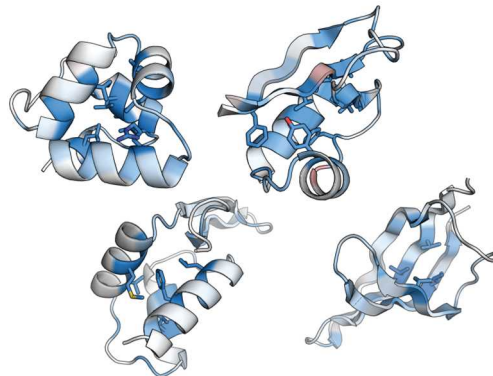


2023年7月20日  
東京大学  
科学技術振興機構（JST）

## タンパク質が「形」を保つ力の超並列測定法 ——タンパク質科学のAI開発にも貢献——

### 発表のポイント

- ◆タンパク質の構造的な安定性を超並列に測定する方法を開発した。
- ◆従来は一度に1種類のタンパク質についての構造安定性しか測定できなかったが、90万種類程度をまとめて測定することに成功した。
- ◆疾患の原因となるアミノ酸変異の特定やタンパク質医薬の効率的な合成を補助するAI開発への貢献が期待される。



タンパク質構造安定性に重要な部位を示した「解剖図」

### 発表概要

東京大学 生産技術研究所 坪山 幸太郎 講師（研究当時、ノースウェスタン大学、ポスドク）と、ノースウェスタン大学 ガブリエル ロックリン 助教 らによる研究グループは、タンパク質の構造安定性を効率よく測定する方法の樹立に成功しました。

タンパク質の構造安定性とは、タンパク質が特定の機能的な構造をどれだけ保ちやすいかを示す指標で、機能を示すタンパク質分子の割合を規定するため、非常に重要な性質の1つです。今までは、一度の実験で1種類のタンパク質の構造安定性しか測定できなかったため、それらと比較、検証するためには多くの時間と費用がかかっていました。そこで、タンパク質のアミノ酸配列情報をDNA配列情報へと変換する工夫などにより、約90万種類までのタンパク質の構造安定性をまとめて一度の実験で測定することに成功しました。今まで測定されたタンパク質の構造安定性を統合したデータベースは約3万種類のタンパク質の情報にとどまっていたので、過去に調べられた構造安定性データの累計に比べても、数十倍のデータを一度の実験で取得することができます（図1）。

近年、ChatGPTをはじめとする深層学習モデル（注1）を基礎としたAIが目覚ましい発達を遂げています。タンパク質科学でもAIの導入が進んでいますが、タンパク質の性質を予測するAIの構築には、膨大なデータを必要とします。今回の研究で取得された、構造安定性についての大規模なデータは、タンパク質科学におけるAI開発のための基盤情報として役立つと考えられます。そのようなAIは、例えば疾患の原因となるアミノ酸変異の特定やタンパク質医薬のより効率的な合成を補助することが期待されます。

本研究成果は、7月20日（日本時間）に「Nature」で公開されます。

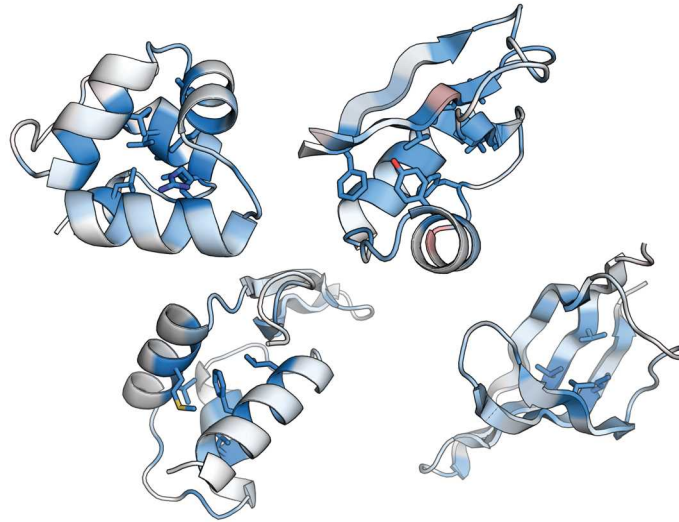


図1：タンパク質構造安定性に重要な部位を示した「解剖図」

タンパク質における、それぞれの位置の構造安定性に対する重要性を示す（青色ほど重要で、赤色ほど重要でない）。タンパク質における構造安定性に重要な位置を示すこのような図を1つ作るために、おおよそ1000-2000種類のアミノ酸配列の構造安定性を測定する必要があるため、従来の手法ではこのような構造安定性の解剖図を作成することは困難であった。

