

クラス A G タンパク質共役受容体専用エンコーダを用いた タンパク質-化合物相互作用予測

山根 永暉^{†1,a)} 石田 貴士^{†1}

概要：G タンパク質共役受容体 (G protein-coupled receptor, GPCR) は生理学および構造的特徴からクラス A~F の 6 クラスに分類される。その中でも最大の GPCR クラスを構成するクラス A GPCR は、リガンドとの結合部位がヘリックス部分であることが判明しており、そのリガンドを同定することは創薬や機能的役割の新たな発見につながる。近年、深層学習を用いたタンパク質-化合物相互作用予測の研究が盛んだが、多くの研究では様々なタンパク質ファミリーに対応できるようタンパク質配列全長を入力として使用している。しかしながら、クラス A GPCR の様にヘリックス部分が化合物との結合に重要と判明している場合、末端領域や細胞内、細胞外ループ領域はノイズとしてエンコーディングされる可能性がある。そこで本研究ではタンパク質をクラス A GPCR に絞った化合物相互作用データセットを作成し、クラス A GPCR のヘリックス部分のアミノ酸配列に注目したエンコーダーモデルを提案する。作成したエンコーダーモデルを、タンパク質-化合物相互作用予測において優れた性能を発揮した TransformerCPI のエンコーダーとして使用した場合、クラス A GPCR データセットにおいてオリジナルのモデルより優れた性能を示した。

キーワード：クラス A GPCR, タンパク質-化合物相互作用予測, TransformerCPI

1. 序論

1.1 G タンパク質共役受容体

G タンパク質共役受容体 (G protein-coupled receptor, GPCR) はホルモンや神経伝達物質、感覚刺激といった様々なリガンドによって活性化される 7 回膜貫通型タンパク質である。[1] その特性によって多くの疾患と関与しているため、承認されている薬剤の約 35% は GPCR をターゲットとしている。[2] これは GPCR が細胞外環境での分子相互作用を容易にする細胞膜で発現することや、化合物との結合部位が明確であることに起因しており、GPCR のリガンドを新規で発見することは創薬や GPCR の特性の解明に非常に有用である。[1] GPCR は生理学的特徴や構造的特徴により、ロドプシン様のクラス A、セクレチン様のクラス B、代謝型グルタミン酸受容体のクラス C、フェロモン受容体のクラス D、cAMP 受容体のクラス E の 6 クラスに分類される。[3][4] それぞれのクラスに分類される GPCR は、7 回膜貫通領域 (7 Transmembrane, 7TM) を持っていることが共通の特徴として挙げられるが、その末

端領域等の構造でクラス特有の特徴が見られ、それにより GPCR のクラス間で化合物との結合領域にも違いがある。例えば、クラス A では化合物の結合領域は 7 回膜貫通領域のみだが、クラス C では 7 回膜貫通領域に加えて Venus flytrap ドメイン (VFT) と呼ばれるサブユニットでリガンドの結合に関与している。(図 1)[1]

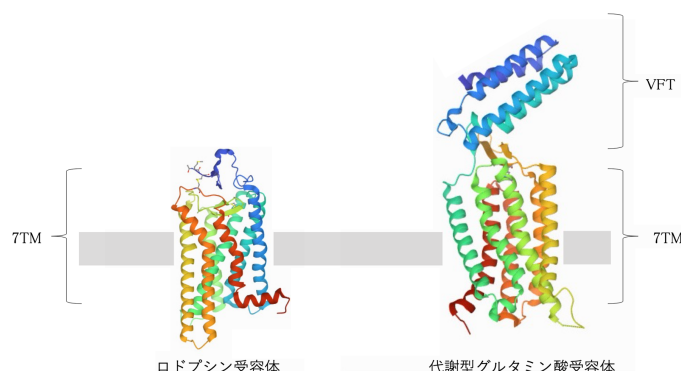


図 1 クラス A GPCR とクラス C GPCR の構造 (chain A)

