

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) 構造生命科学データクラウドの構築運用と高度化 (構造生命科学データクラウドにおけるタンパク質機能推定法の開発と支援)  
(英 語) Development and enhancement of data cloud for structural biology  
( Development and enhancement of protein function prediction methods in data cloud for structural biology )

補助事業担当者 (日本語) 東北大学大学院情報科学研究科 教授 木下賢吾  
所属 役職 氏名 : (英 語) Kengo Kinoshita, Professor,  
Graduate School of Information Sciences, Tohoku 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)

本プロジェクトでは支援と高度化として以下のような成果を得た。

情報拠点としての支援としては、代表機関 (遺伝研) を中心として開発を行った **VaProS** を核として、東北大拠点が高度化の一環として開発を行った天然リガンドデータベースや遺伝子共発現ネットワークデータベースや低分子結合部位データベースの実践的な講習会を定期的に行った。ここでは、特に、ウェット系の研究者が自ら様々なデータベースを渡り歩けるように、**P3K** と **TPRP** の成果の利用推進として我々が開発を行ったデータベースとして、機能部位の分子表面データベース (**eF-site**)、類似性検索による機能部位予測 (**eF-seek**)、遺伝子共発現データベース (**COXPRESdb**, **ATTEDII**) 等も含めた幅広い紹介を行った。また、蛋白質科学会や生物物理学会と言った関連学会でのブースでの説明や、関連研究者へのデータベースの紹介と利用支援も実施した。成果の発信としては、農芸化学会でのシンポジウム、生化学会でのシンポジウムでの発表など、拠点の活動を中心として成果の発表を実施した。分担機関としての個別支

援としては、10 件の個別支援を行い、内 7 件の原著論文の発表を行うことができた。また、分担機関として、独自データベース COXPRESdb, ATTEDII, eF-site の稼働維持と継続的なコンサルテーションを行い、年間計 18 万ユニーク訪問を全世界から得る日本発の存在感の高いデータベースとして利用されるまでになった。

高度化としては、当初計画の目標とした代謝ネットワーク(KEGG データベースを利用)に関わるタンパク質とそれに結合する基質又は生成物(低分子化合物)について、それらの複合体構造が不明のものも含めた対象とした網羅的なモデリングを実施し、「リンク切れ」の問題を 1 つ解決することができた。結果は、既知の複合体構造を含めてデータベースとして公開を開始した (Natural Ligand Database; NLDB)。また、計画の途中から情報拠点が一体となって、ヒトゲノムの中に見られる多型の解釈に資するシステム開発を行うことになったので、NLDB には、一塩基多型や病気に対応する ID や任意のキーワード等による検索が可能な柔軟な検索機能も実装した。さらに情報拠点全体で開発した構造生物科学のデータクラウド VaProS に NLDB だけでなく遺伝子の共発現ネットワークデータベース(タンパク質間相互作用ネットワークも統合済み)も統合することで、当初計画に掲げた複数のネットワークに関わるデータベース横断的な検索を実現することができた。また、遺伝子共発現ネットワーク、タンパク質間相互作用ネットワーク及び代謝ネットワークをデータベース横断的に利用可能にした。以上のように、今後もユーザインターフェースの改良は必要ではあるが、当初想定以上のデータベースとシステムの開発を行うことができたと考えている。その結果、タンパク質間相互作用に大きな影響を与えながら、濃度の調整で変異の影響を回避し高頻度で観察される面白い変異を報告することもでき、様々な低分子を認識することができる特徴的な結合部位(Sociable Binding Site)を同定することができるなど、単なるデータベース構築にとどまらない科学的な知見を得ることもでき、高度化の当初計画は概ね達成できたと考える。

In this project, the following results were obtained as supportive studies and research activities.

As supportive studies as an information team, we have connected our natural ligand database and gene co-expression network database with VaProS developed by National Institute of Genetics. By using VaPros and some other databases such as database of molecular binding sites of the molecular surface database (eF-site), functional site prediction by similarity search (eF-seek), and gene co-expression database (COX PRESdb, ATTED II), we regularly conducted practical workshops for the wet researchers so that they can walk around various databases themselves. In addition, explanation at the booth at the related scientific conference as an introduction of databases to related researchers and support for use were also carried out. For the dissemination of outcomes, the results were presented focusing on the activities of VaProS, in symposiums at the Society of Agricultural Chemistry, the Biochemical Society, and so on. As individual support as Tohoku University Team, we were able to conduct ten supports and published seven original papers. In addition, we maintained our own databases COXPRESdb, ATTEDII, eF - site and these are now used as the databases with a strong presence from Japan, which achieved a total of 180,000 unique visits a year from the whole world.

As for research work, we have solved a "missing link" problem that we proposed at the beginning of this research project, which are often observed when we tried to connect different database. For example, we can use KEGG to find which protein can interact with a ligand, but it is not straightforward to find the complex structure of the protein and the ligand because protein-ligand complex often solved with analogues. As a result, we could launch a new database called Natural

Ligand Database (NLDB), which contained the known ligand complex structures based on the modelling. Also, since the information team tried to integrate genome variation data to VaProS, we have developed a system that will contribute to the interpretation of single nucleotide polymorphisms found in the human genome. Furthermore, by integrating not only NLDB but also gene co-expression network database (protein-protein interaction network has also been integrated) in the VaProS. We still need to improve user interface, but we believe we were able to develop databases and systems beyond initial expectation. As a result, it was possible to report interesting mutations observed at a high frequency while avoiding the influence of mutation by adjusting the concentration in vivo. We can also obtain scientific knowledge beyond database construction, such as being able to identify characteristic binding sites (Sociable Binding Site), we believe that the initial plan for research activities was largely achieved.

