

# 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

新川 栄二<sup>†</sup> 永塚 光一<sup>†</sup> 村田 祐樹<sup>†</sup> 小野 多美子<sup>‡</sup>

細田 正恵<sup>‡</sup> 木下 聖子<sup>‡</sup> 渥美 雅保<sup>†</sup>

<sup>†</sup>創価大学大学院理工学研究科 <sup>‡</sup>創価大学糖鎖生命システム融合研究所

## 1. 研究の背景と目的

近年、新型コロナウイルス感染症の流行が世界中で問題となっているが、日々発見される新たな感染症などにおいて迅速に治療薬を開発することが求められる。こうした背景のもと機械学習を用いた薬剤と標的タンパク質のインタラクションを予測する研究が盛んに行われている。筆者らは先行研究において、タンパク質を修飾する糖鎖の構造情報をインタラクションモデルに利用することで薬剤と標的タンパク質のインタラクション予測精度の向上を確認した[1]。本研究では、糖鎖構造が媒介するインタラクション解析を目的に、新たな薬剤-糖タンパク質間におけるインタラクションモデルを開発する。

## 2. 関連研究

筆者らは先行研究において、薬剤構造情報、アミノ酸配列情報、糖鎖頻度情報を用いて、薬剤と標的糖タンパク質のインタラクションを予測する糖鎖頻度媒介相互アテンションニューラルネットワーク[2]と糖鎖頻度情報の代わりに糖鎖構造情報を用いた糖鎖構造媒介二段相互アテンションモデル[1]を提案している。これらのモデルはタンパク質と薬剤のインタラクションを媒介する糖鎖の働きをモデル化した相互アテンション機構を採用したモデルであり、薬剤エンコーダーと糖タンパク質エンコーダーが相互に作用する構造となっている。薬剤構造情報とアミノ酸配列情報のみを用いる薬剤-タンパク質間インタラクションモデル[3]と比較して、予測精度が向上した実験結果が得られ、薬剤-糖タンパク質間のインタラクション予測における糖鎖情報の有用性が示されている。

## 3. 提案モデル

### 3.1. 糖鎖構造媒介三段相互アテンションモデル

モデルを図1に示す。糖鎖構造媒介三段相互アテンションモデルは、糖鎖構造媒介二段相互アテンションモデルに対してさらに薬剤から糖鎖間へのアテンション機構を追加したモデルである。糖鎖エンコーダーでは、糖鎖構造情報を WURCS[4]を用いて系列変換した後に RoBERTa[5]に基づく糖鎖言語モデルを通して糖鎖特徴ベクトルとして出力する。ここでタンパク質を修飾する糖鎖の数は複数であり、得られる糖鎖特徴ベクトルも複数になるため、薬剤から糖鎖へのアテンション機構において、各糖鎖特徴ベクトルに対してアテンションをとった後に加重平均して最終的な糖鎖特徴ベクトルとしてエンコードする。その後、糖鎖構造媒介二段相互アテンションモデルと同様に糖鎖から薬剤、薬剤からタンパク質へのアテンション機構を経て、薬剤特徴ベクトルとタンパ