1.2 化合物-タンパク質相互作用予測

化合物-タンパク質相互作用 (Compound-Protein Inter-

^{†1} 現在, 東京工業大学
Presently with Tokyo Institute of Technology
^{a)} yamane@cb.cs.titech.ac.jp

action, CPI) を特定することは、新規のリガンドの発見において非常に重要である。[5] CPI 予測の研究は何十年もドッキングシミュレーションを用いたタンパク質の立体構造ベースが主流であったが、立体構造が未知であるタンパク質に関しては適さないスクリーニング方法であることから、立体構造を用いずタンパク質配列と化合物構造情報を使用した機械学習における CPI 予測が行われるようになった。[6] 近年では畳み込みニューラルネットワークを用いてタンパク質配列と化合物をエンコーディングし特徴量の抽出を行う DeepDTA[7] や、畳み込みニューラルネットワークの代わりにグラフニューラルネットワーク [8] を用いた CPI-GNN[9]、GraphDTA[10] といった深層学習を用いた CPI 予測の研究が行われており、2020 年には TransformerCPI[11] と呼ばれる attention 機構を用いた CPI 予測モデルが提案された。TransformerCPI は自然言語処理の分野で提案された Transformer[12] を改良したタンパク質配列情報と化合物構造情報を用いた CPI 予測モデルで、訓練データとテストデータで化合物のラベルが反転されたデータセットを用いて実験を行い、その結果既存モデルよりも非常に高い予測精度を示した。

1.3 クラス A GPCR に特化した CPI 予測

CPI 予測の研究ではタンパク質配列と化合物構造情報を用いた機械学習モデルが提案されてきた。既存の機械学習モデルで使用するタンパク質配列は、一般的にはタンパク質の全配列を用いている。しかしながら、それぞれのタンパク質において化合物との結合に重要な場所はタンパク質配列の中の一部であり、タンパク質の末端等に存在する天然変性領域などは結合に関与しない。そのため既存の CPI 予測モデルは、タンパク質配列のエンコーディングにおいて結合に不必要な情報まで特徴量の抽出を行っていると考えられる。GPCR は化合物との結合領域が 7TM のヘリックス部分やサブユニットに存在することが知られており、その中でもクラス A GPCR は結合領域が 7TM のヘリックス部分のみである。そのため、クラス A GPCR をターゲットとした場合、CPI 予測モデルの入力に使用するタンパク質配列を 7TM のヘリックス部分のみにすることができ、より化合物との結合に重要な特徴量をエンコーディングできると考えられる。クラス A GPCR はロドプシンやアドレナリン受容体、嗅覚受容体を含むヒトで最大の GPCR サブファミリーである。その中にはオーファン受容体も多数あり、クラス A GPCR に特化した CPI 予測モデルを構築することは、今後の新規リガンドの発見や機能解明に非常に有益である。

1.4 本研究の目的

既存の CPI 予測モデルが抱える、タンパク質の全配列を用いることで化合物との結合に不要な部分までエンコー

ディングされるという問題に対して、本研究ではターゲットをクラス A GPCR に絞り、7TM のヘリックス部分のタンパク質配列のみを用いたクラス A GPCR 専用エンコーディングモデル、Helix encoder を提案する。データセットは GLASS DB[13] に登録されているクラス A GPCR の化合物-タンパク質相互作用情報を使用した。作成した Helix encoder を TransformerCPI のエンコーダ部分と置き換えることで、クラス A GPCR に特化した CPI 予測モデルを作成し、TransformerCPI との性能の比較を行う。

2. 提案手法

2.1 提案手法の流れ

クラス A GPCR に特化した CPI 予測モデルの構築にあたって、本章ではまずクラス A GPCR データセットの作成に関して説明を行う。次に、CPI 予測モデルとしてベースとなる TransformerCPI の詳細について説明を行い、最後にクラス A GPCR 配列のエンコーディングを行う Helix encoder について説明する。

