

相互アテンションニューラルネットワークによる 薬剤と糖タンパク質の相互作用解析 Drug-Glycoprotein Interaction Analysis with Mutual Attention Neural Networks

渥美 雅保、新川 栄二
Masayasu Atsumi, Eiji Shinkawa

1 はじめに

医薬データの増加に伴い創薬分野における機械学習を用いた研究が盛んに行われている [1, 2]。このうち、薬剤とタンパク質の相互作用を機械学習により予測することやその部位を解析することは、既存薬剤の新規効能を発見するドラッグリポジショニングのみならず安全で効果的な新薬の発見と開発においても重要である [3, 4, 5]。本報告では、タンパク質を修飾する糖鎖に着目し、薬剤とタンパク質の相互作用を糖鎖の媒介のもとに予測・解析するための深層学習モデルの研究 [6, 7, 8, 9, 10] について述べる。

本研究では、薬剤の化合物情報、タンパク質情報、糖鎖情報の3つの組み合わせを入力してそれらの間の相互作用を予測・解析する相互アテンションニューラルネットワークモデルを提案する。本提案モデルでは、薬剤の化合物情報として分子の化学構造を表記する SMILES 記法 [11] を、タンパク質情報としてアミノ酸配列を、糖鎖情報として糖鎖構造の線形表記法である WURCS [12, 13] 記法を用いる。薬剤の SMILES データは薬剤グラフ構造に、タンパク質のアミノ酸配列データはアミノ酸 n-gram 配列に変換されてモデルに入力され、それぞれグラフニューラルネットワークと畳み込みニューラルネットワークにより薬剤特徴ベクトルとタンパク質特徴ベクトルにエンコードされる。糖鎖の WURCS データは、糖鎖頻度に変換または系列データとしてモデルに入力され、それぞれフィードフォワードニューラルネットワークまたは糖鎖言語モデルにより糖鎖特徴ベクトルにエンコードされる。そして、これら特徴ベクトル間の関連性を相互アテンション機構によりそれぞれの特徴ベクトルに集約したあとで連結してクラシファイアにより相互作用の有無の予測を行う。また、この相互アテンションを解析することにより、薬剤と糖タンパク質の相互作用部位を可視化する。薬剤とタンパク質の特徴エンコーディングには、薬剤とタンパク質間の相互作用を予測する先行研究 [5] で提案されたエンコーダを用いている。この先行研究では、薬剤からタンパク質へのアテンションに基づきそれらの間の相互作用を予測している。

本研究の目的は、提案モデルの相互作用予測における有効性を実験により検証することを通じて、糖鎖に注目した薬剤と糖タンパク質の相互作用予測のための深層学習手法の基礎を確立すること、薬剤標的としての糖鎖の可能性をデータからの機械学習の観点から評価すること、及び薬剤と糖タンパク質の相互作用部位の相互アテンション解析に基づく可視化手法を提供することである。本研究の先行研究に対する新たな独創的な点と貢献は次の3点にまとめられる：

1. 第一に、薬剤とタンパク質の相互作用予測に新たに糖鎖データを組み込んで、糖鎖と薬剤、薬剤とタンパク質間の関連を捉える相互アテンション機構を導入した新しい深層学習モデルを提案する点である。相互アテンション機構は糖鎖がタンパク質と薬剤の反応を仲介するプロセスを表現している。
2. 第二に、糖鎖データの特徴を表現するエンコーダとして、自然言語処理で用いられる表現学習のための言語モデルに基づいて、新たに糖鎖言語モデルを開発した点である。
3. 第三に、提案モデルの相互作用予測における有効性を種々のモデルを比較実験することを通じて検証

するとともに、アテンションの解析により糖鎖と薬剤、薬剤とタンパク質の相互作用部位を可視化する手法を提供する点である。

2 提案モデル

薬剤、糖鎖、タンパク質間の相互作用の予測・解析のための提案モデルに関して、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル [6, 7, 8, 9]、薬剤-糖鎖間アテンションニューラルモデル [6, 7, 9]、及び2つの薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデル [8, 9, 10] について述べる。

2.1 薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル

薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルを図1に示す。本モデルは、薬剤グラフエンコーダ、アミノ酸 n-gram 配列エンコーダ、糖鎖頻度エンコーダの3つのエンコーダと、それらがエンコードする特徴ベクトル間の相互アテンション機構、及び相互作用予測クラシファイアから構成される。糖鎖頻度はタンパク質に付随する糖鎖群の異なる糖鎖の個数を WURCS データから数えたもので、糖鎖頻度ベクトルとして得られる。糖鎖頻度エンコーダは糖鎖頻度ベクトルを入力として糖鎖特徴ベクトルを出力する3層のフィードフォワードニューラルネットワークである。相互アテンション機構は糖鎖の薬剤とタンパク質を仲介する働きをモデル化するために開発したもので、糖鎖特徴をクエリに薬剤サブグラフに対するアテンションを計算して糖鎖との関連重み付きで薬剤特徴をエンコードして、次にその薬剤特徴をクエリにアミノ酸 n-gram 配列特徴に対するアテンションを計算して薬剤との関連重み付きでタンパク質特徴をエンコードする。相互作用予測クラシファイアは、これら3つの特徴ベクトルを連結したベクトルを入力として相互作用の有無を二値分類する。

