



日本学術振興会・産学協力委員会

R022 量子構造生物学委員会

第13回 研究会 資料

2024年3月12日（火）
13:30～17:00

ミーティングスペースAP日本橋（東京）

**R022 量子構造生物学委員会
第 13 回研究会 プログラム**

2024 年 3 月 12 日（火）

会場：ミーティングスペース AP 日本橋
〒103-0027 東京都中央区日本橋 3-6-2
日本橋フロント 6F
(現地会場で開催／オンライン接続はありません)

13:30 – 13:35

はじめに

委員長・三木邦夫（京都大学）

講 演

13:35 – 14:35 （講演 1）

司会：委員・沼本修孝（岡山大学）

機械学習が導くタンパク質の加速進化

東北大学・大学院工学研究科
梅 津 光 央

14:35 – 15:00 休 憩

15:00 – 16:00 （講演 2）

司会：委員・禾 晃和（横浜市立大学）

タンパク質分解医薬 PROTAC の開発の現状と将来展望

国立医薬品食品衛生研究所・有機化学部
出 水 庸 介

16:00 – 17:00 （研究交流討論）

委員・講師間の自由討論

機械学習が導くタンパク質の加速進化

Machine-learning-guided molecular evolution engineering

東北大学 大学院工学研究科 バイオ工学専攻

理化学研究所 革新知能統合研究センター

梅 津 光 央

Mitsuo Umetsu

Department of Biomolecular Engineering, Graduate School of Engineering,
Tohoku University

Advanced Intelligence Project, RIKEN

1. はじめに

タンパク質の設計とは、単純に言えば「20 種類のアミノ酸をどう並べるか」である。言わば、アミノ酸の配列が生み出す「場合の数（配列空間）」の中から目的の機能をもつ配列を見つけ出すことである。しかし、アミノ酸の組合せがつくる配列空間はいまだ膨大で、現在のタンパク質工学技術を用いても、目的の機能・特性をもつ配列を見つけ出すことができるとは限らない。1983 年に Ulmer がタンパク質工学を提唱¹⁾してから 40 年を経た今、ゲノム・タンパク質情報のデータベース化や大規模な変異体ライブラリー作製・スクリーニング操作が発展し、試験管内で進化を模倣する進化分子工学によって目的機能をもつタンパク質を創り出す確率が高まってきている。しかし、これらの技術開発をもってしても、目的の機能・特性をもつタンパク質の確率は現在でも満足のいくものではなく、開発時間や創出確実性はいまだ予測しきれない中で開発を進めざるを得ない。この状況において近年、機械学習を利用することで研究開発の対象となる配列空間もしくは目的解をもつ配列を限定していく試みが行われており、私たちも変異体ライブラリーの作製に人工知能技術のひとつである機械学習を利用して、機能陽性な変異体を高密度に含むライブラリーを設計することで、目的とする機能・物性をもつタンパク質を効率的かつ確実に取得する進化分子工学の開発をおこなっている。

2. 進化分子工学の発展と課題

一般的なタンパク質の進化分子工学では、データベースから目的機能を創出しそうな親タンパク質を選び、その親タンパク質へ変異を導入したライブラリーから目的機能をもった変異体をスクリーニングしていくが、実験的に作製・機能評価できる変異体数は限られているため、ライブラリーの作製では探索する配列空間の設計(「どの位置」に「何のアミノ酸」を出現させるか)が重要になる。変異体ライブラリーからの機能タンパク質の取得は、選択操作とスクリーニング操作に分けることができる²⁾。選択操作の場合は、ライブラリー状態のまま選択操作圧をかけて目的解をもつ変異体を残存させる方法である。そのため、配列空間の設計も実験的に作製可能な範囲($10^{10\sim15}$)内で大規模な変異体群ライブラリーを対象とすることができるが、その規模も 8~10 程度の残基位置を飽和変異できる程度である。また、選択圧をかける手段も限られており、対象となる機能・特性が抗体のような結合機能などに限られてしまう。一方、スクリーニング操作では、ライブラリー中の変異体を一つずつ評価するため、さまざまな機能・特性に対応が可能である。しかし、スクリーニングの対応変異体数には限界($10^{3\sim4}$)があるので、スクリーニングで対応可能な規模での目的解を含む配列空間の設計が難しくなる。

