイオインフォマティクス研究のための、多重配列アラインメントの改良である。

(1) 実験から得られた制約を考慮したタンパク質ドッキング

In silico 変異の計算に基づくフレキシブルドッキングパイプラインを開発し、in vitro のタンパク質間相互作用を再現した。このパイプラインを、$TRAM_2TRIF$ 三量体複合体 (Enokizono et al. 2013)、Regnase-1 のドメイン間相互作用 (Yokogawa et al 2016)、JAK2 キナーゼ (Shan et al. 2014)、JMJD5 と HBx の結合 (Kouwaki et al. 2016) に適用した。さらに、いくつかの挑戦的な問題に、MD 計算を用いたフレキシブルリガンドドッキングを適用した (Kamikawa et al., 2015, Oshima et al., 2015)。

(2) タンパク質-核酸相互作用の予測

タンパク質の立体構造座標から取り出した構造情報を用いて核酸結合座位を予測する方法、aaRNA を開発した (Li et al. 2014)。この方法は平均 85%の予測精度を実現し、この値は既存の方法に比べて約 4%高い。同様の構造情報を用いて、RNA から DNA 結合予測に拡張したもの (aaDNA) も、高い性能を示した。免疫に関連する RNA 結合タンパク質に aaRNA を適用した後、予測された結合残基は後に実験で検証された残基とよく一致した (Masuda et al. 2016; Mino et al. 2015)。タンパク質への核酸のフレキシブルなドッキングに対応するために、aaRNA/aaDNA と共に疎視化分子動力学 (CGMD) 計算を用いた新規手法を開発した。この方法を適用した結果は、多くの系で実験的に実証された。例えば、免疫細胞の活性化に関与する AUF-1 (AU-rich element RNA-binding protein 1) (Nyati et al NAR 2016) や、全身性紅斑性狼瘡 (SLE) における抗一本鎖 DNA 抗体 (Sakakibara et al, submitted)、HIV-1 ゲノム RNA と APOBEC3G の相互作用 (Fukada et al, submitted) などである。

(3) 原子レベル解像度の B 細胞受容体のモデリング

2014 年に、Pfizer、Johnson and Johnson、Scripps Research Institute が行った AMA-II 抗体構造モデリングコンテストに招待され、我々はアステラス製薬、大阪大学蛋白質研究所の研究者からなるチームを組織した。競合するグループ (Schrodinger, Inc., Rosetta, MOE など) は強力であったが、我々のチームが、CDRH-3 (三番目の相補性決定領域) の平均エラーにおいて全体的に最も正確なモデルを与えた (Shirai et al. 2014)。CDRH-3 は抗体のモデリングにおいて最も困難で重要な部分である。続いて、完全に自動化したサーバ、KOTAI Antibody Builder を開発した。これは、AMA-II で用いた半自動的な方法に比べて精度を落とすことなく、数時間で B 細胞受容体配列からモデルを構築することができる (Yamashita et al. 2014)。2016 年に、BCR と TCR の両方に適用可能で、数秒で実行でき、精度を落とさない方法を開発した (Schritt et al, 投稿準備中)。

(4) 多重配列アラインメント

多重配列アラインメントプログラム MAFFT の開発を継続している。最近はより大きなデータを扱うことが必要になっている。また、配列決定技術や、タンパク質コード領域推定手法の限界から、データの質の問題が生じている。すなわち、低品質かつ大規模な配列処理が必要になっている。これらの問題に、(i) 既存データの再利用 (Katoh & Frith 2012; Katoh & Standley 2013)
(ii) 相同な配列に混入した類似性の低い領域をアラインメントから排除するスコアシステム (Katoh & Standley 2016) によって対応することを提案した。(i) について、本プロジェクトにおいては既知の BCR や TCR の配列の多重アラインメントをあらかじめ用意しておくことができ、新しい配列のみを追加することによって高速な解析が可能となった。これらの最近の改良などにより、MAFFT に対する引用数は、2007 年の 228 回/年から 2015 年の 2558 回/年に増加した。累計引用数は、2016 年 11 月現在で、14,000 回である。MAFFT ウェブサービスの普及も進み、ユニークユーザの数は週あたり 3000 に達しつつある。

【英文】

In the PDIS project, we have tried to harmonize “support” and “sophistication” in ways that would directly benefit drug discovery. Based on our initial proposal and on support demands, we focused our sophistication efforts in several areas: (1) Protein docking with experimental constraints; (2) predicting protein-nucleotide interactions; (3) modeling B cell receptors to atomic resolution; (4) improving multiple alignments, which are necessary for topics 1-3, as well as other areas of bioinformatics.