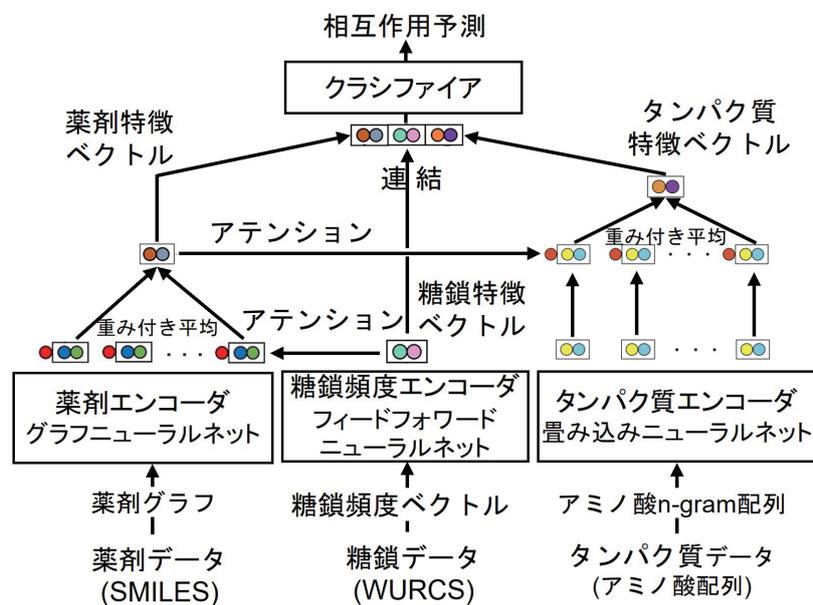


図1：薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル

2.2 薬剤-糖鎖間アテンションニューラルモデル

薬剤-糖鎖間アテンションニューラルモデルを図2に示す。本モデルは、2.1の薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルからタンパク質を扱うサブネットワークを除いて薬剤と糖鎖のネットワークのみを抽出したモデルで、糖鎖の薬剤標的としての役割を特に評価するためのモデルである。

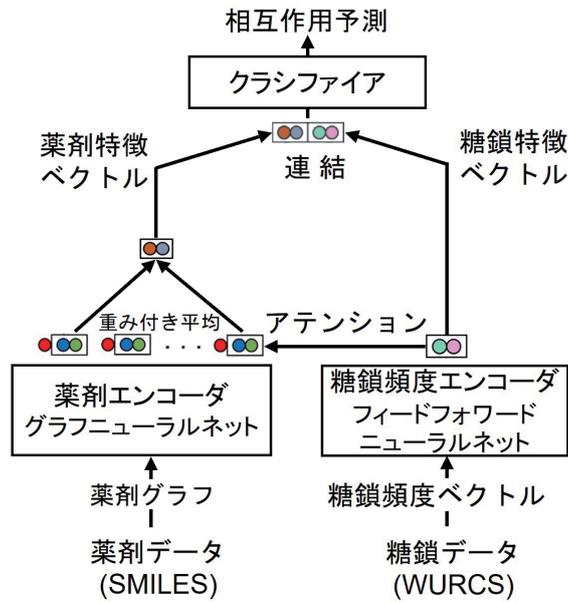


図2：薬剤-糖鎖間アテンションニューラルモデル

2.3 薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデル

薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデルは、糖鎖の単糖の連なり構造を表現し相互作用予測に利用するために、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルの糖鎖頻度エンコーダを糖鎖構造をエンコードする事前学習糖鎖言語モデルに基づくエンコーダに置き換えたモデルである。事前学習糖鎖言語モデルによりエンコードした糖鎖特徴ベクトル系列の相互作用予測への利用の仕方により、糖鎖構造媒介二段階相互アテンションニューラルモデルと糖鎖構造媒介三段階相互アテンションニューラルモデルの2つのモデルを提案する。

2.3.1 事前学習糖鎖言語モデル

糖鎖言語モデルは、WURCS 系列情報を入力として MLM (Masked Language Modeling) 事前学習タスクと下流タスクとしての生物分類タスクにより、糖鎖の単糖の連なり構造を特徴表現学習したモデルである。糖鎖言語モデルの事前学習の過程を図3に示す。

MLM 事前学習では、3.1.1 で述べる全糖鎖構造系列データセットを利用して、自然言語処理において提案されている大規模言語モデル RoBERTa [14] を事前学習糖鎖言語モデルとして訓練する。事前学習に利用可能な全糖鎖構造系列データ数は 4372 で、その内の 9 割を訓練用、1 割をテスト用に分けて訓練に利用する。次に、下流タスクとして 3.1.1 で述べる生物分類データセットを用いて哺乳類と非哺乳類を二値分類する生物分類学習を行う。相互作用予測を行うデータセットはすべてヒトの糖タンパク質のデータであるが、事前学習の段階で哺乳類と非哺乳類に分類する学習を行うことで、相互作用予測の訓練でファインチューニングする際のパラメータを事前学習する。事前学習糖鎖言語モデルは糖鎖構造媒介相互アテンションニューラルモデルに組み込まれて相互作用予測における糖鎖特徴ベクトルのエンコーディングに用いられる。

