

新規アミノ酸配列解析手法の開発とタンパク質工学への応用

中野 祥吾

はじめに

情報技術分野の急速な発展により、膨大な数のデータが日々取得されている。それに伴い今日、これらデータがインターネット上に溢れている。一方で、データそのものだけでは科学的な意味をなさず、データを解析し必要な情報を取り出すことが必要となる。我々は日常的にデータの海にアクセスし、それらを解析し必要な情報を抽出することを行っている。たとえば、スマートフォンを使って地図アプリを起動し、目的地に到達する最短経路を検索することは、多くの読者の方々にも経験があると思われる。この検索過程は、“スマートフォン”というIT端末を用いて、“電子化された地図”というデータにアクセス、“地図アプリ”という解析ツールを用いて解析することで、“最短経路”という情報を得るという形に分解することが可能である。このことは同時に、IT端末、データ、解析ツールのいずれかが欠けると、途端に有用な情報を得ることが難しくなることを示している。以上を考慮に入れたうえで、生物工学分野で扱われるデータ解析について、その現状と問題点について考えてみたい。

生物工学分野において、次世代シーケンサーの利用や各種オミクス解析は今日、広く行われている。これにより莫大なデータがPubMedなどのデータベースに登録されるようになってきている。特にDNA配列決定法の発展によるその進展は目覚ましい。報告によると、現在までにPubMedなどのデータベース上には2億を超えるDNA配列データ（2017年現在）が登録されており、その登録数は現在も伸び続けている。これに同調してタンパク質のアミノ酸配列データの登録数も増えてきている。配列データの中には産業応用上有用な酵素などが隠れており、それらを表舞台に引っ張り出すことは、酵素工学・タンパク質工学分野において、きわめて重要である。一方で、莫大な配列データを解析するツール、解析手法および得られた情報の有効活用についてはまだまだ発展途上にあるのが現状である。

筆者はアミノ酸配列解析を通して、タンパク質工学に資する有用な情報を抽出するため、配列データの新たな解析手法の開発を進めてきた。本稿では筆者らが開発し

たアミノ酸配列解析ツールであるINTMSAlignに関する説明と、そのタンパク質工学への応用例について紹介する。

アンサンブル学習法を用いた

新規アミノ酸配列解析ツール、INTMSAlignの開発

タンパク質工学的手法により酵素の機能を改変・向上させる際に、酵素の立体構造から機能向上に重要なアミノ酸残基を予測し変異導入を行う、いわゆる合理的設計が広く行われている。一方で、合理的設計では耐熱性向上や酵素反応速度パラメータ (k_{cat} や K_m) の改良が難しい場合が多々存在する¹⁾。特に、膜タンパク質などに代表される難解析性タンパク質では、立体構造が利用できず、合理的設計の適用が難しいケースが存在する。このような場合、研究対象のタンパク質（以下STP）とこの相同なタンパク質の配列を、マルチプルシーケンスアライメント法（MSA法）を用いて比較解析することで、どのような変異を導入するか方針を決めることが行われている。MSA法は、ある程度の配列相同性（20%以上）を有する複数のアミノ酸配列を、なるべく一致するように整列する手法である。MSA法を用いることで、ファミリー間で保存されているアミノ酸残基、いわゆるファミリーの機能発現に関わる残基を予測することができる。広く用いられているソフトウェアとして、ClustalΩやMAFFTなどがあげられる。一方で、現在のMSA法の弱点として、アライメントできる配列数に限度がある（1000配列以下）ことがあげられる。

STPと相同なアミノ酸配列が大量に存在する場合（たとえば1000個以上）に、真に保存されているアミノ酸残基を予測するためには、この弱点を克服する必要がある。筆者は大量のアミノ酸配列を扱える新たなMSA法の開発が必要だと考え、図1に示した新たなアミノ酸配列解析ツール、INTMSAlignを考案した²⁾。INTMSAlignはアミノ酸配列の組合せをさまざまに変えてMSAを行い、最終的にその結果を統合することを行う、いわゆるアンサンブル学習法を採用したMSA法である。具体的な解析手順は以下の通りである。INTMSAlignの解析に必要なデータは、STPとその相同な配列からなるライブラリ（配列数の制限はなし。図中ではx個）の二つで