3. 機械学習による配列空間探索

この進化分子工学の課題において、私たちはある一定の配列空間規模において機械学習が最適解探索に有効であることを示している。私たちが用いた機械学習では教師データが必要になるため、ある程度の変異体を作製して各々の機能・特性値を実測する。そして、アミノ酸を適切な記述子で数値表現することで、変異配列と蛍光特性値を紐づけた教師データを作成し機械学習を行っている。

緑色蛍光タンパク質 (GFP) を黄色化した実験では、蛍光色の決定に関与の可能性がある 4 残基に限定し、配列空間中に含まれる変異体数の規模が 16 万 ($20^4 = 1.6 \times 10^5$) の条件で行っている³⁾。教師データとしては、配列空間中に存在する変異体数の 0.1 %にあたる 153 種の GFP 変異体を調製して変異情報とともに蛍光の黄色度と蛍光強度を収集し、弱いながらも黄色蛍光を発する変異体を 4 種含むデータを得た (図 1)。そして、このデータを教師として高速な

ベイズ最適化ソフトウェアである COMBO⁴⁾を用いて「より黄色く・より明るい変異配列」を目標として機械学習を行い、空間配列中の全変異体（16 万）の順位付けを行った。その結果、上位 2 位の変異体はこれまで報告されている黄色蛍光変異体の中でもっとも強い黄色蛍光を示す Venus であり、機械学習が上位に予測した変異体を 60 種程度作製したところ、その多くが教師データ中の変異体よりも強い黄色蛍光を示した。これは、配列空間中の 0.1 %程度の変異体数で、かつ、低いながらも目的機能を示す陽性率が数%程度の教師データであれば、適切な機械学習モデルは最適解に近い変異体を含む集団を予測することができ得ることを示唆している。

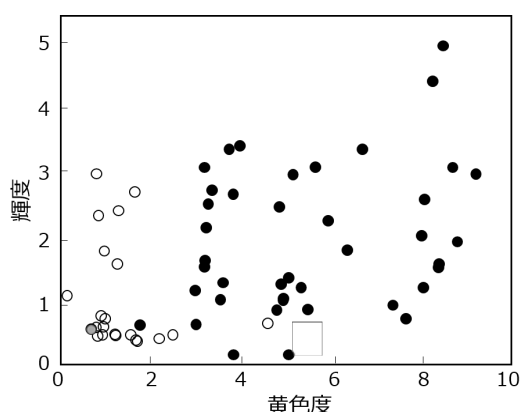


図1 GFP の黄色化実験における変異体の機能値.

灰色：野生型，白：教師データの変異体，黒：機械学習提案変異体.

より広い配列空間を対象とした実験として、ペプチド転移酵素 sortaseA の酵素活性を向上させている⁵⁾。sortaseA では、5 残基に変異を導入することで高い酵素活性を示す変異体（5M 変異体）が報告されており⁶⁾、この 5 残基がつくる配列空間中の全変異体数 ($20^5 = 320$ 万) の $2 \times 10^{-3} \%$ にあたる 80 程度の変異体から教師データを作成し、「より酵素活性が高い変異配列」を目標として機械学習をおこなった。その際に、5M 変異体を教師データに加える有無を変えた 2 ケースでおこなったところ、両ケースにおいて、上位 50 変異体中の 7 割は野生型よりも高い活性を示し、強い陽性変異体が存在しなくても機械学習は進化の指向性を示すことができることが示唆された (図 2)。しかし一方、そのほとんどは 5M 変異体の活性まで至っておらず、GFP の時よりも進化の度合いが小さかった。そこでさらに、機械学習予測上位変異体の結果も含めて再度機械学習を実施したところ、より高い変異体集団が同定され、最終的には 5M 変

異体よりも活性の高い変異体を複数取得できた。このように、機械学習を用いることによって、進化の速度に差はあれど、設定した配列空間中から目的の機能・特性をもつ変異体を高い確度で同定することができる可能性が示され始めている。

