

戦略的創造研究推進事業 AIP 加速研究  
(AIP 加速 PRISM 研究)  
研究課題「ターゲット検証用ツール抗体の開発」

## 研究終了報告書

研究期間 2018年8月～2020年3月

研究代表者：津田 宏治  
(東京大学大学院新領域創成科学研究科、教授)

## § 1 研究実施の概要

### (1) 実施概要

本研究では、ウェットラボにおけるバイオテクノロジーによる抗体改良技術に機械学習を組み合わせることによって、従来よりも高速に有用な抗体を設計する機械学習支援抗体開発プロトコル Point-and-Optimize を構築した。このプロトコルは、ファージ提示法と呼ばれるバイオテクノロジー手法と、機械学習の一手法であるベイズ最適化を組み合わせたものである(下図)。ファージ提示法では、 $10^8$  種類程度のランダムに生成した抗体 DNA をファージのゲノムに組み込み、ターゲットに結合する抗体のみを濃縮できるパニング操作を繰り返すことで、望ましい抗体が濃縮された溶液を得る。濃縮された溶液には、 $10^5$ 種類程度の抗体が存在しており、含まれている数が多いほど、有望な抗体である。Illumina 社の Next-generation sequencer を用いて、溶液中のタンパク質の配列と、その出現頻度(濃度)のペアを作成し、機械学習用の訓練データとする。これを用いて、ベイズ推定法の一つであるガウス過程を学習し、ベイズ最適化を用いれば、原則的には新たな抗体配列を推薦することができる。しかし、抗体配列の可能性は膨大であり、ベイズ最適化をそのまま適用するには、計算量上の問題がある。また、膨大な空間から、少ないデータセットで探索を行うには、情報量の不足が懸念される。そこで、濃縮溶液をモノクローナル化し、200 種類程度の抗体を単離し、ターゲットとの結合強度を計測することで、一点から数点の有望な抗体(Positive Control)を得る(Pointing)。その上で、ベイズ最適化の探索空間を、この抗体の周辺(3変異以内)に限ることで、より狙いを絞った推薦を可能にしている。本プロトコルを、肺ガンを含む多くのガンのターゲットとして知られている EGFR に適用し、機械学習により推薦された抗体を含むスマートライブラリを作成した。

本研究のもう一つの目的は、ベイズ最適化よりも高度で、より性能の高い機械学習手法を設計することである。その目的を達するため、深層学習に基づいたペプチド自動設計ツール BOPP (Blackbox Optimization for Peptides and Proteins)を作成し、一般公開した。このツールは、分子動力学法などによるペプチドの評価と、深層強化学習を用いたペプチドの提案を繰り返すことで、次第にペプチドを最適化することが可能である。本手法の能力を確かめるため、細胞膜を透過するペプチドの設計を行なった。この題材は、データベースが充実しており、また、分子動力学計算による膜透過シミュレーションが可能のため、ベンチマークに適している。その結果、シミュレーション上で膜透過できるペプチドを多数設計し、数個に関しては検証実験を行なって、有望なペプチドの設計に成功した。本手法を抗体開発に用いるには、分子動力学計算が確立していないなど、まだハードルがあるが、今後に向けて、興味深い研究テーマである。

