

機械学習を用いた 環状ペプチドの膜透過性予測手法の開発

山田 雄太¹ 吉川 寧^{2,3} 和久井 直樹^{2,3} 大上 雅史^{2,3} 秋山 泰^{1,2,3,a)}

概要: 薬剤が腸管から吸収され、また細胞内の標的へと到達するためには、膜透過性が高いことが必要とされる。近年、環状ペプチドを用いた薬剤設計が新たに注目されているが、適切な膜透過性を持たせるように設計することは低分子と比べて難しい。本研究は、環状ペプチド医薬品の膜透過性を初期に予測できる技術の開発を目的とし、機械学習を用いた環状ペプチドの膜透過性予測手法を提案した。提案手法では、環状ペプチドを構成する残基に着目した膜透過性予測を行うため、*a.* 環状ペプチド全体、*b.* 環状ペプチドを構成する各残基、から特徴量を計算する2つの方法を検討した。また、得られた特徴量に対して訓練データによる特徴選択を行い、検証データに対して膜透過性予測を行った。結果、残基毎のモデル *b.* は全体モデル *a.* と比較して高い実験値との相関が得られる予測モデルを生成することに成功し、環状ペプチドの膜透過性予測において部分構造に着目するアプローチの有用性を示した。

キーワード: 環状ペプチド, 細胞膜透過性, 機械学習, 特徴選択

Development of membrane permeability prediction method for cyclic peptides with machine learning

YUTA YAMADA¹ YASUSHI YOSHIKAWA^{2,3} NAOKI WAKUI^{2,3} MASAHIITO OHUE^{2,3}
YUTAKA AKIYAMA^{1,2,3,a)}

Abstract: For a drug to be absorbed from the intestinal tract and to reach the intracellular target, higher cell membrane permeability is required. In recent years, cyclic peptides have attracted attention as new modality, but designing them to have appropriate membrane permeability is difficult as compared with small molecules. This study proposes a method for predicting membrane permeability of cyclic peptides using machine learning with the purpose of developing technology that can predict membrane permeability of cyclic peptide drugs. In this study, to predict membrane permeability focusing on residues constituting cyclic peptides, two methods for calculating features were considered based on; *a.* whole structure of a cyclic peptide, and *b.* each residue of a cyclic peptide, respectively. Then, feature selection with training data was performed on the obtained features. Membrane permeability prediction was performed on test data. As the results, the model *b.* showed better performance in generating a predictive model that correlates with the experimental value compared with the whole model *a.* We showed that the approach focusing on the local structures is useful in the membrane permeability prediction of the cyclic peptides.

Keywords: cyclic peptide, membrane permeability, machine learning, feature selection

¹ 東京工業大学 工学部 情報工学科
Department of Computer Science, Faculty of Engineering,
Tokyo Institute of Technology
² 東京工業大学 情報理工学院 情報工学系
Department of Computer Science, School of Computing,
Tokyo Institute of Technology
³ 東京工業大学 中分子 IT 創薬研究推進体

1. 序論

薬剤が胃や腸管を通じて体内へと吸収され、また細胞内

Middle Molecule IT-based Drug Discovery Laboratory
(MIDL), Tokyo Institute of Technology

^{a)} akiyama@c.titech.ac.jp

の薬物標的に到達するためには、細胞膜を透過することが必要となる。そのため、薬剤が経口投与可能であるか、細胞内標的を狙えるかを判断する上で、細胞膜透過性を知ることが重要である [1]。膜透過性を実験的に測定することで適切でない薬剤候補化合物を候補から除外することが可能となるが、実験コストがかさむため、開発初期に候補化合物の膜透過性を計算機で予測する技術が求められている。

近年注目されているモダリティである環状ペプチドは、アミノ酸鎖が環状構造を形成しているものを指す。すでに医薬品として上市されているシクロスポリンは、11 残基のアミノ酸から構成される天然物由来の環状ペプチドである [2]。シクロスポリンは比較的大きな化合物（中分子化合物）でありながら膜透過性を有することが知られているが、一般に分子量が大きくなってくると膜透過能が低くなる傾向があるため、中分子、環状ペプチド医薬品開発において膜透過能が高い分子の設計技術の確立が課題となっている [3]。