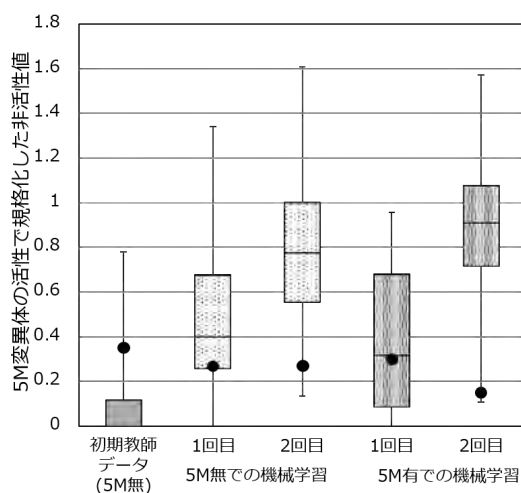


図2 機械学習予測と実測検証の繰り返しによるペプチド連結酵素の活性向上。
黒：野生型。

また、抗体様分子においては、11 残基位置(配列空間規模： $20^{11} = 2 \times 10^{14}$)を変異対象とした配列空間を対象として、次世代シーケンサー(NGS)を利用して配列空間の規模を 10^9 程度まで絞り込み機械学習をおこなうことで、標的分子に結合する変異体を同定している⁷⁾。まず、11 残基をランダム変異した抗体様分子をファージへ提示したライブラリーを 10^8 程度で調製して標的分子に対して選択操作をおこなったところ、実験的な操作だけでは標的に特異的に結合する変異体を取得することはできなかった。そこで、選択操作におけるライブラリー中の変異体群の分布変化を NGS 解析し、「残存するリード数が多い変異体=標的結合性が高い」とした教師データから機械学習器を作成した。そして、各残基位置で有望なアミノ酸種を絞り込むことで機械学習を使って探索する配列空間を 10^9 程度に絞り込み、その対象配列空間中の全変異体の標的結合性を予測した(図 3)。その結果、機械学習により提案された予測上位 10,000 配列をアミノ酸配列の類似度でクラスタリング処理し、予測上位 1,000 配列を含んでいたグループについてファージライブラリーを再度作製し、再び生体外選択操作をおこなった結果、標的分子選択的に 80 nM の EC_{50} を示す変異体を取得することができた。このことから、標的結合機能については、次世代シーケンシ

ング解析データを探索対象となる配列空間の限定と教師データに活用することによって、機械学習により目的機能分子を取得する確度を向上できることが示されたと考えている。

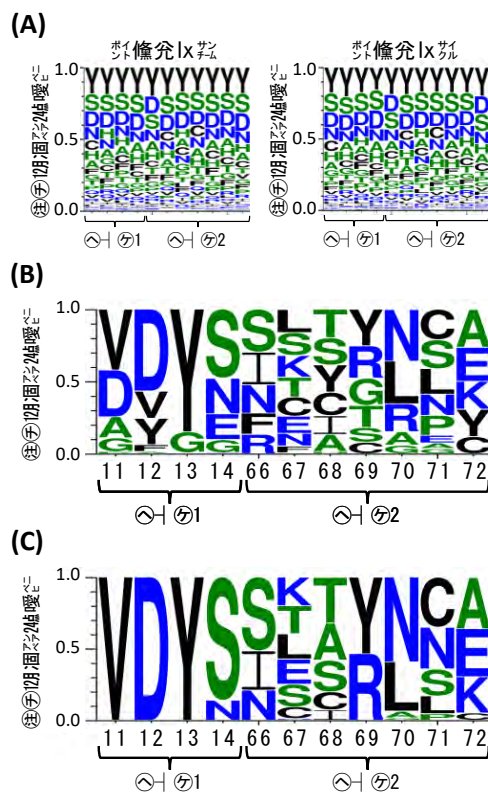


図 3. 抗体様分子の変異導入残基におけるアミノ酸出現頻度. ファージ提示法による選択操作前(A 左), 選択操作後(A 右), 機械学習が提案した上位 10,000 配列(B), そして予測上位 1,000 配列を含むグループ(C) のシーケンスロゴ. 各文字は出現するアミノ酸を表し, サイズが大きいほど出現頻度が高いことを示す.

4. 最後に

以上のように, 私たちは, 変異導入による機能タンパク質の開発における, 「『どの位置』に『何のアミノ酸』を出現させるか」という課題に対して, ある程度「どの位置」が限定された場合での機械学習の有効性を示してきた. また現在では, 「どの位置」と「何のアミノ酸」の両設計をまとめて行う取り組みもおこなっており, 本講演ではその内容についても一部紹介したい.

