



名古屋大学
NAGOYA UNIVERSITY



岡山大学
OKAYAMA UNIVERSITY

配布先: 文部科学記者会、科学記者会、名古屋教育記者会、岡山大学記者クラブ

報道の解禁日(日本時間)

(テレビ, ラジオ, インターネット) : 2024 年 6 月 21 日(金) 午前 9 時

(新聞) : 2024 年 6 月 21 日(金) 付夕刊

2024 年 6 月 19 日

報道機関 各位

全タンパク質構造への薬の結合親和性から薬効と副作用を予測 ～シミュレーションと AI・機械学習で薬のメカニズムを理解する～

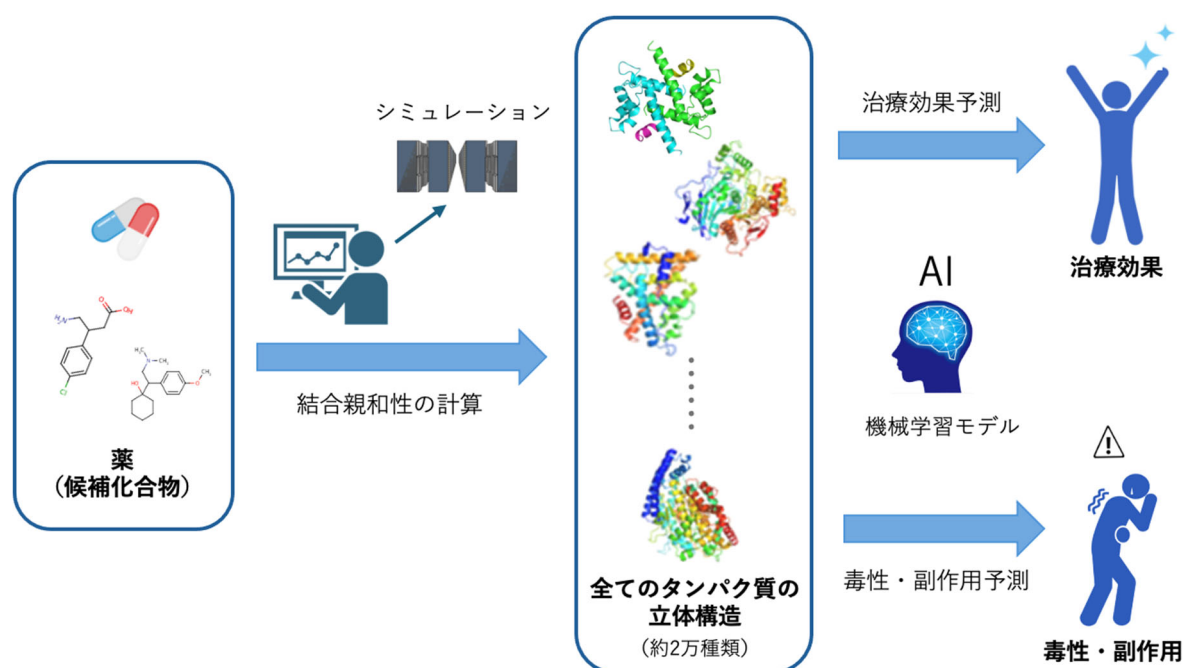
【本研究のポイント】

- ・ヒトの全てのタンパク質の立体構造を、ドッキングシミュレーション^{注 1)}と機械学習^{注 2)}で解析し、薬の効能や毒性・副作用を予測する新しい計算手法を開発した。
- ・全タンパク質に対する薬の結合親和性を計算することで、薬の結合パターンを明らかにし、薬の副作用の発現に関与するタンパク質の抽出を可能にした。
- ・従来は考慮できなかったタンパク質群の影響を提案手法は考慮できるため、薬のメカニズムの理解や医薬品開発の期間短縮や費用削減に貢献することが期待される。

