

日本医療研究開発機構 創薬基盤推進研究事業 事後評価報告書

I 基本情報

研究開発課題名：機械学習による中分子医薬の多要素複合デザイン技術の開発

Development of AI-based multi-functional design technology for middle molecule drugs

研究開発実施期間：令和元年 12 月 1 日～令和 4 年 3 月 31 日

研究開発代表者 氏名：齋藤 裕

SAITO Yutaka

研究開発代表者 所属機関・部署・役職：

国立研究開発法人産業技術総合研究所・人工知能研究センター・主任研究員

Senior Researcher, Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST).

II 研究開発の概要

バイオ医薬品の開発では、標的との結合親和性を向上するだけでなく、免疫原性、構造安定性、溶解性などの多数の要素を最適化する必要がある。これまでに申請者らは、機械学習を利用してペプチド医薬の結合親和性を向上するデザイン技術の開発を進めてきた。本研究開発では、この技術を発展させて、結合親和性を含む多数の要素の同時改善を可能とする「多要素複合デザイン技術」を開発した。また、多要素複合デザイン技術の有用性を実証するために、悪性リンパ腫マーカー分子 X を標的とするペプチドをモデルとして結合親和性・溶解性・免疫原性の同時改善を試み、実験による検証まで行った。さらに、バイオ医薬品開発における多要素複合デザイン技術の実証例として、移植片対宿主病 (GVHD) の創薬ターゲット分子 Y を標的とするペプチドの開発を行った。これらのペプチドの開発においては、構造ゆらぎの制御に立脚した独自の創薬モダリティである Fluctuation-regulated Affinity Protein (FLAP) を使用した。

・マルチタスク学習によるペプチド医薬の多要素複合デザイン技術の開発

機械学習分野の手法であるマルチタスク学習を利用して、ペプチド医薬の多要素複合デザイン技術を開発した。対象の要素としては、結合親和性に加えて、免疫原性、構造安定性、溶解性を取り扱った (図 1)。

まず、ペプチドのアミノ酸配列情報から結合親和性などの多要素を同時に予測するマルチタスク予測手法を開発した。与えられた教師データに対して最適な特徴量と学習モデルを自動的に選択するシステムを開発した。特に、自然言語処理分野の最新の深層学習手法である BERT を応用した特徴量を新規に考案して導入した。本研究で開発した特徴量と学習モデルの最適化システムは、ペプチド以外のバイオ医薬品 (抗体、核酸など) にも応用可能であり、機械学習をもちいたバイオ医薬品デザインにおける重要な基盤技術となる。また、本研究で新規に考案した BERT にもとづく特徴量は、自然言語処理分野の最新の深層学習手法をタンパク質に応用した独創的な技術であり、論文採択 (Yamaguchi et al. *Brief Bioinform.* 2021)、国際会議 ISMB / ECCB 2021 での Best Talk Award を含む学会発表賞 5 件の成果をあげた。

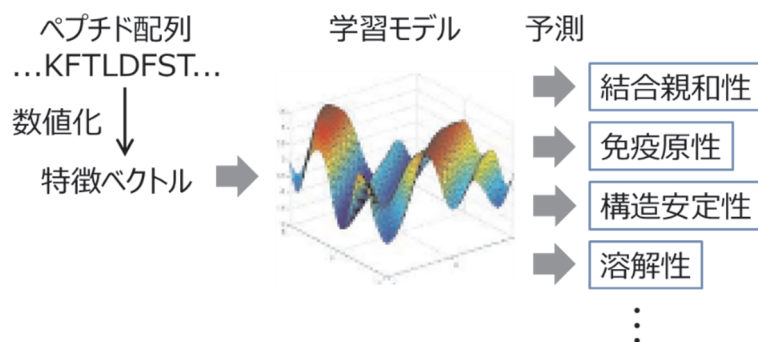


図 1. マルチタスク予測手法

次に、マルチタスク予測手法を遺伝的アルゴリズムやマルコフ連鎖モンテカルロ法などの探索アルゴリズムと組み合わせることで、膨大な配列空間にも対応可能にした。探索アルゴリズムの性能を評価するために、先行研究で取得した VHH 抗体の結合親和性データセット（配列数 12,723）をもちいた計算機実験を行った。VHH 相補性決定領域の 39 残基を全通りに変異させた場合の配列空間 20^{39} ($>10^{50}$) から、結合親和性予測値の高い配列を探索した。その結果、従来のランダム探索による手法と比較して、結合親和性予測値の高い変異体を効率よく発見できることが明らかになった。本研究で開発した探索アルゴリズムは、ペプチド以外のバイオ医薬品（抗体、核酸など）にも適用可能であり、特に変異残基数が多く配列空間が膨大となる場合の配列探索において必要不可欠な基盤技術である。

