

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 15 日現在

機関番号：14501

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25440023

研究課題名(和文) New insights into protein folding based on bioinformatics analysis of cell-free protein synthesis

研究課題名(英文) New insights into protein folding based on bioinformatics analysis of cell-free protein synthesis

研究代表者

TOKMAKOV A・A (TOKMAKOV, ALEXANDER)

神戸大学・遺伝子実験センター・研究員

研究者番号：20301278

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：本研究では無細胞タンパク質合成のバイオインフォマティクス手法を開発した。この手法は無細胞系を用いた実験科学的なタンパク質発現データと計算機上で予想されるポリペプチドの諸性質との比較検証データを解析する基盤を提供する。発現レベル低下と可溶性発現度の上昇がタンパク質の構造不全を起こすことが明らかとなった。この傾向は構造不全領域に特異的な諸状態(疎水性の低下、表面アクセシビリティの上昇、ユビキチン化モチーフやPEST配列の露出などが関係することも明らかとなった。また、リン酸化、グリコシル化、アセチル化が構造全領域で高頻度に見られること、逆にメチル化は低頻度であることも明らかとなった。

研究成果の概要(英文)：In the present study, a bioinformatics approach for identification of multiple physicochemical and structural properties associated with cell-free soluble protein expression was developed. It is based on categorical data analysis of correlations between experimentally observed expression scores and multiple features of amino acid sequences calculated/predicted by bioinformatics. Using this approach, we found that protein disorder is associated with low propensity for detectable cell-free protein expression and elevated ratio of soluble expression. These tendencies are rooted in the distinct features of intrinsically disordered regions, such as low hydrophobicity, elevated surface accessibility and high abundance of sequence motifs for proteolytic degradation. We also found that the eukaryotic PTMs, such as phosphorylation, glycosylation and acetylation, display a clear preference for occurrence in disordered regions of plant proteins, whereas methylation tends to avoid disorder.

研究分野：生物学

キーワード：protein folding cell-free synthesis bioinformatics

1. 研究開始当初の背景

(1) 構造的・機能的ゲノミクス研究の分野で無細胞タンパク質合成系が多用されはじめている。この系が、生細胞抽出液をベースとするハイスループットなタンパク質合成・生産のニーズにかなったものであるからである。しかしながら、無細胞タンパク質合成系メカニズムにはまだ未解明の部分が多く、これまでにこの系によって大量生産されたタンパク質の種類も生体内の全タンパク質のほんの一部にすぎない。

(2) 大腸菌を宿主として可溶性のタンパク質を生産する場合、そのアミノ酸配列にともなう様々な物理化学的性状、すなわちポリペプチドの長さ、電荷、等電点、アミノ酸残基の種類、その他諸々が収量に影響する。問題は、これら様々な性状を正確に計算・予想してタンパク質の合成とフォールディングをうながす方法論がまだ確立していないということである。

(3) これまでの研究で私たちは、原核生物システムを使った無細胞タンパク質合成に対してもっとも影響のあるアミノ酸配列上の物理化学的および構造生物学的性状のリスト化をおこなった(Kurotani et al. 2010)。また、タンパク質合成の収量と予想される翻訳後修飾反応(PTMs)の有無とのあいだに統計学的に有意な相関がある例を多く見出すことに成功した(Tokmakov et al. 2012)。

2. 研究の目的

(1) 本研究の目的は、無細胞抽出液系におけるタンパク質発現と当該タンパク質のアミノ酸配列から予想される各種性状パラメータとの相関を明らかにするバイオインフォマティクス手法を確立し、良好な発現を示すタンパク質グループの特徴を特定するためのアルゴリズムを明らかにすることである。この研究の成果はタンパク質フォールディング研究の新たな展開に貢献することが期待される。

(2) 大腸菌をつかった以前の研究課題と同様に、今回の真核生物細胞を使った無細胞タンパク質合成系のバイオインフォマティクス研究においても、アミノ酸配列がもつ物理化学的および構造生物学的性状がタンパク質の良好な発現とフォールディングにどう影響するかを明らかにするための実験計画を立案した。

(3) 構造機能予測のためのバイオインフォマティクス手法を使い、タンパク質構造とアミノ酸配列上の特定の性状、たとえば翻訳後修飾反応の有無、の2つの事項の相関をゲノムワイドに解析する。

(4) 本研究ではさらに、本研究者がこれまで

に研究・構築してきたアフリカツメガエル *Xenopus laevis* 卵母細胞・未受精卵のタンパク質無細胞発現系および細胞内発現系を用いた実験を続行し、その生物学的意義の解明や機能改善に取り組む。