【研究概要】

名古屋大学大学院情報学研究科の坂尻 由子 研究員、山西 芳裕 教授、岡山大学学術研究院医歯薬学域の澤田 隆介 助教、九州工業大学院情報工学研究院の柴田 友和 研究員の研究グループは、AlphaFold^{注 3)}に代表される AI 技術で得られた全てのヒトタンパク質の立体構造をドッキングシミュレーションと機械学習モデルで解析し、薬の効能や毒性・副作用を予測する新しい計算手法を開発しました。約 8 千種類の既存薬と約 2 万種類のヒトタンパク質の結合親和性のシミュレーションから、様々な疾患に対する既存薬の新規効能を網羅的に予測できることを示しました。さらに、数百種類の毒性に対する薬の副作用を機械学習モデルで予測し、副作用の発現に関与するタンパク質を抽出することにも成功しました。本研究の提案手法は、薬のメカニズムの理解や医薬品開発の期間短縮や費用削減に貢献することが期待されます。

本研究成果は、2024 年 6 月 21 日午前 9 時(日本時間)付 Cell Press 社「iScience」で公開されます。



【研究背景と内容】

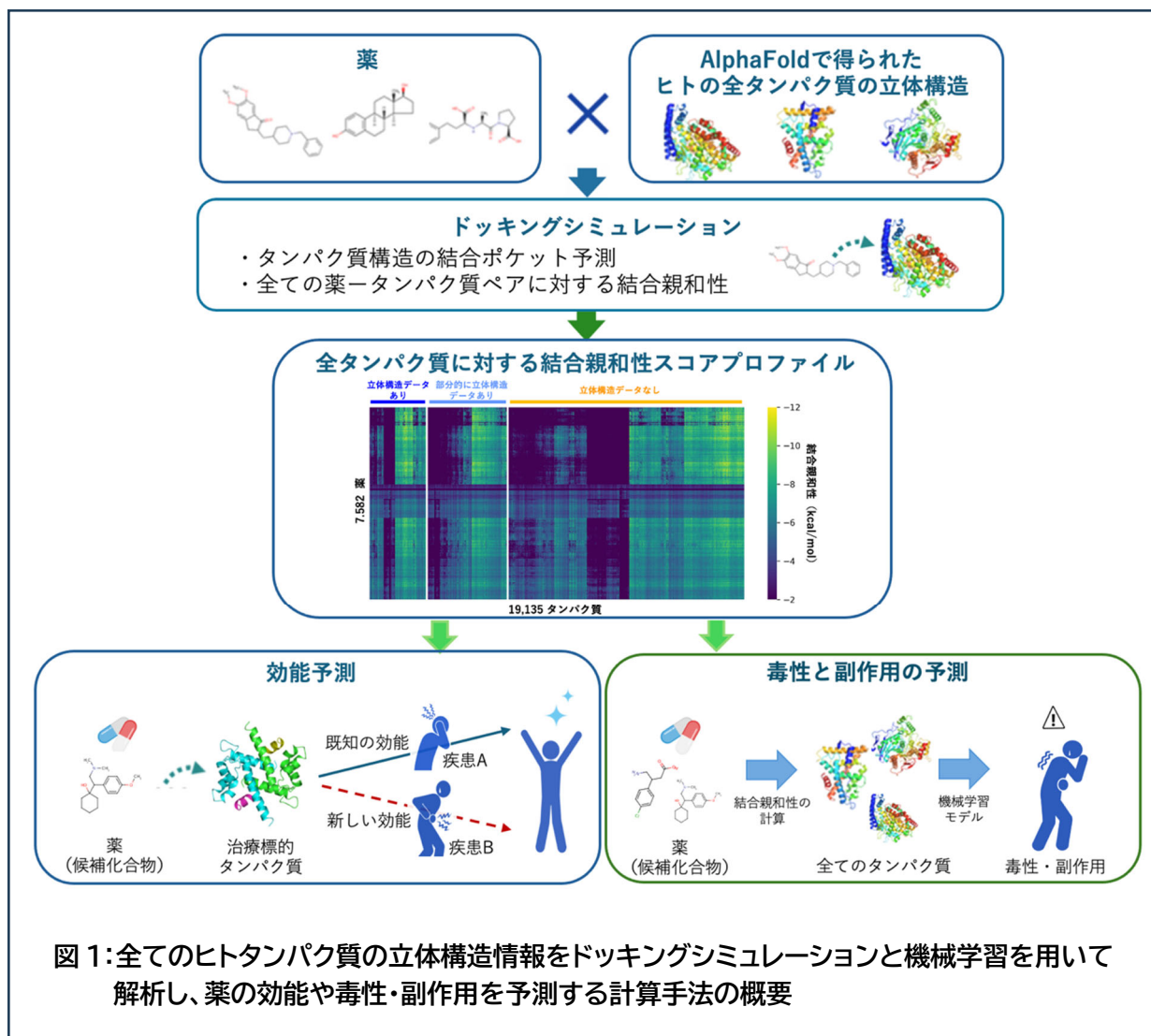
創薬プロセスにおいて、医薬品候補化合物とタンパク質の結合親和性を定量的に評価することは非常に重要です。医薬品候補化合物が相互作用するタンパク質が分かれば、潜在的な効能や毒性・副作用の予測につながります。ドッキングシミュレーションは、医薬品候補化合物とタンパク質の結合親和性を評価する代表的な情報技術ですが、その実行には実験的に決定されたタンパク質の立体構造情報が必要です。