## 発表内容

### 〈研究の背景〉

タンパク質は、20種類のアミノ酸が多数連なった物質で、視覚や聴覚、運動機能などあらゆる生物学的な活動の基礎を担っています。生体内のタンパク質に関する最も基本的な概念として、「アミノ酸配列がそのタンパク質の構造を決め、さらにその機能を決定する」という「アンフィンセンのドグマ（注2）」が知られています。この原理で示されているように、ほぼ全てのタンパク質の機能は、構造によって決まっています。また、ほとんどのタンパク質は、構造を持たない解けた状態ときちんと折り畳まれた構造を持つ状態を含む複数の状態を行き来しています。したがって、きちんと折り畳まれた構造をもつ分子の割合を示す「タンパク質の構造安定性」という特徴量は、機能的なタンパク質の割合を示すことになり、タンパク質の最も重要な特徴量の1つです。実際に、タンパク質の構造安定性が低下すると、タンパク質同士が意図しない相互作用をしてしまったり、機能的なタンパク質が不足してしまったりすることにより、がんなど様々な疾患を引き起こすこともあります。

上記のような重要性にも関わらず、タンパク質の構造安定性を正確に理解し、タンパク質のアミノ酸配列や構造から予測することは困難でした。その理由として、タンパク質の構造安定性測定にかかる多大な時間と費用が挙げられます。例えば、これまでタンパク質の構造安定性を測定するには、大腸菌などを使って1つ1つのタンパク質を作製し、精製、そして、タンパク質の構造を壊す変性剤を様々な濃度に加え、その偏光スペクトルを測定するなど、時間のかかる実験工程を経る必要がありました。しかも、それぞれのタンパク質について、全てのステップを独立に行わなくてはならず、多数のタンパク質の構造安定性を一度に測定することができませんでした。

そのため、過去数十年に渡って測定されているタンパク質の構造安定性のほとんどを統合したデータベースですら、約3万種類のタンパク質の情報にとどまっています。また、それらは様々な論文データを統合していることから、その測定条件やデータの質は様々であり、タンパク質の構造安定性を統合的に理解することは困難でした。

### 〈研究の内容〉

そこで、本グループは、タンパク質の構造安定性を効率よく測定する手法を開発しました。

本手法の開発にあたり、大きく 2 つの工夫を凝らしました。1 つ目は、タンパク質のアミノ酸配列情報を DNA の核酸情報に変換することです。次世代 DNA シーケンス（注 3）の発達により、アミノ酸配列を読み取るよりも DNA 配列のほうがより正確かつ大量に解析することが可能になりました。そのため、タンパク質とそれに対応する cDNA をそれぞれ結びつけることが可能な cDNA display 法（注 4）を用いています。この方法で cDNA の配列を読み取ることにより、アミノ酸配列を特定することができます。cDNA display 法と次世代 DNA シーケンスの 2 つの技術を組合せることにより、非常に多くの種類のタンパク質のアミノ酸配列を一度にまとめて解析することが可能となりました。

2 つ目は、タンパク質の構造安定性を定量するために、タンパク質の切断酵素であるプロテアーゼ酵素を用いたことです。プロテアーゼは、ほどけてしまった不安定な状態のタンパク質を切断することができるため、構造安定性が高いタンパク質はゆっくりと、構造安定性が低いタンパク質については、素早く切断することができます。すなわち、タンパク質が切断される速度を測定することにより、タンパク質の構造安定性を測定することが可能となります。

以上のような工夫により、一度の実験で約 90 万種類のタンパク質の構造安定性を測定することに成功しました。このような実験を数回繰り返し、その中から質の高いデータを選定し、約 80 万種類のタンパク質の構造安定性のデータセットを取得し、公開しました。過去の文献を統合したタンパク質の構造安定性に関するデータベースは、約 3 万種類の情報に留まっていたのに対し、本手法ではその約 25 倍以上の大きさのデータベースを新たに提供したことになります。大きさもさることながら、測定条件が同一であることも、本データベースの有用性を高めています。

さらに本研究では、天然タンパク質や人工タンパク質約 500 種類において、それぞれのタンパク質についての構造を保つ仕組みについて調べました。そのために、1 つ 1 つのタンパク質において、それぞれの位置におけるアミノ酸を別の 20 種類のアミノ酸へと置換した場合とそれぞれのアミノ酸を欠損させた場合、そしてグリシンもしくはアラニンを挿入した場合の構造安定性を測定しました。これにより、それぞれのタンパク質において、どの部位が構造維持に重要なのかなど、構造安定性を可視化することが可能となりました（図 2）。