計算機による膜透過性予測の研究は、これまで低分子化合物を中心に、分子動力学法（molecular dynamics, MD）や機械学習による試みがなされてきた。MD による予測では溶液拡散モデルやその改良モデルによる膜透過率の推定が主に行われてきた。Orsi ら [4] は、MD によって 8 つの有機小分子の膜透過率の計算を行い、Sun ら [5] は transition-tempered metadynamics を用いて低分子化合物の膜透過率の計算を行った。また、Lee らの研究 [6] では、Steered MD 法を用いた低分子化合物の膜透過率計算が行われた。しかしながら、MD に基づく推定法は限られた低分子にしか適用されておらず、多数の（数百種以上の）分子に対して適用しようとすると計算時間の面で困難がある。

Leung ら [7] は機械学習によって、分子量 800 以下の化合物の PAMPA アッセイ [8], [9] で得られた膜透過率実験値に対する非線形回帰予測を行った。Rossi ら [10] は低分子と環状ペプチドの 2 つに対して低分子データを用いた重回帰分析による Caco-2 アッセイ [11], [12] 実験値の細胞膜透過性予測を行った。また、橋田 [13] は分子量 1,000 以下の低分子化合物の Caco-2 アッセイで得られた膜透過率実験値に対する GA-PLS 法による QSAR モデリングを行った。これらはいずれも MD を用いた研究に比べて規模の大きいデータを扱っており、複数の分子に対しての適用という面では機械学習の方が計算時間の面で有利である。

しかしながら、規模の大きいデータを扱った環状ペプチドの膜透過性に関する機械学習の研究はこれまでに行われていない。環状ペプチドでは、先に挙げたシクロスポリンも含めて、一般の L 体アミノ酸 20 種以外に、非天然型の側鎖を有するアミノ酸や主鎖 N-メチル化修飾されたアミノ酸、D 体アミノ酸など、数百以上の人工アミノ酸残基を含んでいる [14]。そのため、通常のアミノ酸配列におけるバイオインフォマティクス解析とは異なるアプローチが必

須となる。本研究では、このような特徴を有する環状ペプチドの医薬品開発を加速するため、機械学習による膜透過性予測を行うことを目的とする。機械学習において環状ペプチドの特徴ベクトルの設計はきわめて重要であり、本研究では環状ペプチド全体の 2D 構造のほか、環状ペプチドを構成する各残基に着目した特徴ベクトル設計について検討した。

2. 提案手法

2.1 提案手法の概要

本研究では、環状ペプチドに対して 2D 記述子による特徴ベクトルを生成し、教師あり学習手法による予測モデル構築を行った。以下に、具体的な提案手法の手順を記すとともに図 1 にフローを示す。

提案手法

手順 1 訓練データセットから統合計算化学システムの MOE[15] を用いて、*a.* 環状ペプチド全体、および *b.* 環状ペプチドを分割した各残基（図 2）から、2D 記述子を計算して特徴ベクトルとする。*b.* では残基数分の記述子の値が得られるため、それらの平均値を用いる（図 3）。206 個の記述子による 206 次元の特徴ベクトルが生成される。それぞれの特徴次元は訓練データを用いて標準化を行い、また訓練データの平均と分散を用いて検証データにおける標準化を行う。

手順 2 *a.* と *b.* それぞれの特徴次元から、訓練データおよび検証データ内にて分散が 0 となるものを除去する。また、訓練データを用いて検証データを標準化した際に -3 以下または $+3$ 以上となる値が現れる特徴次元は外れ値として除去する。

手順 3 残った特徴次元に対して Random Forest[16] を用いて重要度を算出し、特徴選択を行う。

手順 4 特徴選択によって選ばれた特徴次元による特徴ベクトルを用いて、Support Vector Regression (SVR) [17] による非線形回帰予測を行う。