しかしながら、ヒトのタンパク質の大半は立体構造が未知であり、創薬研究の障害になっていました。一方で、近年の AI 技術の進歩とともに、タンパク質の立体構造がアミノ酸配列から高精度で予測できるようになってきており、なかでも Google DeepMind 社が発表した AlphaFold は、実験同定と遜色ないレベルでタンパク質の立体構造情報を提供できるまでになってきています。

名古屋大学大学院情報学研究科の坂尻 由子 研究員と山西 芳裕 教授、岡山大学学術研究院医歯薬学域の澤田 隆介 助教、九州工業大学院情報工学研究院の柴田 友和 研究員の研究グループは、AlphaFold に代表される AI 技術で得られた全てのヒトタンパク質の立体構造情報をドッキングシミュレーションと機械学習で解析し、薬の効能や毒性・副作用を予測する新しい計算手法を開発しました(図 1)。

まず、AlphaFold によって得られた全てのヒトタンパク質の立体構造(立体構造が実験的に未決定のタンパク質を含む)に対して、薬の結合親和性を調べるために大規模なドッキングシミュレーションを行いました。そして、各薬の全ヒトタンパク質に対する結合親和性スコア(PBAS:Proteome-wide binding affinity score)を計算して、薬の結合パターンを反映する PBAS プロファイルを構築しました(図 2)。

この PBAS プロファイルを機械学習モデルで解析することによって、様々な疾患に対する薬の新規効能や数百種類の毒性に対する潜在的な副作用の予測を可能にしました。



提案手法を 7,582 種類の既存薬(KEGG DRUG データベースに収納されている日本や欧米の低分子医薬品)に適用し、19,135 種類のヒトタンパク質に対して大規模ドッキングシミュレーションを行いました。各薬の全ヒトタンパク質に対する結合親和性スコアを計算し、PBAS プロファイルを構築しました。

例えば、分娩誘発剤であるクロプロステノールは標的であるプロスタグランジン受容体(PTGFR)と-10.2 kcal/mol で高い結合親和性(図 3-A)、2 型糖尿病の薬であるエルツグリフロジンは標的である SLC5A2 と-10.43 kcal/mol で高い結合親和性(図 3-B)、鎮痛薬であるブプレノルフィン(ブプレノルフィン)は標的である μ オピオイド受容体(OPRM1)と-10.1 kcal/mol で高い結合親和性(図 3-C)を示しました。

このように、立体構造が実験的に決定されていないタンパク質に対しても、コンピュータ上でアミノ酸配列から推定されたタンパク質立体構造に対して薬の相互作用が上手く再現されていることが確認できました。