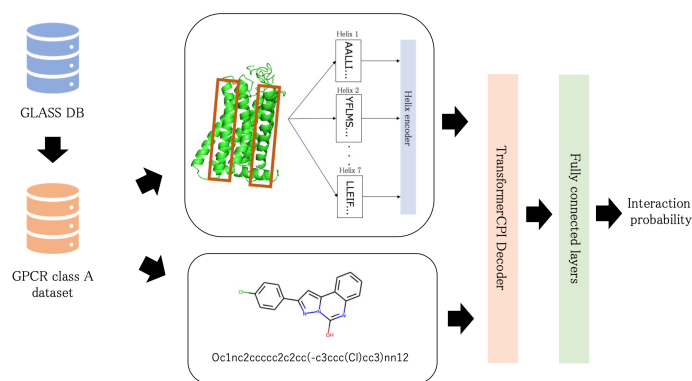


図 2 本研究の流れ

2.2 データセット

クラス A GPCR データセットの作成には GLASS DB を使用した。GLASS DB は実験的に検証された GPCR-リガンド相互作用情報が登録されている。データセットの作成にあたって、まず、GLASS DB に登録されている相互作用のうちクラス A GPCR に関するものを抽出する。GLASS DB には GPCR のクラスについての情報がないことから、抽出には GPCR DB[14] に登録されているクラス A GPCR の UniProt ID[15] に該当するものをクラス A GPCR として使用した。最終的なクラス A GPCR データセットは以下の手順で作成される。

- (1) データベースソースが ChEMBL[16] であるリガンドの相互作用を抽出。
- (2) 結合親和性に関して、実験値が IC50、EC50、Ki で評価されている相互作用を抽出。それぞれの実験値に対

して負の対数を取り、IC50、EC50 は閾値を 4、Ki は閾値を 6 にして 0-1 に分類を行う。

- (3) データセットの質の向上のため。相互作用情報が 1 つしかないリガンドの削除、片方のクラスにしか存在しないリガンドの削除を行う。

完成したクラス A GPCR のデータセットの詳細は表 1 の通りである。

表 1 GPCR クラス A データセットの詳細

タンパク質の数	化合物の数	相互作用の数	正例	負例
424	22,175	101,608	41,265	60,343

2.3 TransformerCPI

TransformerCPI は自然言語処理の分野で提案された Transformer のアーキテクチャをベースにした CPI 予測モデルである。本節では TransformerCPI のアーキテクチャについて説明を行う。

2.3.1 入力

TransformerCPI は word2vec[17] を用いて作成されるタンパク質配列の特徴ベクトルとグラフ畳み込みニューラルネットワークを用いて作成される化合物の特徴ベクトルを入力とする。word2vec は構文的、意味的な単語関係に関する分散ベクトル表現を学習する教師なし手法である。タンパク質配列の特徴ベクトルの作成に使用される word2vec モデルは、UniProt に登録されているヒトタンパク質配列をコーパスとし、30 エポックで事前学習されている。学習は一定長のアミノ酸配列 (3-mers) を単語として分割、配列全体を文書とみなして行われており、最終的にタンパク質配列は事前学習された word2vec モデルを用いて実数値の 100 次元で表現される。化合物の特徴ベクトルの作成には、まず化合物の SMILES を RDKit パッケージを用いて各原子に対して 34 次元のベクトルを生成する。次に式 1 で表されるグラフ畳み込みニューラルネットワークを用いて、近傍の原子特徴を統合した各原子の表現を学習する。

$$\mathbf{H}^{(l+1)} = f(\mathbf{H}^{(l)}, \mathbf{A}) = \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}_3^{(l)}) \quad (1)$$

ここで、 a を化合物の原子数とすると $\mathbf{A}$ は $\mathbf{A} \in R^{a \times a}$ の隣接行列であり、 $\tilde{\mathbf{A}}$ は単位行列 $\mathbf{I}$ を用いて $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ と表現される。 $\tilde{\mathbf{D}}$ は $\tilde{\mathbf{A}}$ の次数行列であり、グラフニューラルネットワークは $\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$ で正規化された隣接行列 $\tilde{\mathbf{A}}$ と l 番目のグラフニューラルネットワーク層の出力 $\mathbf{H}^{(l)} \in R^{a \times f}$ の積に重み $\mathbf{W}_3^{(l)} \in R^{f \times f}$ を掛けたものを非線形活性化関数 $\sigma(\cdot)$ で処理する。ここで f は各原子の特徴ベクトルであり、先行研究では $f = 34$ である。また、グラフニューラルネットワークの層は $l = 1$ とされている。グラフ畳み込みニューラルネットワークで処理された化合物は最終的に a 次元の特徴ベクトルとして表現される。