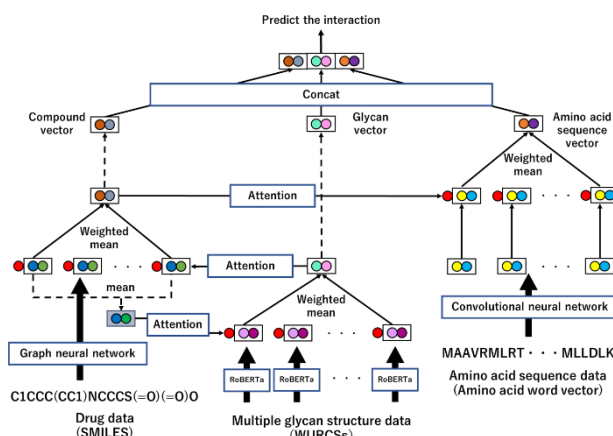


図1. 糖鎖構造媒介三段相互アテンションモデル

ク質特徴ベクトルがそれぞれのエンコーダーにより得られ、糖鎖特徴ベクトルと合わせて連結した後に最終的にインタラクションの有無を分類する。追加した薬剤から糖鎖への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インタラクションを分析することができるため、糖鎖構造媒介二段相互アテンションモデルと比較して糖鎖単位での新薬開発の情報産出やモデルの説明性の向上が可能になる。

## 4. データセット

薬剤-糖タンパク質間にインタラクションがあるペアサンプルをタンパク質とのインタラクション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[6]から、インタラクション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[7]から集める。また薬剤-糖タンパク質間のペアサンプル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[8]から抽出する。インタラクションありデータを521、インタラクションなしデータを692の計1213サンプル用意した。

## 5. 実験

### 5.1. インタラクション予測

#### 5.1.1. 実験設定

本実験では、糖鎖構造媒介三段相互アテンションモデルのインタラクション予測性能を薬剤-タンパク質間インタラクションモデル、糖鎖頻度媒介相互アテンションモデル、糖鎖構造媒介二段相互アテンションモデルと比較する。性能評価指標としては AUC を採用する。また、データ数が両実験ともに少数であるため、10-分割交差検定を行い、結果の平均をとる。データセットの内訳は訓練用データセットを全体の8割、開発用データセットを1割、テスト用データセットを残り1割とする。

Interaction Analysis between Glycoproteins and Drugs by Glycan Structure-Mediated Mutual Attention Neural Network

Eiji Shinkawa<sup>†</sup>, Koichi Nagatsuka<sup>†</sup>, Yuki Murata<sup>†</sup>, Tamiko Ono<sup>‡</sup>, Masae Hosoda<sup>‡</sup>, Kiyoko Kinoshita<sup>‡</sup>, Masayasu Atsumi<sup>†</sup>

<sup>†</sup>Graduate School of Science and Engineering, Soka University

<sup>‡</sup>Glycan & Life System Integration Center, Soka University

表 1. 実験結果比較

| モデル                  | AUC          |
|----------------------|--------------|
| 薬剤-タンパク質間インタラクションモデル | 0.852        |
| 糖鎖頻度媒介相互アテンションモデル    | 0.915        |
| 糖鎖構造媒介二段相互アテンションモデル  | <b>0.924</b> |
| 糖鎖構造媒介三段相互アテンションモデル  | 0.915        |

5.1.2. 実験結果と考察

全モデルのそれぞれの AUC を表 1 に示す。これらの実験結果から、糖鎖構造媒介三段相互アテンションモデルは 91.5% で、糖鎖頻度媒介相互アテンションモデルと同等の予測性能が得られた。92.4% と最も高い予測性能が得られた糖鎖構造媒介二段相互アテンションモデルとの差は 0.9% であった。また、これらの薬剤-糖タンパク質間インタラクションモデルの予測性能は薬剤-タンパク質間インタラクションモデルよりも高く、約 6~7% の向上が確認された。これらの結果から、糖鎖構造情報を薬剤-タンパク質間のインタラクション予測に用いると糖鎖頻度情報を用いた場合よりも同等以上の予測精度が得られることがわかり、また糖鎖構造媒介三段相互アテンションモデルも糖鎖構造媒介二段相互アテンションモデルと同程度の予測性能があることが確認された。

5.2. アテンション分析

5.2.1. 実験設定

本実験では、新薬開発のための糖鎖に関連した情報産出やモデルの説明性の向上を目的として、糖鎖構造媒介三段相互アテンションモデルで追加した薬剤から糖鎖へのアテンション機構におけるアテンション分析を行う。