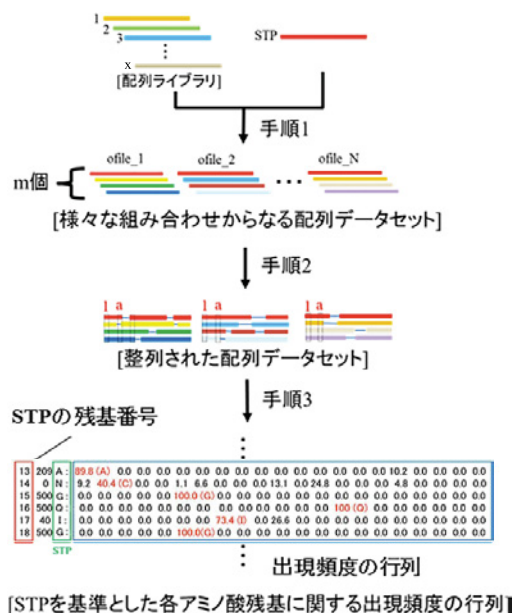


図1. INTMSAlignのアルゴリズム. 解析したいターゲットとなるタンパク質のアミノ酸配列 (STP) とそのファミリー配列からなる配列ライブラリを基に解析を行う. 手順1から3の過程を経て, STPの各アミノ酸残基に関する出現頻度を計算する. ※学会HPのPDFではカラー表示されます.

ある (図1). ここで, STPのアミノ酸残基数を n , ライブラリ中から選択する配列の数を m と定義する (n, m はいずれも整数). INTMSAlignは, ①STPと m 個 (10個未満) の配列をライブラリからランダムに選択した配列データセット (dataN) を複数個 (ここでは N 個) 作製し (手順1, 図1), ②すべてのデータセットに関して, ClustalWなどの既存のMSA法を用いてアラインメントを行い (ResultN) (手順2, 図1), ③STPの a 残基目 ($1 \leq a \leq n$) について, 20種類のアミノ酸およびギャップに関する出現頻度を計算するソフトウェア (手順3, 図1) である. なお, 出現頻度はResult1からResultNのすべてのファイルをまとめたうえで計算される. 最終的に出現頻度は 21 (20種類アミノ酸とgapの出現頻度) $\times n$ の行列として出力される.

このように, INTMSAlignのアルゴリズムはきわめて単純ではあるが, 多数の相同な配列が存在するライブラリ内において, STPを基準とした各アミノ酸残基に関する保存性を, 出現頻度という2次元の行列で出力できるため, さまざまな数値解析・タンパク質配列設計法に応用しやすい利点がある. 以下, INTMSAlignの応用例についていくつか紹介する.