-
1. Kevin M. Ulmer, "Protein Engineering", *Science*, **219**, 666-671 (1983).
 2. Yajie Wang, Pu Xue, Mingfeng Cao, Tianhao Yu, Stephan T. Lane, and Huimin Zhao, "Directed Evolution: Methodologies and Applications", *Chemical Reviews*, **121**, 12384-12444 (2021).
 3. Yutaka Saito, Misaki Oikawa, Hikaru Nakazawa, Teppei Niide, Tomoshi Kameda, Koji Tsuda, and Mitsuo Umetsu, "Machine-Learning-Guided Mutagenesis for Directed Evolution of Fluorescent Proteins", *ACS Synthetic Biology*, **7**, 2014–2022 (2018).
 4. Tsuyoshi Ueno, Trevor David Rhone, Zhufeng Hou, Teruyasu Mizoguchi, and Koji Tsuda, "COMBO: An efficient Bayesian optimization library for materials science", *Materials Discovery*, **4**, 18-21 (2016).
 5. Yutaka Saito, Misaki Oikawa, Takumi Sato, Hikaru Nakazawa, Tomoyuki Ito, Tomoshi Kameda, Koji Tsuda, and Mitsuo Umetsu, "Machine-learning-guided library design cycle for directed evolution of enzymes: the effects of training data composition on sequence space exploration", *ACS Catalysis*, **11**, 14615–14624 (2021).
 6. Irwin Chen, Brent M. Dorr, and David R. Liu, "A general strategy for the evolution of bond-forming enzymes using yeast display", *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 11399- 11404 (2011).
 7. Tomoyuki Ito, Thuy Duong Nguyen, Yutaka Saito, Yoichi Kurumida, Hikaru Nakazawa, Sakiya Kawada, Hafumi Nishi, Koji Tsuda, Tomoshi Kameda, and Mitsuo Umetsu, "Selection of target-binding proteins from the information of weakly enriched phage display libraries by deep sequencing and machine learning", *mAbs*, **15**, 2168470 (2023).

タンパク質分解医薬 PROTAC の開発の現状と将来展望

Current status and future prospects of PROTAC

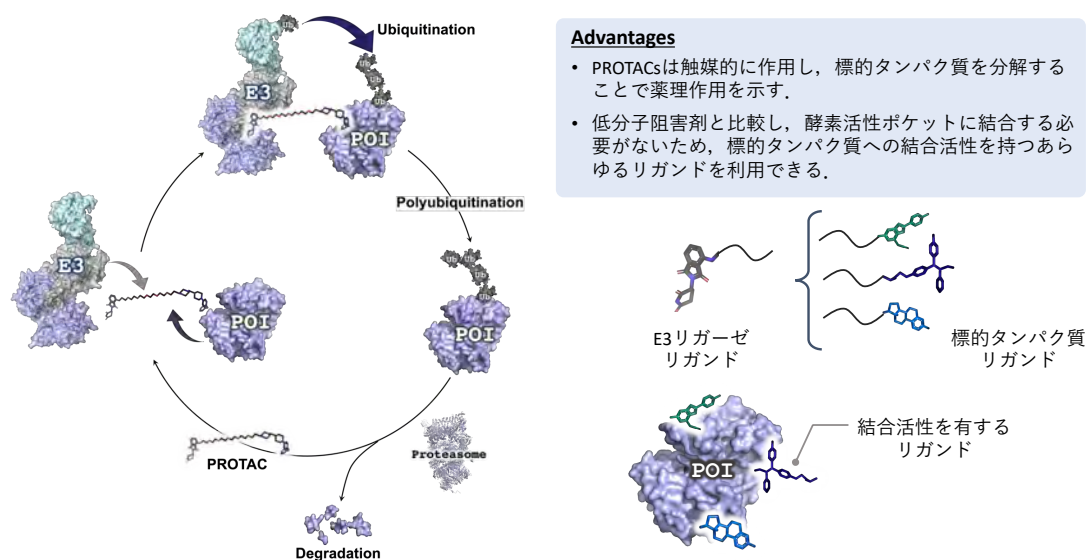
国立医薬品食品衛生研究所 有機化学部

出水 庸 介

Yosuke Demizu

Division of Organic Chemistry, National Institute of Health Sciences

PROTACs (Proteolysis Targeting Chimeras) は、ユビキチンリガーゼ (E3 リガーゼ) リガンドと標的タンパク質 (Protein of Interest : POI) リガンドを適切なリンカーを介して連結したキメラ分子である。PROTAC は、生体内のタンパク質分解機構であるユビキチン・プロテアソームシステムを利用することで、標的タンパク質を分解し、その機能を抑制できる [1]。具体的には、PROTAC がそれぞれのリガンドを介して E3 リガーゼと標的タンパク質を近接させることで、標的タンパク質がポリユビキチン化され、プロテアソームに認識され分解される (図 1)。PROTAC は、触媒的に作用することで標的タンパク質を分解し薬理活性を示すこと、低分子阻害剤のように酵素活性ポケットにリガンドが結合する必要がなく、標的タンパク質の様々な領域 (ドメイン) に対してリガンドを設計できること等の特徴を有していることから、革新的な創薬モダリティのひとつとして期待されている。