(1) Protein docking with experimental constraints

We developed a flexible docking pipeline based on in silico mutagenesis calculations that reproduced in vitro protein-protein interactions. We applied the pipeline to the trimeric complex of $TRAM_2TRIF$ (Enokizono et al., 2013), Regnase-1 domain-domain interactions (Yokogawa et al., 2016), JAK2 kinase (Shan et al., 2014) and JMJD5-HBx binding (Kouwaki et al., 2016). We also carried out flexible ligand docking for several challenging systems using molecular dynamics (Kamikawa et al., 2015; Oshima et al., 2015).

(2) Predicting protein-nucleotide interactions

We developed tools for prediction of nucleotide binding sites in proteins using structural features extracted from 3D coordinates of proteins. The resulting tool called aaRNA achieved 85% predication accuracy on average, which was ~4% better than what had been reported previously (Li et al., 2014). The extension from RNA- to DNA-binding prediction using the same structural features (aaDNA) gave similarly satisfying results. After applying aaRNA to immune-related RBPs, predicted binding residues were often in agreement with those verified later by experiments (Masuda et al., 2016; Mino et al., 2015). To address the problem of flexibly docking nucleotides to protein structures, we developed a novel method using coarse-grained molecular dynamics (CGMD) simulations with aaRNA or aaDNA as an attractive potential. Applications of our method were verified experimentally for a range of systems, including the AU-rich element RNA-binding protein 1 (AUF-1) responsible for immune cell activation (Nyati et al NAR 2016), a single-stranded DNA targeting antibody in systemic lupus erythematosus (SLE) (Sakakibara et al, submitted), and APOBEC3G interacting with HIV-1 genomic RNA (Fukada et al submitted).

(3) Modeling B cell receptors to atomic resolution

In 2014, our lab was invited to join the AMA-II antibody structural modeling contest sponsored by Pfizer, Johnson and Johnson and the Scripps Research Institute. Our team included researchers from Astellas Pharma and the Institute for Protein Research at Osaka University. In spite of stiff competition (Schrodinger, Inc., Rosetta, MOE etc.), our team produced the most accurate models overall: average errors in the third heavy chain complimentary region (CDRH-3), which is the most difficult and important part of an antibody to model, were lower than in any other group (Shirai et al., 2014). We subsequently developed KOTAI Antibody Builder, a fully automated server that can model B cell receptor sequences in hours with no loss in accuracy compared to the semi-automated method used in AMA-II (Yamashita et al 2014). In 2016, we developed an accelerated version of the software that can render models for both BCRs and TCRs within seconds with no loss in accuracy (Schritt et al, in prep).

(4) Multiple sequence alignment

We are continuously developing the popular multiple sequence alignment (MSA) program, MAFFT. Along with advances in sequencing technologies, the need for handling larger data is increasing. There is also a problem with

data quality, due to limitations in current sequencing technologies and identification of protein-coding regions. Thus it is becoming necessary to handle large low-quality data in MSAs and other sequence analyses. We have taken two steps to remedy this problem in MAFFT: (i) Reuse of pre-compiled data (Katoh and Frith, 2012; Katoh and Standley, 2013); (ii) A new scoring system to automatically exclude unrelated regions in otherwise homologous sequences (Katoh and Standley, 2016). Regarding the first point, in many cases, a reliable MSA of known sequences is available and one needs only to identify the relationship of new sequences to the known sequences (such as is the case for BCRs and TCRs). Due in part to these improvements, the number of citations per year for MAFFT has increased from 228 in 2007 to 2558 in 2015. The cumulative number of citations is about 14,000 as of Nov. 2016. The MAFFT web server is also very popular: The number of unique users is approaching 3000 per week.

