**研究成果報告書：変異体探索 人工酵素設計PJT**

目次

1. **はじめに2**

1.1　本書の位置づけ2

1.2　研究の背景2

1. **理論、システム構成2**

2.1　最適化問題2

2.2　最適化手法3

1. 実験結果**4**
2. 考察**4**
3. まとめ**4**
4. 参考文献**4**
5. 付録**4**
6. **はじめに**
   1. **本書の位置づけ**

本書は、人工酵素設計PJT（2020年3月～2022年10月）の要素技術の一つである「変異体探索」の調査研究の概要・実施結果について報告を行うものである。

* 1. **研究の背景**

本PJTでは、酵素の産業応用上の課題として、特許発明に基づく酵素利用制限や、天然物の機能利用制限だと捉え、その解決のために、酵素の探索だけでなく酵素の人工設計が鍵となると考えた。さらに、有用な酵素を人工的に設計する上で、天然酵素のアミノ酸残基の改変によって変異体を生成するアプローチが期待できる。本PJTでは、このアプローチに従って、天然酵素の特徴を継承しながら、アミノ酸配列の類似度が一定以下の新規の変異体を生成することを目標としていた。

一方、変異数が多いと、天然タンパク質（Wild-Type）から大きく離れるため、タンパク質立体構造のフォールドが不安定になる、本来の機能を失うなどの可能性が高くなる。よって、変異数を制限しながら、優良なWild-Typeの変異体候補を効率的に生成する方法が必要だと考えられる。

Rosettaは、酵素設計のためのRosetta Enzyme Designソフトウェアである。その机上設計プロトコルとして、FastDesignやCartesian DDGがある。FastDesignは、Wild-Typeのアミノ酸配列の配列長や残基を変更し、エネルギースコア（Rosettaスコア）が最小となるような変異体を生成する。また、Cartesian DDGは、Wild-Typeのアミノ酸配列の変異位置・残基を入力すると、それに整合するように最小化されたエネルギースコア（Rosettaスコア）と構造データを出力する。しかしながら、FastDesignには変異数を制限する機能がない、あるいはCartesian DDGはユーザが変異位置・残基を指定する必要があることが要因で、これらの設計プロトコルは、変異数を制限しながら、Wild-Typeの変異体候補を効率的に生成することを実現しにくい。

以上から、本研究では、変異数制約を満たすWild-Typeの変異体候補を探索・生成する方法を確立した。具体的には、Cartesian DDGに変異数制約を加えた上位最適化を組み込むことで、全探索やランダム探索よりも効率的に変異体を生成する。

1. **理論、システム構成**
   1. **最適化問題**

類似度が一定以下の変異体生成のために検討した技術について説明する。図1に今回検討した技術の概要を示す。指定した変異数内の変異位置・残基の組を、上位最適化の最適化変数に設定し、目的関数値であるエネルギースコア（Rosettaスコア）を最小化するような、最適化問題を構築した。この最適化問題を全探索やランダム探索よりも効率的なアルゴリズムで解くことで、類似度が一定以下かつエネルギースコアが良い変異体を得ることが期待できる。

以降は、最適化問題と最適化手法について説明する。構築した組合せ最適化問題は式として定式化される。

ただし、は目的関数、は離散的な実行可能解の集合、は組合せ的な構造を有する最適化変数（解）である。本研究では、アミノ酸残基の20種・配列長から構成される各アミノ酸配列が解である。また、は、任意のアミノ酸配列が、基準とする天然アミノ酸配列と異なる残基数を出力する関数であるため、式の制約条件は、基準とする天然アミノ酸配列から変異以下のアミノ酸配列に制限することを意味する。したがって、全組合せは通りだが、制約条件を満たす組合せのみで構成したのが探索空間である。また、目的関数は、基準とする天然アミノ酸配列が任意のアミノ酸配列に変異したとき、Cartesian DDGで計算したときに得るエネルギースコアである。

　なお、天然配列からの変異に着目した表現として、変異位置と残基の組で表現する方法がある。例えば、2番目の残基をセリンSに置換、10番目の残基をアラニンAに置換する変異はと表現できる。以降ではこの変異表現を解の代わりに使用する。

