

# タンパク質デザインの統計力学

高橋 智栄 (東京大学大学院理学系研究科知の物理学研究センター takahashi-tomoei@g.ecc.u-tokyo.ac.jp)

千見寺 浄慈 (名古屋大学大学院工学研究科 chikenji@bmp.ap.pse.nagoya-u.ac.jp)

時田 恵一郎 (名古屋大学大学院情報学研究所 tokita@i.nagoya-u.ac.jp)

「デザイン」と書くと、美術作品や工業製品を創造する作業を思い浮かべる方がおられるかもしれない。しかしここではデザインをより一義的な意味、すなわち、何らかの目的や機能を持つものの設計という意味で用いる。

ではタンパク質デザインとは何か？ タンパク質とは、アミノ酸がひも状につながったものがぐにゃぐにゃと折れ曲がり3次元的な構造をなすことで様々な生物学的機能を果たす高分子である。各タンパク質が持つ特異的な3次元構造のことをタンパク質立体構造という。タンパク質が立体構造をとった状態は、生理条件下ではアミノ酸のつながりのパターンであるアミノ酸配列のみで決まる熱平衡状態であると考えられており、これを**アンフィンセンのドグマ**という。したがってタンパク質デザインとは、タンパク質が何らかの機能や構造を持つようにアミノ酸配列を決定することである。

コンピュータによるタンパク質デザインには2024年のノーベル化学賞が授与されており、世界的にも注目される極めて重要なテーマである。

タンパク質デザインはタンパク質立体構造予測問題の逆問題にもなっている。タンパク質立体構造予測とは、ある与えられたアミノ酸配列でできたひもが折れたたむ立体構造の形を予測することである。また逆問題とは、出力から逆向きに入力を推論する問題である。例えば、M. Kacの“Can one hear the shape of drum?”で議論された太鼓の音からその形を推論することなどが挙げられる。

我々は、粗視化タンパク質モデルである格子タンパク質に対して、タンパク質デザイン問題を、ベイズ学習によって定式化し、統計力学の確率モデルと解析手法を用いて

解くことを試みた。ベイズ学習とは、ベイズ統計の方法に基づく機械学習の方法である。ベイズ統計では観測データの確率分布に含まれるパラメータ(あるいは潜在変数)さえも確率変数であると考え、そして、観測データが得られた上でのパラメータの事後的な確率分布に基づいて推論を行う。パラメータの先験的な確率分布のことを事前分布といい、データが得られた上でのパラメータの事後的な確率分布を事後分布という。

そして我々は、我々が独自に提案する、**配列重率仮説**という進化したアミノ酸配列が満たすべき性質に基づく事前分布を仮定することで、先行研究において最大の計算量的困難であった立体構造パターン探索をスキップした事後分布を導出する。配列重率仮説とは筆者らが提唱するアミノ酸配列の進化に関する新しい仮説である。この仮説は、タンパク質を構成するアミノ酸の配列、つまり進化を遂げた配列は、淘汰された配列に比べてある自由エネルギーをより低減させ、対応するボルツマン重率をより大きくすると考える。さらに、確率分布を確率変数の効率的なサンプリングによって近似する手法であるマルコフ連鎖モンテカルロ法と、Cavity法と呼ばれる平均場近似を拡張した計算手法によって、導出した事後分布から、与えられた格子タンパク質構造に折れたたむアミノ酸配列を推定し、2次元の比較的小さな**格子HPモデル**に対しては高精度でデザインが成功すること、またマルコフ連鎖モンテカルロ法とCavity法による結果がほぼ同等となることも見出した。我々が提案する手法はタンパク質モデルの詳細に依存しない手法であり、より現実的なタンパク質モデルに対しても拡張可能である。

## 用語解説

**アンフィンセンのドグマ：**  
タンパク質の立体構造(天然構造)は生理条件下においてアミノ酸配列だけで決まるとするテーゼ・教義。1960年代のAnfinsenによる、あるタンパク質の構造を変性させても変性物質を取り除くと元の立体構造に戻るという実験に基づく。

**配列重率仮説：**



アミノ酸配列で決まるタンパク質の自由エネルギーが低い領域

配列重率仮説において想定されるアミノ酸配列空間の概念図。灰色で塗られた空間がアミノ酸配列のあらゆるパターンに対応する空間(配列空間)である。その中の一部分である緑色および濃い緑色の領域のアミノ酸配列ほど、アミノ酸配列によって決まるある自由エネルギーが低く、それらが生物の体の中で進化してきたと考える。

**格子HPモデル：**  
タンパク質の構造を格子上の自己排除ウォークで表現し、アミノ酸を格子点で表すタンパク質の粗視化モデル。ここで自己排除ウォークとは格子(あるいはグラフ上)の一度通った点を通らない経路のことである。アミノ酸の種類も水分子に反発する疎水性と水分子と結合する親水性の二つに簡単化する。