2.3.2 エンコーダ

word2vec によって作成されたタンパク質配列はエンコーダによってより抽象的な表現に変換される。TransformerCPI のエンコーダは Conv1D と gated linear unit (GLU)[18] を用いた Gated 畳み込みニューラルネットワークを使用する。隠れ層 $h_0; \dots; h_L$ は式 gated で計算される。

$$h_l(\mathbf{X}) = (\mathbf{X} * \mathbf{W}_1 + \mathbf{s}) \otimes \sigma(\mathbf{X} * \mathbf{W}_2 + \mathbf{t}) \quad (2)$$

$\mathbf{X} \in R^{n \times m_1}$ は層 h_1 の入力であり、 $\mathbf{W}_1 \in R^{k \times m_1 \times m_2}$ 、 $\mathbf{s} \in R^{m_2}$ 、 $\mathbf{W}_2 \in R^{k \times m_1 \times m_2}$ 、 $\mathbf{t} \in R^{m_2}$ は学習パラメータである。ここで L は隠れ層の数、 n は配列の長さ、 m_1 、 m_2 は入力次元と隠れ層の出力次元である。また、 k はパッチサイズで $\sigma(\cdot)$ はシグモイド関数である。先行研究では、 $L = 3$ 、 $m_1 = 64$ 、 $m_2 = 128$ 、 $k = 8$ となっている。エンコーダの出力は、タンパク質の長さを b とした時 $p_1, \dots, p_b$ と表現される。

2.3.3 デコーダ

エンコーダによって生成されたタンパク質の特徴ベクトルとグラフニューラルネットワークで生成された化合物の特徴ベクトルは、デコーダによって相互作用特徴ベクトルに変換される。TransformerCPI のデコーダは Transformer のデコーダと同様のアーキテクチャを使用しており、multiheaded self-attention 層と feed forward 層からなる。self-attention 層は、キー $\mathbf{K}$ 、値 $\mathbf{V}$ 、クエリ $\mathbf{Q}$ の 3 つを入力とし、式 3 のように attention を計算する。

$$\text{attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (3)$$

ここで d_k は $\mathbf{K}$ 、 $\mathbf{Q}$ の次元である。デコーダはエンコーダで表現されるタンパク質特徴ベクトルの中で重要な部分に注目し、二つのベクトルから相互作用の特徴を捉えることができる。

2.3.4 出力層

デコーダの出力である相互作用特徴ベクトル $x_1, \dots, x_a$ は式 4 によって計算される重みを用いて、64 次元の相互作用特徴ベクトルを生成する。(式 5)

$$\alpha_i = \frac{e^{x'_i}}{\sum_{i=1}^a e^{x'_i}} \quad (4)$$

$$\mathbf{y}_{\text{interaction}} = \sum_{i=1}^a \alpha_i \mathbf{x}_i \quad (5)$$

ここで x'_i は式 6 で計算される。

$$x'_i = \|x_i\|_2^2 \quad (i = 1, 2, 3, \dots, a) \quad (6)$$

最後に相互作用特徴ベクトル $\mathbf{y}_{\text{interaction}}$ は全結合層によって化合物とタンパク質の結合確率 $\tilde{y}$ として計算される。

2.4 クラス A GPCR に特化した CPI 予測モデル

提案するクラス A GPCR に特化した CPI 予測モデルは

TransformerCPI のデコーダと出力層はそのまま、エンコーダ部分を 7 つのヘリックスのみに注目したエンコーディングモデルに置き換えたものである。エンコーダの入力は、word2vec によって生成された 7 つのヘリックスの特徴ベクトルである。ヘリックス部分の配列は、クラス A GPCR のタンパク質配列から UniProt に登録されている 2 次構造情報を用いて 7 つのヘリックス配列を抽出する。ここでそれぞれのヘリックスを $h_i (i = 1, 2, 3, \dots, 7)$ 、各ヘリックスの長さを b_{h_i} とすると、抽出した 7 つのヘリックス配列は、先行研究で使用された word2vec を用いて $b_{h_i} \times 100$ 次元にエンベディングされる。

