

2-way prediction 法による GPCR リガンドの結合予測

百石 弘 澄^{†1} 杉原 稔^{†2} 諏訪 牧子^{†2,†1}
加藤 毅^{†3,†2} 山名 早人^{†1} 藤 渕 航^{†2,†1}

G タンパク質共役型受容体 (GPCR) は創薬研究の中心となっており, GPCR と相互作用する未知のリガンドを予測することは重要な課題である. しかしながら, 相互作用するリガンドが全く知られていない GPCR も多く存在し, これらの GPCR と相互作用するリガンドを予測することは学習用サンプルの不足から困難となる. そこで, 我々はサポートベクターマシンを利用した 2-way prediction 法を提案する. この方法では, リガンド, GPCR, 双方から予測を行い, 相互作用するリガンドの情報と GPCR の情報が全くない場合の予測に対応した. 実験により提案手法の有効性を示す.

Prediction of GPCR ligands by 2-way prediction method

HIROTO HYAKKOKU,^{†1} MINORU SUGIHARA,^{†2}
MAKIKO SUWA,^{†2,†1} TSUYOSHI KATO,^{†3,†2}
HAYATO YAMANA^{†1} and WATARU FUJIBUCHI^{†2,†1}

G-protein coupled receptors (GPCRs) are important pharmacological targets and to predict unknown interactions between GPCRs and ligands is one of the most interesting topics in the current computational biology. However, ligands of many GPCRs are experimentally not identified yet and it is difficult to predict unknown ligands of GPCRs because of insufficiency of training data set. We have developed a 2-way prediction method based on the support vector machine. In this method, the prediction is performed by using both information of ligands and GPCRs and one can apply this method to the case where information of the interaction between ligands and GPCRs is completely missing. We discuss the validity of the 2-way prediction method.

1. はじめに

G タンパク質共役型受容体 (GPCR) は細胞膜を貫通するタンパク質であり, 細胞外からのリガンドとの相互作用によって, 細胞外からの信号を細胞内へ伝える役割を持っている. この外部の信号を内部に伝える役割は, 嗅覚, 視覚, 味覚など感覚器でも重要な役割を果たしている. その重要性から創薬研究の中心になることが多く, 近年, GPCR を計算機で同定する研究¹⁾ や, GPCR に相互作用するリガンドをコンピュータ上で予測する研究が盛んに行われている²⁾³⁾. しかし, 数種類しか立体構造が分かっているものがないことや, 相互作用するリガンドがほとんど分かっていない GPCR が多数存在するなど, 予測の役に立つと思われる実験データが非常に少ない. そのため, GPCR に相互作用するリガンドを予測することは困難である.

実際にサポートベクターマシンを用いて, GPCR に相互作用するリガンドを予測する研究が行われている³⁾. リガンドと GPCR の特徴を組合せ, リガンド情報のない GPCR と相互作用するリガンドを予測しているが, その予測は難しく, 予測精度はよい場合で 78% であった.

本研究では, 相互作用する GPCR の情報がないリガンドの結合予測, リガンド情報のない GPCR に対する予測, さらに, リガンドと GPCR 両者についての情報が全て欠けている状態での予測方法である 2-way prediction 法を提案する. 図 1, 図 2, 図 3 にそれぞれの場合の例を示す.

本稿の構成は以下のようになっている. 第 2 節ではデータセットについて述べる. 第 3 節では提案する予測方法と予測に使う特徴を示す. 第 4 節では実験方法, 実験結果, 考察について述べる. 第 5 節で結論をまとめる.

2. データセット

データセットには, 既存研究との比較を容易にするため, Jacob らと同じデータセットを用いた³⁾. Jacob らは GPCR とリガンドのデータベース GLIDA⁴⁾ から, 創薬ターゲット

†1 早稲田大学

Waseda University

†2 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

†3 お茶の水女子大学

Ochanomizu University

	G1	G2	G3	G4	G5	G6	G7	...	Gn
L1			○		○				
L2	○				○		○		
L3	○	○		○					○
L4				○			○		
L5	○	○			○	○			
L6									
L7			○						
L8			○				○		
L9					○	○			
⋮									
⋮									
Lm			○	○					

図 1 相互作用する GPCR が分かっていないリガンド結合予測の例。○は正の相互作用を表す。○記号のない部分は相互作用未知である。網掛け部分であるリガンド L6 の行は、相互作用既知の GPCR が存在しない。

Fig.1 An example of predicting a ligand with unknown GPCRs. The ○ and blank represent positive and unknown interactions, respectively. The information in shaded area (the L6 row) is absent.