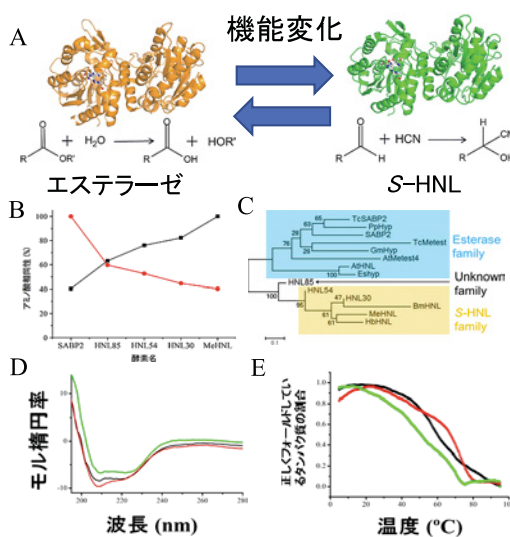


図2. エステラーゼとS-HNLは20–30%ほどの配列相同性を示し, かつ類似した立体構造を有するが, 触媒する反応は大きく異なる (A). これらの酵素の機能変化がどのようにして起こったのか, 人工S-HNL (HNL85, HNL54, HNL30) を設計し, その機能解析を通して解明を目指した. 人工設計したS-HNLの配列相同性 (B) と系統解析 (C). 設計した三つのS-HNLの中で, HNL30がもっとも自然界由来のS-HNLに近く, 逆にHNL85はもっとも離れた配列を持つ (BおよびC). CD測定の結果, 設計したS-HNLはいずれも自然界由来のS-HNLと同様の構造を有し (D), かつ耐熱性 (T_m 値) は $HNL54 > HNL85 > HNL30$ の順に優れていた (E). ※学会HPのPDFではカラー表示されます.

S選択的ヒドロキシニトリルリアーゼの人工設計とその機能解析

MSA法を応用したタンパク質工学的手法として, 完全コンセンサ設計法 (FCD) と祖先型設計法 (ASR) が広く利用されている. FCDでは, 機能に重要なアミノ酸残基はファミリー間においても高度に保存されているという前提のもと, 研究対象となるタンパク質の全配列をもっとも高く保存されているアミノ酸残基 (コンセンサ残基) に置換することで人工タンパク質を設計する方法である³⁾. 一方で, ASRでは, MSAと系統解析の結果を基に, 系統樹上の節に位置する祖先型タンパク質の配列を設計する方法である⁴⁾. 両者はMSAの結果を使うという点では類似しているが, 出力される配列は異なるものになる. 本項ではまず, INTMSAlignを用いてFCDにより, S-ヒドロキシニトリルリアーゼ (S-HNL) の人工設計とその機能解析を行った例について記述する.

ヒドロキシニトリルリアーゼ (以下, HNL) はケトンやアルデヒド化合物にニトリル基を付加する反応を触媒可能な酵素 (図2A) であり, 今日, 光学活性なシアノヒドリン化合物の合成に応用されている⁵⁾. S-HNLは

S体のシアノヒドリン化合物の合成に利用されており、キャッサバなど特定の植物において発現することが報告されている⁵⁾。S-HNLは、各種生物において広く発現している、 α/β -hydrolaseフォールドを有する一部のエステラーゼ（たとえばSABP2、図2B）と類似した構造・配列を有することから、エステラーゼが分子進化を経てS-HNL活性を獲得するよう、機能変化したと考えられている（図2A）⁶⁾。しかし機能変化の過程で、どの部位に変異が導入されて、どのような機能変化が生じたのか、明らかになっていない。

筆者らは、エステラーゼがS-HNLに変化する過程で、どのような機能変化が生じたのか実験的に推測するために、INTMSAlignを用いて三つの人工S-HNL（HNL85, HNL54, HNL30）を設計し、その機能解析を行った。詳細な設計手順は紙面の都合上割愛するが、簡単に説明すると、図1で示したSTPをキャッサバHLN（MeHNL）の配列に、Blastp解析で得られたMeHNLと相同な配列をライブラリとして設定したのち、INTMSAlignを用いて解析、FCD法を用いて設計することで、SABP2（エステラーゼ）に近い人工タンパク質（HNL85）とMeHNL（S-HNL）に近いHNL30、その中間に位置するHNL54を設計した（図2C）。これら三つの人工タンパク質について、円二色性分散計（CD）を用いた測定の結果、三つの人工タンパク質はいずれも溶液中で正しくフォールディングしていることを確認した（図2D）。耐熱性について、加温による222 nmにおけるCDスペクトル変化から、 T_m を算出することで評価した（図2E）。結果、耐熱性に関してはHNL85がもっとも高く、逆にHNL30がもっとも低いという結果になった（図2E）²⁾。