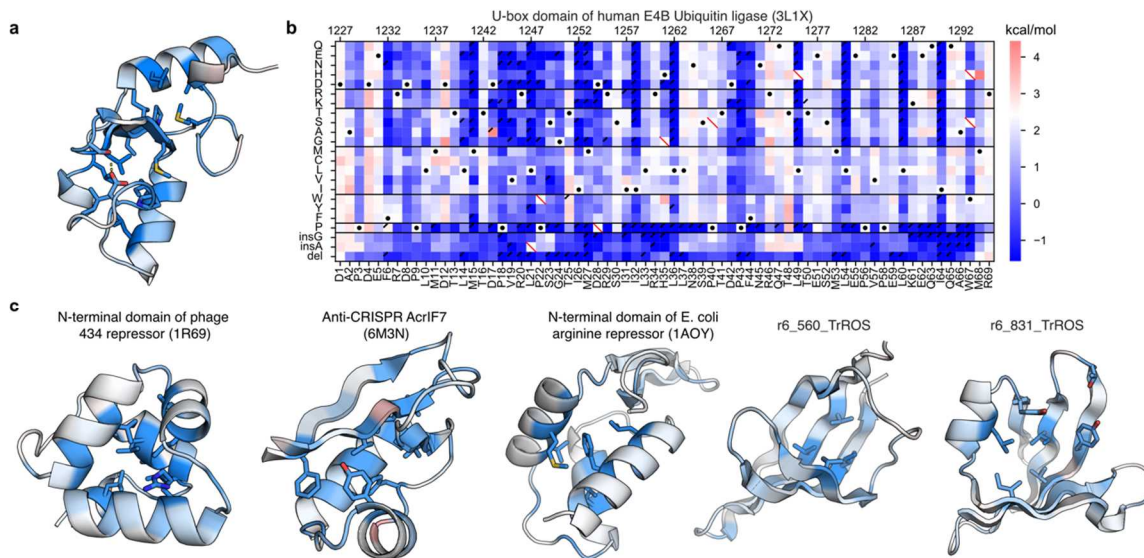


図 2: タンパク質の構造安定性の解剖図

(a, c) タンパク質のそれぞれの位置の構造安定性に対する重要性を示す（青色ほど重要で、赤色ほど重要でない）。(b) (a) のタンパク質の構造図の元データ。野生型のタンパク質の構造安定性（白・黒点）と比べてそれよりも構造安定性が低下する（青）、もしくは安定化する（赤）ことを示す。1 種類のタンパク質の「解剖図」を作成するために、千種類以上のアミノ酸配列の構造安定性を測定する必要がある。横軸はタンパク質中の位置を、縦軸はアミノ酸の置換や欠損、挿入を示す。

また、生物内のタンパク質での各アミノ酸の使われやすさと、アミノ酸がもたらす構造安定性との関係性についても調べてみました。予想通り、構造を安定化させるアミノ酸ほど、よく使われるという傾向があることがわかり、その傾向を定量化することができました。加えて、構造の安定性による影響を排除した際のアミノ酸の使われ方についても調べてみたところ、水に溶けやすい親水性のアミノ酸はより使用されやすい一方で、水に溶けにくい疎水性アミノ酸、特に芳香環を含む芳香族アミノ酸は使われにくい傾向にあることがわかりました(図 3)。これは、芳香族アミノ酸の合成コストが高いこと、また、芳香族アミノ酸を含む疎水性アミノ酸は可溶性を損なうことなどに起因すると考えられます。

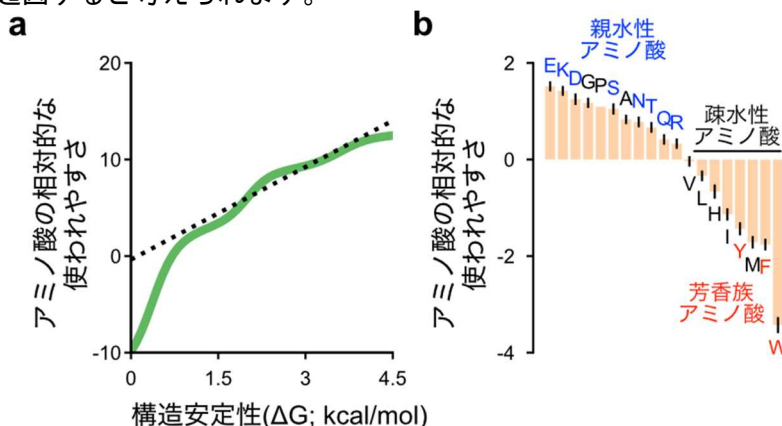


図 3: 天然タンパク質におけるアミノ酸使用のルール

(a) 構造安定性 (X 軸) とアミノ酸の相対的な使われやすさ (Y 軸) との関係性を示す。緑色の線が左下から右上へ観察されていることから、構造安定性を高めるアミノ酸ほど (X 軸右向き)、より使われやすい (Y 軸上向き) ことが見て取れる。(b) (a) のデータを用いて、構造安定性による影響を排除した際の、それぞれのアミノ酸の使われやすさを示す。アルファベットは 20 種類のアミノ酸を示す。青色は親水性アミノ酸、赤色は疎水性アミノ酸であり、右側に存在するアミノ酸は全て疎水性アミノ酸である。