になり得る分子量 450 未満程度のリガンドを選び、選ばれたリガンドのいずれかと相互作用する GPCR のみを抽出した。表 1 に利用したデータの概要を示す。

データベース GLIDA には、相互作用する GPCR とリガンドの組合せは登録されているが、相互作用しない組合せは登録されていない。例として図 3 のように○のみとなっている。ここで、○は正の反応を表しており、記号の無い部分は相互作用未知であることを表している。相互作用しない負のデータについても、Jacob らと同様に以下の方法で選択する。ある GPCR に対して、リガンドの特徴を用いて相互作用するリガンドを予測する実験をする際には次のように選択する。あるリガンドがある GPCR と相互作用するが、他の GPCR とは相互作用するという情報が GLIDA に含まれていない場合、そのリガンドはその GPCR とは相互作用しないと仮定する。ただし、この手順で正例と同数だけ負例を選択する。負例を選択した結果は例として図 4 のようになる。×は負の反応を表す。また、あるリガンドに対して、GPCR の特徴を用いて相互作用する GPCR を予測する実験をする際には、相互作用しない GPCR の負例を正例と同数だけ同様に選択する。

	G1	G2	G3	G4	G5	G6	G7	...	Gn
L1			○						
L2	○						○		
L3	○	○		○					○
L4				○			○		
L5	○	○				○			
L6			○			○	○		
L7			○						
L8			○				○		
L9						○			
⋮									
⋮									
Lm			○	○					

図 2 相互作用するリガンドが分かっていない GPCR 結合予測の例。網掛け部分である GPCR G5 の列は、相互作用既知のリガンドが存在しない。

Fig.2 An example of predicting a GPCR with unknown ligands. The information in shaded area (the G5 column) is absent.

3. 手 法

今回、我々は以下の 3 つの方法によって GPCR とリガンドの相互作用の予測を行った。

- (1) 各 GPCR について、相互作用既知のリガンドの特徴を用いて、未知のリガンドを予測する方法。
- (2) 各リガンドについて、相互作用既知の GPCR の特徴を用いて、未知の GPCR を予測する方法。
- (3) 相互作用する GPCR、リガンド、の情報がない場合に、方法 (1) と (2) を組み合わせてリガンドと GPCR の結合を予測する 2-way prediction 法。

この節では、方法 (1) で利用するリガンドの特徴の取り方、方法 (2) で利用する GPCR

表 1 使用データの概要
Table 1 Summary of the data set.

リガンド数	GPCR 数	相互作用の組合せ数	GPCR あたりの相互作用の数
2445	80	4051	50.6

の特徴の取り方，方法（3）の 2-way prediction 法について述べる．

3.1 リガンドの特徴

化学構造式で表現されるリガンド物質を，辺と頂点の集合で表現されるグラフで表す．頂点にはそれぞれの原子，辺には結合の多重度に応じた重みを対応づける．グラフで表現された全てのリガンドから頻出部分グラフを抽出する⁵⁾．頻出部分グラフとは，任意の閾値を s としたとき，ある部分グラフが s 個以上のリガンドに含まれるグラフのことをいう．この s をサポート値と呼ぶことにする．得られた頻出部分グラフの構造をリガンドが持つならば 1，持たないならば 0，という 1 と 0 のみの値を持つ特徴ベクトルを予測に用いる．