3. 研究の方法

(1) 無細胞タンパク質合成は以前報告した方法にもとづき、PCR増幅させたDNA断片を用いるプログラムによりおこない(Kurotani et al. 2010)、最終容量20-50 mlのタンパク質合成反応液を得た。反応液中には可溶化促進剤や分子シャペロンを加えなかった。可溶性および不溶性のタンパク質は10,000 g、10分間の遠心分離により分離し、5-mlに分注した分離前試料と遠心分離上清とを12.5%分離ゲルを用いてSDS-PAGEで解析した。泳動後のSDS-PAGEゲルはCoomassie Brilliant Blue染色と抗Hisタグ抗体をつかった免疫ブロッティングにより分析し、得られたデータからタンパク質の発現レベルと可溶性を評価した。その結果、解析対象としたタンパク質の発現傾向を次の3つのカテゴリに分類した：可溶性、不溶性、および発現不良性である。タンパク質は個々にこれら3つのカテゴリのどれかに分類され、その他のカテゴリの性質を示すことはなかった。そのため、これから説明するカテゴリ別のデータ解析を行うこととした。

(2) 既存のバイオインフォマティクス手法を使い、発現タンパク質がもつ様々な性状を計算・予測した。Grand average of hydropathy (GRAVY) と等電点は、Expasyサーバ(<http://cn.expasy.org/>)が提供する無料ソフトを使い計算した。二次構造(DSC_EH%ftp://ftp.dcs.aber.ac.uk/pub/users/rdk/dsc/)や破綻構造[regional order neural network (RONN) parameter : <http://www.strubi.ox.ac.uk/RONN>]の割合も、それぞれにオンライン供与されている手法を使い計算した。Low-complexity regionsの割合はSEGアルゴリズムにより計算した。オンライン供与されている既存のツールにより、このほかシグナル配列(<http://www.cbs.dtu.dk/services/SignalP/>)、膜貫通配列(TM : http://bp.nuap.nagoya-u.ac.jp/sosui/sosuisignal/sosuisignal_submit.html)、PEST配列(プロリン、グルタミン酸、セリン、トレオニンに富む配列 <http://emboss.bioinformatics.nl/cgi-bin/emboss/pepfind>)、コイルドコイル構造(pepCoil : <http://emboss.sourceforge.net/apps/cvs/emboss/apps/pepcoil.html>)、ドメイン間リンカー配列(DomCut : <http://www.bork.embl.de/suyamadomcut/>)、ジスルフィド結合部位(Dipro :

<http://contact.ics.uci.edu/brige.html>), さらには溶媒に対するタンパク質のアクセシビリティ (Acc_ProE : <ftp://ftp.dcs.aber.ac.uk/pub/users/rdk/dsc/>) も計算した。以上あげた以外のタンパク質性状については、ProteoMix ソフトウェアを使い計算・予測した。リン酸化、グリコシル化、アミド化、Asx ヒドロキシル化、硫酸化、プレニル化などの翻訳後修飾反応の予想は PROSITE スキャニングツール PS_SCAN (http://www.hpa-bioinfotools.org.uk/cgi-bin/ps_scan/ps_scanCGI.pl) を用いておこなった。ユビキチン化と SUMO 化の予測は UbPred と SUMOsp2.0 を使いそれぞれおこなった。

(3) 個々のタンパク質がもつ性状と発現スコアの相関は、YES/NO、離散変数、連続変数の3通りで示した。3カテゴリ性質(可溶性、不溶性、および構造不良性)の度合いは、タンパク質それぞれで調べたパラメータごとに確定した。得られた描画データはエクセル・チャート・スムージング・アルゴリズムを使い直線形化し、データセット中のタンパク質の各種パラメータ値に基づく分布を解析することで、情報を整理統合した。

(4) 上記説明した3つのカテゴリに基づく解析データを使い、実験的に得られたタンパク質の各種性状データと無細胞系発現の従順性の相関における統計学的な有意性を3カテゴリ性質(可溶性、不溶性、および構造不良性)ごとに評価した。YES/NO型の評価軸で統計学的な有意性を検証するときは、二方向性分割表を用いた。オンライン供与されているツール

(<http://statpages.org/ctab2x2.html>) を使い、フィッシャー直接検定法も計算に用いた。得られる値が限られる離散変数型および連続変数型の評価軸で統計学的な有意性を検証するときは、ピアソン対積率相関係数を求めた。積率相関係数の系統学的優位性を検証するときは、積率相関係数の実験値(r)と試料サイズを計算式に与えた上で片側確率値を計算した。積率相関係数と p 値の計算は、ともにオンラインで供与された統計計算ソフトを用いた

(<http://www.danielsoper.com/statcalc3/>)。

4. 研究成果

(1)