5.2.2. 実験結果と考察

糖鎖構造媒介三段相互アテンションモデルにおける薬剤から糖鎖へのアテンションのヒートマップを図 3 に示す。また、アテンションの例として、図 4 の薬剤構造情報 ID:42 の、表 2 における糖鎖構造情報 ID:1,3 に対するアテンションを表 3 に示す。この例では、糖鎖構造情報 ID:1 のアテンションは 0.226、糖鎖構造情報 ID:3 のアテンションは 0.021 であり、薬剤構造情報 ID:42 に対して、糖鎖構造情報 ID:1 のほうが ID:3 よりインタラクションにおける関連度が高いことが示される。このことから図 4 のような薬剤は表 2 における ID:1 のような構造を持つ糖鎖が修飾するタンパク質に対して有効である可能性がある。

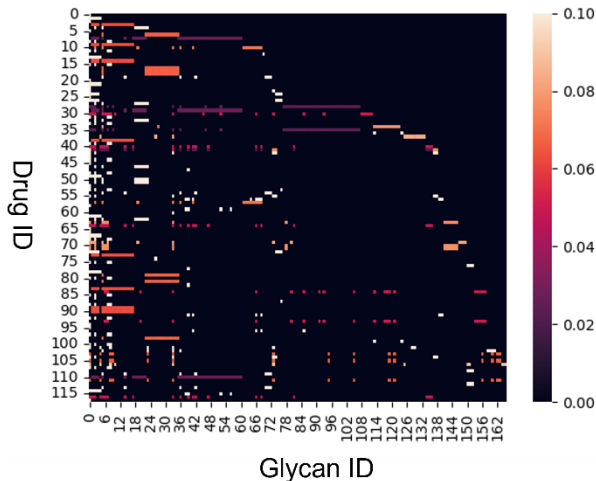


図 3. 薬剤構造情報

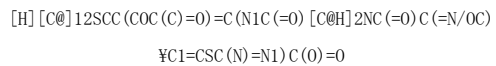


図 4. 薬剤構造情報 (ID:42)

表 2. 糖鎖構造情報

| 糖鎖構造情報 |   |
|--------|---|
| ID:1   | axh*NCC3=O  axh*NCC3=O  axh5 axh5 axh*NCC3=O <br> axh5  Aadxh*NCC3=O  axh5 axh*NCC3=O  axh5  Aadxh*NCC3=O |
| ID:3   | axh*NCC3=O  axh*NCC3=O  axh5 axh5 axh*NCC3=O <br> axh5 axh*NCC3=O  axh5 axh5 axh*NCC3=O  axh5             |

表 3. 薬剤-糖鎖間アテンション比較

| 糖鎖構造情報 | アテンション       |
|--------|--------------|
| ID:1   | <b>0.226</b> |
| ID:3   | 0.021        |

6. 結論

本研究では、糖鎖構造が媒介するインタラクション解析を目的に、新たな薬剤-糖タンパク質間におけるインタラクションモデルである糖鎖構造媒介三段相互アテンションモデルを開発した。実験の結果、まず糖鎖構造媒介二段相互アテンションモデルと同程度の予測性能があることを確認した。また、薬剤-糖鎖間におけるアテンションを解析することで、インタラクションにおける糖鎖単位の分析例を示し、新薬開発のための情報産出に対する可能性を示した。

参考文献

- [1] 新川栄二, 他, 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, 第 36 回人工知能学会全国大会論文集, JSAI2022 巻, p. 4C3GS1002, 2022.
- [2] 新川栄二, 他, 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, 第 83 回全国大会講演論文集, Vol.2021, No.1, pp.469-470, 2021.
- [3] 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.
- [4] 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.
- [5] Liu, Y., et al., RoBERTa: A Robustly Optimized BERT Pretraining Approach, arXiv:1907.1169, 2019.
- [6] 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.
- [7] 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.
- [8] Yamada, I., The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences. Nat Methods. 649-650, 2020.