我々は、様々な疾病関連タンパク質を標的とした PROTAC 開発を行ってきた。特に標的タンパク質に適した低分子、ペプチド、核酸等をリガンドとして活用する PROTAC デザインを行ってきた [2-6]。本研究では、1) 造血器型プロスタグランジン D 合成酵素 (H-PGDS) を標的とした低分子型 PROTAC の in silico デザイン [3]、2) デコイ核酸を利用した PROTAC 開発 [5]、について紹介する。

1) H-PGDS を標的とした PROTAC 開発

H-PGDS は、デュシェンヌ型筋ジストロフィー (DMD) 治療の新しい標的として注目されている。この遺伝性筋障害は、X 染色体上のジストロフィン遺伝子の変異が原因であり、患者の多くは 30 歳前に呼吸不全や心不全で死亡する。DMD の治療法には、ジストロフィンの量を回復させる治療法 (エクソン・スキップ療法) や、病気の進行を抑え筋肉の質を保持する方法がある。損傷した DMD 患者の骨格筋では、H-PGDS が過剰に発現し、プロスタグランジン D₂ (PGD₂) の増加と PGD 受容体を介した炎症を引き起こす。H-PGDS の活性を抑制する治療法は有望であるが、現在の阻害剤は臨床試験での効果が十分でないため、新しい作用機序を有する医薬品開発が求められている。そこで本研究では、分子モデリングを利用して H-PGDS を標的とする PROTAC をデザインし、活性の向上と DMD 治療薬としての可能性を検討した。

当初、我々が開発した **PROTAC(H-PGDS)-1** は、H-PGDS 阻害剤の TFC-007 と E3 リガーゼである CRBN に結合する Pomalidomide が PEG5 リンカーを介して結合した構造であり、細胞レベルにおいて 10 nM で H-PGDS 分解活性を有していた [7]。先ず始めに、化学計算ソフトウェア MOE (MOLSYS 社) の PROTAC Modeling Tool を利用して、**PROTAC(H-PGDS)-1** と H-PGDS、CRBN の三者複合体解析を行った。その結果、これらの分子が安定な三者複合体を形成する可能性が確認できた。しかしながら、PEG5 リンカー部位が両タンパク質間で構造を形成しており、リンカーの長さを短縮することが可能である可能性が示唆された。そこで、リンカー長が異なる (PEG4~PEG0) PROTAC をデザインし、それらの三者複合体の安定構造を解析した。その結果、PEG4~PEG1 リンカーを持つ PROTAC も三者複合体を形成できる可能性があること、さらに、リンカー構造を持たない **PROTAC(H-PGDS)-6** においても安定な三者複合体形成が示唆された (図 2)。