2.2 残基の分割および変換

環状ペプチドから残基毎に分割して特徴量を計算する際、ペプチド結合からカルボキシ基とアミド基が生成されることで特徴量の値が大きく変化する。この特徴量の大幅な変化を防ぐため、本研究ではアミド基を N-メチル基へ、カルボキシ基をアルデヒド基へと変換した上で、特徴量計算に用いた。（図 2）

2.3 特徴選択

回帰モデル構築において、モデルの解釈性や汎化性能を

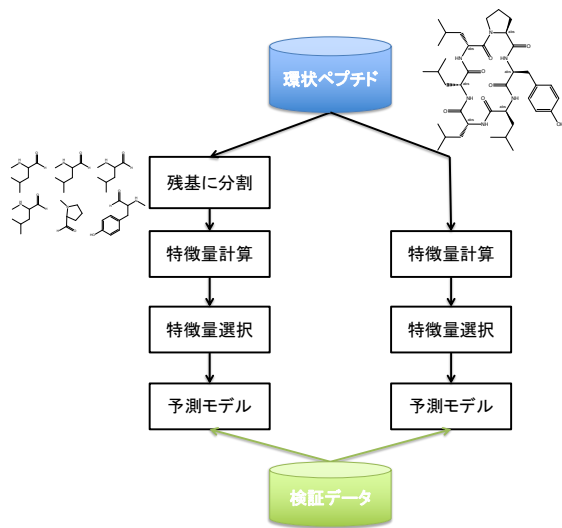


図 1 予測モデル構築のフロー

Fig. 1 The flow of prediction model construction

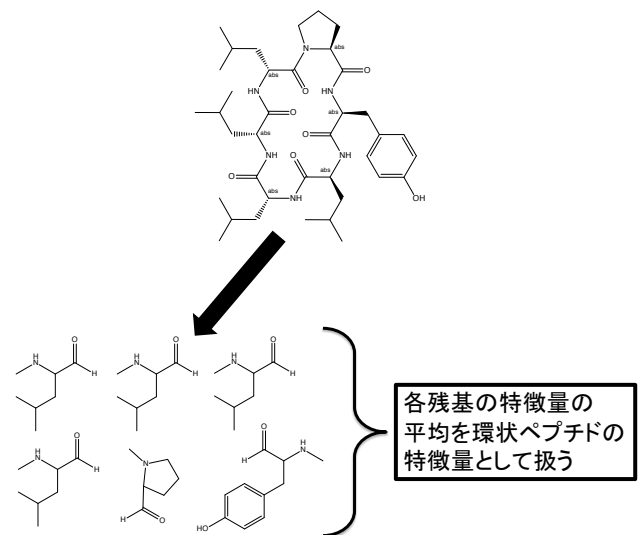


図 3 残基毎の特徴量の計算

Fig. 3 Calculating feature value of each residue

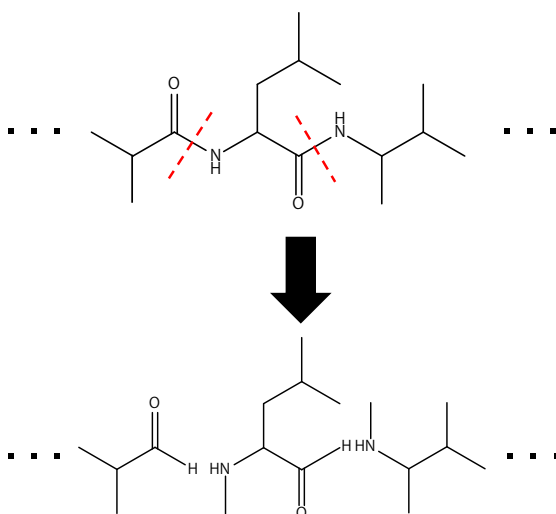


図 2 環状ペプチドからの残基の分割方法

Fig. 2 The method of separating residue from cyclic peptide

高めるために特徴選択を行うことが重要である。本研究では Random Forest[16] の重要度 (feature importances) を用いて特徴次元に順位を割り当て、訓練データ内における最適な特徴量の次元数を求めた。以下に特徴選択の手順を示す。