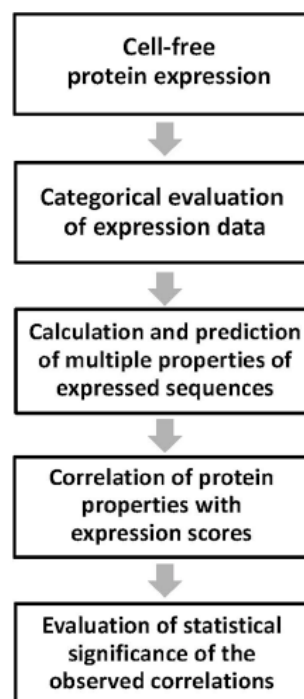


図1 本研究で開発・運用した解析手法のフロー図データ解析手法を時系列順に示す Tokmakov et al. 2014)。

図1に本研究で開発し実験作業に用いたバイオインフォマティクス手法、すなわち無細胞タンパク質発現とその物理化学的・構造生物学的性状の解明をカップルさせた研究プラットフォームのフローを示す。この手法は以下の5つの要素を含んでいる。(1) 同一タンパク質の小スケール無細胞系発現実験、(2) タンパク質発現結果のカテゴリ分類、(3) コンピュータ計算・予測による様々な物理化学的・構造生物学的性状の特定、(4) 個々のタンパク質が示した発現状況と前述のコンピュータ解析結果の対応づけ、(5) 実験的に得られた相関データの統計学的有意差の検証。これら各要素の詳細については、すでに査読付きの英文学術論文にて発表済みである (Tokmakov 2014; Tokmakov et al. 2014)。

(2) 3.(2)で説明したバイオインフォマティクス手法を用いて、単子葉植物および双子葉植物由来のタンパク質の構造不良と主な翻訳後修飾反応(セリン/トレオニン/チロシンのリン酸化、プロリンのOグリコシル化、アスパラギンのNグリコシル化、リジンのアセチル化、リジン/アルギニンのメチル化)との関係についてプロテオームスケールで解析した。その結果、リン酸化、アセチル化、およびOグリコシル化の三者は、タンパク質の構造不良領域に起こりやすい傾向にあることがわかった。逆に、メチル化は同領域に起こりにくいこともわかった。Nグリコシル化は、この

ような傾向性のどちらもが見られなかった以上の結果をまとめたのが図2である。

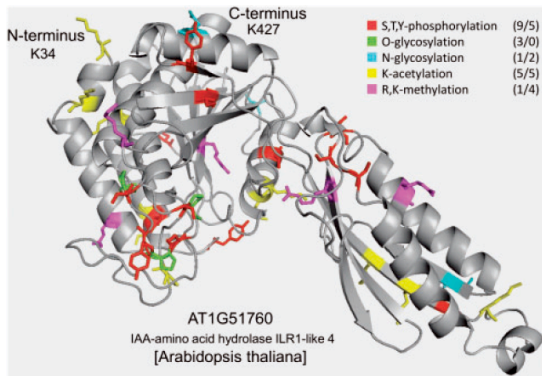


図2 シロイヌナズナのヨード酢酸-アミノ酸ヒドロラーゼの三次元構造およびその翻訳後修飾反応部位。図の右上に示す項目と数値は、各翻訳後修飾反応がタンパク質構造の不良化または正常化の出現頻度に与える影響の度合いを示している (Kurotani et al. 2014)。

(3) 本研究では真核生物タンパク質の無細胞系での発現を新たに開発したバイオインフォマティクス手法により解析することを目的とした。そのため、真核生物に加え原核生物も対象に、それぞれにおける内在性タンパク質構造不良と無細胞系タンパク質発現とを比較検証した。その結果、構造不良は可溶性発現の度合いが高いことと同調的に高頻度となること、一方で検出レベルを超えたタンパク質発現がある場合には構造不良の度合いが減少することが明らかとなった。さらには、これらの傾向の元となる要因として、低疎水性、ユビキチン化やPEST配列のようなタンパク質分解シグナル配列モチーフの表面アクセシビリティや含有率の上昇が構造不良を示すタンパク質領域に見られることが重要であることも明らかとなった (図3)。

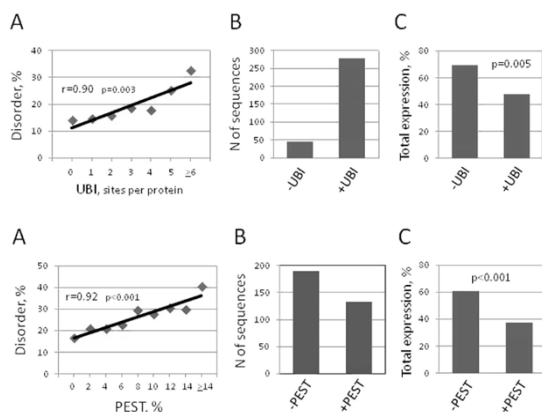


図3 ユビキチン化やPEST配列があると、タンパク質に構造不良が起こる傾向が高くなる (Tokmakov et al. 2015)。

(4) 真核生物タンパク質の無細胞発現系および細胞を用いた発現系をより高性能化し生物学研究に資するものにするため、アフリカツメガエルの卵母細胞と未受精卵をつかった1細胞遺伝子発現のモニタリング解析 (細胞質サンプリング、定量的RT-PCRを含む) をおこなった。ルシフェラーゼ遺伝子をT7プロモータ駆動性のプラスミドでT7-RNAPタンパク質と共発現させる実験をおこない、ルシフェラーゼを十数時間にわたり発現し続ける卵母細胞システムも構築に成功した (図4)。