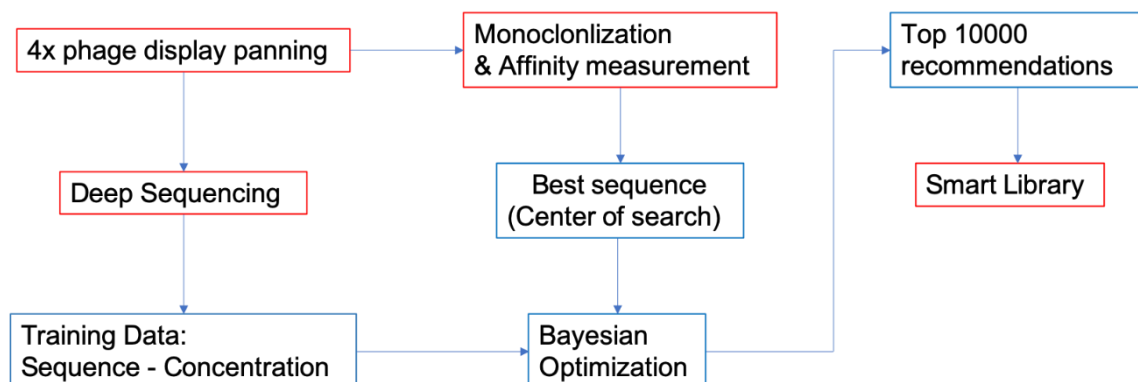


図: 機械学習支援抗体開発プロトコル Point-and-Optimize の概要。赤い四角が、ウェットラボによる作業であり、青い四角が、計算機上での機械学習に関わる部分である。

## (2) 顕著な成果

### < 優れた基礎研究としての成果 >

#### 1. 機械学習支援抗体開発プロトコル Point-and-Optimize の開発と検証

概要: 機械学習の一種であるベイズ最適化と、バイオテクノロジーを組み合わせた抗体設計プロトコル Point-and-Optimize を構築した。ファージ提示法によって、機械学習の訓練データを得ると同時に、モノクローナル化によって、一点から数点の Positive Control 配列も取得する。ラクダ由来抗体 VHH の39アミノ酸可変領域に対して、EGFR に結合する Positive Control を発見し、スマートライブラリを構築した。

#### 2. 深層学習と分子動力学計算によるペプチド自動設計ツール BOPP の開発

概要: 計算機上でタンパク質を設計することができる、新たな手法 BOPP を開発した。BOPP では、正例から学習された Recurrent Neural Network を用いてペプチド候補を多数生成し、正負例から学習された分類器によって数を絞りこんだ後、分子動力学計算によって検証するという手順をたどる。ベンチマーク問題として、細胞膜を透過する長さ 20 アミノ酸の設計に取り組んだ結果、一つのペプチドに関して実験で細胞膜透過を確かめることができた。

### < 科学技術イノベーションに大きく寄与する成果 >

#### 1. 複数の企業との共同研究開始

概要: 今回の課題に開発した機械学習によるタンパク質設計は、革新的なアプローチであり、多くの医療・材料・酵素関係の企業の関心を集めている。産総研においては、分子動力学計算と人工知能を組み合わせた蛋白質高機能化について、企業2社と共同研究契約を締結し、研究を開始している。東北大学では、2社と共同研究契約、1社と守秘義務契約を結んだ。

### < 代表的な論文 >

Teppei Niide, Noriyoshi Manabe, Hikaru Nakazawa, Kazuto Akagi, Takamitsu Hattori, Izumi Kumagai, and Mitsuo Umetsu, “Complementary Design for Pairing between Two Types of Nanoparticles Mediated by a Bispecific Antibody: Bottom-Up Formation of Porous Materials from Nanoparticles”, *Langmuir*, 3067–3076 (2019).

## § 2 研究実施体制

### (1) 研究チームの体制について

#### ① 津田グループ

研究代表者: 津田 宏治 (東京大学大学院新領域創成科学研究科 教授)

研究項目: 高性能実験計画法の開発

#### ② 梅津グループ

主たる共同研究者: 梅津 光央 (東北大学大学院工学研究科 教授)

研究項目: 機械学習・分子動力学計算支援による抗体機能化法の開発

#### ③ 亀田グループ

主たる共同研究者: 亀田 倫史 (産業技術総合研究所人工知能研究センター 主任研究員)

研究項目: 分子動力学計算による変異体開発支援

### (2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

機械学習によるタンパク質の設計については、各地での招待講演などを通して、多くの研究者に周知を試みている。また、企業からも大きな関心を得ており、産総研では、分子動力学計算と人工知能を組み合わせた蛋白質高機能化について、企業2社と共同研究契約を締結し、研究を開始している。東北大学では、本研究における機械学習支援を受けた進化分子工学操作について7社の企業から問い合わせがあり面談を行った。その中で2社は共同研究契約、1社は守秘義務契約を結び、さらに1社は共同研究契約へ向けて検討を行うことになった。また、医療関係の研究者とは、製薬企業へ導出が期待できるタンパク質について共同で研究を進めていくことになった。