2.3.2 事前学習糖鎖言語モデルを組み込んだ相互アテンションニューラルモデル

薬剤-糖タンパク質間糖鎖構造媒介二段階相互アテンションニューラルモデルを図4に示す。このモデルは、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルの糖鎖頻度エンコーダを糖鎖言

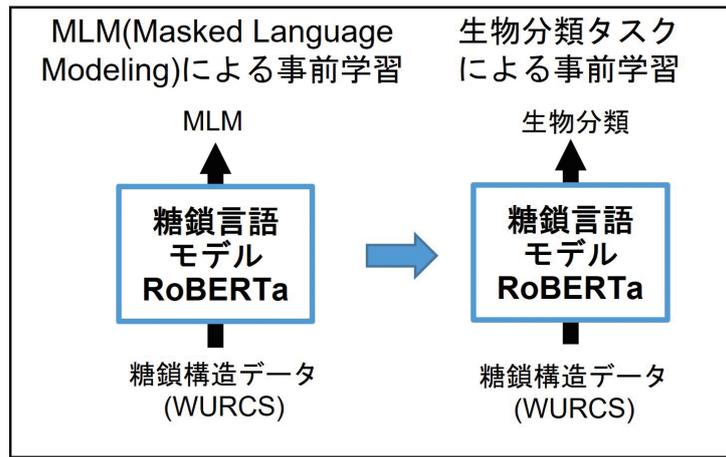


図3：糖鎖言語モデルの事前学習

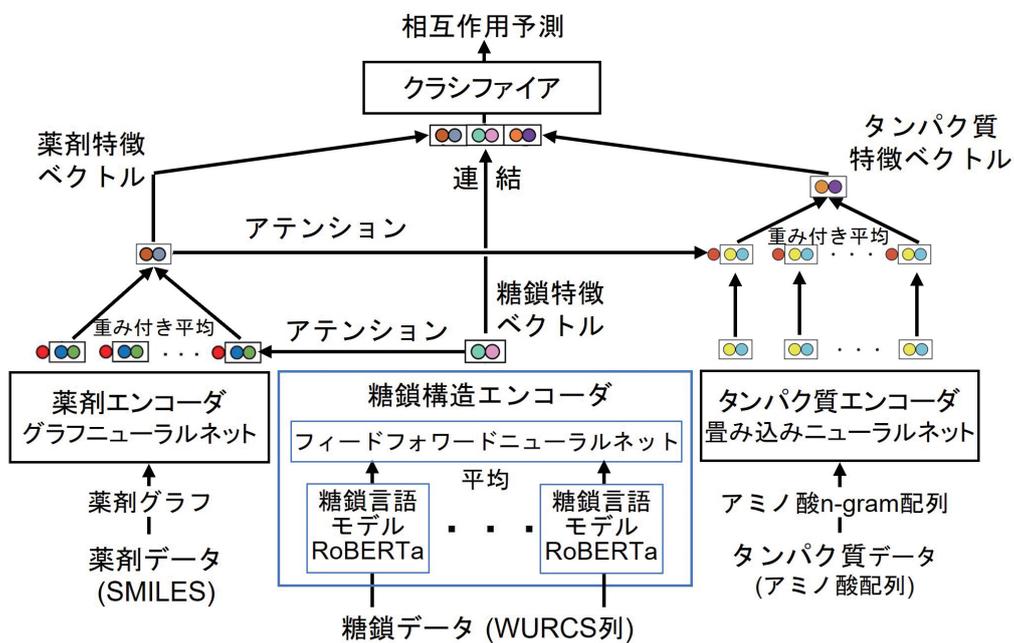


図4：薬剤-糖タンパク質間糖鎖構造媒介二段階相互アテンションニューラルモデル

語モデルに基づくエンコーダに置き換えたモデルで、薬剤グラフエンコーダ、アミノ酸 n-gram 配列エンコーダ、糖鎖言語に基づく糖鎖構造エンコーダの3つのエンコーダと、それらがエンコードする特徴ベクトル間の相互アテンション機構、及び相互作用予測クラシファイアから構成される。糖鎖構造エンコーダでは、まず、糖鎖言語モデルによりタンパク質に付随する各糖鎖をエンコードする。タンパク質に付随する糖鎖は複数ありこれら糖鎖から得られる特徴ベクトルも複数になるため、次にそれらを平均化して集約し、フィードフォワードニューラルネットワークにより最終的な糖鎖特徴ベクトルが得られる。糖鎖構造エンコーダを組み込んだモデルの学習においては、事前学習済み糖鎖言語モデルの最終層近くの層がファインチューニングされる。

このモデルの糖鎖と薬剤間のアテンションによる関連重み付き特徴集約を拡張したモデルが、図5に示す薬剤-糖タンパク質間糖鎖構造媒介三段階相互アテンションニューラルモデルである。このモデルでは、糖鎖構造エンコーダにおいて、糖鎖言語モデルによりエンコードされた各糖鎖の特徴を平均化して集約する前に、薬剤特徴ベクトルからこれら特徴ベクトルへのアテンションが追加されている。ここで、アテンション