次に人工タンパク質の酵素学的諸性質の検討を行った。HNL85に関しては、(S)-マンデロニトリル((S)-Man)および(R)-マンデロニトリル((R)-Man)を基質としたときの酵素効率 (k_{cat}/K_m) は、決定できなかった。一方で、(R)-Manに対する k_{cat}/K_m はHNL54が、(S)-Manに対してはHNL30が最大であった。酵素のエナンチオ選択性を評価する指標であるE値を算出したところ、HNL30は133という高い値 (S選択性) を有していたが、HNL54は4.9と低い値を有していた。以上の結果から、エステラーゼからS-HNLへの機能変化の過程を以下のように予測した。まず、エステラーゼとS-HNLの活性のトレードオフが2点の変異で生じることが予測されている (StepA, 図3)。ただ、2点変異ではS-HNL活性がきわめて微弱であり、かつR-HNL活性も有する。そこで次の過程では、変異導入によりS-HNL活性を最大化することを旨とする一方で、R-HNL活性の減少や耐熱性

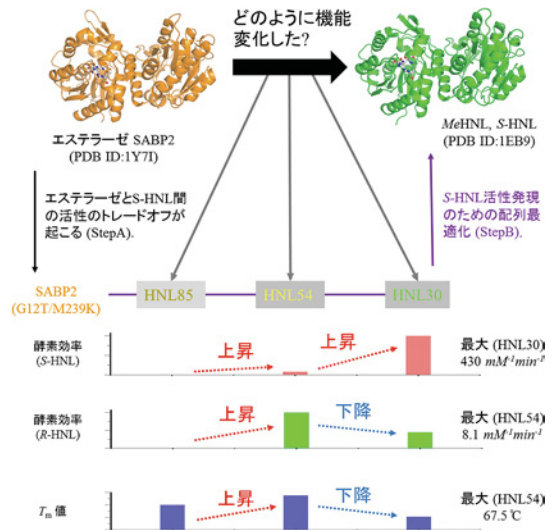


図3. 変異蓄積によるエステラーゼからS-HNLへの機能変換の過程に関する概略図。まず、エステラーゼに二点変異(G12T, M239K)が起こることで、エステラーゼとS-HNLの活性のトレードオフが起こる。その後、主にタンパク質表面に変異が蓄積することで、S-HNL活性が上昇する。この過程でR-HNL活性や T_m 値は増減する。※学会HPのPDFではカラー表示されます。

の減少といった機能のトレードオフをもって、エステラーゼはS-HNLへと機能変化してきたと予測した (StepB, 図3)。立体構造予測の結果より、HNL85からHNL30へ配列が変遷する際、変異導入は主にタンパク質表面で起こっており、かつ活性中心から10 Å半径内では変異が導入されないことを確認している。以上の結果は酵素の種類にも依存するが、ある酵素を別の活性を示すようにデザインする際には、活性中心だけでなく、タンパク質表面にも合理的に変異を導入する必要があることを示唆している²⁾。

短鎖型L-スレオニン脱水素酵素の人工設計と構造機能解析に基づくFCD法とASR法の比較解析⁶⁾

INTMSAlignを用いてFCD法により、S-HNLの人工設計に成功した。しかしS-HNLの成功例だけでは、設計が偶然うまくいった可能性を否定できず、本手法がその他のタンパク質の設計に広く応用できるとは言えない。そこで、本手法を適用してさまざまなタンパク質の設計とその機能評価を行うことを通して、その有用性を証明することを目指した。まず、短鎖型L-スレオニン脱水素酵素 (SDR-TDH) を研究対象とし、INTMSAlignとその応用法を用いてSDR-TDHの人工設計を行った。SDR-TDHはL-スレオニンの β -ヒドロキシ基の脱水素反応を触媒し、2-アミノ-3-ケトブチル酸を生成する、 NAD^+ 依存型酵素である⁷⁾。本酵素は血中L-スレオニン