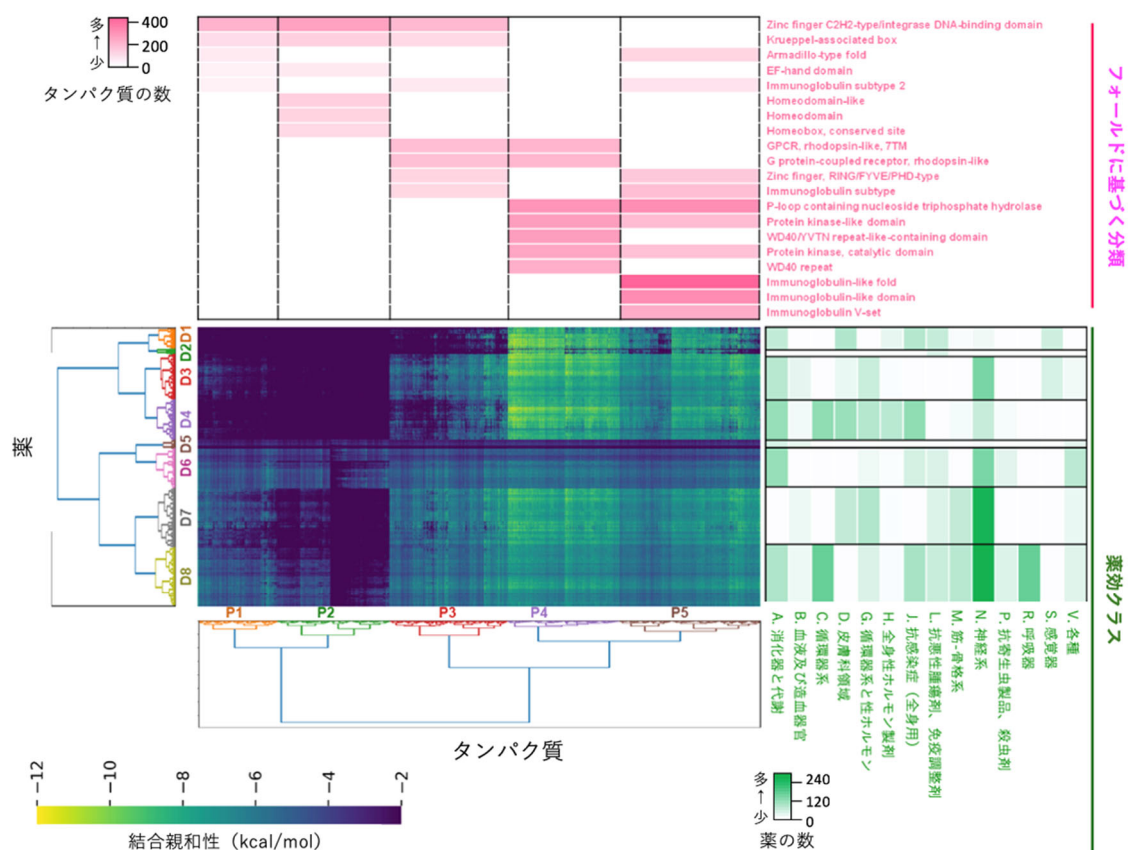


図 2. 全ての既存薬の PBAS プロファイルのクラスタリング解析

7,582 種類の既存薬と 19,135 種類のヒトタンパク質のペアで大規模ドッキングシミュレーションを行い、各薬の全ヒトタンパク質に対する結合親和性スコア (PBAS: Proteome-wide binding affinity score) のプロファイルを構築した。全ての既存薬の PBAS プロファイルをヒートマップで中央に示す。行は薬を表し、列はタンパク質を表す。行列の要素の色は結合親和性の強さを表し、青色から黄色になるに伴い高くなることを示す。各行の相互作用パターンで薬のクラスタリングを行ったところ、グループ化された薬のクラスターは薬効クラスに対応する傾向があり、8 つのクラスター (D1-D8) に分かれた。分類クラスでは、濃い緑になるほど薬の数が増えることを示す。各列の相互作用パターンでタンパク質のクラスタリングを行ったところ、グループ化されたタンパク質のクラスターはタンパク質のフォールドに対応する傾向があり、5 つのクラスター (P1-P5) に分かれた。分類クラスでは、濃いピンクになるほどタンパク質の数が増えることを示す。

次に、各薬の PBAS プロファイルを基に、559 種類の疾患に対する新規効能を予測しました。ある疾患の治療標的タンパク質の立体構造に対して薬の結合親和性が高ければ、その薬は対象疾患に対して効能があると考えられます。