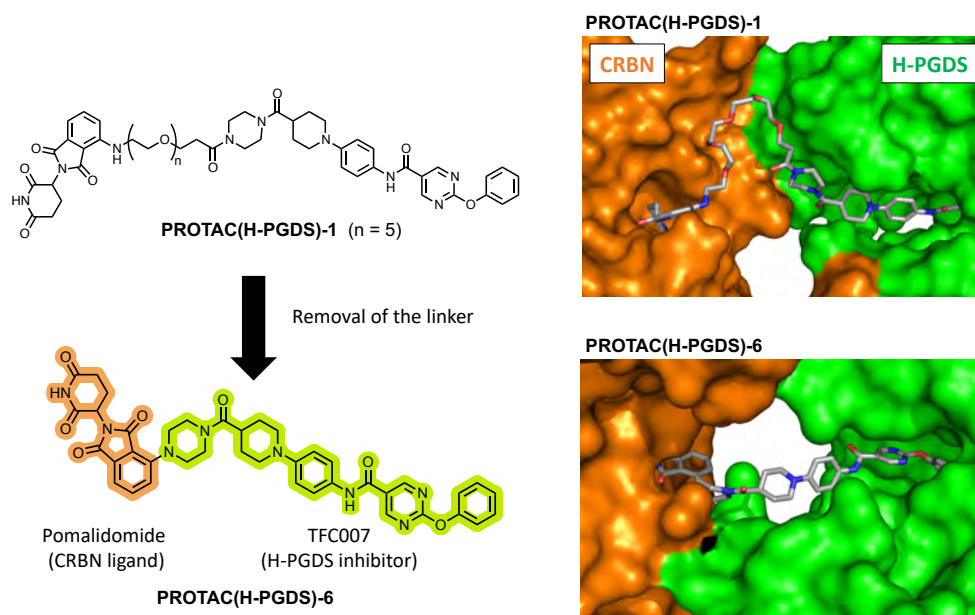


図2 **PROTAC(H-PGDS)-1** 及び **PROTAC(H-PGDS)-6** の構造と
MD シミュレーションによる三者複合体構造

そこで、分子モデリングの妥当性を検証するために、**PROTAC(H-PGDS)-1**～**PROTAC(H-PGDS)-6** を合成し、それらの H-PGDS 分解活性をウェスタンブロットで評価した。その結果、PEG リンカーの長さが短い PROTAC ほど H-PGDS 分解活性が高まることが明らかになった (図 3)。特に、リンカーを持たない **PROTAC(H-PGDS)-6** は、最も高い分解活性を示した (タンパク質分解の 50% を引き起こす濃度 $DC_{50} = 17.3$ pM)。また、タンデムマスタグを用いたプロテ

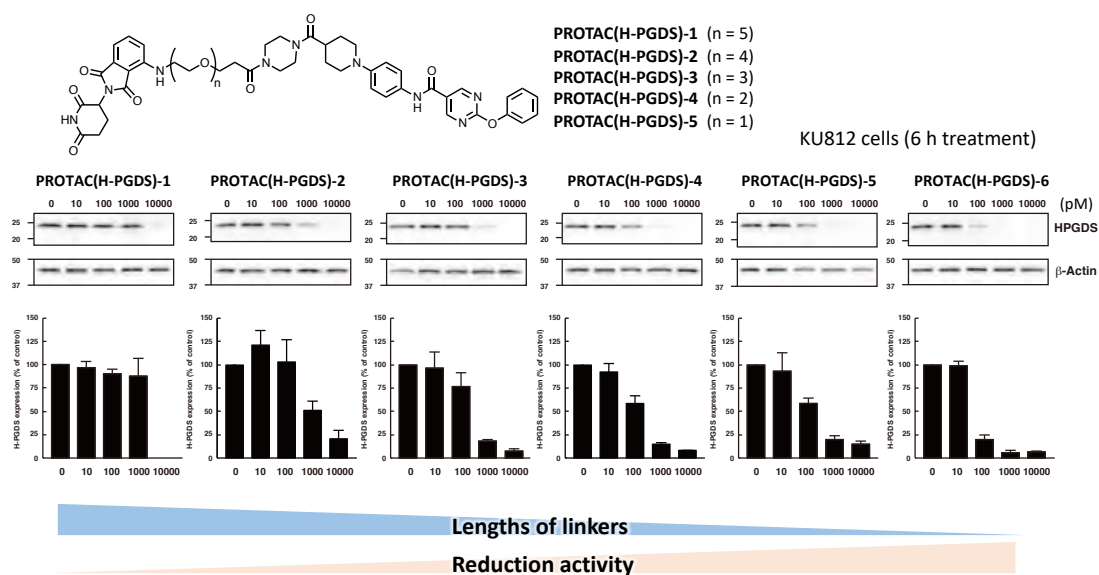


図3 **PROTAC(H-PGDS)-1**～**6** の H-PGDS 分解活性 (KU812 細胞)

オーム解析を行った結果、8,184 個のタンパク質が検出された中で H-PGDS 以外のタンパク質の減少は観測されず、**ROTAC(H-PGDS)-6** は高選択的に H-PGDS タンパク質を分解することも明らかとなった。さらに、DMD のモデルマウスを用いた in vivo 評価においても、H-PGDS の分解を確認することができた。

一般に PROTAC は、二つのリガンドをリンカーで結合させているため分子量が大きく、回転可能結合数が多く、薬剤としての特性に欠けることが多い。しかしながら、リンカーを排除した **PROTAC(H-PGDS)-6** は、分子量や回転可能結合数が小さく、より薬剤らしい性質を持つ。ケーススタディではあるが、本アプローチは PROTAC のリンカーを取り除くことの有効性を示し、医薬品候補としての可能性を高めるものである [3]。