3.2 GPCR の特徴

まず立体構造が分かっている GPCR4 種類 (β_1 アドレナリン受容体⁶⁾， β_2 アドレナリン受容体⁷⁾⁸⁾，アデノシン受容体⁹⁾，ロドプシン¹⁰⁾) について，リガンド結合部位を決定する．リガンド結合部位は，リガンドに最も近接するアミノ酸残基を選び，それらのアミノ酸をリガンド結合部位とした．具体的には β_1 アドレナリン受容体， β_2 アドレナリン受容体，ロドプシンについてはリガンドから 4.5Å 以内にあるヘリックス中のアミノ酸を，アデノシン受容体については 5.0Å 以内にあるヘリックス中のアミノ酸を選んだ．選んだアミノ酸を表 2 に示す．

立体構造既知の各 GPCR と予測に利用する GPCR のアミノ酸配列をアライメントし，立体構造既知 GPCR のリガンド結合部位に対応する部分のアミノ酸のみを抽出し，特徴とする．

各リガンド結合部位を 21 次元のベクトルで表現する．ここで 21 次元としたのは，アライメントの際に例外的にギャップを含む可能性があるため，20 種類のアミノ酸に加えてギャップという情報を特徴に取り入れるためである．そして対応するアミノ酸の部分のみ 1，他は 0 といった特徴ベクトルを生成する．

3.3 2-way prediction 法

2-way prediction 法では，ある GPCR とあるリガンドについて，その GPCR に相互作用するリガンドが全て未知であり，そのリガンドに反応する GPCR も全て未知である，という GPCR とリガンドの組合せが相互作用するかどうかを予測する．ここで，相互作用するリガンドが全て未知の GPCR を orphan GPCR と呼び，相互作用する GPCR が全て未知のリガンドを orphan リガンドと呼ぶことにする．図 4 にこの例を示す．網掛け部分は GPCR (G5) について全てのリガンドが未知の場合と，リガンド (L6) について全ての GPCR が未知の場合を表している．G5 が orphan GPCR であり，L6 が orphan リガンド

である．この網掛け部分が交差する部分が 2-way prediction 法で予測する対象である．この交差する部分を double orphan と呼ぶことにする．

2-way prediction 法には，各 GPCR について orphan リガンドが相互作用するかどうかを予測した後，予測結果と GPCR の特徴から orphan GPCR と orphan リガンドの相互作用を予測する方法と，各リガンドについて orphan GPCR が相互作用するかどうかを予測した後，予測結果とリガンドの特徴から orphan GPCR と orphan リガンドの相互作用を予測する方法がある．前者を 2-way prediction 法 (A) とし，後者を 2-way prediction 法 (B) とする．

2-way prediction 法 (A) では，まず相互作用するリガンドを k 個以上含む GPCR それぞれについて，リガンドの特徴から，orphan リガンドがそれぞれの GPCR に相互作用するかどうかを予測する．その結果，orphan リガンドに相互作用する複数の GPCR の情報が得られる．得られた GPCR の特徴を用いて最終的に予測したい double orphan の相互作用を予測する．このように 2 段階の予測を行う．例えば図 4 の場合には，G1 の GPCR について，L2, L3, L5, L7, L8 などの情報を用いて L6 と G1 が相互作用するかどうかを予測する．この予測を G2, G3 など他の GPCR についても行う．その結果図 5 のように空白であった L6 の行が ○ もしくは × で埋められる．次に L6 の行について，予測された G1, G2, G3 などの GPCR の特徴を用いて G5 の部分を予測する．

逆に 2-way prediction 法 (B) では，先に相互作用する GPCR を k 個以上含むリガンドそれぞれについて，GPCR の特徴から orphan GPCR がそれぞれのリガンドに相互作用するかどうかを予測する．その結果を用いて，リガンドの特徴から最終的に予測したい double orphan の相互作用を予測する．図 4 の場合には，先に G5 の列の L6 以外の空白を予測し，予測された L6 以外の L1, L2, L3 などの情報から L6 を予測する方法である．

4. 実 験

4.1 実験方法

各 GPCR について，リガンドの特徴を用いて各 GPCR に相互作用する未知リガンドの予測をする方法 (1)，各リガンドについて，GPCR の特徴を用いて各リガンドに相互作用する未知 GPCR の予測をする方法 (2)，2-way prediction 法による未知 GPCR と未知リガンドの相互作用の予測する方法 (3)，の 3 通りの実験を行った．

• 方法 (1)

既知の正例と同数の負例を前節のようにランダムに生成する．データが少ないとうまく

表 2 選択したアミノ酸
 Table 2 Selected amino acids.