特徴選択の手順

- 手順 1 特徴量に対して Random Forest を用いて 5-fold Cross Validation (CV) を行い、それぞれの特徴量の重要度 (feature importances) を算出し特徴量に順位を割り当てる。
- 手順 2 訓練データ内にて、特徴量を 5 次元刻みで増やしながら SVR で 5-fold CV を行い、予測結果の R^2 が最大値を示す特徴量の次元数を求める。
- 手順 3 求めた特徴量の次元数の ± 4 次元の範囲で R^2 の最大値を示す次元数がないかを調べるために、同様の 5-fold CV を行う。
- 手順 4 最終的に最も高い R^2 を示したときの特徴量セットを用いて特徴ベクトルとする。

Random Forest のパラメータチューニングは、パラメータである決定木の数 (n estimators) と決定木に用いる特徴次元の数 (max features) を表 1 に示す範囲とし、訓練データに対して 5-fold CV における R^2 値が最適となるパラメータセットをグリッドサーチで求めた。特徴次元の重要度の値は決定木の作成方法により変動するため、特徴次元の重要度に信頼性を持たせる目的で、特徴次元の順位付けには 5-fold CV で求めた 5 回分の重要度の総和を用いた。また、RBF カーネルを用いた SVR のパラメータチューニング (ソフトマージンパラメータ C および RBF カーネルパラメータ γ) は、評価指標を決定係数 R^2 とした 5-fold CV のグリッドサーチによって最適なパラメータを求めた。探索したパラメータの範囲を表 2 に示す。

表 1 パラメータの範囲 (Random Forest)
Table 1 Range of parameters (Random Forest)

n estimators	10, 10^2 , 10^3 , 10^4
max features	auto (全特徴量を用いる), sqrt (全特徴量数の平方根値を用いる), log2 (全特徴量数の \log_2 対数値を用いる)

表 2 パラメータの範囲 (SVR)
Table 2 Range of parameters (SVR)

C	2^{-3} , 2^{-2} , 2^{-1} , ..., 2^9 , 2^{10}
γ	2^{-10} , 2^{-9} , 2^{-8} , ..., 2^{-1} , 1

3. 実験方法

3.1 使用データセット

本研究では、膜透過アッセイの実験値が公開されている Furukawa ら [18] の研究と Wang ら [19] の研究で用いられた環状ペプチドデータセットを用いた。これらのデータセット中の環状ペプチドは人工アミノ酸を含む。データセットには、環状ペプチドの構造式および実験的に計測された膜透過係数値 $\log P_{app}$ が記載されている。以下にそれぞれのデータセットの詳細を示す。

3.1.1 Furukawa データセット

Furukawa ら [18] の論文で公開された環状ペプチドのデータセットである。6 残基の環状ペプチド 400 種類が記載されているが、内 7 種類の化合物は膜透過係数の実験値が記載されていなかったため、本研究ではそれらを除いた 393 種類の環状ペプチドを用いた。本研究では、Furukawa データセットを訓練データセットとして用いた。

3.1.2 Wang データセット

Wang ら [19] の論文で公開された環状ペプチドのデータセットである。6 残基の環状ペプチド 62 種類が記載されているが、内 8 種類は溶解度の低さにより膜透過係数の実験値 P_{app} が計測不可能であったため、本研究ではそれらを除いた。また、本研究で扱う特徴量は 2D 記述子であるため、化合物の光学異性体を区別することができない。すなわち、L 体と D 体のアミノ酸からは同一の記述子が計算される。Wang データセットの中には光学異性体を含んでおり、同じ特徴ベクトルで実験値が異なる化合物同士がデータセットに混在することを避けるため、光学異性体に関してはランダムに 1 つを選択しそれ以外は除いた。最終的に、本研究では 36 種類の環状ペプチドを Wang データセットとして用いた。本研究では、Wang データセットを検証データセットとして用いた。Furukawa データセットと揃える目的で、膜透過係数値 P_{app} は常用対数 \log_{10} を取った値である $\log P_{app}$ に変換して用いた。

3.2 実験方法

本研究で開発した予測システムは、scikit-learn ライブラリ [20] を用いて Python で実装した。

3.2.1 非線形回帰予測

選ばれた特徴量を用いて、検証データに対して SVR で非線形回帰予測を行った。パラメータチューニングは、訓練データ内にて 5-fold CV のグリッドサーチを行った。