本データベースの規模が従来に比べて圧倒的に大きいため、タンパク質の構造安定性を規定する重要な性質や、構造安定性に関わる一般則などを提唱することが可能となりました。

#### 〈今後の展望〉

近年、タンパク質科学において、深層学習モデルを基礎とする AI が積極的に導入され、大きな成功を収めています。しかし、そのような AI を構築するためには、タンパク質の性質に関する膨大なデータが事前に必要です。本研究において取得された、高品質かつ大規模なタンパク質の構造安定性データは、アミノ酸配列からタンパク質の構造安定性を正確に予測するような AI の構築に役立つと期待されます。タンパク質の構造安定性は、タンパク質の発現量や細胞内の存在量などの様々な重要な性質に影響を与えるため、タンパク質の合成量やがんなどの疾患とも密接な関係があります。本データベースを利用して、タンパク質の構造安定性を正確に予測することができれば、そのようなタンパク質の合成量の正確な予測やがんなどの疾患の原因の特定がより容易になることが期待されます。したがって、本研究成果は、病気の原因となるタンパク質のアミノ酸変異を予測するための AI 開発に繋がることから、工学や臨床医学におけるタンパク質の応用を下支えする可能性を秘めています。

#### 発表者

東京大学

生産技術研究所

坪山 幸太郎 (講師) <研究当時: ノースウェスタン大学 ポスドク>

ノースウェスタン大学

ファインベルグ医学校

ガブリエル ロックリン (助教)

#### 論文情報

〈雑誌〉 Nature

〈題名〉 Mega-scale experimental analysis of protein folding stability in biology and design

〈著者〉 Kotaro Tsuboyama, Justas Dauparas, Jonathan Chen, Elodie Laine, Yasser Mohseni Behbahani, Jonathan J. Weinstein, Niall M. Mangan, Sergey Ovchinnikov, Gabriel J. Rocklin

〈DOI〉 10.1038/s41586-023-06328-6

#### 研究助成

本研究は、Northwestern University Startup Funding (Gabriel Rocklin Assistant Professor)、JST さきがけ「人工タンパク質による、高次構造体の自由自在な解体・分解 (課題番号: JPMJPR21E9)」の支援により主に実施されました。

## 用語解説

### （注1）深層学習モデル

ニューラルネットワークと呼ばれる脳の神経回路を模倣したような構造を持つ機械学習モデル。ネットワークの層を多層化、複雑化させることで、囲碁などを含む多様な課題を高精度で解くことが可能になっている。

### （注2）アンフィンセンのドグマ

アンフィンセンによって、約半世紀前に提唱されたタンパク質の基本概念。「タンパク質のアミノ酸配列が構造を決め、その構造が機能を決定する」という法則であり、生物学にてごく一部の例外を除き広く受け入れられている。

### （注3）次世代 DNA シーケンス

一度の解析で、百億程度までの DNA 分子の配列を同定可能な基盤技術。各 DNA 配列の数を正確に定量できるため、下記の cDNA display 法と組合せることで、多数のタンパク質の存在量を正確に定量することが可能となる。

### （注4）cDNA display 法

Yamaguchi NAR 2009 で提案された、試験管内でタンパク質と対応する cDNA を結びつける方法。cDNA 配列を解読すれば、そのタンパク質のアミノ酸配列を解読できる。

## 問合せ先

### 〈研究に関する問合せ〉

東京大学 生産技術研究所

講師 坪山 幸太郎（つぼやま こうたろう）

Tel : 03-5452-6333 E-mail : ktsubo[at]iis.u-tokyo.ac.jp

### 〈報道に関する問合せ〉

東京大学 生産技術研究所 広報室

Tel : 03-5452-6738 E-mail : pro[at]iis.u-tokyo.ac.jp

科学技術振興機構 広報課

Tel : 03-5214-8404 E-mail : jstkoho[at]jst.go.jp

### （JST 事業に関すること）

科学技術振興機構 戦略研究推進部 ライフイノベーショングループ

保田 睦子（やすだ むつこ）

Tel : 03-3512-3524 E-mail : presto[at]jst.go.jp