

報道解禁 7月19日(土) 5:00時(日本)、7月18日(金) 15:00時(米東海岸)

## プレスリリース

2025年7月18日

長浜バイオ大学

### 天然変性領域をターゲットにした創薬を AI で可能に

#### 【ポイント】

天然変性領域(IDR) \*<sup>1</sup>とはタンパク質分子中で構造をとらないペプチド領域であり、従来の創薬研究ではターゲット(標的)として見過ごされあるいは避けられてきました。本研究はこれらの IDR をターゲットとした創薬に利用できる AI(機械学習器) IDRdecoder を作成し、ドラッグ相互作用部位と相互作用するドラッグ分子の構造を予測することを可能にしました。

#### 【研究の概要】

天然変性領域(IDR)をターゲットとした創薬のための合理的な方法論は、主に参照実験データの不足のために、未開発のままです。長浜バイオ大学バイオサイエンス学部の白井 剛教授らのグループは、IDR の機能、ドラッグ(医薬品分子)相互作用部位、およびドラッグ分子の部分構造を予測するための機械学習器 IDRdecoder を提案しました。この機械学習では、データギャップ(IDR の実験データ、すなわち学習データの圧倒的な不足)を解決するために、段階的転移学習を採用しました。その結果、畳み込みニューラルネットワークモデルである IDRdecoder は、既存のいくつかの手法と比較して性能向上を示しました。この解析から、例えば、IDR 上の Tyr(チロシン)と Ala(アラニン)残基はドラッグ標的部位として好まれること、アルキル基のような柔軟な部分構造はドラッグ分子として好まれることも示されました。

背景：医薬品の開発効率(承認された医薬品の数/投資した研究開発費)は年々低下していることが指摘されています。この原因の一つに、Low hanging fruit と呼ばれる容易な創薬ターゲットの枯渇が指摘されています。現在見過ごされがちな創薬ターゲットに天然変性領域(IDR)があります。IDR とは、決まった立体構造をとることが常識とされたタンパク質

中で、例外的に立体構造を取らない領域です。しかし近年、この IDR が疾患を含む生物の機能に重要な役割を持つことが明らかになりつつあります。例えば、ヒトの疾患の原因となる変異のうち 29% は IDR で起こります。こうして、IDR は創薬の有望な新ターゲットとして浮上してきました。しかしながら、これまでの創薬は主として立体構造をとるタンパク質をターゲットとしており、現状、論理的(データサイエンス的)に IDR をターゲットとした創薬に使える計算方法はほとんどありません。

**方法：** IDR をターゲットとした創薬に使える方法がない理由は明らかで、それは機械学習 (AI) の訓練に使える実験データがほとんど蓄積していない(データギャップ)ことです。そこで我々はこのデータギャップを埋める方法を考えました。具体的には、あえて構造をとる(しかし IDR 傾向が比較的高い)タンパク質のデータを訓練データに使うこと、および、転移学習<sup>\*2</sup> という技術によって、段階的に IDR の分類・ドラッグの標的部位の予測・ドラッグの部分構造を予測するシステムにしたことです。この方法でつくられたニューラルネットワーク型機械学習器を IDRdecoder と名付けました。

**結果：** 転移学習によってトレーニングされた IDRdecoder は、IDR を機能(DNA 結合など)に基づいて分類(アミノ酸配列の類似性に依存せず、機能の似た IDR をクラスタリング)できることが示されました。また、ドラッグ結合部位の予測については、従来提案されていた方法、特に ChatGPT などにも利用されている Transformer<sup>\*3</sup> を上回る予測性能を示しました(AUC<sup>\*4</sup> IDRdecoder = 0.616, Transformer = 0.538)。またドラッグ分子の部分構造の予測においては、Transformer と同等の性能(AUC IDRdecoder = 0.702, Transformer = 0.694)を示しました。

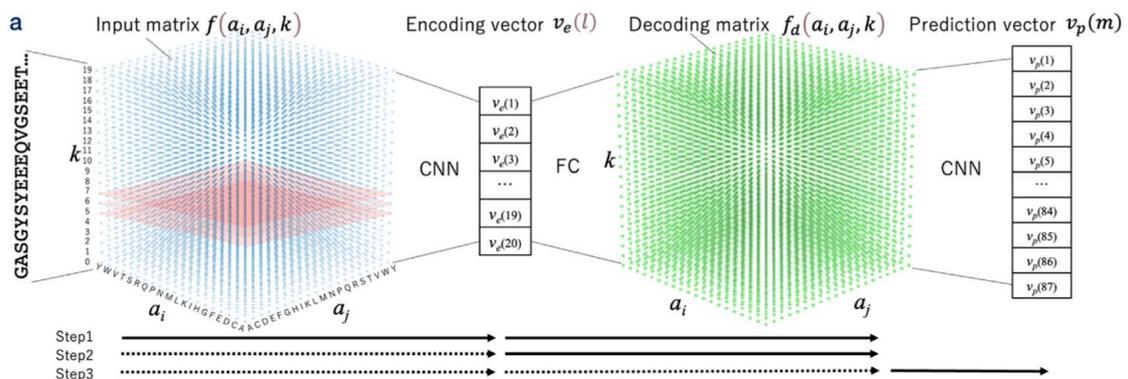


図1 IDRdecoder の模式図 左が入カアミノ酸配列、右が予測の出力で、 $v_e(l)$  が IDR の分類、 $f_d(a_i, a_j, k)$  が相互作用部位、 $v_p(m)$  がドラッグ分子構造の予測値となります。