3.2.2 予測精度の比較指標

特徴選択の結果を用いて検証データを予測した時の $\log P_{app}$ の相関係数 R 、決定係数 R^2 、RMSE を評価指標として算出し、環状ペプチドに対する機械学習を用いた膜透過性の予測モデルの応用が可能であるか考察する。式 (1), (2), (3) にそれぞれの計算方法を示す。 \hat{y}_i はテストデータの目的変数、 y_i はその予測値を指す。 $\bar{\hat{y}}$ と \bar{y} は、それぞれ \hat{y}_i と y_i の平均値を表す。

$$R = \frac{\sum_{i=0}^n (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=0}^n (\hat{y}_i - \bar{\hat{y}})^2} \sqrt{\frac{1}{n} \sum_{i=0}^n (y_i - \bar{y})^2}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=0}^n (\hat{y}_i - y_i)^2}{\sum_{i=0}^n (\hat{y}_i - \bar{\hat{y}})^2} \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=0}^n (\hat{y}_i - y_i)^2}{n}} \quad (3)$$

4. 結果

4.1 環状ペプチド全体的特徴情報を用いた予測

4.1.1 特徴量選択結果 (環状ペプチド全体)

206 次元の特徴量から 2.1 節にて述べた外れ値を除いた結果、85 次元まで削減された。そこから、訓練データ内において特徴量を 5 次元刻みで 5-fold CV の非線形回帰予測を行った結果、80 次元にて決定係数 R^2 が最大を示した (図 4)。しかし、図 4 では、30 次元周辺において正規分

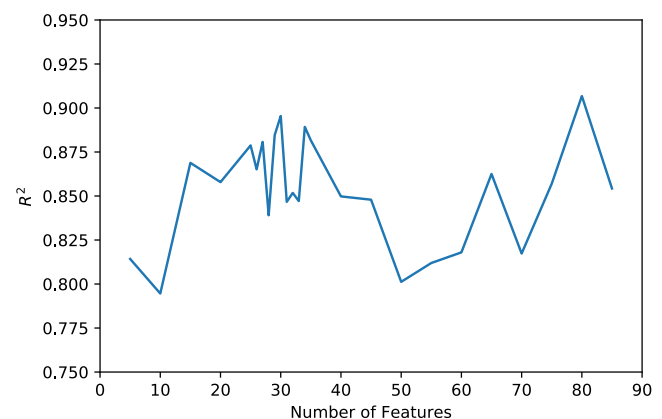


図 4 訓練データ内にて特徴量 85 次元から 5 次元刻みで予測を行い、30 次元の ± 4 の範囲で再び予測を行った時の決定係数 R^2 の最大値の推移 (環状ペプチド全体)

Fig. 4 Changes in the maximum value of the determination coefficient R^2 (whole cyclic peptide) depending on the number of features

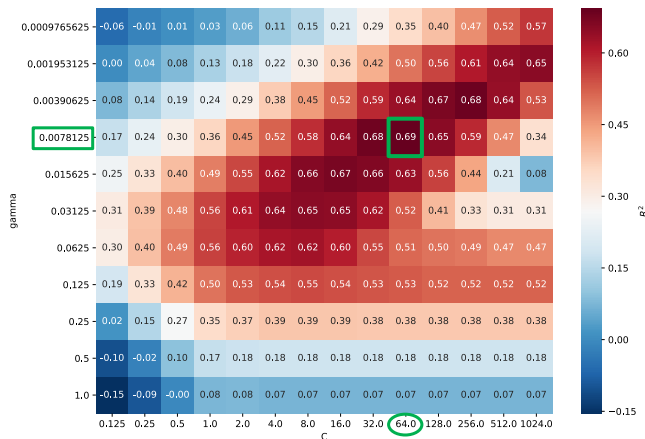


図 5 パラメータチューニングのヒートマップ (環状ペプチド全体)
Fig. 5 The heatmap of parameter tuning (whole cyclic peptide)