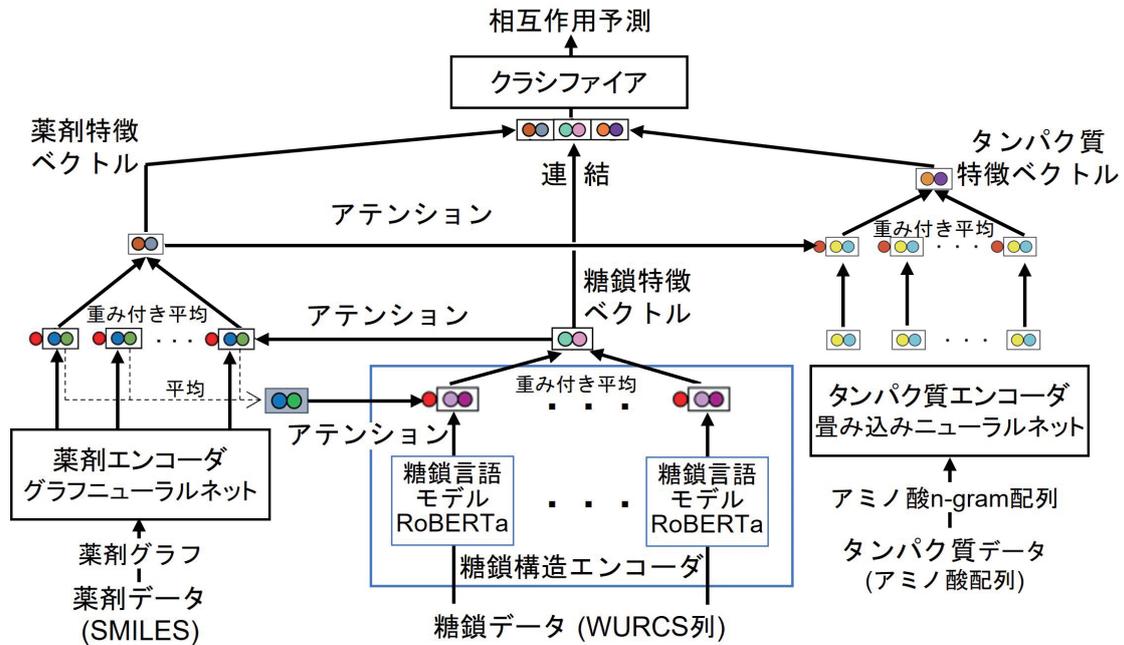


図 5：薬剤 - 糖タンパク質間糖鎖構造媒介三段階相互アテンションニューラルモデル

のクエリとなる薬剤特徴としては、薬剤サブグラフ特徴を平均した特徴ベクトルが用いられる。最終的な糖鎖特徴ベクトルは、各糖鎖の特徴ベクトルをこのアテンションを関連度重みとして重み付き平均することにより得られる。この薬剤特徴ベクトルをクエリにして糖鎖の特徴ベクトルに対して計算したアテンションを解析することにより、糖鎖単位での薬剤との相互作用を解析することが可能となるため、糖鎖構造媒介二段階相互アテンションモデルと比較して相互作用に関するモデルの説明性の向上を図ることができる。

2つの糖鎖構造媒介相互アテンションニューラルモデルでは、糖鎖頻度媒介相互アテンションニューラルモデルと同様に、この得られた糖鎖特徴をクエリに薬剤サブグラフに対するアテンションを計算して糖鎖との関連重み付きで薬剤特徴をエンコードして、次にその薬剤特徴をクエリにアミノ酸 n-gram 配列特徴に対するアテンションを計算して薬剤との関連重み付きでタンパク質特徴をエンコードする。そして、これら3つの特徴ベクトルを連結して相互作用予測クラシファイアに入力して相互作用の有無の二値分類がなされる。

3 実験

実験では、相互作用の予測性能をモデル間で比較して糖鎖情報を予測に用いることの効果を検証し薬剤標的としての糖鎖の可能性を評価する。また、相互アテンションの解析に基づき糖鎖と薬剤、薬剤とタンパク質の相互作用部位の可視化を行う。

3.1 データセット

データセットは、2.3.1 で述べた糖鎖言語モデルの MLM による事前学習に用いる全糖鎖構造系列データセットと生物分類タスクによる糖鎖言語モデルの追加事前学習に用いる生物分類データセット、及び相互作用の予測性能を比較評価するために用いる相互作用予測データセットからなる。

3.1.1 糖鎖言語モデル学習用データセット

糖鎖言語モデルの MLM による事前学習に用いる全糖鎖構造系列データセットを、糖鎖に関連する遺伝子

やタンパク質などの情報を提供する GlyCosmos [15] から抽出する。GlyCosmos から抽出した事前学習に利用可能なデータ数は 4372 である。

また、糖鎖言語モデルの追加事前学習に用いる生物分類データセットを、全糖鎖構造系列データセットと同様に GlyCosmos から抽出する。ここで、生物種は哺乳類と非哺乳類の2つであり、それぞれ 1590 ずつの合計 3180 のデータからデータセットは構成される。