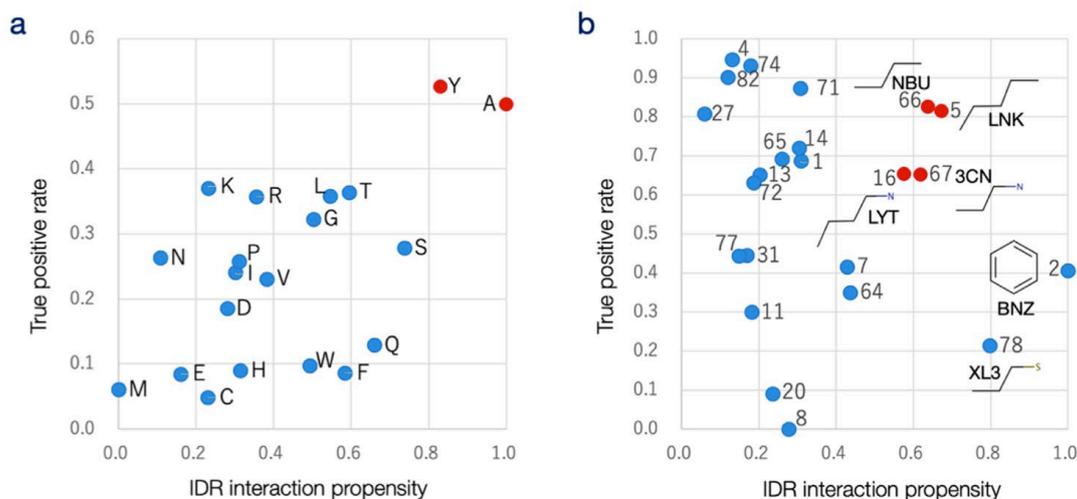


図2 IDRcdecoder による分析

(a) Y(チロシン)とA(アラニン)が相互作用傾向が高く(横軸)かつ IDRdecoder が高く評価(縦軸)します。(b) アルキル分子(炭素 C が直線状に結合した分子)が相互作用傾向が高く(横軸)かつ IDRdecoder が高く評価(縦軸)します。

研究成果の意義：これまで存在しなかった、IDR 配列をターゲットとした創薬に利用できる機械学習器(AI)が開発されました。特に、ドラッグの分子構造を予測できる方法はこれまで存在しませんでした。これは、IDR 標的創薬を大きく前進させる結果であり、新規の創薬ターゲットの探索に応用可能です。

この研究は、日本医療研究開発機構 (AMED) 革新的先端研究開発支援事業(AMED-CREST 感染症創薬基盤)の支援(課題番号：JP24gm1610009)を得て行われたものです。

## 用語解説

### \*1 天然変性領域(IDR)・天然変性タンパク質(IDP)

タンパク質の中で、特定の立体構造を形成しない領域(ペプチド)。少数のアミノ酸の繰り返し配列を持つなど、特徴のあるアミノ酸配列をとっていることが多い。一部の IDR は特定の相互作用分子(リガンド)が存在する場合に限り立体構造をとることがある。

### \*2 転移学習

機械学習の手法の一つで、まず本来の目的とはある程度異なるが、大量に存在する学習(トレーニング)データで機械学習を行なったのちに、その機械学習器を本来の学習データで訓練する方法。

### \*3 Transformer

ニューラルネット型の機械学習器の一つで、言語などの(時)系列データを文脈にそって(時)系列処理することなく高速に並列学習できるという特徴を持つ。この特徴のため、ChatGPT などの言語生成 AI に利用されて大きな成功をおさめている。

\*4 AUC

AUC は Area Under Curve の略称で、予測の性能を統計的に評価する ROC(Receiver operating Characteristic)曲線から算出される数値。0.5 以下で予測性能なし、0.5 以上で 1.0 に近づくほど予測性能が高いと評価される。

#### 【論文情報】

掲載誌： Frontiers in Bioinformatics (Frontiers Media S.A.発行)

タイトル： IDRdecoder: A machine learning approach for rational drug discovery toward intrinsically disordered regions (邦訳：IDRdecoder:天然変性領域をターゲットとした合理的創薬のための機械学習アプローチ)

著者：塩生くらら(しおにゅう くらら) 1, 大森 聡(おおもり さとし) 1、平田 昴(ひらた すばる) 2、石田裕一(いしだ ひろかず) 1、白井 剛(しらい つよし) 1,2\*  
(\*印は連絡著者)

1. 長浜バイオ大学バイオサイエンス学部, 2. 滋賀大学データサイエンス学部

DOI： 10.3389/fbinf.2025.1627836

#### 【本研究に関するお問い合わせ】

長浜バイオ大学 バイオサイエンス学部

教授 白井 剛(しらい つよし)

E-mail: t\_shirai@nagahama-i-bio.ac.jp

TEL: 0749-64-8117

#### 【報道担当】

長浜バイオ大学 アドミッション・オフィス 広報担当

Tel : 0749-64-8100

E-mail : kouhou@nagahama-i-bio.ac.jp