



## タンパク質設計問題を解決する数学公式を発見

～圧倒的に高速な設計により、創薬・新素材・新デバイス開発への応用が期待～

### 【本研究のポイント】

- ・生物学、情報学、物理学の理論を融合させることにより、タンパク質設計問題の解を与える数学公式を発見した。
- ・この数学公式を用いることにより、タンパク質設計を含む難しい逆問題に必須だった膨大な計算量のシミュレーションの必要がなくなり、圧倒的に高速な設計が可能になった。
- ・一般的な逆問題、新素材・新デバイスの設計問題への応用も期待される。

### 【研究概要】

国立大学法人東海国立大学機構 名古屋大学大学院情報学研究科の高橋 智栄 博士後期課程学生 (JST 融合フロンティア次世代リサーチャー)、時田 恵一郎 教授、大学院工学研究科の千見寺 浄慈 助教らの研究グループは、シミュレーションや実験をせずにタンパク質設計をするための数学公式を世界で初めて導出することに成功しました。

創薬の分野において、必要な機能や薬効を発揮する新しいタンパク質の設計は、タンパク質の設計図であるアミノ酸配列を少しずつ変えながら、実験や構造探索シミュレーションを実行し、出来上がったタンパク質が必要な機能や薬効を発揮するかをチェックするという、非常に時間のかかる作業が繰り返されてきました。

本研究で導出した数学公式を用いることで、膨大な計算量を必要とする配列探索シミュレーションや構造探索シミュレーションの必要がなくなり、従来手法に比べて計算コストが圧倒的に低いタンパク質設計が可能になることが期待されます。さらに、本手法はタンパク質の詳細な性質に依存しないので、新素材や新デバイスの設計問題などへの応用も期待されます。

本研究成果は、2022年10月12日付イギリス物理学会誌「Journal of Statistical Mechanics: Theory and Experiment」に掲載されました。

## 【研究背景と内容】

創薬の分野では、必要な機能や薬効を発揮する新しいタンパク質の設計が行われています。これは、与えられたアミノ酸配列をもつタンパク質がどのような立体構造に折りたたむかを考える、いわゆるタンパク質フォールディング問題とは反対向きの、与えられた立体構造に折りたたむアミノ酸配列を探索するという問題です。

従来は、タンパク質の設計図であるアミノ酸配列を少しずつ変える「逆問題ループ」シミュレーションと、実験や立体構造探索シミュレーションの「順問題ループ」シミュレーションの「二重ループ」を実行して、与えられた立体構造が熱力学的に安定になるかをチェックするという作業を繰り返していたため、正確な設計には膨大な計算量を必要としていました。

順問題シミュレーションでは、最近では AlphaFold (タンパク質の構造予測を実行する人工知能プログラム) などの AI を用いた手法により短時間で実行することが可能になりつつありますが、タンパク質設計問題においては、アミノ酸配列を探索する逆問題ループの中で順問題ループを膨大な回数繰り返す必要があり、今なお数百、数千残基のアミノ酸を持つ巨大なタンパク質の精確な設計はほぼ不可能といった状況です。

しかし、昨年本グループが発表した論文 (Takahashi, Chikenji & Tokita, Phys. Rev. E, 104:014404, 2021) (名古屋大学研究成果発信サイト 2021 年 8 月 25 日付、<https://www.nagoya-u.ac.jp/researchinfo/result/2021/08/ai.html>) により、タンパク質進化に関する新しい仮説とベイズ学習<sup>注1)</sup> を組み合わせることで、順問題ループのシミュレーションが不要となり、従来手法に比べて圧倒的に高速なタンパク質設計を実現することに成功しました。そこでは、逆問題ループの計算にマルコフ連鎖モンテカルロ法<sup>注2)</sup> を用いりましたが、本研究では、キャビティ法 (空洞法)<sup>注3)</sup> という複雑系物理学・情報統計力学の理論を適用することにより、マルコフ連鎖モンテカルロ法によるシミュレーションも不要なアミノ酸配列の推定のための数学公式を導出することに成功しました。

また、本論文においては、昨年発表した論文で導入されたタンパク質進化の仮説 (ベイズ学習における事前分布<sup>注4)</sup>) が、複雑系物理学や情報統計力学において知られる、いわゆる「西森ライン」<sup>注5)</sup> に相当すること、さらにタンパク質設計問題が情報理論の誤り訂正符号に対応し、本手法によるアミノ酸配列の推定がベイズ最適になっていることも示されました。