3.1.2 相互作用予測用データセット

相互作用の予測性能の比較評価用に薬剤-糖タンパク質間に相互作用があるペアサンプルをタンパク質との相互作用情報などを含む薬物に関する臨床レベルのデータを集めたデータベース DrugBank [16] から、相互作用がないペアサンプルを信頼できるネガティブサンプルを作成した研究 [17] から集める。また、これらペアサンプルの WURCS データを糖タンパク質などの複合糖質に関する情報を含んだ各種のリソースを整理した統合データベース Glygen [18] から抽出する。

そして、次の2つの実験用に表1のようなサイズの2つのデータセットを作成した。

- 実験1：薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル、薬剤-糖鎖間アテンションニューラルモデル、薬剤-タンパク質間アテンションニューラルモデルの比較
- 実験2：薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル、2つの薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデル、薬剤-タンパク質間アテンションニューラルモデルの比較

ここで、薬剤-タンパク質間アテンションニューラルモデルとは、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルから糖鎖を扱うサブネットワークを除いて薬剤とタンパク質のネットワークのみを抽出したモデルで、薬剤とタンパク質間の相互作用を予測する先行研究 [5] のモデルに相当するモデルである。実験1のデータセットは、相互作用ありデータが1174、相互作用なしデータが2348の計3522サンプル、実験2のデータセットは、相互作用ありデータが521、相互作用なしデータが692の計1213サンプルからなる。

表1：相互作用予測用データセット

データセット	相互作用あり	相互作用なし	合計
実験1 データセット	1174	2348	3522
実験2 データセット	521	692	1213

3.2 相互作用予測性能の比較評価

実験では、相互作用予測用データセットの8割を訓練用、1割を検証用、残りの1割をテスト用とする10分割交差検定によりモデルの予測性能の比較評価を行った。評価指標としては、AUC (Area Under the ROC Curve)、適合率 (Precision)、再現率 (Recall)、及び適合率と再現率の調和平均であるF値 (F-measure)を用いる。また、アミノ酸 n-gram としては3-gram のサブ配列を用いる。

実験1では、糖鎖情報の相互作用予測における効果、及び相互アテンション機構の有効性を示すために、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル、薬剤-糖鎖間アテンションニューラルモデル、薬剤-タンパク質間アテンションニューラルモデルの相互作用予測性能を比較する。表2に実験結果を示す。薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルが AUC が 0.891、F

値が0.752でともに最も高い性能を示し、薬剤-糖鎖間アテンションニューラルモデルの性能をそれぞれ0.024、0.012ポイント、薬剤-タンパク質間アテンションニューラルモデルの性能をそれぞれ0.065、0.056ポイント上回った。また、薬剤-糖鎖間アテンションニューラルモデルが薬剤-タンパク質間アテンションニューラルモデルより高い性能を示した。これら実験結果より、糖鎖情報を用いることでより高い相互作用予測性能が得られること、及び相互アテンション機構がその効果を引き出すために適切に機能していることが確かめられた。特に、薬剤と糖タンパク質の相互作用予測において、糖鎖、タンパク質を単独で用いるよりも両方を共に用いることが予測性能の向上に寄与することがわかった。また、薬剤-糖鎖間アテンションニューラルモデルの方が薬剤-タンパク質間アテンションニューラルモデルよりも相互作用予測性能が高いことから、糖鎖自体の薬剤標的としての有効性が示唆された。

表2：実験結果1

モデル名	AUC	Precision	Recall	F 値
薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル	0.891	0.733	0.773	0.752
薬剤-糖鎖間アテンションニューラルモデル	0.867	0.725	0.756	0.740
薬剤-タンパク質間アテンションニューラルモデル	0.826	0.653	0.745	0.696

実験2では、糖鎖構造情報の相互作用予測における効果、及び事前学習糖鎖言語モデルの有効性を示すために、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル、2つの薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデル、薬剤-タンパク質間アテンションニューラルモデルの相互作用予測性能を比較する。表3に実験結果を示す。2つの薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデルが薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルと薬剤-タンパク質間アテンションニューラルモデルの性能を上回った。このうち薬剤-糖タンパク質間糖鎖構造媒介二段階相互アテンションニューラルモデルがAUCが0.924、F値が0.832で最も高い予測性能を示し、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルの性能をAUCで0.009、F値で0.014ポイント上回り、薬剤-タンパク質間アテンションニューラルモデルの性能をAUCで0.072、F値で0.075ポイント上回った。また、実験1と同様に、糖鎖頻度を用いる薬剤-糖タンパク質間相互アテンションニューラルモデルが薬剤-タンパク質間アテンションニューラルモデルの性能を上回ることも再確認された。これら実験結果より、糖鎖構造情報を用いることでより高い相互作用予測性能が得られること、及び事前学習糖鎖言語モデルがその効果を引き出すために適切に機能していることが確かめられた。一方、薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデルにおいて、三段階の相互アテンションは2.3.2で述べたように糖鎖単位での薬剤との相互作用を解析するために導入されたが、相互アテンションを二段階から三段階にすることの相互作用予測における効果は見られなかったことから、相互アテンション機構の設計にさらなる工夫が必要なことも示唆された。