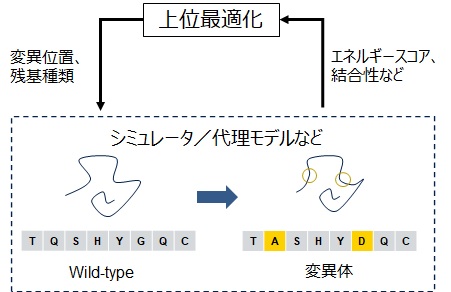


図1：変異体探索プロトコルの概要

* 1. **最適化手法**

図2に、配列空間における変異体探索のイメージを示す。最適化手法はLocal Searchを使用する。Local Searchは、単点探索アルゴリズムで、探索空間（配列空間）内の暫定解の近傍を生成し、その近傍内の目的関数値が改善する解に移動していく。近傍内の改善位置にしか移動しないため、初期位置に依存して、いずれかの局所解に収束する。本研究では、近傍で最も目的関数値が優れた解を移動先として選択するbest改善を使用した。一様乱数で与えた解から、複数並列のLocal Searchを適用した。

近傍の定義について説明する。本研究で使用した近傍は、Hamming距離が1以内かつ変異数が以下を満たす解の集合とした。これにより、近傍解は制約条件を満たす解に限定することを保証する。このとき、近傍のサイズ（最適化問題の次元数）は、摂動を加える解自身を除き、となる。例えば、最大4点変異、配列長36の場合、次元である。実際の近傍生成では、上記に加えて実応用上好ましくない解を除外することで、次元数を削減することが可能である。

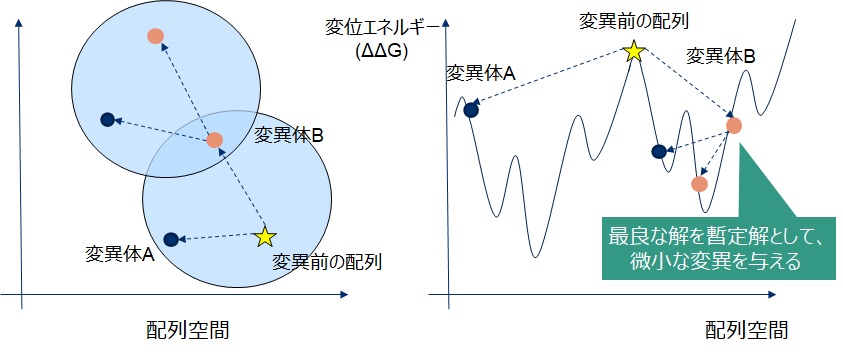


図2：配列空間における変異体探索のイメージ

1. **実験結果**

図3に、1CBHのWild-Typeに三点変異を指定した場合の探索で得た、変異体の立体構造例を示す。三点変異体のアミノ酸残基に変異した部位を紫色で示している。Wild-TypeのRosettaスコアは-127.9だが、三点変異体のRosettaスコアは-157.8であるため、大きく改善した変異体が得られていることが確認できる。

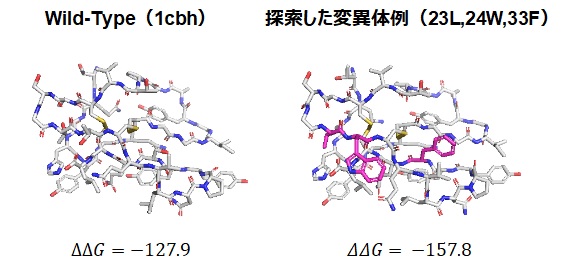


図2：1CBHの三点変異体の主鎖構造（紫色は変異部分）

1. **考察**
2. **まとめ**
3. **参考文献**
4. J. Jumper et al.: “Highly accurate protein structure prediction with AlphaFold”, Nature (2021)
5. M. Mirdita et al.:“ColabFold: Making protein folding accessible to all”, Nature (2021)
6. T. Kosugi et al.:“Solubility-Aware Protein Binding Peptide Design Using AlphaFold”, Biomedicines, Vol.10, No.1626 (2022)
7. 中林：「2019年度 共同研究最終報告書「人工セルラーゼ設計手法の開発に向けた要素技術の調査研究」」（2020）
8. **付録**