例えば、分娩誘発剤であるクロプロステノールは、PTGFR に対する結合親和性が高かったため、眼科的高血圧と緑内障に対する効能が予測されました。実際に眼圧下降におけるクロプレステノールの効能は動物で報告されており、その類似体はヒトの眼圧上昇の

治療に使用されていることが確認できました。2 型糖尿病薬であるエルツグリフロジンは、FDFT1 に対する結合親和性が高かったため、心血管疾患およびアテローム性動脈硬化性心血管疾患に対する効能が予測されました。エルツグリフロジンは、第 3 相試験で心血管疾患の治療に有効であると報告されており、今回の予測結果の妥当性を示唆する結果となりました。鎮痛薬であるブプレノルフィン、ITGA4 に対する結合親和性が高かったため、腸管における炎症性疾患になどに対する効能が予測されました。ブプレノルフィンは、興味深いことに、過去の研究でマウスモデルで腸の炎症を抑えることが報告されており、それを反映する結果となりました。

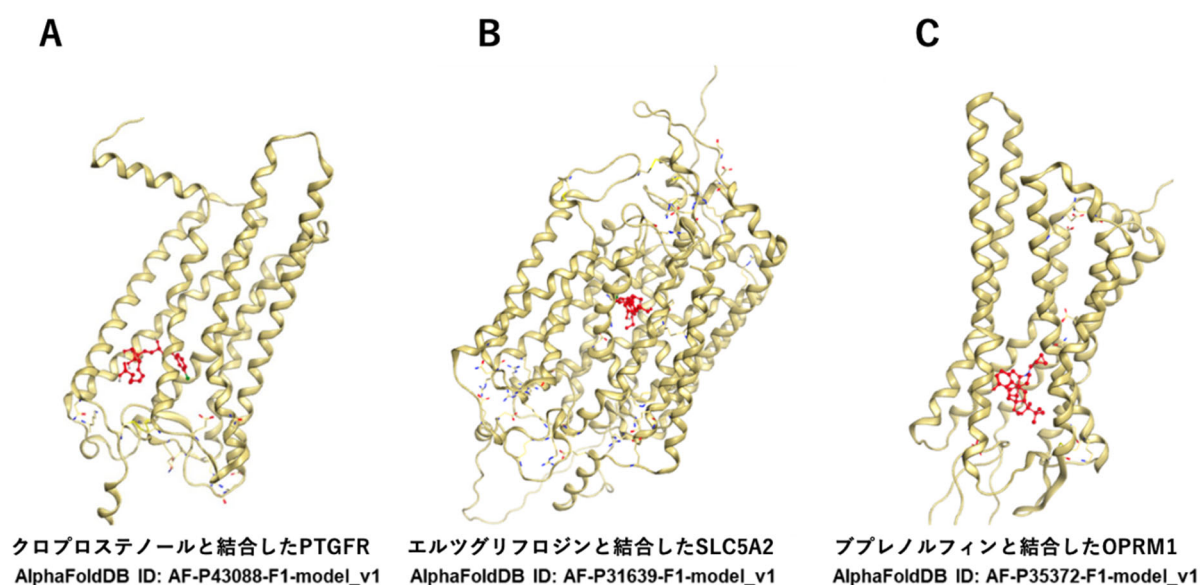


図 3:ドッキングシミュレーションによって推定されたタンパク質と薬の結合様式

薬は赤色のグラフ、タンパク質は黄色のリボンモデルで示した。(A)クロプロステノールと PTGFR の結合様式。(B)エルツグリフロジンと SLC5A2 の結合様式。(C)ブプレノルフィンと OPRM1 の結合様式。これらのタンパク質の立体構造は AlphaFold を用いてアミノ酸配列から推定したものである。

次に、これらの PBAS プロファイルを用いて、薬の毒性・副作用を予測する機械学習モデルの構築を行いました。薬の添付文書に記載されている副作用情報を収集したデータベースである SIDER を、副作用の学習データとして用いました。