2) デコイ核酸を利用した PROTAC 開発

これまでに報告されている PROTAC 研究の多くは、主にがん関連のキナーゼをターゲットとしており、これらのキナーゼの低分子阻害剤がターゲットリガンドとして用いられている。一方で、キナーゼに加えて転写因子も様々ながんに関連しているが、それらの多くは阻害剤開発に適した構造を持っていないため、最適なりガンドが不足しており、転写因子を対象とした低分子型 PROTAC をデザインすることは困難である。

ヒトでは約 1,600 種類の転写因子が報告されており、そのうちおよそ 96% の DNA 結合領域の配列が明らかにされ、データベース化されている [8]。そこで我々は、転写因子をターゲットにするための新しいアプローチとして、転写因子が結合する DNA 配列を模したデコイ核酸を標的リガンドとする核酸型 PROTAC 開発を検討した。本研究では、転写因子の一つであるエストロゲン受容体 α (Estrogen Receptor α : ER α) をモデル標的としたデコイ核酸型 PROTAC について紹介する。

ER α リガンドには、ER α タンパク質とエストロゲン応答配列 (ER α 結合 DNA 配列) の 21 塩基のデコイ核酸を選択しその 5'末端をアルキン修飾した **ER(dec)**を設計した。また、PROTAC に汎用されている 3 種類の E3 リガーゼリガンド (LCL161 for IAPs、VH032 for VHL、Pomalidomide for CRBN) に PEG3 リンカーを介してアジド化した分子を設計した。これらをクリック反応で連結し、3 種類の PROTAC (**LCL-ER(dec)**、**VH-ER(dec)**、**POM-ER(dec)**) を合成した (図 4)。

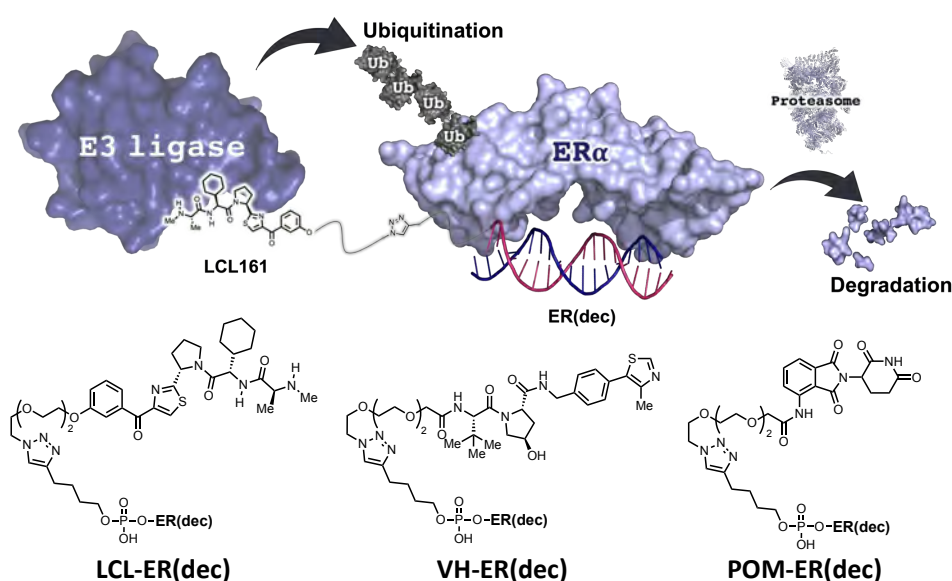


図4 デコイ核酸型 PROTAC のデザイン

合成した PROTAC は、強い ER α 結合活性 ($IC_{50} = 30 \sim 50$ nM) を持ち、ER α 分解活性を有していた (MCF-7 細胞)。その中でも **LCL-ER(dec)** が最も高い分解活性を持ち、転写因子を含む他のタンパク質 (AR、AhR、p65、BRD4、p300、GAPDH、CRABP2、 β -actin) を分解せず ER α 選択性を示した (図 5) [5]。さらに、**LCL-ER(dec)** の核酸配列末端をヘアピン化した **LCL-ER(dec)-H46** は、ER α に対する選択性を維持しながら **LCL-ER(dec)** よりも長時間の分解活性を有することが明らかとなった [9]。