表3：実験結果2

モデル名	AUC	Precision	Recall	F 値
薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル	0.915	0.827	0.813	0.820
薬剤-糖タンパク質間糖鎖構造媒介二段階相互アテンションニューラルモデル	0.924	0.812	0.858	0.834
薬剤-糖タンパク質間糖鎖構造媒介三段階相互アテンションニューラルモデル	0.915	0.825	0.836	0.830
薬剤-タンパク質間アテンションニューラルモデル	0.852	0.743	0.775	0.759

3.3 相互アテンション解析に基づく相互作用部位の可視化

薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルと薬剤-糖タンパク質間糖鎖構造媒介三段階相互アテンションニューラルモデルにおけるアテンションを解析することにより、糖鎖-薬剤間と薬剤-タンパク質間の相互作用部位を可視化することを通じて、相互作用に対するモデルの説明性を向上させる。

薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルの糖鎖から薬剤サブグラフへと薬剤からタンパク質アミノ酸サブ配列への2つのアテンションを解析して可視化したアテンションヒートマップをそれぞれ図6aと図6bに示す。これらアテンションヒートマップから、糖鎖頻度ベクトルがどの薬剤サブグラフと、また、薬剤がどのアミノ酸サブ配列と、どの程度の強さで関連しているかが色の明るさでわかる。例えば、図6aのアテンションヒートマップ上の点に対して、図7aのあるIDの糖鎖頻度情報が図7bまたは図7cのIDの薬剤サブグラフとそれぞれ0.217と0.003のアテンションを持つこと、即ち図7bの薬剤サブグラフの方が図7cの薬剤サブグラフより強い関連を持つことを知ることができる。これより、薬剤サブグラフ7bを含むような薬剤は7aのような糖鎖頻度情報を持つタンパク質に対して有効であると推測することが可能となる。

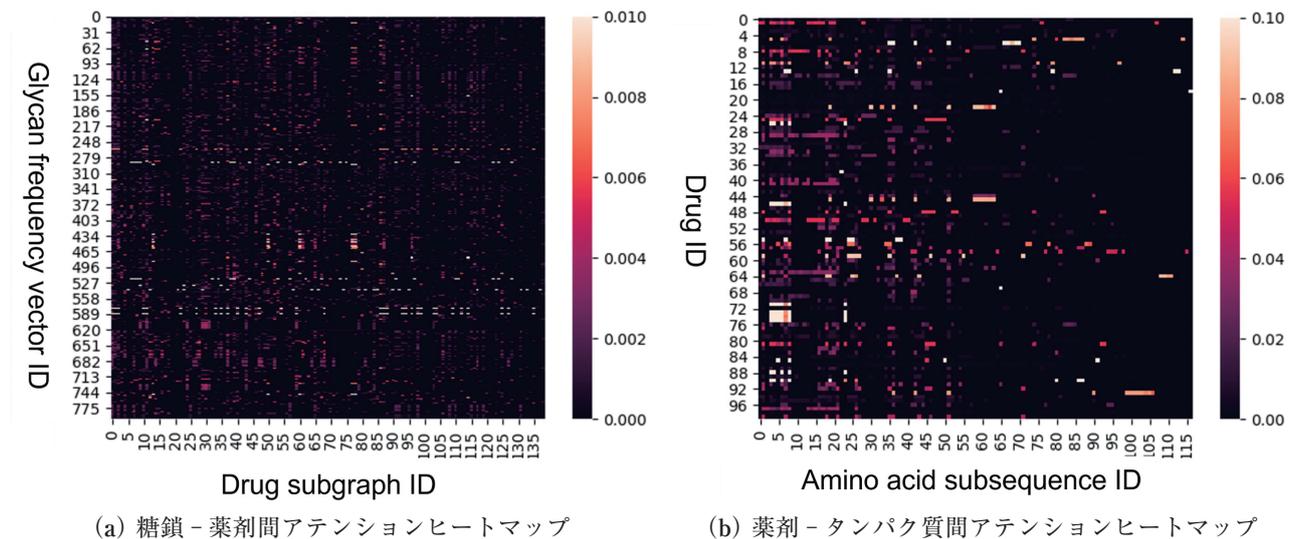
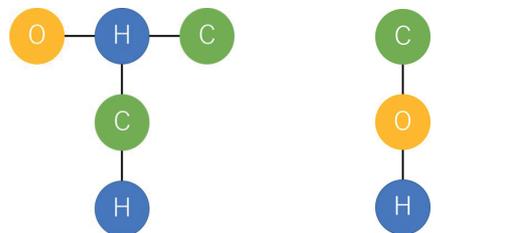