2.4.1 Helix encoder

Helix encoder はそれぞれのヘリックスに対して用意された 7 つの TransformerCPI のエンコーダと、それぞれのエンコーダの出力を入力とする全結合層からなる。それぞれのヘリックスの長さは異なるため、入力ヘリックスの最大長 b_{max} をとり $b_{max} \times 100$ 次元に変換されたものとなる。 b_{max} より短いヘリックスに関しては 0 パディングを行う。それぞれのエンコーダでは、TransformerCPI のエンコーダを用いて $b_{max} \times 64$ 次元に圧縮され、出力を連結した $b_{max} \times 448 (= 64 \times 7)$ 次元のベクトルを全結合層で $b_{max} \times 64$ に変換する。Helix encoder の出力は最終的に TransformerCPI のデコーダの入力として扱われる。Helix encoder の全体図は図 3 の通りである。

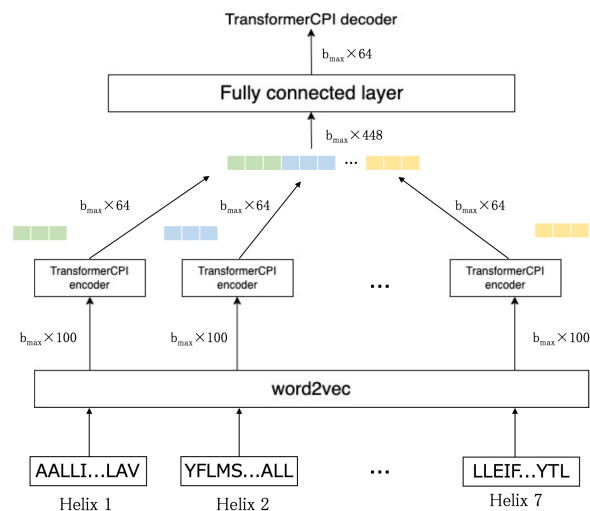


図 3 Helix encoder の概要

2.4.2 学習詳細

最適化手法には先行研究に沿って RAdam[19] による LookAhead[20] 最適化手法を使用した。モデルの学習の際に用いる損失関数は式 7 で表されるバイナリクロスエントロピーを用いた。

$$\text{Loss} = -\{d \log y + (1 - d) \log(1 - y)\} \quad (7)$$

ここで、 y は任意の出力値、 d はそれに対応する正解値で

ある。また、学習に用いたハイパーパラメータ先行研究と同様のものを使用した。

3. 結果

3.1 実験手法

本研究では、作成した Helix encoder の効果を確認するために、TransformerCPI のエンコーダを Helix encoder に置き換えたモデル (Helix encoder) とオリジナルアーキテクチャの TransformerCPI の性能の比較を ROC 曲線 (Receiver Operating Characteristic curve) 下面積の AUC (Area Under the receiver operating characteristic Curve) にて行う。また、CPI 予測にあたってタンパク質配列をヘリックス部分に使用することがどれほどの効果があることを確認するために、オリジナルアーキテクチャの TransformerCPI の入力を 7 つのヘリックス配列を連結したものとする TransformerCPI (helix sequences) を作成し、精度の比較を行う。CPI 予測において学習データにないタンパク質に関して予測できるかといった観点が重要になるため、学習データとテストデータに同じタンパク質が入らないように学習データ : テストデータ = 8 : 2 に分割を行い (表 2)、3 つのモデルに対して学習データで学習を行い、テストデータを用いて AUC による精度比較を行なった。モデルの選定に関しては学習データに対して、学習データ : 検証データ = 8 : 2 にランダムに分割をし、各モデルに対して 100 エポックの学習の中で最も高い検証データの AUC を達成したモデルを最終的なモデルとした。