β_1 アドレナリン受容体	W117	T118	D121	V122	A208	S211	S215	W303	F306	F307	N310	N329	Y333							
β_2 アドレナリン受容体	W109	D113	V114	V117	T118	Y199	A200	S203	S204	S207	W286	F289	F290	N293	Y308	N312	Y316			
アデノシン受容体	L85	F168	E169	M177	N181	W246	L249	H250	N253	M270	Y271	I274								
ロドプシン	Y43	M44	L47	T94	E113	A117	T118	G120	G121	E122	M207	H211	F212	F261	W265	Y268	A269	F293	A295	K296

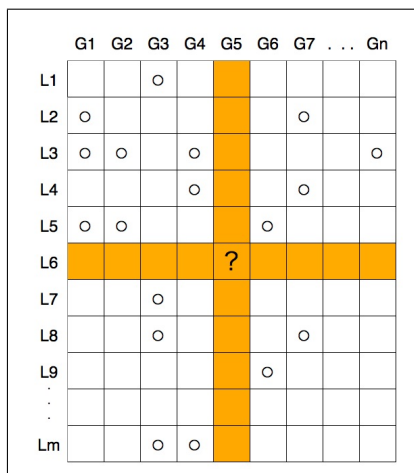


図 3 リガンドと GPCR 両者についての情報が全て欠けている予測の例。また、正例のみのデータセットの例。
 Fig. 3 An example of prediction with neither known ligands nor known GPCRs. And also the example of the data set containing only positive samples.

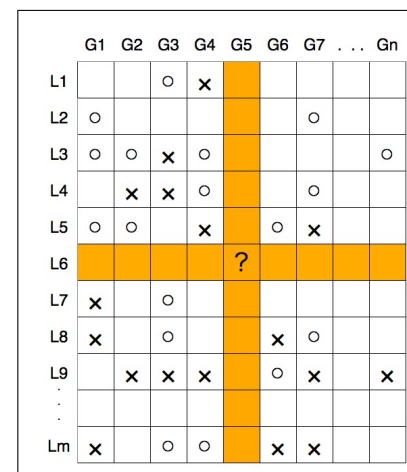


図 4 負例生成済みデータセット。× は負の相互作用を表す。網掛け部分は G5 に対して全てのリガンドとの相互作用が未知であることと、L6 に対して全ての GPCR との相互作用が未知であることを意味している。網掛け部分が交差する部分が double orphan であり、2-way prediction 法の予測する対象である。
 Fig. 4 An example of the data set containing generated negative samples. The × represent negative interactions. The information in shaded area is absent. The crossing position is the double orphan, which is the target of the 2-way prediction method.

	G1	G2	G3	G4	G5	G6	G7	...	Gn
L1			○	×					
L2	○						○		
L3	○	○	×	○					○
L4		×	×	○			○		
L5	○	○		×		○	×		
L6	○	×	○	×	?	○	×		
L7	×		○						
L8	×		○			×	○		
L9		×	×	×		○	×		×
⋮									
Lm	×		○	○		×	×		

図 5 2-way prediction 法による予測. G1, G2, G3, などについて, L6 以外のリガンドの特徴を用いて学習し, L6 の行を予測値で埋めた例. L6 の行の予測値を用いて, 最終的に?の部分予測する.

Fig.5 Prediction by the 2-way prediction method. The interaction of GPCRs (except G5) with L6 is predicted with the training data set except L6. By using the predicted interaction of L6 with GPCRs, the interaction between G5 and L6 is determined.