図6：糖鎖頻度媒介相互アテンションのアテンションヒートマップ

糖鎖情報	頻度
uxxxxh, uxxxxh_2*NCC\3=O	2
uxxxxh_2*NCC\3=O	1
a1122h-1a.1-5, a1122h-1b.1-5, a1221m-1a.1-5, a2112h-1b.1-5, a2122h-1b.1-5_2*NCC\3=O	16

(a) 糖鎖頻度情報



(b) 薬剤サブグラフ 1 (c) 薬剤サブグラフ 2

図7：図6aのアテンションヒートマップの点に対応する糖鎖頻度情報と薬剤サブグラフの例

薬剤-糖タンパク質間糖鎖構造媒介三段階相互アテンションニューラルモデルの薬剤からタンパク質に付随する複数の糖鎖へのアテンションを解析して可視化したアテンションヒートマップを図8に示す。このアテンションヒートマップから、薬剤がどの糖鎖とどの程度の強さで関連しているかが色の明るさでわかる。例えば、図8のアテンションヒートマップ上の点に対して、図9aの構造を持つあるIDの薬剤が図9bの糖

鎖1または糖鎖2のIDの糖鎖構造とそれぞれ0.226と0.021のアテンションを持つこと、即ち糖鎖1の糖鎖構造の方が糖鎖2の糖鎖構造より強い関連を持つことを知ることができる。これより、図9aのような構造を持つ薬剤は図9bの糖鎖1のような構造を持つ糖鎖が付随するタンパク質に対して有効であると推測することが可能となる。

薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデルと薬剤-糖タンパク質間糖鎖構造媒介二段階相互アテンションニューラルモデルでは、糖鎖から薬剤へのアテンションと薬剤からタンパク質へのアテンションの解析により、それぞれタンパク質に付随する糖鎖群の薬剤サブ構造との相互作用部位の可視化と薬剤のアミノ酸サブ配列との相互作用部位の可視化が提供された。また、薬剤-糖タンパク質間糖鎖構造媒介三段階相互アテンションモデルで追加した薬剤からタンパク質に付随する複数の糖鎖へのアテンションの解析により、薬剤の各糖鎖との糖鎖単位での相互作用部位の可視化が提供された。

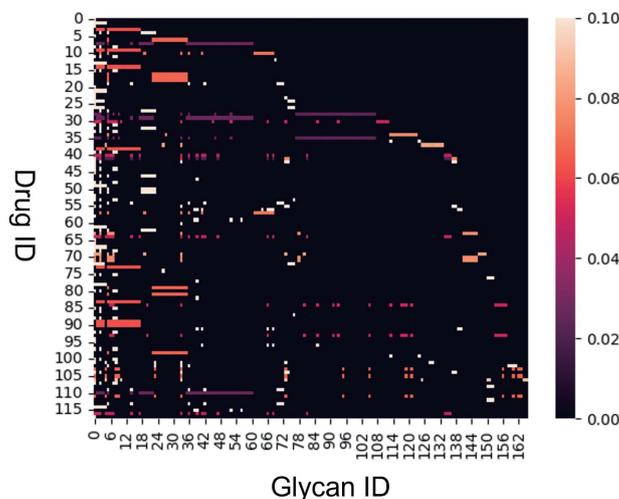


図8：糖鎖構造媒介三段階相互アテンションの薬剤-糖鎖間アテンションヒートマップ

	糖鎖構造情報
[H][C@]12SCC(COC(C)=O)=C(N1C(=O)[C@H]2NC(=O)C(=N/OC)¥C1=CSC(N)=N1)C(O)=O	糖鎖1 axh*NCC3=O axh*NCC3=O axh5 axh5 axh*NCC3=O axh5 Aadxh*NCC3=O axh5 axh*NCC3=O axh5 Aadxh*NCC3=O
(a) 薬剤情報	糖鎖2 axh*NCC3=O axh*NCC3=O axh5 axh5 axh*NCC3=O axh5 axh*NCC3=O axh5 axh5 axh*NCC3=O axh5
	(b) 糖鎖情報

図9：図8のアテンションヒートマップの点に対応する薬剤と糖鎖の例

4 むすび

本報告では、薬剤とタンパク質の相互作用を媒介する糖鎖の働きに着目し、薬剤・糖鎖・タンパク質間の相互作用を予測・解析する相互アテンション機構に基づく深層学習モデルについて述べた。深層学習モデルとして、薬剤-糖タンパク質間糖鎖頻度媒介相互アテンションニューラルモデル、薬剤-糖鎖間アテンションニューラルモデル、及び事前学習糖鎖言語モデルを組み込んだ2つの薬剤-糖タンパク質間糖鎖構造媒介相互アテンションニューラルモデルを提案し、相互作用予測実験を通じて、(1) 相互アテンション機構に基づき糖鎖情報をモデルに組み込むことでより高い相互作用予測性能が得られること、(2) 事前学習糖鎖言語モデルに基づき糖鎖構造情報をエンコーディングすることでより高い相互作用予測性能が得られること、を示して、糖鎖の薬剤標的としての働きを評価するとともに、相互アテンション機構と事前学習言語モデルの