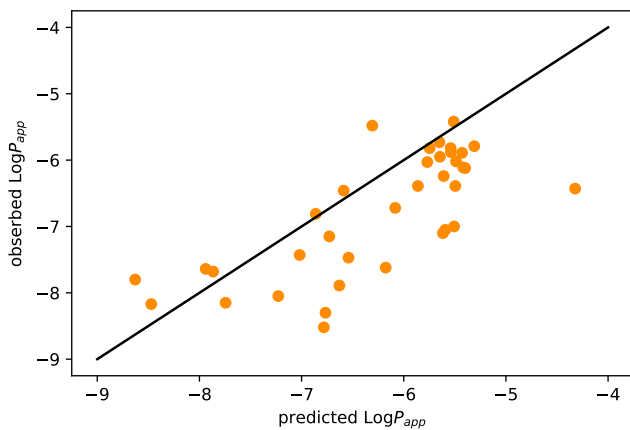


図 6 検証データに対する予測結果 (環状ペプチド全体)
Fig. 6 Prediction results for test data (whole cyclic peptide)

布のように増減の山が確認できたのに対して、80次元の周辺では次元数の変化に対して性能が不安定に上下していた。そのため、本研究では80次元周辺ではなく、30次元周辺のほうが最良であると判断し、30次元を最良な次元数として扱った。30次元から±4次元で同様のCVを行った結果、30次元の時の最大の R^2 を示したため、このモデルの最良な次元数を30次元と定めた。

4.1.2 予測結果 (環状ペプチド全体)

SVRのパラメータチューニングの結果をヒートマップにしたものを図5に示す。図5より、パラメータは $C = 2^6, \gamma = 2^{-7}$ のとき最良の結果を得た。これを用いてSVRで予測を行った結果、 $R = 0.734, R^2 = 0.051, RMSE = 0.877$ の予測精度を得た(図6)。

4.2 残基毎の情報をを用いての予測

4.2.1 特徴量選択結果 (残基毎)

206次元の特徴量から外れ値を除いた結果、64次元まで削減された。そこから、訓練データ内において特徴量を5

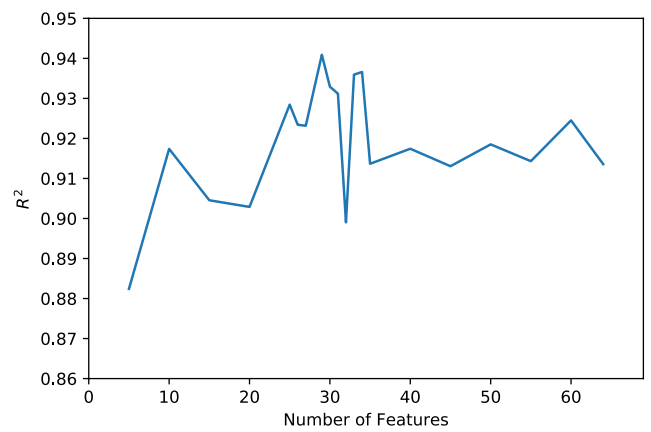


図 7 訓練データ内にて特徴量 85 次元から 5 次元刻みで予測を行い、30次元の±4の範囲で再び予測を行った時の決定係数 R^2 の最大値の推移 (残基毎)

Fig. 7 Changes in the maximum value of the determination coefficient R^2 (each residue) depending on the number of features

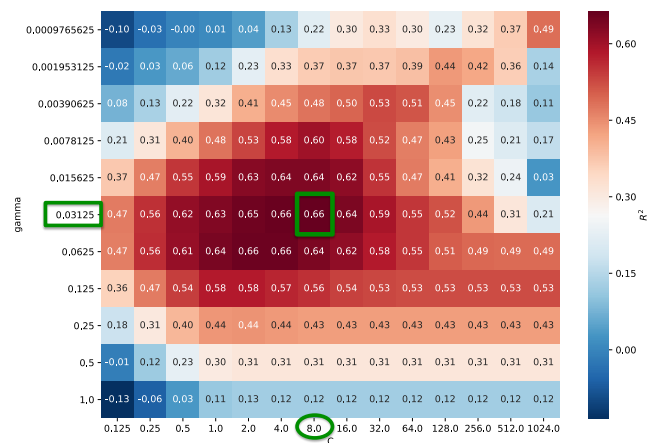


図 8 パラメータチューニングのヒートマップ (残基毎)
Fig. 8 The heatmap of parameter tuning (each residue)