学習できないため, 今回は正例が 5 以上ある 62 個の GPCR に対して予測実験を行った. 予測にはサポートベクターマシン (SVM) を利用した. SVM を利用する際には, 種々のカーネルの使用が考えられるが, 今回は線形カーネル, 多項式カーネル, ガウスカーネル, の 3 種類のカーネルと, 各パラメータを変化させ予測精度の比較を行い, 最もよい精度をその GPCR における予測精度とした. 予測精度は以下で定義される Accuracy を使用し, 5-fold のクロスバリデーションによって精度を計算した.

$$\text{Accuracy} = \frac{\text{正負を正しく予測できたサンプル数}}{\text{全てのサンプル数}}$$

• 方法 (2)

各リガンドに対して, 相互作用既知 GPCR の特徴を用いて未知 GPCR との相互作用を予測する. 予測方法は方法 (1) の場合と同様に行い, 正例が 5 以上ある 83 個のリガンドに対して予測実験を行った. SVM の利用や, カーネルや各パラメータの選択方法も方法 (1) と同様にし, 予測精度も同様に計算した.

• 方法 (3)

2-way prediction 法 (A) と 2-way prediction 法 (B) はどちらも手順は似ているので, 2-way prediction 法 (A) の手順を以下に述べる. まず, すでに正例か負例かが分かっているデータに対して, その GPCR 以外で, 安定した予測結果を得るために正例を 10 以上 ($k = 10$) 含んでいる GPCR を選ぶ (48 個). 選ばれた GPCR それぞれに対して, 方法 (1) と同じようにクロスバリデーションを行い, 最も高い精度を得られたカーネルとパラメータを利用して, orphan リガンドの予測を行う. その結果, orphan リガンドに対して, 複数の GPCR に対して相互作用するかどうかという複数のデータが得られる. orphan リガンドの相互作用を予測した後, GPCR の特徴と得られた予測結果を用いて, クロスバリデーションを行い最も高い精度を得られたカーネルとパラメータを用いて, orphan GPCR との相互作用の予測を行う. 2-way prediction 法の予測性能を評価するために, 正の double orphan 50, 負の double orphan 50 を仮定し最終的に 2-way prediction 法による予測精度を計算する.

4.2 結果

• 方法 (1) の実験結果. サンプル数の違いによる予測精度を表 3 に示す. この表 3 は辺 4 から 9, サポート値 61 から 70 の場合である. ここでは 5 から 200 以上の正例を含む GPCR の精度の平均を計算した. 予測精度にあまり変化はなく, 学習リガンド数にあまり依存しないと考えられる. 5 より小さい場合に関しては, リガンドが少なすぎ, 予測精度に大きなばらつきがあったため今回は 5 以上を比較している.

特徴として部分グラフの取り方のサポート値を変化させた性能の変化を図 6 に示す. 棒グラフは各サポート値の範囲に対する予測精度を表しており, 折れ線グラフは各サポート値に含まれる特徴数 (部分グラフ数) である. 部分グラフの総数は非常に多く, 特徴としてすべての部分グラフを利用することは難しいため, サポート値と辺の数に範囲を設定し, その範囲に含まれる部分グラフのみを特徴として利用した. 辺の数は 4 から 9 の範囲を設定し, それぞれについてサポート値を 31 から 100 まで 10 刻みで範囲を設定した. その結果が図 6 である.

サポート値の違いによる予測精度の違いは大きくても 1%前後であり, あまり変化がなかったといえる. サポート値を小さくすると, 部分グラフの数が大きくなりすぎ, 特徴がうまくとれない. 逆にサポート値を大きくすると, ほぼ全てのリガンドに含まれる一般的すぎる部分グラフとなり, 予測の役に立たない. そのため中程度のサポート値を選んでいるが, 中程度のサポートを選んだ場合には, あまり大きな予測精度の差はなかつ

たことがわかる．すなわち，中程度のサポート値を持つ部分グラフのもつ特徴に大きな差がないことになる．

● 方法 (2) の実験結果．

GPCR 数の違いによる予測精度の変化を表 4 に示す．今回のデータセットでは各リガンドに相互作用する GPCR の数が少ないので 5 から 20 以上の GPCR を含むリガンドのみの性能の平均を計算した．あまり違いがみられないため，学習 GPCR 数にはあまり依存しないことがわかる．リガンドの情報を用いたときと同様に，5 より小さい場合には予測精度におおきなばらつきがあったため，5 以上のサンプルを含むリガンドのみを比較した．