表 2 学習データとテストデータの詳細

	タンパク質の数	相互作用の数	正例	負例
学習データ	408	82,199	34,156	48,043
テストデータ	16	19,409	7,109	12,300

3.2 実験結果

テストデータにおいて Helix encoder はオリジナルの TransformerCPI より高い AUC を達成した。(図 4) また、オリジナルの TransformerCPI のアーキテクチャで入力をヘリックス部分の配列を用いた場合、最も高い AUC となった。よって、クラス A GPCR においてタンパク質配列をヘリックス部分に限定することは非常に有益であるが、ヘリックスごとに TransformerCPI のエンコーダを用意するよりも、ヘリックス部分の配列全体でエンコーディングの方が重要であることが確認できた。

各モデルのテストデータにおける ROC 曲線は図 5 のようになった。TransformerCPI と helix sequences に関しては曲線特徴に大きな差は見られなかったが、Helix encoder に関しては他のモデルと比べて偽陽性率が低い段階で感度が高い結果となった。そのため、生物学的実験の前段階で

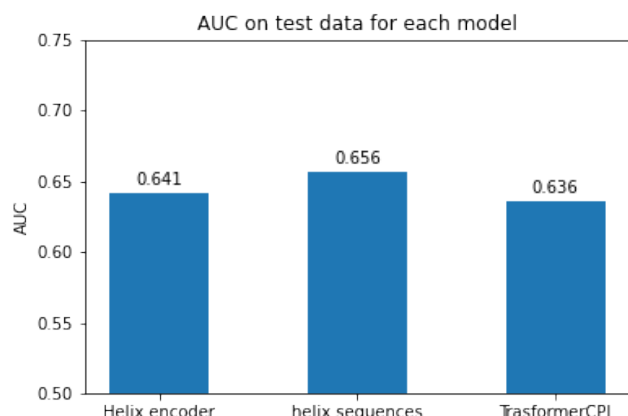


図 4 テストデータにおける各モデルの AUC

ヒット化合物候補を多く見つけたい場合は、Helix encoder を用いることが適していると考えられる。

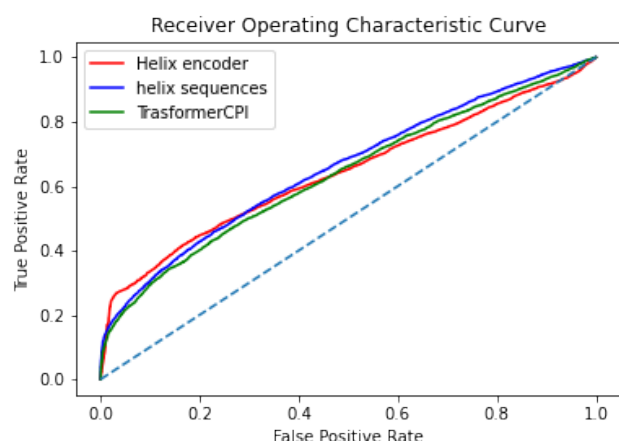


図 5 テストデータにおける各モデルの ROC 曲線

4. 考察

実験で用いた各モデルに関して、テストデータに含まれる 16 個のタンパク質毎の AUC は表 3 のようになった。TrasformerCPI で元々 AUC が 0.5 を切っていた 7 つのタンパク質のうち、Q9Y5N1 以外の 6 つのタンパク質が Helix encoder や helix sequences で 0.5 以上の AUC を示しており、一方で P29274 (Adenosine receptor A2a) や P30542 (Adenosine receptor A1)、P08588 (Beta-1 adrenergic receptor) といったタンパク質では TrasformerCPI の方が高い AUC を達成した。これは P29274 のような細胞外ループが化合物との結合に関係するタンパク質 [21] では使用する配列をヘリックス部分に絞らない方がいいと考えられるが、Helix encoder や helix sequences の方が AUC が高い P0DMS8 (Adenosine receptor A3) でも細胞外ループが結合に関係していることが報告されており [22]、入力配列だけでなく Helix encoder のアーキテクチャに改善の余地が