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　　件、国際誌　　件）

1. Yokogawa M, Tsushima T, Noda NN, Kumeta H, Enokizono Y, Yamashita K, Standley DM, Takeuchi O, Akira S, Inagaki F. Structural basis for the regulation of enzymatic activity of Regnase-1 by domain-domain interactions. Sci Rep 2016, 6:22324.
2. Masuda K, Ripley B, Nyati KK, Dubey PK, Zaman MM, Hanieh H, Higa M, Yamashita K, Standley DM, Mashima T, et al. Arid5a regulates naive CD4+ T cell fate through selective stabilization of Stat3 mRNA. J Exp Med 2016, 213:605-619.
3. Kouwaki T, Okamoto T, Ito A, Sugiyama Y, Yamashita K, Suzuki T, Kusakabe S, Hirano J, Fukuhara T, Yamashita A, Saito K, Okuzaki D, Watashi K, Sugiyama M, Yoshio S, Standley DM, et al. Hepatocyte Factor JMJD5 Regulates Hepatitis B Virus Replication through Interaction with HBx. J Virol 2016, 90:3530-3542.
4. Katoh K, Standley DM. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. Bioinformatics 2016, 32:1933-1942.
5. Kamikawa Y, Hori Y, Yamashita K, Jin L, Hirayama S, Standley DM, Kikuchi K. Design of a protein tag and fluorogenic probe with modular structure for live-cell imaging of intracellular proteins. Chemical Science 2016, 7:308-314.
6. Onishi M, Ozasa K, Kobiyama K, Ohata K, Kitano M, Taniguchi K, Homma T, Kobayashi M, Sato A, Katakai Y, Yasutomi, Y. Wijaya E, Igarashi Y, Nakatsu N, Ise W, Inoue T, Yamada H, Vandenbon A, Standley DM, et al. Hydroxypropyl-beta-Cyclodextrin Spikes Local Inflammation That Induces Th2 Cell and T Follicular Helper Cell Responses to the Coadministered Antigen. J Immunol 2015, 194:2673-2682.
7. Ohshima J, Sasai M, Liu J, Yamashita K, Ma JS, Lee Y, Bando H, Howard JC, Ebisu S, Hayashi M, Takeda K, Standley DM, et al. RabGDIalpha is a negative regulator of interferon-gamma-inducible GTPase-dependent cell-autonomous immunity to Toxoplasma gondii. Proc Natl Acad Sci U S A 2015, 112:E4581-4590.

8. Mino T, Murakawa Y, Fukao A, Vandenbon A, Wessels HH, Ori D, Uehata T, Tartey S, Akira S, Suzuki Y, Vinuesa CG, Ohler U, Standley DM, et al. Regnase-1 and Roquin Regulate a Common Element in Inflammatory mRNAs by Spatiotemporally Distinct Mechanisms. Cell 2015, 161:1058-1073.
9. Yamashita K, Ikeda K, Amada K, Liang S, Tsuchiya Y, Nakamura H, Shirai H, Standley DM. Kotai Antibody Builder: automated high-resolution structural modeling of antibodies. Bioinformatics 2014, 30:3279-3280.
10. Vandenbon A, Teraguchi S, Takeuchi O, Suzuki Y, Standley DM. Dynamics of enhancers in myeloid antigen presenting cells upon LPS stimulation. BMC Genomics 2014, 15 Suppl 10:S4.
11. Tartey S, Matsushita K, Vandenbon A, Ori D, Imamura T, Mino T, Standley DM, Hoffmann JA, Reichhart JM, Akira S, et al. Akirin2 is critical for inducing inflammatory genes by bridging IkappaB-zeta and the SWI/SNF complex. EMBO J 2014, 33:2332-2348.
12. Takemura N, Kawasaki T, Kunisawa J, Sato S, Lamichhane A, Kobiyama K, Aoshi T, Ito J, Mizuguchi K, Karuppuchamy T, Matsunaga K, Miyatake S, Mori N, Tsujimura T, Satoh T, Kumagai Y, Kawai T, Standley DM et al. Blockade of TLR3 protects mice from lethal radiation-induced gastrointestinal syndrome. Nat Commun 2014, 5:3492.
13. Shirai H, Ikeda K, Yamashita K, Tsuchiya Y, Sarmiento J, Liang S, Morokata T, Mizuguchi K, Higo J, Standley DM, et al. High-resolution modeling of antibody structures by a combination of bioinformatics, expert knowledge, and molecular simulations. Proteins 2014, 82:1624-1635.
14. Morikawa H, Ohkura N, Vandenbon A, Itoh M, Nagao-Sato S, Kawaji H, Lassmann T, Carninci P, Hayashizaki Y, Forrest AR, Standley DM, et al. Differential roles of epigenetic changes and Foxp3 expression in regulatory T cell-specific transcriptional regulation. Proc Natl Acad Sci U S A 2014, 111:5289-5294.
15. Li S, Yamashita K, Amada KM, Standley DM. Quantifying sequence and structural features of protein-RNA interactions. Nucleic Acids Res 2014, 42:10086-10098.
16. Lensink MF, Moal IH, Bates PA, Kastritis PL, Melquiond AS, Karaca E, Schmitz C, van Dijk M, Bonvin AM, Eisenstein M, Jimenez-Garcia B, Grosdidier S, Solernou A, Perez-Cano L, Pallara C, Fernandez-Recio J, Xu J, Muthu P, Praneeth Kilambi K, Gray JJ, Grudinin S, Derevyanko G, Mitchell JC, Wieting J, Kanamori E, Tsuchiya Y, Murakami Y, Sarmiento J, Standley DM, et al. Blind prediction of interfacial water positions in CAPRI. Proteins 2014, 82:620-632.
17. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. Methods Mol Biol 2014, 1079:131-146.
18. Hall D, Li S, Yamashita K, Azuma R, Carver JA, Standley DM. RNA-LIM: A Novel Procedure for Analyzing Protein/Single-Stranded RNA Propensity Data with Concomitant Estimation of Interface Structure. Anal Biochem 2014.
19. Hall D, Li S, Yamashita K, Azuma R, Carver JA, Standley DM. A novel protein distance matrix based on the minimum arc-length between two amino-acid residues on the surface of a globular protein. Biophys Chem 2014.
20. Furukawa Y, Teraguchi S, Ikegami T, Dagliyan O, Jin L, Hall D, Dokholyan NV, Namba K, Akira S, Kurosaki T, Baba Y, Standley DM Intrinsic Disorder Mediates Cooperative Signal Transduction in STIM1. J Mol Biol 2014.