最後に、公開されているバイオ医薬品の物性データをもちいて、多要素複合デザイン技術の性能を評価した。ベンチマークデータとして、2016 年時点で臨床試験の第 2 相試験を突破している 137 個の抗体医薬について、12 種類の物性値（構造安定性、発現量、非特異的結合活性、凝集性など）を測定した公開データ（JPNAS. 114, 5, 944-949, 2017）をもちいた。このデータを教師データとして、配列から 12 種類の物性値を同時予測するマルチタスク予測手法を構築した。その結果、マルチタスク予測手法はシングルタスク予測手法よりも高い予測精度を達成した。これにより、多要素複合デザイン技術の有効性が示された。創薬におけるマルチタスク学習の利用は、これまで低分子医薬品を対象とした毒性予測などにおいて行われてきたが、バイオ医薬品への適用例は本研究が世界初である。この成果は、2021 年の生命医薬情報学連合大会において口頭発表しており、2022 年内に論文投稿を行う予定である。

・多要素複合デザイン技術による分子 X 標的 FLAP の結合親和性・溶解性・免疫原性の同時改善

悪性リンパ腫マーカー分子 X を標的とする FLAP の結合親和性・溶解性・免疫原性の同時改善をモデルとして、多要素複合デザイン技術の有用性を実証した。

FLAP 足場ペプチドのうち 14 残基をランダム化したライブラリを構築して、ほ乳類細胞をもちいたフェノタイプスクリーニングと次世代シーケンス解析によって、分子 X に結合する FLAP 群を回収した。この中で、リード数が多くパニングのラウンド間で顕著な濃縮率を示した配列 762 個を分子 X 標的 FLAP 群として、多要素複合デザイン技術の教師データとした。

取得した教師データをもちいて予測手法を構築して、探索アルゴリズムと組み合わせることで、 8.0×10^{17} という膨大な配列空間から探索を行った。探索された配列のうち結合親和性予測値の高かった上位 10,000 配列の中から、溶解性予測値の高い配列（X-H1、X-H2）および溶解性予測値の低い配列（X-L1、X-L2）を後の実験による評価のために選択した。これらの配列の免疫原性予測値は、高いものから低いものまで様々であった。このように、本研究で開発した多要素複合デザイン技術は、膨大な配列空間から所望の性質を有すると予測される配列を柔軟に探索することができた。

デザインした FLAP の結合親和性・溶解性・免疫原性を実験で評価したところ、溶解性は全て予測通りの結果となった。マウス脾臓細胞の IL-2 放出量を指標とする評価においては、いずれも免疫原性は認められなかった。また、溶解性の高い FLAP のうち 1 つは分子 X に強く結合することが明らかになった。以上の結果から、本研究で開発した多要素複合デザイン技術によって、バイオ医薬品として望ましい性質を有するペプチドをデザインできることが示唆された。

・多要素複合デザイン技術による分子 Y 標的 FLAP の開発

多要素複合デザイン技術をもちいて、移植片対宿主病 (GVHD) の創薬ターゲット分子 Y を標的とする FLAP を開発した。

分子 X の場合と同様に、ほ乳類細胞をもちいたフェノタイプスクリーニングと次世代シーケンス解析によって、分子 Y に結合する FLAP 群を回収した。この中で、リード数が多くパニングのラウンド間で顕著な濃縮率を示した配列 369 個を分子 Y 標的 FLAP 群として、多要素複合デザイン技術の教師データとした。

取得した教師データをもちいて予測手法を構築して、探索アルゴリズムと組み合わせることで、 8.0×10^{17} という膨大な配列空間から探索を行った。探索された配列のうち結合親和性予測値の高かった上位 10,000 配列の中から、溶解性予測値の高い配列 (Y-H1、Y-H2) および溶解性予測値の低い配列 (Y-L1、Y-L2) を後の実験による評価のために選択した。これらの配列の免疫原性予測値は、高いものから低いものまで様々であった。このように、本研究で開発した多要素複合デザイン技術は、膨大な配列空間から所望の性質を有すると予測される配列を柔軟に探索することができた。

デザインした FLAP の結合親和性・溶解性・免疫原性を実験で評価したところ、溶解性は全て予測通りの結果となった。マウス脾臓細胞の IL-2 放出量を指標とする評価においては、いずれも免疫原性は認められなかった。また、溶解性の高い FLAP は分子 Y に強く結合することが明らかになった。以上の結果から、本研究で開発した多要素複合デザイン技術によって、バイオ医薬品として望ましい性質を有するペプチドをデザインできることが示唆された。

・今後の課題と方向性

本研究で開発した多要素複合デザイン技術を FLAP 以外の中分子ペプチドや、抗体などの様々な機能性タンパク質に適用する。本技術は、バイオ医薬品の開発における「多要素の同時改善」という普遍的な問題に切り込むものであり、将来、様々な創薬モダリティの開発プロセスに大きなブレイクスルーをもたらすだろう。また、特徴量や学習モデルとして、日々進歩している最新の機械学習技術を随時取り入れることにより、予測精度のさらなる向上を目指す。

Biopharmaceutical development requires not only improving binding affinity to the target, but also optimizing a number of factors such as immunogenicity, structural stability, and solubility. The applicants previously developed a design technique to improve the binding affinity of peptide drugs using machine learning. In the present study, by extending this technology, we have developed a "multi-functional design technology" that enables simultaneous improvement of many factors including binding affinity. In order to demonstrate the usefulness of the multi-functional design technology, we attempted to simultaneously improve binding affinity, solubility, and immunogenicity using a peptide targeting the malignant lymphoma marker (the molecule X) as a model, and verified the results through experiments. In addition, peptides targeting the molecule Y, a drug target for graft-versus-host disease (GVHD), were developed as a demonstration example of multi-functional design technology in biopharmaceutical development. In the development of these peptides, we used fluctuation-regulated affinity protein (FLAP), a unique drug modality based on the control of conformational fluctuations.