GPCR の特徴であるリガンド結合部位の取り方の種類を変え，その違いを表 5 に示す．今回リガンド結合部位と予測される部分として 6 種類のリガンド結合部位を比較した． β_1 アドレナリン受容体 (adre β_1)， β_2 アドレナリン受容体 (adre β_2)，アデノシン受容体 (adeno)，ロドプシン (rho)，のそれぞれのリガンド結合部位に加え，2 種類のアドレナリンレセプタのリガンド結合部位の和集合 (adre)，Jacob ら³⁾ (Jacob) と同じリガンド結合部位を比較した．2 種類のアドレナリンレセプタの和集合，アデノシンレセプタの結合部位の予測精度がよくなっている． β_1 アドレナリンレセプタのみの場合には予測精度が下がっているが，このレセプタがヒトではなく七面鳥のものであり，今回対象としている GPCR がヒトであることを考えると納得できる．

● 方法 (3) の実験結果．

リガンドの特徴の取り方と GPCR の特徴の取り方を変えて実験を行った結果を表 6，表 7 に示す．表 6 は，各 GPCR に関して orphan リガンドの予測を先に行った場合であり，表 7 は，各リガンドに関して orphan GPCR の予測を先に行った場合である．ともに辺の数は 6 から 10 の場合であり，サポート値の違いによる予測精度を示している．各リガンドに関して orphan GPCR の予測を先に行った場合はランダム予測の場合 (50%) とほぼ同等であり，この方法は実用的でないことがわかる．理由として推測されることは，各リガンドについて相互作用する GPCR の数が十分に存在するものが少ないためと考えられる．この数が少ないため，2 段階目に利用するための予測値の数が少なくなる．すると誤って予測してしまった結果に影響されやすくなり，最終的な精度があがらないと考えられる．逆に，各 GPCR に関して orphan リガンドの予測を先に行った場合の予測は，よい場合で 70%を超えて，ある程度予測に成功している．

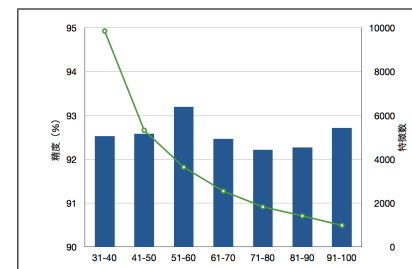


図 6 辺の数の範囲を 4 から 9 に設定したときの，サポート値の変化による予測精度の変化．左の縦軸は予測精度 (%) であり棒グラフに対応している．右の縦軸は特徴数であり折れ線グラフに対応している．横軸は設定したサポート値の範囲である．

Fig. 6 Dependence of the prediction accuracy on the support value. The number of edges is set between 4 and 9. The vertical axes on the right and left sides correspond to the number of features for the line graph and the prediction accuracy for the bar graphs (%), respectively. The horizontal axis corresponds to the support value.

4.3 考 察

リガンド情報を用いて各 GPCR に関して未知リガンドを予測する方法はリガンドが 5 つ以上ある場合には平均で 93%となり予測に成功している．また，リガンドが 5 つ以上ある GPCR は全体の 77.5%であり，今回はサポート値や辺の数によって特徴を選択したが，こ

表 3 GPCR のリガンド数の違いによる予測精度 (%) の変化．
Table 3 Prediction accuracy with different number of GPCRs.

リガンド数	5	10	20	50	100	200
予測精度	93.1	92.4	91.9	91.4	91.4	92.6
リガンド数以上の正例を含む GPCR 数	62	48	41	38	10	4

表 4 GPCR 数の違いによる予測精度 (%) の変化．
Table 4 Prediction accuracy with different number of ligands.