あると考えられる。

表 3 テストデータにおける各タンパク質の AUC

uniprot id	Helix encoder	helix sequences	TrasformerCPI
P0DMS8	0.515	0.541	0.467
P08908	0.502	0.601	0.598
P50406	0.446	0.53	0.491
P28223	0.542	0.514	0.495
P07550	0.559	0.526	0.468
P61169	0.524	0.529	0.527
P29274	0.494	0.629	0.69
P21453	0.523	0.843	0.792
P30542	0.641	0.625	0.648
Q9Y5N1	0.497	0.473	0.494
P08588	0.482	0.542	0.557
P21917	0.55	0.554	0.448
Q99500	0.825	0.74	0.68
P33533	0.495	0.571	0.483
P19327	0.599	0.469	0.547
P13945	0.637	0.606	0.585

また、各タンパク質におけるヘリックスの長さとの関係 (図 6) を見ると、Helix encoder を用いることで AUC が向上したタンパク質や、ヘリックス部分のみを入力としたことで AUC が向上したタンパク質のヘリックス配列の長さには明確な特徴はなかった。性能が向上した要因として、配列をヘリックス部分に絞ることでテストデータに保存されているタンパク質との配列相溶性がより高くなったものが学習データに存在するようになった可能性が考えられる。

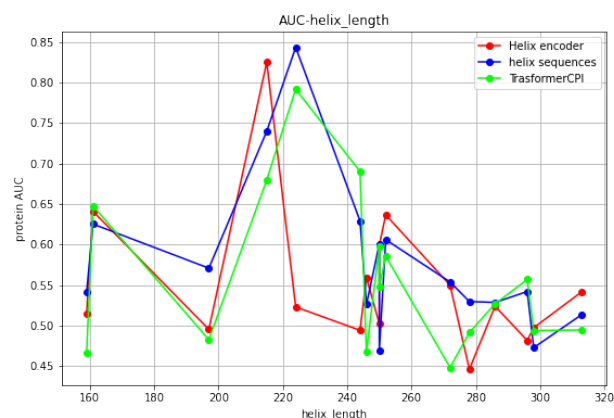


図 6 各タンパク質の AUC とヘリックス配列の長さの関係

5. 結論

本研究は、タンパク質の配列のうち結合に重要と考えられる部分を用いて CPI 予測を行うことで精度の向上を試みた。結果として、結合に重要な部分が明確であるクラス A

GPCRにおいて、提案手法は全てのタンパク質配列を用いた場合より高いAUCを達成した。それぞれのヘリックスに対してエンコーダを用意するヘリックスのエンコーディング手法は先行研究よりも高い性能を示したが、7つのヘリックス配列を連結したものを先行研究のエンコーダの入力としたモデルの方が高いAUCを示した。そのため、ヘリックスのエンコーディング手法にはまだ改善の余地があると考えられる。また、タンパク質配列のうち、あらかじめ化合物との結合に重要な場所を抽出することでCPI予測の改善が見られることは、クラスA GPCRに限らない可能性があるため、今後他のファミリーに対しても効果を確認していく必要がある。