21. Vandenbon A, Kumagai Y, Teraguchi S, Amada KM, Akira S, Standley DM. A Parzen window-based approach for the detection of locally enriched transcription factor binding sites. BMC Bioinformatics 2013, 14:26.
22. Uehata T, Iwasaki H, Vandenbon A, Matsushita K, Hernandez-Cuellar E, Kuniyoshi K, Satoh T, Mino T, Suzuki Y, Standley DM, et al. Malt1-induced cleavage of regnase-1 in CD4(+) helper T cells regulates immune activation. Cell 2013, 153:1036-1049.
23. Moretti R, Fleishman SJ, Agius R, Torchala M, Bates PA, Kastritis PL, Rodrigues JP, Trellet M, Bonvin AM, Cui M, Rooman M, Gillis D, Dehouck Y, Moal I, Romero-Durana M, Perez-Cano L, Pallara C, Jimenez B, Fernandez-Recio J, Flores S, Pacella M, Praneeth Kilambi K, Gray JJ, Popov P, Grudinin S, Esquivel-Rodriguez J, Kihara D, Zhao N, Korkin D, Zhu X, Demerdash ON, Mitchell JC, Kanamori E, Tsuchiya Y, Nakamura H, Lee H, Park H, Seok C, Sarmiento J, Liang S, Teraguchi S, Standley DM, et al. Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. Proteins 2013.
24. Liang S, Zheng D, Standley DM, Guo H, Zhang C. A novel function prediction approach using protein overlap networks. BMC Syst Biol 2013, 7:61.
25. Kawasaki T, Takemura N, Standley DM, Akira S, Kawai T. The second messenger phosphatidylinositol-5-phosphate facilitates antiviral innate immune signaling. Cell Host Microbe 2013, 14:148-158.
26. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013, 30:772-780.
27. Hobro AJ, Standley DM, Ahmad S, Smith NI. Deconstructing RNA: optical measurement of composition and structure. Phys Chem Chem Phys 2013, 15:13199-13208.
28. Enokizono Y, Kumeta H, Funami K, Horiuchi M, Sarmiento J, Yamashita K, Standley DM, Matsumoto M, Seya T, Inagaki F. Structures and interface mapping of the TIR domain-containing adaptor molecules involved in interferon signaling. Proc Natl Acad Sci U S A 2013, 110:19908-19913.

（2）学会・シンポジウム等における口頭・ポスター発表

1. 口頭 Standley DM Nov 14, 2015 Australian National University. Protein Structure and Function “A structural view of immune repertoires”国外
2. 口頭 Standley DM Jan 1. 2016 The 9th Adjuvant Research Conference “Next-generation biomarkers from Big Data”国内
3. 口頭 Standley DM March 6th, 2015, Osaka Univ, Institute for Protein Research, Analysis and prediction of protein assembly structures by bioinformatics “NGS-driven docking of protein-RNA complexes” 国内
4. 口頭 Standley DM Sept 26, 2014, Biophysical Society of Japan, “Intrinsic disorder mediates cooperative signal transduction in STIM1”国内
5. 口頭 Standley DM January 14-15, 2014 Antibody Design, Modeling, and Applications “Kotai antibody builder: a pipeline for antibody modeling and design”国内

6.　　口頭　Standley DM August 26th, 2013 Kyoto University, Institute for Viral Research, Colloquium “3D Molecular Modeling of antibodies and protein-RNA interactions 国内

（３）「国民との科学・技術対話社会」に対する取り組み

該当なし

（４）特許出願

Antibody clustering software: patent pending 2016-181520 (2016)