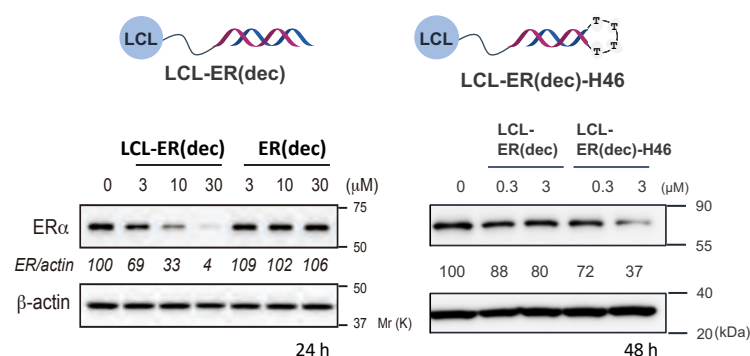


図5 デコイ核酸型 PROTAC の ER α 分解活性 (MCF-7 細胞)

近年、いくつかの研究グループがデコイ核酸を利用した PROTAC の開発に成功し、NF κ B、E2F、LEF1、ERG、cMyc、 α -brachyury、Nrf2/MafG 等の分解が実現されている [10-13]。このように、さまざまな転写因子の分解を実現する

ことで、がん治療などにおけるデコイ核酸型 PROTAC の応用範囲が拡大している。しかしながら、デコイ核酸型 PROTAC の開発には、細胞内導入効率や標的特異性の向上といった課題が残されている。将来的には、核酸医薬品のドラッグデリバリー技術の応用により、核酸型 PROTAC の開発がさらに進むことが期待される。

-
1. Crews C. M. *et al.*, PROTAC targeted protein degraders: the past is prologue. *Nat Rev Drug Discov.* **21**, 181-200 (2022).
 2. Okitsu K, Demizu Y. *et al.*, Development of a Small Hybrid Molecule That Mediates Degradation of His-Tag Fused Proteins. *J Med Chem.* **61**, 576-582 (2018).
 3. Yokoo H, Demizu Y. *et al.*, Discovery of a Highly Potent and Selective Degradator Targeting Hematopoietic Prostaglandin D Synthase via In Silico Design. *J Med Chem.* **64**, 15868-15882 (2021).
 4. Yokoo H, Demizu Y. *et al.*, Peptide Stapling Improves the Sustainability of a Peptide-Based Chimeric Molecule That Induces Targeted Protein Degradation. *Int J Mol Sci.* **22**, 8772 (2021).
 5. Naganuma M, Demizu Y. *et al.*, Development of Chimeric Molecules That Degrade the Estrogen Receptor Using Decoy Oligonucleotide Ligands. *ACS Med Chem Lett.* **13**, 134-139 (2022).
 6. Tsujimura H, Demizu Y. *et al.*, Development of DNA aptamer-based PROTACs that degrade the estrogen receptor. *ACS Med Chem Lett.* **14**, 827-832 (2023).
 7. Yokoo H, Demizu Y. *et al.*, Development of a Hematopoietic Prostaglandin D Synthase-Degradation Inducer. *ACS Med Chem Lett.* **12**, 236-241 (2021).
 8. Lambert S A. *et al.*, The Human Transcription Factors. *Cell* **172**, 650-665 (2018).
 9. Naganuma M, Demizu Y. *et al.*, Structural optimization of decoy oligonucleotide-based PROTAC that degrades the estrogen receptor. *Bioconjug Chem.* **34**, 1780-1788 (2023).
 10. Liu J. *et al.*, TF-PROTACs Enable Targeted Degradation of Transcription Factors. *J Am Chem Soc.* **143**, 8902-8910 (2021).
 11. Shao J. *et al.*, Destruction of DNA-Binding Proteins by Programmable Oligonucleotide PROTAC (O'PROTAC): Effective Targeting of LEF1 and ERG. *Adv Sci.* e2102555 (2021).
 12. Samarasinghe K T G. *et al.*, OligoTRAFTACs: A generalizable method for transcription factor degradation. *RSC Chem Biol.* **3**, 1144-1153 (2022).
 13. Ji J. *et al.*, ARE-PROTACs Enable Co-degradation of an Nrf2-MafG Heterodimer. *J Med Chem.* **66**, 6070-6081 (2023).