参考文献

- [1] Antonella Di Pizio, Maik Behrens, and Dietmar Krautwurst. Beyond the flavour: The potential drug-gability of chemosensory g protein-coupled receptors. *International Journal of Molecular Sciences*, Vol. 20, No. 6, 2019.
- [2] Krishna Sriram and Paul A. Insel. G protein-coupled receptors as targets for approved drugs: How many targets and how many drugs? *Molecular Pharmacology*, Vol. 93, No. 4, pp. 251–258, 2018.
- [3] Secker A. Halling-Brown M. et al. Davies, M.N. Gpcrtree: online hierarchical classification of gpcr function. *BMC Res Notes*, Vol. 1, p. 67, 2008.
- [4] Simon D Harding, Joanna L Sharman, Elena Faccenda, Chris Southan, Adam J Pawson, Sam Ireland, Alasdair J G Gray, Liam Bruce, Stephen P H Alexander, Stephen Anderton, Clare Bryant, Anthony P Davenport, Christian Doerig, Dorian Fabbro, Francesca Levi-Schaffer, Michael Spedding, Jamie A Davies, and NC-IUPHAR. The IUPHAR/BPS Guide to PHARMACOLOGY in 2018: updates and expansion to encompass the new guide to IMMUNOPHARMACOLOGY. *Nucleic Acids Research*, Vol. 46, No. D1, pp. D1091–D1106, 11 2017.
- [5] Clark D. Czodrowski-P. et al. Vamathevan, J. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*, Vol. 18, p. 463–477, 2019.
- [6] Jacoby E. Bredel, M. Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat Rev Genet*, Vol. 5, pp. 262–275, 2004.
- [7] Hakime Öztürk, Arzucan Özgür, and Elif Ozkirimli. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics*, Vol. 34, No. 17, pp. i821–i829, 09 2018.
- [8] Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, Vol. 20, No. 1, pp. 61–80, 2009.
- [9] Masashi Tsubaki, Kentaro Tomii, and Jun Sese. Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences. *Bioinformatics*, Vol. 35, No. 2, pp. 309–318, 07 2018.
- [10] Thin Nguyen, Hang Le, and Svetha Venkatesh. Graphdta: prediction of drug–target binding affinity using graph convolutional networks. *bioRxiv*, 2019.
- [11] Lifan Chen, Xiaoqin Tan, Dingyan Wang, Feisheng Zhong, Xiaohong Liu, Tianbiao Yang, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and Mingyue Zheng. TransformerCPI: improving compound–protein interaction prediction by sequence-based deep learning with self-attention mechanism and label reversal experiments. *Bioinformatics*, Vol. 36, No. 16, pp. 4406–4414, 05 2020.
- [12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc., 2017.
- [13] Wallace K. B. Chan, Hongjiu Zhang, Jianyi Yang, Jeffrey R. Brender, Junguk Hur, Arzucan Özgür, and Yang Zhang. GLASS: a comprehensive database for experimentally validated GPCR–ligand associations. *Bioinformatics*, Vol. 31, No. 18, pp. 3035–3042, 05 2015.
- [14] Gáspár Pándy-Szekeres, Mauricio Esguerra, Alexander S Hauser, Jimmy Caroli, Christian Munk, Steven Pilger, György M Keserű, Albert J Kooistra, and David E Gloriam. The G protein database, GproteinDb. *Nucleic Acids Research*, Vol. 50, No. D1, pp. D518–D525, 09 2021.
- [15] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, Vol. 49, No. D1, pp. D480–D489, 11 2020.
- [16] Anna Gaulton, Anne Hersey, Michał Nowotka, A Patrícia Bento, Jon Chambers, David Mendez, Prudence Mutowo, Francis Atkinson, Louisa J Bellis, Elena Cibrián-Uhalte, Mark Davies, Nathan Dedman, Anneli Karlsson, María Paula Magariños, John P Overington, George Papadatos, Ines Smit, and Andrew R Leach. The ChEMBL database in 2017. *Nucleic acids research*, Vol. 45, No. D1, January 2017.
- [17] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc., 2013.
- [18] Yann N. Dauphin, Angela Fan, Michael Auli, and David Grangier. Language modeling with gated convolutional networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70 of *Proceedings of Machine Learning Research*, pp. 933–941. PMLR, 06–11 Aug 2017.
- [19] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond, 2021.
- [20] Michael R. Zhang, James Lucas, Geoffrey Hinton, and Jimmy Ba. Lookahead optimizer: k steps forward, 1 step back, 2019.
- [21] Byron Carpenter and Guillaume Lebon. Human adenosine a2a receptor: Molecular mechanism of ligand binding and activation. *Frontiers in Pharmacology*, Vol. 8, , 2017.
- [22] Kerry Barkan, Panagiotis Lagarias, Margarita Stampelou, Dimitrios Stamatis, Sam Hoare, Dewi Safitri, Karl-Norbert Klotz, Eleni Vrontaki, Antonios Kolocouris, and Graham Ladds. Pharmacological characterisation of novel adenosine A3 receptor antagonists. *Sci. Rep.*, Vol. 10, No. 1, p. 20781, November 2020.