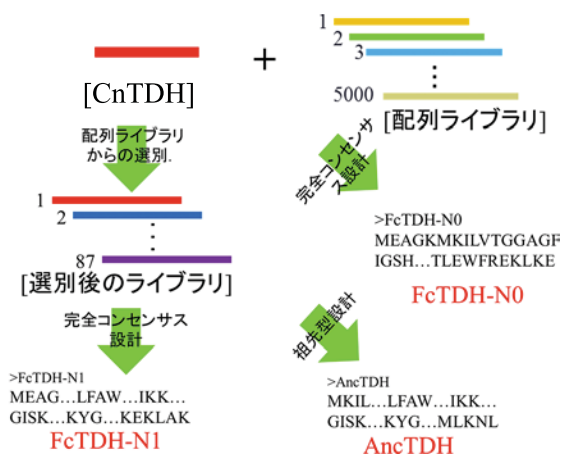


図4. INTMSAlignを用いたライブラリ選別と三つの人工TDH配列の設計 (FcTDH-N0, FcTDH-N1, AncTDH) について。まず、自然界由来TDHであるCnTDHをBlastpで解析し、5000個の相同な配列を選別した。これを使ってFCD法によりFcTDH-N0を設計した。次にINTMSAlignの新機能を用いて配列ライブラリを分類し、87個の相同な配列からなるライブラリを選別した。このライブラリを用いて、FCDおよびASR法により、FcTDH-N1およびAncTDHをそれぞれ設計した。三つの設計した人工TDHに関して、pET15bにサブクロニングを行い、発現・精製・各種生化学的・構造生物学的実験を行った。※学会HPのPDFではカラー表示されます。

濃度定量への応用が期待されている⁸⁾。現在までに10程度のSDR-TDHが報告されており、その反応機構についてはさまざまなグループにより解明が進みつつある⁹⁻¹²⁾。

まず設計に用いるアミノ酸配列情報の収集を行った。STPの配列として*Cupriavidus necator*由来のSDR-TDH (以下CnTDH) を用いた。CnTDHの配列をBlastpで解析して、CnTDHと類似した配列を持つ5000個の配列の取得し、これを配列ライブラリとした(図4)。これらのデータをINTMSAlignにより解析し、FCD法により人工タンパク質、FcTDH-N0を設計した(図4)。しかし実験の結果、FcTDH-N0は可溶性画分に得ることができず、かつSDR-TDH活性は確認できなかった。そこで配列ライブラリの中から、人工設計に適した配列を選別し、87個の配列からなる新たなライブラリを作製した(図4)。配列の選別は、CnTDHと5000個の配列から構成されるライブラリ中の配列を一つずつ比較し、CnTDHを基準に143位がVal, 174位がLeu, 188位がTyr, そして214位がMetの配列のみを選ぶことで行った。選別に使用したアミノ酸残基(筆者らは相関振動残基と呼んでいる)の同定法に関しては、公表していないので記述できないが、これも理論的に同定することが可能である。この選別した配列ライブラリを用いて、FCD法を用いてFcTDH-N1を、ASR法を用いてAncTDHを設計した(図4)。次にこれら二つのタンパク質に関して各種生化学的、構造生物学的手法により、その性質検討を行った。