次元刻みで5-fold CVの非線形回帰予測を行った結果、30次元にて決定係数 R^2 が最大を示した。さらに、30次元から±4次元で同様のCVを行った結果、29次元の時の最大の R^2 を示したため、このモデルの最良な次元数を29次元と定めた。図7には、次元数を変えてCVを行った際の各次元数における最大の R^2 をプロットしたものを示す。

4.2.2 予測結果 (残基毎)

RBFカーネルを用いたSVRのパラメータチューニングの結果をヒートマップにまとめたものを図8に示す。パラメータは $C = 2^3, \gamma = 2^{-5}$ のとき最良の結果を得た。これを用いてSVRで予測を行った結果、 $R = 0.868, R^2 = 0.230, RMSE = 0.784$ の予測精度を得た(図9)。

5. 考察

5.1 予測モデル

環状ペプチドを構成する残基の情報をを用いたモデルは、

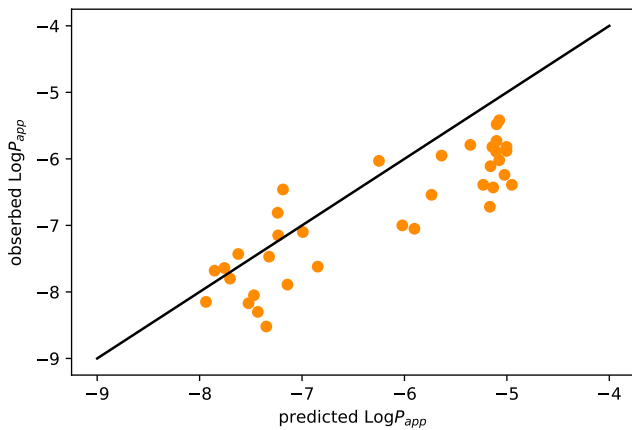


図 9 検証データに対する予測結果 (残基毎)

Fig. 9 Prediction results for test data (each residue)

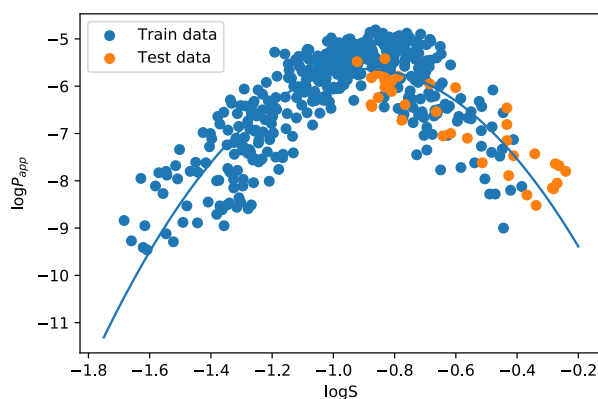


図 10 $\log S$ と $\log P_{app}$ の関係図

Fig. 10 Relationship between $\log S$ and $\log P_{app}$

環状ペプチド全体を用いたモデルと同等もしくはそれ以上に実験値と相関のあるモデルを生成することができた。

残基毎の情報を用いたモデルでは、図 9 にあるように決定係数 $R^2 = 0.2$ と低い値であるものの、相関係数 $R = 0.8$ と強い相関を示す予測結果を得られた。図 10 に本研究で用いたデータの実験値 $\log P_{app}$ と特徴量 $\log S$ の間の関係をプロットしたものを示した。訓練データにおいて上に凸な 2 次曲線に類似した関係が見られたため、図 10 には訓練データを 2 次曲線でフィッティングした曲線も示した。図 10 からわかるように、検証データに対してもプロットした 2 次曲線に類似した分布が確認できた。そのため、残基の特徴情報を用いたモデルでは、訓練データと検証データ間に類似した相関が確認できる特徴量をモデルに含んでいたことで、膜透過性の予測結果もまた相関が見られたと考えられる。