GPCR 数	5	10	15	20	30
予測精度	97.0	94.0	94.4	95.7	96.6
GPCR 数以上の正例を含むリガンド数	83	13	6	3	1

表 5 相互作用既知の GPCR が 5 以上存在するリガンドについて，リガンド結合部位の違いによる予測精度の変化．
Table 5 Prediction accuracy with different ligand binding sites.

結合部位	adre	adre β_1	adre β_2	adeno	rho	Jacob
予測精度	94.4	90.5	93.1	94.4	93.3	92.1

の選択の違いによって大きな予測精度の変化はみられなかった。

GPCR の情報を用いて各リガンドに関して未知 GPCR を予測する方法は、GPCR が 5 つ以上ある場合には平均で約 97% となり予測に成功している。しかし、GPCR が 5 つ未満のリガンドは多く、全体の 96.5% である。GPCR の情報が豊富に存在する場合には非常に有効な方法であることが分かったが、実際には GPCR の情報が少ない場合が多く、この方法を適用できない場合もある。十分にデータが存在し、予測が可能な場合には、リガンド結合部位の選び方によって予測精度が変わることが分かった。このことから、リガンド結合部位は結合するリガンドに係り、リガンド結合部位の定義がリガンド予測に重要であることを示唆している。

また、2-way prediction 法では、リガンド、GPCR、両方の情報が無い場合に、最大で精度が 72% となり、全く情報の無い相互作用であっても、予測可能であることが分かる。

Jacob ら³⁾ も GPCR とリガンドの相互作用の予測を SVM によって行っており、この研究では彼らと同じデータセットを使ったので、性能の比較を行った。今回使った 3 つの方法について比較した。その比較をまとめた表を表 8 に示した。

Jacob らはリガンド予測の際に他の GPCR からの情報も使っているが、我々は他の GPCR からの情報は全く使っていない。それに関わらず、ほぼ同等の性能を発揮することができ、我々の特徴の取り方が有効であることが言える。また、我々は一般的なカーネルのみを利用していることから、複雑なカーネルなどの設計もそれほど意味がないことも分かる。orphan GPCR の予測については、我々は他の GPCR に相互作用するリガンドの情報

が豊富にある場合には予測に成功しているが、ない場合には予測自体が不可能である。よってこのようなリガンドに相互作用する GPCR の情報が少ないデータセットでは不利になってしまう。他の GPCR に相互作用するリガンドがある場合において、orphan GPCR の予測は、Jacob らの方法に比べて我々の方法は非常に有効であると言える。リガンド情報がない場合にも我々は 2-way prediction によって予測を可能にしている。Jacob らはこの最も難しい問題についての結果を出していないため直接比較することはできなかった。

擬似的に比較をするため以下の方法をとった。Jacob らと同様に orphan GPCR を予測する場合に、double orphan となる相互作用は 1723 あり、全体の相互作用数に対する割合は 70.4% である。すなわち、70.4% は 2-way prediction 法の予測の対象となる。残りの 29.6% の相互作用については、Jacob らは予測するリガンドと同じリガンドを学習に利用したことになる。orphan GPCR に相互作用するリガンドを予測する場合に、他の GPCR に相互作用するリガンドが分かっている場合、我々は方法 (2) によって予測する。2-way prediction の予測精度を 72%、方法 (2) の予測精度を 97% として、予測精度の重み付き平均を求めると、 $72 \times 0.704 + 97 \times 0.296 = 79.6\%$ となり、Jacob らとわずかではあるがよい精度が期待できる。

5. おわりに

方法 (1)、方法 (2) によって、学習サンプルが豊富に存在する場合には非常に高い精度で予測できることを示した。情報の少ない GPCR の問題に対応するために、方法 (3) である 2-way prediction 法を開発し、全くの情報の無い GPCR とリガンドに対する予測が可能であることを示した。GPCR リガンドの予測の問題は情報が少なく非常に難しく、まだ改善すべき部分は多くある。

謝辞 本研究に対しての議論、ご意見を頂きました、情報数理研究所 小野幸輝さんに心より感謝いたします。

本研究は一部科研費 (20300104) の助成を受けたものである。

表 6 orphan リガンド予測を先に行った場合の 2-way prediction 法による予測の結果 (%)。

Table 6 Prediction accuracy with 2-way prediction. In this prediction, orphan ligands are predicted first.