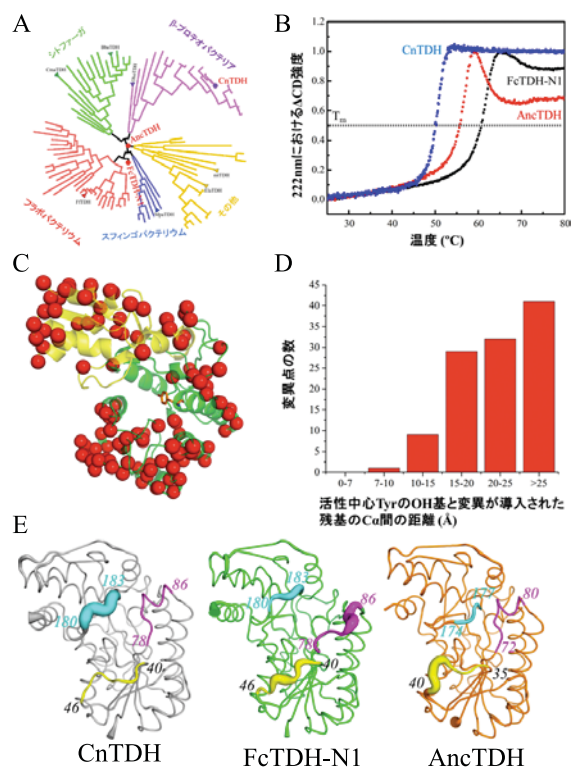


図5. 設計したFcTDH-N1とAncTDH, およびTDH活性を確認できた自然界由来TDHの系統解析(A)。本研究ではCnTDH(自然界由来), FcTDH-N1, AncTDHを用いて実験を行った(filled circleで表示)。CD測定による耐熱性の確認(B)。耐熱性はFcTDH-N1がもっとも高く、CnTDHがもっとも低かった。人工設計により変異が導入された箇所解析(C)。CnTDHと比べて、活性中心から7Å以上離れた位置に変異が導入されていた(D)。分子動力学シミュレーションによる溶液中での動きの違いに関する解析(E)。基質特異性や活性発現に重要なループ領域のみ、動きが異なることが判明した。※学会HPのPDFではカラー表示されます。

学的、構造生物学的手法により、その性質検討を行った。なお、FcTDH-N1およびAncTDHはいずれも可溶性画分に50 mg/L以上の収量で得ることが可能である。

第一に、FcTDH-N1とAncTDHの系統解析を行った(図5A)。その結果、両方のタンパク質はいずれも系統樹の中心付近に位置しており、既知の自然界由来SDR-TDHと異なる配列を有することを確認した。次に、CnTDHとFcTDH-N1, AncTDHをCD測定により解析し、 T_m 値を基にその耐熱性を評価した(図5B)。結果、FcTDH-N1およびAncTDHはCnTDHと比べてそれぞれ10, 5°C高い T_m 値を有していた(図5B)。酵素活性測定の結果、CnTDHとFcTDH-N1は同等の k_{cat}/K_m 値を有しており、AncTDHは高温領域(50°C)でもっとも高い k_{cat}/K_m 値を有することを確認した。次に人工SDR-TDHの設計に際して、CnTDHのどの領域に変異が導入されたか、X線結晶構造解析を通してその確認を行い、

FcTDH-N1 および AncTDH は CnTDH と同様の立体構造を有することを確認した (図 5C). 一方で, 変異導入箇所については, 主にタンパク質表面に集中していることが判明した (図 5D). 事実, FcTDH-N1 および AncTDH の設計により全部で 112 点の変異が CnTDH 上に導入されていたが, 活性中心から 10 Å 以内に変異が導入されていたのは, わずか 2 残基であった. タンパク質表面に導入された変異が SDR-TDH に及ぼす影響を考察するため, CnTDH, FcTDH-N1 および AncTDH に関してそれぞれ 100 nsec の分子動力学シミュレーションを行った (図 5E). その結果, SDR-TDH の活性発現において, NAD⁺ 結合に重要な 40–46 loop, 基質である L-スレオニンの認識に重要な, 80–86 および 180–183 loop の動きが 3 者で異なることが判明した (図 5E). 以上の結果より, 人工設計により導入された変異は, SDR-TDH の動きを変化させ, 耐熱性や酵素活性の発現などに影響を及ぼすことが示唆された.

本結果は INTMSAlign を用いた設計により, 自然界の酵素と比べて, 活性を維持しつつ耐熱性を向上させた人工タンパク質の配列デザインが可能であることを示唆していた. 一方, 今回用いた FCD 法と ASR 法に関して, その優位性に関する議論が以前から行われている. 報告によると ASR 法が FCD 法に比べて, より優れた人工タンパク質を設計できるという記述が多い^{13,14)}. しかし今回の SDR-TDH の設計法を通して, 配列ライブラリの選別に注意すれば, FCD 法でも ASR 法で設計したものと同色のない機能を持ち, かつ自然界のタンパク質よりも優れた人工タンパク質を設計できることを示すことができた. 現在, 更なる設計精度向上に向けた手法の改良を進めている.