スパースモデリングという機械学習技術を用いて構築した副作用予測モデルを、7,582 種類の全ての既存薬に適用し、285 種類の毒性に対する薬の潜在的な副作用を予測しました。予測結果の妥当性の検証には、薬の販売後に報告された薬害事象を収集している米国食品医薬品局の有害事象報告システムである FAERS (FDA Adverse Event Reporting System) を用いました。FAERS は学習に使用した SIDER とは独立したデータセットであり、モデルの学習には用いていない薬に対して予測された副作用が実際に報告されているかを検証しました。

提案した PBAS プロファイルの性能を、これまでの副作用予測の先行研究で提案され

ている他のプロファイル: Fingerprint (化学部分構造プロファイル)、TESS (Target estimation with similarity search プロファイル)、TELR (Target estimation with logistic regression プロファイル) と比較しました (図 4)。

性能評価指標としては、AUC (Area under the ROC curve) と AUPR (Area under the Precision-Recall curve) を用いました。どちらも値が高ければ性能が高いことを意味する精度の指標になります。本研究で提案している PBAS プロファイルの精度は、AUC と AUPR の両方の指標において、統計的に有意なレベルで先行研究のプロファイルの精度を上回り、PBAS が最も良い性能を示すことが確認できました。

これらの結果は、薬の副作用を予測する上で、全てのヒトタンパク質に対する薬の結合パターンを考慮することが有用であることを示唆しています。

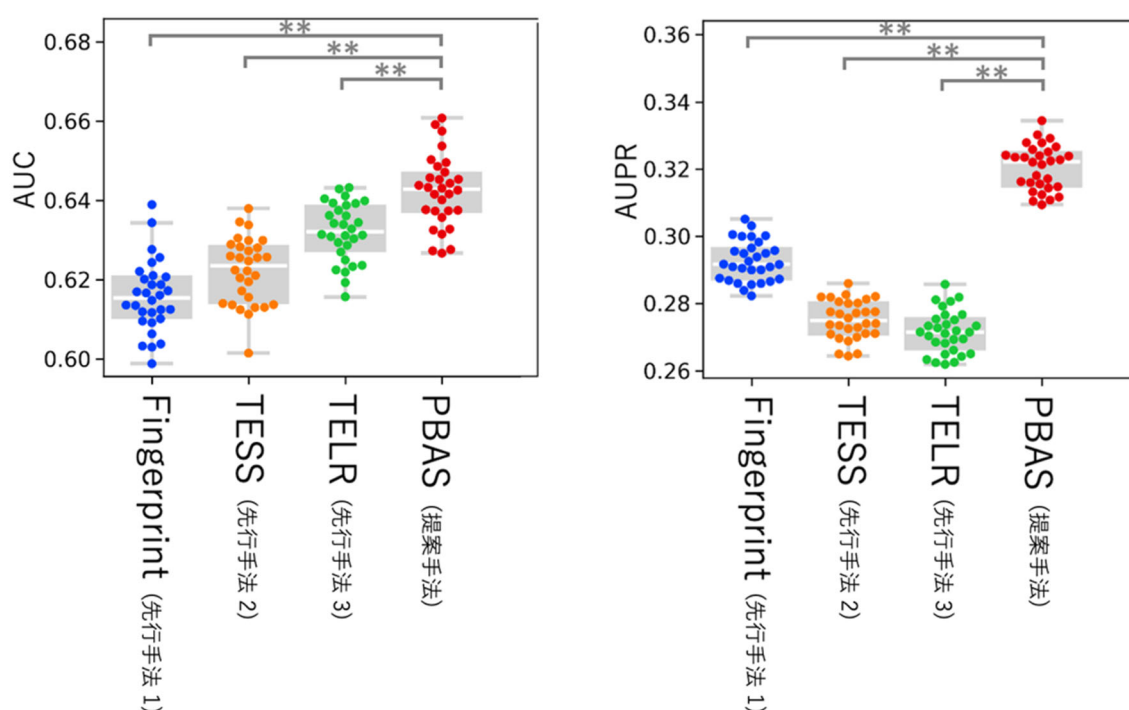


図 4. 副作用予測の性能評価

*は統計的有意水準を示す: ** $p < 0.01$. FARES データセットで性能評価した時の AUC (左) と AUPR (右)。4 つのプロファイルそれぞれについての数値実験を繰り返した ($n = 30$)。Fingerprint (先行手法 1) では、青色、TESS (先行手法 2) はオレンジ、TELR (先行手法 3) は緑色、PBAS (提案手法) は赤色で示す。Fingerprint は薬の化学構造フィンガープリントに基づく部分構造プロファイル、TESS (Target estimation with similarity search) プロファイルは構造類似性検索を用いて推定された薬-タンパク質相互作用に基づく標的タンパク質プロファイル、TELR (Target estimation with logistic regression) プロファイルは教師あり学習により推定された薬-タンパク質相互作用に基づく標的タンパク質プロファイルを表す。