サポート値	adre	adre β_1	adre β_2	adeno	rho	Jacob
61-70	69	66	70	62	59	68
81-90	69	71	72	65	66	61

表 7 orphan GPCR 予測を先に行った場合の 2-way prediction 法による予測の結果 (%)

Table 7 Prediction accuracy with 2-way prediction. In this prediction, orphan GPCRs are predicted first.

サポート値	adre	adre β_1	adre β_2	adeno	rho	Jacob
61-70	47	55	42	44	51	48
81-90	40	50	54	45	47	48

表 8 既存研究との比較

Table 8 The comparison with the previous research

	リガンドの予測	GPCR の予測	全ての情報がないときの予測
Ours	93	97	72
Jacob	93	78	

参 考 文 献

- 1) Ono, Y., Fujibuchi, W. and Suwa, M.: Automatic gene collection system for genome-scale overview of G-protein coupled receptors in Eukaryotes, *Gene*, Vol.364, pp.63 – 73 (2005). Beyond the Identification of Transcribed Sequences: Functional, Expression and Evolutionary Analysis.
- 2) Rognan, D.: Chemogenomic approaches to rational drug design, *British Journal of Pharmacology*, Vol.152, No.1, pp.38–52 (2007).
- 3) Jacob, L., Hoffmann, B., Stoven, V. and Vert, J.-P.: Virtual screening of GPCRs: An in silico chemogenomics approach, *BMC Bioinformatics*, Vol.9, No.1, pp.363–378 (2008).
- 4) Okuno, Y., Yang, J., Taneishi, K., Yabuuchi, H. and Tsujimoto, G.: GLIDA: GPCR-ligand database for chemical genomic drug discovery, *Nucleic Acids Research*, Vol.34, pp.D673–D677 (2006).
- 5) Yan, X. and Han, J.: gSpan: Graph-Based Substructure Pattern Mining, *Data Mining, IEEE International Conference on*, Vol.0, p.721 (2002).
- 6) Warne, T., Serrano-Vega, M.J., Baker, J.G., Moukhametzianov, R., Edwards, P.C., Henderson, R., Leslie, A. G.W., Tate, C.G. and Schertler, G. F.X.: Structure of a [bgr]1-adrenergic G-protein-coupled receptor, *Nature*, Vol.454, No.7203, pp.486–491 (2008).
- 7) Rosenbaum, D.M., Cherezov, V., Hanson, M.A., Rasmussen, S. G.F., Thian, F.S., Kobilka, T.S., Choi, H.-J., Yao, X.-J., Weis, W.I., Stevens, R.C. and Kobilka, B.K.: GPCR Engineering Yields High-Resolution Structural Insights into 2-Adrenergic Receptor Function, *Science*, Vol.318, No.5854, pp.1266–1273 (2007).
- 8) Cherezov, V., Rosenbaum, D.M., Hanson, M.A., Rasmussen, S. G.F., Thian, F.S., Kobilka, T.S., Choi, H.-J., Kuhn, P., Weis, W.I., Kobilka, B.K. and Stevens, R.C.: High-Resolution Crystal Structure of an Engineered Human 2-Adrenergic G Protein Coupled Receptor, *Science*, Vol.318, No.5854, pp.1258–1265 (2007).
- 9) Jaakola, V.-P., Griffith, M.T., Hanson, M.A., Cherezov, V., Chien, E. Y.T., Lane, J.R., IJzerman, A.P. and Stevens, R.C.: The 2.6 Angstrom Crystal Structure of a Human A2A Adenosine Receptor Bound to an Antagonist, *Science*, Vol.322, No.5905, pp.1211–1217 (2008).
- 10) Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Trong, I.L., Teller, D.C., Okada, T., Stenkamp, R.E., Yamamoto, M. and Miyano, M.: Crystal Structure of Rhodopsin: A G Protein-Coupled Receptor, *Science*, Vol.289, No.5480, pp.739–745 (2000).