### 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 7 件)

1. Tsugita M, Morimoto N, Tashiro M, Kinoshita K, Nakayama M. SR-B1 is a Silica Receptor that Mediates Canonical Inflammasome Activation. Cell Rep 18(5), 1298-1311, 2017
2. Gojobori T, Ikeo K, Katayama Y, Kawabata T, Kinjo AR, Kinoshita K, Kwon Y, Migita O, Mizutani H, Muraoka M, Nagata K, Omori S, Sugawara H, Yamada D, Yura K. VaProS: A Database-Integration Approach for Protein/Genome Information Retrieval. J Struct Funct Genomics 17(4), 69-81, 2016
3. Murakami Y, Omori S and Kinoshita K. NLDB: a database for 3D protein-ligand interactions in enzymatic reactions. J Struct Funct Genomics, 2016, 17(4), 101-110
4. Fujiwara Y, Kondo H. X, Shirota M, Kobayashi M, Takeshita K, Nakagawa A, Okamura Y and Kinoshita K. Structural basis for the membrane association of ankyrinG via palmitoylation. Sci Rep, 2016, 6, 23981
5. Murakami Y, Mizuguchi K. Homology-based prediction of interactions between proteins using Averaged Open-Dependence Estimators. BMC Bioinformatics.2014,15, 213.
6. Murakami Y, Kinoshita K, Kinjo AR and Nakamura H. Exhaustive comparison and classification of ligand-binding surfaces in proteins. Protein Sci, 2013, 22(10), 1379-1391.
7. Shirota M, Kinoshita K. Analyses of the general rule on residue pair frequencies in local amino acid sequences of soluble, ordered proteins. Protein Sci. 2013, 22(6),725-733.

(2) 学会・シンポジウム等における口頭・ポスター発表

1. ゲノム情報の理解に向けたタンパク質立体構造情報の活用, 口頭, 木下賢吾, 第4回創薬等支援基盤プラットフォーム事業公開シンポジウム, 2016/12/7, 国内.

2. 天然リガンドデータベースの構築とタンパク質低分子相互作用解析への応用, 口頭, 木下賢吾, 日本農芸化学会 2016 年度大会シンポジウム, 2016/3/28, 国内.
3. 機能未知遺伝子の分子機能・細胞機能の推定, 口頭, 木下賢吾, 創薬等支援基盤プラットフォーム事業講演会, 2015/9/25, 国内.
4. Prediction of the biological and biochemical functions of uncharacterized genes, 口頭, 木下賢吾, 日本バイオインフォマティクス学会 北海道地域部会セミナー, 2015/8/21, 国内.
5. ゲノム情報解析の現状と課題, 口頭, 木下賢吾, 次世代 PGS フォーラム設立記念シンポジウム, 2014/12/4, 国内.
6. Prediction of biological and biochemical function of uncharacterized genes, 口頭, Kengo Kinoshita, 台湾国立清華大学ワークショップ, 2014/11/21, 国内.
7. A Study of the Structural and Physicochemical Properties of Sociable Ligand-Binding Sites in Proteins, ポスター発表, 村上洋一, 木下賢吾, 第 52 回 日本生物物理学会年会, 2014/9/25-27, 国内.
8. 構造生命科学データクラウドの構築運用と高度化, ポスター発表, 木下賢吾, 第二回創薬等支援技術基盤プラットフォーム公開シンポジウム, 2014/8/27, 国内.
9. Analysis of Physicochemical Properties of Multi-Ligand Binding Sites in Proteins, ポスター発表, 村上洋一, 木下賢吾, 第 14 回 蛋白質科学会年会, 2014/6/25-27, 国内.
10. タンパク質機能解析に向けた立体構造情報の活用基盤の構築, 口頭, 木下賢吾, 日本農芸化学会 2014 年度大会, 2014/3/30, 国内.
11. Protein Function Prediction based on Similarity of Molecular Surfaces of Ligand-Binding Sites, ポスター発表, Murakami Yoichi, Kinoshita Kengo, Kinjo R Akira, Nakamura Haruki, The 201 A3nnual Conference of the Japanese Society for Bioinformatics, 2013/10/29-31, 国内.
12. 生命情報科学の現状と今後の課題, 口頭, 木下賢吾, 第一回創薬等支援技術基盤プラットフォーム公開シンポジウム, 2013/9/24, 国内.
13. An Automatic and Semi-automatic Approach for Predicting Protein-Protein Complex Structures, ポスター発表, Murakami Y, Kanamori E, Sarmiento J, Liang S, Standley D.M, Shirota M, Kinoshita K, Tsuchiya Y, Nakamura H, CAPRI Meeting 2013, 2013/4/17, 国外.

### (3) 「国民との科学・技術対話社会」に対する取り組み

1. 身近な例で学ぶバイオインフォマティクス, 木下賢吾, 生命医薬情報学連合大会 2015 年大会特別講演, 2015/10/29, 国内.
2. ゲノム配列から見るヒトの多様性, 木下賢吾, コスモス・カレッジ・インターンシップ (宮城県宮城第一高等学校), 2015/10/2, 国内.
3. 個人ゲノムと創る未来, 木下賢吾, 第三回生命医薬情報学連合大会 公開セッション, 2014/10/4, 国内.

### (4) 特許出願

該当無し