## 【成果の意義】

本研究において導出された数学公式を用いることにより、タンパク質設計にかかる時間を昨年発表した効率的な設計手法を用いた場合と比べても、さらに 10 分の 1 程度に短縮することができました。さらに、得られた数学公式はアミノ酸残基ごとの並列計算が可能な形式になっているため、実際の創薬がターゲットとするような巨大なタンパク質の設計に応用する際に計算量の爆発の問題が起こりにくいというメリットもあります。また、本手法はタンパク質の詳細な性質に依存しないので、新素材・新デバイ

スの設計問題などへの応用も期待されます。

このように、タンパク質フォールディングのような複雑な生命現象とその設計問題を、ベイズ学習という最新の情報学的手法や、昨年ノーベル物理学賞の複雑系物理学・情報統計力学的手法を用いて定式化・解析し、数学公式を発見したことは、生物学、情報学、物理学の理論の融合とそのシナジー効果としても意義深く、2017年に新設された名古屋大学大学院情報学研究科複雑系科学専攻が目指す「情報学と自然科学の融合」の最新成果のひとつであるとも考えられます。

本研究は、JST 次世代研究者挑戦的研究プログラム (JPMJSP2125)、東海国立大学機構融合フロンティア次世代研究事業 (融合フロンティア次世代リサーチャー) および科学研究費補助金 (19H03166, 19K03650) の支援のもとで行われたものです。

### 【用語説明】

注 1) ベイズ学習:

ベイズ統計に基づくシンプルかつ効率的な機械学習の手法。ベイズ統計学は、古典的な頻度論主義の統計学と異なり、統計モデルのパラメータ (タンパク質設計の場合はアミノ酸配列) までも何らかの確率分布 (事前分布) に従うとする。そして、観測されたデータ (タンパク質設計の場合はタンパク質の立体構造) を固定した上で、パラメータのデータによる条件付き確率 (事後分布) を求め、事後分布をもとにパラメータの推定や予測を行う。マルコフ連鎖モンテカルロ法などの効率的なサンプリング手法の発達により、事後分布からの推定や予測が容易になり、ベイズ学習はその応用範囲を拡大させている。本研究はタンパク質設計問題にベイズ学習を応用した最初の例である。

注 2) マルコフ連鎖モンテカルロ法:

マルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo methods, 通称 MCMC) は、求める確率分布を均衡分布として持つマルコフ連鎖 (確率過程の一種であるマルコフ過程のうち、とりうる状態が離散的なもの) を生成することにより確率分布のサンプリングを行うアルゴリズムの総称である。同時事後分布に従う乱数を継時的に生成して実行する。代表的な MCMC としてメトロポリス・ヘイスティングス法やギブスサンプリングがある。

注 3) キャビティ法 (空洞法):

統計物理学における平均場モデルを解くために、マーク・メザールとジョルジョ・パリージ (2021 年度ノーベル物理学賞受賞) が 1987 年に発表した数学的手法。多くの無秩序系の物性理論的問題や組合せ最適化問題における基底状態やその特性を計算するために使われ成功を収めてきた。本研究は、タンパク質設計問題にキャビティ法を応用した最初の例である。

注4) 事前分布:

統計モデルのパラメータが従うとする確率分布のこと。データを観測して比較的客観的に仮定できる統計モデル(尤度関数)に対し、データが観測される以前に利用可能な情報をもとに仮定する必要がある。ドメイン知識からの知見やそれに基づく分析者の直感、あるいは事後分布の計算のしやすさなどから決定される場合が多い。あるいは考えている状況や経験事実から明らかな場合などもある。本研究では、アミノ酸配列の事前分布は、フォールディング時の温度と同じ「進化温度(アミノ酸残基の変異のしやすさ)」をもつボルツマン分布に従うと仮定する。このことは、対応する自由エネルギーを最小にするアミノ酸配列が進化しやすいというタンパク質進化の仮説に対応する。

注5) 西森ライン:

1980年代に西森秀稔が発見した、あるランダムスピン系における内部エネルギーの厳密解が得られるパラメータ平面上の曲線。この四半世紀ほどの間に、西森ラインが情報理論においても重要な役割を果たすことが示されており、特に、誤り訂正符号問題において西森ライン上での復号がベイズ最適を達成することなどが証明されている。本研究は、西森ライン上でタンパク質設計を行った最初の例である。

**【論文情報】**

雑誌名: Journal of Statistical Mechanics: Theory and Experiment

論文タイトル: The cavity method to protein design problem

著者: Tomoei Takahashi, George Chikenji and Kei Tokita

DOI: 10.1088/1742-5468/ac9465

URL: <https://iopscience.iop.org/article/10.1088/1742-5468/ac9465>