今回構築した副作用予測モデルは、スパースモデリングという機械学習技術を用いているため、予測に重要な特徴を抽出することが可能です。予測モデルにおける重みが大

図 5 は、薬、タンパク質、副作用の3者間のネットワークの一部を示しています。薬とタンパク質のリンクは、薬とタンパク質の結合親和性に対応します。タンパク質と副作用のリンクは、モデルにおける重みに対応します。例えば、デラヒルジンという薬は副作用として肺炎を惹起することが予測され、肺炎に関連するタンパク質の一つとしてキチナーゼ 3 様タンパク質 1 (CHI3L1) が抽出されました。興味深いことに、血液中の CHI3L1 レベルが肺炎重症度と関連があることが実際に報告されていることが確認できました。これらは、機械学習モデルが副作用の発現に関与するタンパク質を上手く抽出していることを意味します。



薬は青丸、タンパク質は赤四角、副作用は緑菱形で示す。水色の実線は、ドッキングシミュレーションから推定された薬とタンパク質の結合親和性を表す。オレンジの点線は、機械学習モデルで重みの大きい副作用とタンパク質の関係を表す。

本研究では、ヒトの全てのタンパク質の立体構造をドッキングシミュレーションと機械学習で情報解析し、薬の効能や毒性・副作用を予測する新しい計算手法を開発しました。ドッキングシミュレーションにより、各薬の全てのヒトタンパク質への結合親和性を計算したPBAS プロファイルを構築することによって、各薬が体内でどのようなタンパク質と結合する可能性があるかを明らかにした点が特色です。実際に、ごく一部のタンパク質としか結合しない特異性の高い薬だけでなく、多くのタンパク質と結合する特異性の低い薬を

同定できました。これは、低分子の医薬品化合物は従来考えられていたよりも多くのタンパク質と相互作用し、様々な副作用の発現に関わっている可能性を示唆しています。

本研究の提案手法は薬の主な治療標的だけでなくオフターゲット^{注 4)}のタンパク質も考慮したポリファーマコロジー^{注 5)}の理解や様々な創薬工程への応用に役立つことが期待されます。

本研究は、JSPS 科学研究費補助金・基盤研究 A 21H04915「医療ビッグデータから難治性疾患の創薬標的を予測する革新的 AI 手法の開発」(研究代表者:山西芳裕)の支援を受け行われました。

【用語説明】

注 1)ドッキングシミュレーション:

タンパク質と薬のドッキングシミュレーションは、結合様式と結合親和性を予測する情報技術である。

注 2)機械学習:

データからパターンを学習し、新しい予測や分類を自動化する情報技術である。AI を支える技術として活用されている。

注 3)AlphaFold:

DeepMind 社(現在の Google DeepMind 社)が開発した AI モデルで、タンパク質のアミノ酸配列からタンパク質の立体構造を高い精度で予測することができる。

注 4)オフターゲット:

本来目標とした標的タンパク質以外に作用してしまうタンパク質。

注 5)ポリファーマコロジー:

薬物の標的分子として単一のタンパク質だけでなくすべてのタンパク質(プロテオーム全体)を考慮する考え方。

【論文情報】

雑誌名:「iScience」

論文タイトル: Predicting therapeutic and side effects from drug binding affinities to human proteome structures

著者:Sawada, R.(岡山大学), Sakajiri, Y. (名古屋大学), Shibata, T.(九州工業大学), and Yamanishi, Y.(名古屋大学)

DOI:10.1016/j.isci.2024.110032

URL: <https://www.sciencedirect.com/science/article/pii/S2589004224012574>