

## 平成 27 年度 全体研究開発報告書

1. 補助事業名：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）
2. 補助事業課題名：構造バイオインフォマティクス・リテラシーの浸化と深化（ゲノムスケール構造・変性領域アノテーションを用いた支援と高度化）
3. 研究開発代表者：前橋工科大学 准教授 福地佐斗志
4. 研究開発の成果

【支援】：「構造バイオインフォマティクス・リテラシーの浸化と深化」（代表：名古屋大学）の分担機関として、主にプロテーム・配列・天然変性領域の比較・予測・解析を担当し、代表機関と密接に連携し適切な支援技術を提供した。

【課題番号 2021】「シェディング標的タンパク質の認識配列探索」  
慶応大学（現東京医科歯科大学）・白壁研究室では、マウスのシェディング標的タンパク質の切断配列近辺に変異体を作成し、認識配列の解明を進めている。腫瘍壊死因子(TNF $\alpha$ )、上皮成長因子(EGF)などきわめて重要な働きをする分泌タンパク質は、まず膜タンパク質として発現し、シェディングによる切断を受けて分泌タンパク質になることから、シェディングは創薬の観点から重要である。実験で判明した切断される配列と切断されない配列について、配列プロフィールを作成し切断の有無を予測した。実験でシェディングの有無を検証した結果、シェディング標的タンパク質の認識配列の一部が明らかになった。得られた成果は、現在執筆中の論文にまとめ投稿する予定である。

【課題番号 2051】「霊長類と共進化した乳酸菌の特徴」  
京都府立大学・牛田研究室では、アフリカに住む野生のゴリラより単離したビフィズス菌ゲノムをヒト由来菌と比較することで、ビフィズス菌に起こった進化を理解することを目指している。牛田研究室により決定されたゲノム配列から得られたアミノ酸配列に対し、UniProt のバクテリアサブセットに対し psi-blast 検索を行い、機能分類として GO タームを付加した。さらに、NCBI の COG データベースの機能分類もアノテーションした。これらの機能分類を用い菌種間で比較したところ、アミノ酸代謝、及び糖代謝タンパク質に、ヒト由来菌とゴリラ由来菌の間で差があることが示唆された。このことを統計的に示すため、検定のためのスクリプトを作成した。これにより、ヒトの腸内環境、すなわち食生活の変化がビフィズス菌の多様化に関与していることが示唆された。本課題は、現在論文執筆中である。

【課題受付番号 2188】「キナーゼが認識するリン酸化サイト配列の特徴抽出」  
名古屋大学・貝渕研究室・天野グループでは、膨大なキナーゼ・リン酸化部位情報のセットが蓄積されている。キナーゼとリン酸化部位の双方が対になった大量の情報を用い、より高精度のモチーフの抽出およびリン酸化部位以外の相互作用部位（ドッキングモチーフ）の抽出を目指す。現在、既知ドッキングモチーフを持つ MAPK・そのリン酸化酵素・基質タンパク質のアミノ酸配列を収集し、ドッキングモチーフの抽出プログラムの作成に取りかかっている。また、本課題は事業終了後も継続し技術提供してゆく予定である。

【高度化】：「構造バイオインフォマティクス・リテラシーの浸化と深化」（代表：名古屋大学）の分担機関として、天然変性タンパク質の計算機を用いた解析法の高度化を行った。具体的な成果を以下に示す。

【天然変性領域の機能部位予測】  
天然変性タンパク質の天然変性領域には機能領域が存在する。特に、天然変性領域を介したタンパク質との相互作用は、転写調節・シグナル伝達等に重要である。これらの相互作用部位は、名古屋大学と共同開発したデータベース IDEAL で ProS (Protean segment) と呼び収集されている。この情報を基に、天然変

性領域の機能部位予測法を開発している。まず、天然変性領域予測DICHOTと類似したSVMを用いた機械学習モデルを構築した。IDEALよりProSを持つタンパク質を抜き出しホモログとアライメントした後、保存度・組成等の特徴量を抽出し、学習モデルを構築した。現在、【DICHOTモデルのアップデート】で述べたモデル評価システムを用いチューニングを行っている。また、DICHOTオンデマンドシステムに、この予測を組み込んだ新たなインターフェースを開発した。この予測システムは28年度中に公開する。次に、二次構造予測を基にした方法を開発している。IDEALで収集しているProSを解析すると、その40%弱が $\alpha$ ヘリックス、5%が $\beta$ シート、残りがコイルである。これらは他のタンパク質と相互作用する際に形成される二次構造である。そこで、これらの二次構造が既存の二次構造予測プログラムでどの程度予測可能かを、7個の予測プログラムで検証した。検証の結果、 $\alpha$ ヘリックスについてはすべてのプログラムで7割以上予測可能であり、特にPHDが良い結果を示した。そこで、UniProtに収録されたヒトタンパク質のDICHOTによる予測天然変性領域に対してPHDを適用した。天然変性領域は全体の38%であり、その中に10万近いヘリックス領域が予測された。これは、すべての予測ヘリックス領域の3割弱である。天然変性領域で形成されるヘリックスの予測精度が7割程度であることを考えると、少なくとも数万程度のヘリックスProS領域が天然変性領域中に存在することが示唆される。これらの予測ヘリックスには、UniProtにより他のタンパク質との相互作用を示唆されるものや、典型的なProSのモチーフであるLxxLLモチーフを持つもの等が含まれている。現在、より確からしいProSヘリックス領域を絞り込むため、配列保存度の定量化・モチーフの抽出を行っている。

#### 【タンパク質の効率的合成を目指した翻訳速度見積もりプログラム】

新規データ解析法の開発として、翻訳速度分布解析の実用性の評価を始めた。タンパク質の翻訳速度分布が適切でないと、タンパク質の正常な折り畳みが終了する前に凝集してしまう。タンパク質のどの領域に翻訳エラーが発生しやすいか分析することにより、正常タンパク質の効率的合成が可能になると期待される。コドンの使用頻度及び、次世代シーケンサーを用いたリボゾーム・プロファイリングを用いた最新の実験結果を基に、翻訳速度見積もりプログラムを開発している。各タンパク質を球状構造領域と天然変性領域に分けて解析したところ、球状構造領域に比べ天然変性領域ではレアコドンが高頻度で使われていることが分かった。さらに球状構造領域との境界から離れるほど天然変性領域のレアコドン使用頻度が高くなる傾向も検出された。天然変性領域では、レアコドンの使用により翻訳エラーが生じて、タンパク質機能に影響しない場合が多いと考えられる。天然変性領域のDNA配列を解析し、特徴を見いだした数少ない例として、Nucleic Acids Research誌に投稿した。

#### 論文発表

An optimized Npro-based method for the expression and purification of intrinsically disordered proteins for an NMR study. *Intrinsically Disordered Protein*. Goda, N., Matsuo, N., Tenno, T., Ishino, S., Ishino, Y., Fukuchi, S., Ota, M., and Hiroaki, H. *Intrinsically disordered protein*. 2015, Vol. 3, 1 – 6.

Investigation of the fatty acid transporter-encoding genes SLC27A3 and SLC27A4 in Autism. Maekawa, M. Iwayama, I., Ohnishi, T., Toyoshima, M., Shimamoto, C., Hisano, Y., Toyota, T., Balan, S., Matsuzaki, H., Iwata, Y., Takagai, S., Yamada, K., Ota, M., Fukuchi, S., Okada, Y., Akamatsu, W., Tsujii, M., Owada, Y., Okano, H., Mori, N., and Yoshikawa, T. *Scientific Reports*, 2015, 5, Article number:16239.