Development of multi-functional design technology using multitask learning

We developed a multi-factorial complex design technique for peptide drugs using multitask learning, a method in the field of machine learning. In addition to binding affinity, immunogenicity, structural stability, and solubility were treated as target elements (Figure 1).

First, we developed a multi-task prediction method that simultaneously predicts multiple factors such as binding affinity from peptide amino acid sequence information. We developed a system that automatically selects the best features and learning model for given training data. In particular, we devised and introduced new features based on BERT, a deep learning method in the field of natural language processing. The optimization system for features and learning models developed in this study can be applied to biopharmaceuticals other than peptides (antibodies, nucleic acids, etc.), and is an important fundamental technology for biopharmaceutical design using machine learning. The BERT-based features newly developed in this study are an original technology that applies a deep learning method in the field of natural language processing to proteins, and have been accepted for publication (Yamaguchi et al. *Brief Bioinform.* 2021) and received the 5 conference presentation awards, including Best Talk Award at the international conference ISMB / ECCB 2021.

Simultaneous improvement of binding affinity, solubility, and immunogenicity of FLAPs targeting the molecule X by the multi-functional design technology

We demonstrated the utility of a multi-functional design technique by the simultaneous improvement of binding affinity, solubility, and immunogenicity of FLAPs targeting the malignant lymphoma marker (the molecule X) as a model.

We constructed a library of FLAP scaffold peptides where 14 residues were randomized, and selected a set of FLAPs that bind to the molecule X by phenotypic screening using mammalian cells and next-generation sequencing analysis. Among them, 762 sequences with a large number of reads and a significant enrichment rate between rounds of panning were selected as FLAPs targeting the molecule X, and used as training data for the multi-functional design technology.

Using the obtained training data, we constructed a prediction method and combined it with a search algorithm to search from a huge sequence space of 8.0×10^{17} . Among the top 10,000 sequences with high predicted binding affinity, sequences with high predicted solubility (X-H1, X-H2) and sequences with low predicted solubility (X-L1, X-L2) were selected for later experimental evaluation. The predicted immunogenicity of these sequences varied from high to low. Thus, the multi-functional design technology developed in this study was able to flexibly search for sequences predicted to have the desired properties from a vast sequence space.

The binding affinity, solubility, and immunogenicity of the designed FLAPs were evaluated experimentally, and the solubility was all as predicted. No immunogenicity was observed in any of the evaluations using IL-2 release from mouse spleen cells as an indicator. In addition, one of the highly soluble FLAPs was found to bind strongly to the molecule X. These

results suggest that the multi-functional design technology developed in this study can be used to design peptides with desirable properties for biopharmaceuticals.

Development of FLAPs targeting the molecule Y using multi-functional design technology

We developed FLAPs targeting molecule Y, a drug target for graft-versus-host disease (GVHD), using multi-component composite design technology.

As in the case of the molecule X, a set of FLAPs binding to the molecule Y was selected by phenotypic screening using mammalian cells and next-generation sequencing analysis. Among these, 369 sequences with a large number of reads and a significant enrichment rate between rounds of panning were selected as FLAPs targeting the molecule Y, and used as training data for the multi-functional design technology.

Using the obtained training data, we constructed a prediction method and combined it with a search algorithm to search from a huge sequence space of 8.0×10^{17} . Among the top 10,000 sequences with high predicted binding affinity, sequences with high predicted solubility (Y-H1, Y-H2) and sequences with low predicted solubility (Y-L1, Y-L2) were selected for later experimental evaluation. The predicted immunogenicity of these sequences varied from high to low. Thus, the multi-functional design technology developed in this study was able to flexibly search for sequences predicted to have the desired properties from a vast sequence space.

The binding affinity, solubility, and immunogenicity of the designed FLAPs were evaluated experimentally, and the solubility was all as predicted. No immunogenicity was observed in any of the evaluations using IL-2 release from mouse spleen cells as an indicator. In addition, it was found that FLAP with high solubility bound strongly to the molecule Y. These results suggest that the multi-component composite design technique developed in this study can be used to design peptides with desirable properties as biopharmaceuticals.

Future directions

The multi-functional design technology developed in this study will be applied to middle molecule peptides other than FLAP and various functional proteins such as antibodies. This technology will address the major problem of "simultaneous improvement of multiple factor" in biopharmaceutical development, and will bring a breakthrough in the development process of various drug modalities in the future. In addition, we aim to further improve the prediction accuracy by incorporating the latest machine learning technologies, which are advancing day by day, as features and learning models.