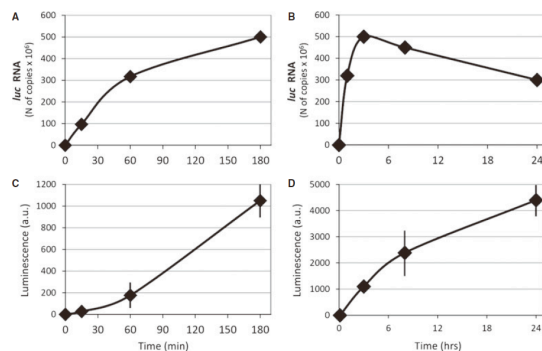


図4 アフリカツメガエルの単一卵母細胞におけるルシフェラーゼ発現のモニタリング mRNA 発現 (A, B) およびタンパク質発現 (C, D) の時間推移を示す (Tokmakov et al. 2014)。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 5 件)

1. Tokmakov AA, Kurotani A, Ikeda M, Terazawa Y, Shirouzu M, Stefanov V, Sakurai T, Yokoyama S. Content of intrinsic disorder influences the outcome of cell-free protein synthesis, *Scientific Reports*, peer-reviewed, 5:14079, 2015. doi: 10.1038/srep14079.

2. Tokmakov AA, Hashimoto T, Hasegawa Y, Iguchi S, Iwasaki T, Fukami Y. Monitoring gene expression in a single *Xenopus* oocyte using multiple cytoplasmic collections and quantitative RT-PCR, *FEBS Journal*, peer-reviewed, v. 281, pp. 104-114, 2014. doi: 10.1111/febs.12576.

3. Tokmakov A. Identification of multiple physicochemical and structural properties associated with soluble expression of eukaryotic proteins in cell-free bacterial extracts, *Frontiers in Microbiology*, peer-reviewed, 5:295, 2014. doi: 10.3389/fmicb.2014.00295.

4. Kurotani A, Tokmakov A, Kuroda Y, Fukami Y, Shinozaki K, Sakurai T. Correlations of protein disorder with post-translational modifications in plants. *Bioinformatics*, peer-reviewed, v. 30, pp. 1095-1103, 2014. doi:10.1093/bioinformatics/btt762.

5. Tokmakov A, Kurotani A, Shirouzu M, Fukami Y, Yokoyama S. Bioinformatics analysis and optimization of cell-free protein expression. *Methods in Molecular Biology*, peer-reviewed, v. 1118, pp. 17-33, 2014. doi: 10.1007/978-1-62703-782-2_2.

[学会発表] (計 4 件)

1. Tokmakov A. Bioinformatics approach to identify physicochemical and structural properties associated with successful cell-free protein synthesis. *17th International Conference on Bioinformatics and Biotechnology*, 2015.04.08, (Marrakech, Morocco).

2. Tokmakov A, Kurotani A, Shirouzu M, Fukami Y, Yokoyama S. The presence of multiple sites of ubiquitination and SUMOylation is associated with increased cell-free production of soluble protein. *Ubiquitin and UBLs: At the crossroads from chromatin to protein EMBO Conference*, 2014.10. 20, (Buenos Aires, Argentina).

3. Tokmakov A. Bioinformatics analysis of cell-free protein synthesis. *ICPAC 2014 International Conference on Pure and Applied Chemistry*, 2014.06.24, (Flic en Flac, Mauritius).

4. Tokmakov A, Hashimoto Y, Hasegawa Y, Iguchi S, Iwasaki T, Fukami Y. (2013.09) Gene expression profiling in single *Xenopus* oocytes and eggs. *ISPROF 2013. 1st International Symposium on Profiling*, 2013.09.02, (Costa da Caparica, Lisbon, Portugal).

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
出願年月日 :
国内外の別 :

○取得状況 (計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

[その他]
ホームページ等

6. 研究組織
(1) 研究代表者

TOKMAKOV Alexander (TOKMAKOV A.A.)
神戸大学 遺伝子実験センター 研究員
研究者番号 : 20301278

(2) 研究分担者
()

研究者番号 :

(3) 連携研究者
()

研究者番号 :