有効性を確かめた。また、薬剤からタンパク質に付随する複数の糖鎖へのアテンション、糖鎖群から薬剤サブグラフへのアテンション、薬剤からタンパク質アミノ酸サブ配列へのアテンションを解析してアテンションヒートマップとして可視化することを通じて、薬剤がどの糖鎖と、糖鎖群がどの薬剤サブグラフと、薬剤がどのアミノ酸サブ配列とそれぞれどの程度の強さで関連しているか表示することを可能とした。今後の課題としては、提案モデルで検出された相互作用を生物学的な実験を通じて検証すること、深層学習に基づく手法と生物学的実験に基づく手法を融合させること、それらの過程の中で相互作用予測モデルの改良をすること、などが挙げられる。

謝辞

本研究は、創価大学理工学部の渥美研究室と木下研究室の共同研究として、また、新川栄二君の卒業研究、及び修士研究として行われた。本研究の議論に参加してくれた渥美研究室の村田祐樹君、永塚光一君、実験で用いたデータセットの準備に協力いただいた木下研究室の小野多美子さん、細田正恵さん、そして、本共同研究の機会を与えていただいた木下聖子先生に感謝いたします。

参考文献

- [1] Askar, H., et al.: Deep learning in drug discovery: an integrative review and future challenges, *Artificial Intelligence Review*, 2022.
- [2] Vamathevan, J., et al.: Applications of machine learning in drug discovery and development, *Nature Reviews Drug Discovery*, Vol.18, pp.463-477, 2019.
- [3] Bagherian, M., et al.: Machine learning approaches and databases for prediction of drug-target interaction: a survey paper, *Briefings in Bioinformatics*, Vol.22, Issue 1, pp.247-269, 2021.
- [4] Wang, P., et al.: Structure-Aware Multimodal Deep Learning for Drug-Protein Interaction Prediction, *Journal of Chemical Information and Modeling*, Vol.62, No.5, pp.1308-1317, 2022.
- [5] Tsubaki, M., et al.: Compound-protein interaction prediction with end-to-end learning of neural networks for graphs and sequences, *Bioinformatics*, Vol.35, No.2, pp.309-318, 2019.
- [6] 新川栄二, 永塚光一, 村田祐樹, 小野多美子, 細田正恵, 木下聖子, 渥美雅保: 糖鎖情報を組み込んだニューラルネットワークによる薬剤とタンパク質の相互作用予測, *情報処理学会第83回全国大会論文集*, 4ZC-02, 第4分冊, pp.469-470, 2021.
- [7] 新川栄二, 永塚光一, 村田祐樹, 小野多美子, 細田正恵, 木下聖子, 渥美雅保: 相互アテンションニューラルネットワークによる糖タンパク質と薬剤のインタラクション解析, *人工知能学会第35回全国大会論文集*, 2Xin5-07, 4p., 2021.
- [8] 新川栄二, 永塚光一, 村田祐樹, 小野多美子, 細田正恵, 木下聖子, 渥美雅保: 事前学習言語モデルを用いた糖鎖エンコーディングに基づく糖タンパク質と薬剤のインタラクション予測, *人工知能学会第36回全国大会論文集*, 4p., 2022.
- [9] Eiji Shinkawa, Koichi Nagatsuka, Yuki Murata, Tamiko Ono, Masae Hosoda, Kiyoko F. Aoki-Kinoshita, Masayasu Atsumi: Prediction and Analysis of Drug-Glycoprotein Interactions with Mutual Attention Neural Networks. *Proc. 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems*, F-2-B-1, 2022.

- [10] 新川栄二, 永塚光一, 村田祐樹, 小野多美子, 細田正恵, 木下聖子, 渥美雅保: 糖鎖構造媒介相互アテンションニューラルネットワークによる糖タンパク質と薬剤間のインタラクション解析, 情報処理学会第85回全国大会論文集, 6ZG-07, 第4分冊, pp.545-546, 2023.
- [11] Weininger, D.: SMILES, a chemical language and information system. I.Introduction to methodology and encoding rules, Journal of Chemical Information and Modeling, Vol.28, No.1, pp.31-36, 1988.
- [12] Tanaka, K., et al.: WURCS: The Web3 Unique Representation of Carbohydrate Structures, Journal of Chemical Information and Modeling, Vol.54, No.6, pp.1558-1566, 2014.
- [13] Matsubara, M., et al.: WURCS 2.0 Update To Encapsulate Ambiguous Carbohydrate Structures. Journal of Chemical Information and Modeling, Vol.54, No.4, pp.632-637, 2017.
- [14] Liu, Y., et al.: RoBERTa: A Robustly Optimized BERT Pretraining Approach, arXiv:1907.11692, 2019.
- [15] Yamada, I., et al.: The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences, Nature Methods, Vol.17, pp.649-650, 2020.
- [16] Wishart, D. S., et al.: DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Res., Vol.46, No.D1, D1074-D1082, 2018.
- [17] Liu, H., et al.: Improving compound-protein interaction prediction by building up highly credible negative samples, Bioinformatics, Vol.31, No.12, pp.i221-i229, 2015.
- [18] York, W.S., et al.: GlyGen: computational and informatics resources for glycoscience, Glycobiology, Vol.30, No.2, pp.72-73, 2020.