まとめ

以上, 新規アミノ酸配列解析ツール INTMSAlign のアルゴリズム開発, 加えてそのタンパク質工学への応用例として, 人工 S-HNL の設計とその機能評価, および L-スレオニン脱水素酵素の人工設計を通して FCD 法と ASR 法のベンチマーク解析を行った例を紹介した. INTMSAlign のその他の応用例として, 不溶化するタンパク質を可溶化させる変異点を推測するための指標として HiSol スコアを定義, そのスコアを基にいくつかの不溶性タンパク質を可溶化させることにも成功している¹⁵⁾.

余談になるが, INTMSAlign を用いたタンパク質設計に関する論文投稿・成果発表を進める中で強く感じる

ことがある. それは, “筆者らが都合の良いタンパク質だけを選択して設計しているのだから, うまく設計できているように見える” という懸念である. このような懸念を払しょくするため, 外部研究機関 (複数の企業や大学) との共同研究を通して高機能化した人工タンパク質の設計に取り組んでいる. 特許や共同研究の都合上詳細は記載できないが, 現在までに膜タンパク質や抗体を含む複数の人工タンパク質の設計・高機能化に成功しており, 4 件の特許出願を行っている. 共同研究を通して客観的な視点から本手法を厳しく評価していただくことで, 設計精度の更なる向上を目指している. もし本手法を試してみたいという読者の方がおられたら, ご一報頂けると幸いです.

謝 辞

本研究は独立行政法人科学技術振興機構 (JST) の戦略的創造研究推進事業 (ERATO) および日本学術振興会の科学研究費助成事業 (16K18688) を受けて行われたものです. ご指導をいただいた ERATO 浅野酵素活性プロジェクト研究総括の富山県立大学工学部・浅野泰久教授, 立教大学理学部・常盤広明教授, ならびに静岡県立大学食品栄養科学部・伊藤創平准教授に心よりお礼申し上げます. また, 本研究にご協力いただいた多くの共同研究者の方々に感謝申し上げます.

文 献

- 1) Krista, L. M. and Romas, J. K.: *Trends Biotechnol.*, **23**, 231 (2005).
- 2) Nakano, S. and Asano, Y.: *Sci. Rep.*, **5**, 8193 (2015).
- 3) Benjamin, T. P. and Ashley, M. B.: *Protein Eng. Des. Sel.*, **29**, 245 (2016).
- 4) Gumulya, Y. and Gillam, E. M.: *Biochem. J.*, **474**, 1 (2017).
- 5) Dadashpour, M. and Asano, Y.: *ACS Catal.*, **1**, 1121 (2011).
- 6) Nakano, S. *et al.*: *Biochemistry*, **57**, 3722 (2018).
- 7) Kazuoka, T. *et al.*: *J. Bacteriol.*, **185**, 4483 (2003).
- 8) Ueatrongchit, T. and Asano, Y.: *Anal. Biochem.*, **410**, 44 (2011).
- 9) Yoneda, K. *et al.*: *FEBS J.*, **277**, 5124 (2010).
- 10) Yoneda, K. *et al.*: *J. Biol. Chem.*, **287**, 12966 (2012).
- 11) Nakano, S. *et al.*: *J. Biol. Chem.*, **289**, 10445 (2014).
- 12) Motoyama, T., Nakano, S. *et al.*: *Biochemistry*, **56**, 5758 (2017).
- 13) Risso, V. A. *et al.*: *Proteins*, **82**, 887 (2014).
- 14) Gumulya, Y. and Gillam, E. M.: *Biochem. J.*, **474**, 1 (2017).
- 15) Matsui, D. and Nakano, S. *et al.*: *Sci. Rep.*, **7**, 9558 (2017).