また、2 つのモデルにおいて各特徴量を用いた主成分分析 (principal component analysis, PCA) を行った (図 11, 図 12)。環状ペプチド全体のモデルでは検証データは訓練データの範囲内に分布していた。対して、残基毎のモデルにおいては検証データの多くは訓練データの分布から大

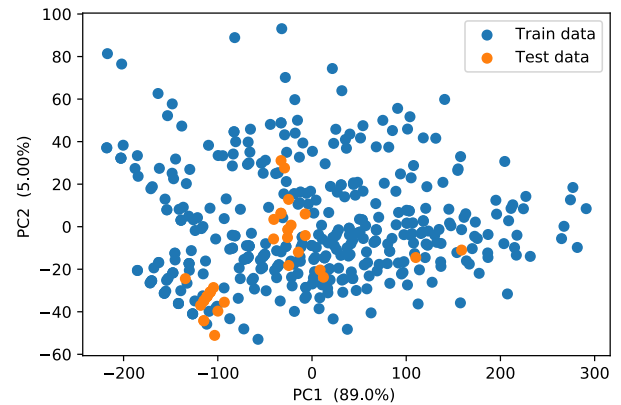


図 11 PCA の結果と寄与率 (環状ペプチド全体)

Fig. 11 PCA analysis of feature space (whole cyclic peptide)

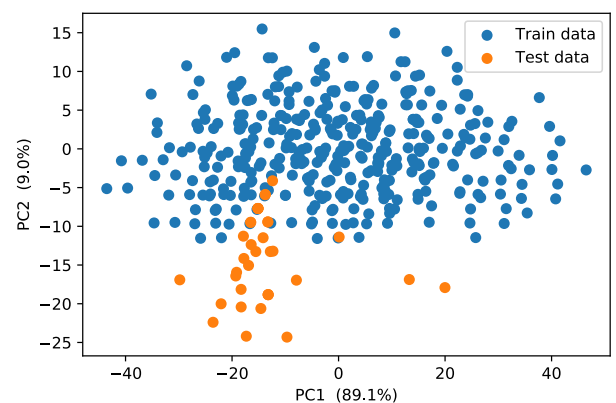


図 12 PCA の結果と寄与率 (残基毎)

Fig. 12 PCA analysis of feature space (each residue)

きく外れていた。そのため、より公平なモデルの評価を行うのであれば、残基毎のモデルにおいて検証データは訓練データ内の範囲内から選択するべきであったと考えられる。今後訓練データが増え、検証データを包含することができれば、さらなる精度の向上が見込めると考えられる。

5.2 特徴量選択

本研究で提案したモデルにおいて、選択された特徴量の上位の一部を表 3 に示す。上位に選ばれた特徴量の多くは物性値に関連し、水溶性や脂溶性、電荷、ファンデルワールス表面積といった特徴量が確認された。これらの特徴量は、過去の膜透過の研究 [10], [21] から関係があるとされてきた特徴量と一致する。ゆえに、本研究で行った Random Forest は解釈性の面でも環状ペプチドの特徴量を順位付けするには妥当な手法だと考える。また、次元数についても 29 次元、30 次元とモデルの構造を単純化し理解しやすくしたという点において Random Forest を用いた特徴量選択は妥当なものであったと考えている。

一方で、図 4, 図 7 からわかるように、一部の次元数では次元数が 1 変化するだけで決定係数 R^2 が極端に上下する。そのため、5-fold CV の R^2 の最大値を選択するだけ

表 3 上位に選ばれた特徴量の一部
Table 3 Most contributing features

Features	説明
$\log S$	水に対する溶解度（水溶性）
$h\text{-}\log P$	オクタノール/水分分配係数（脂溶性）
PEOE_RPC+	正の部分電荷
PEOE_VSA_+1	ファンデルワールス表面積
SMR_VSA	モル屈折率

では最良な次元数と断言することができない。試行回数を増やすことや R^2 の平均値を採用する等の改善を加え、次元数の変化に対して比較的ロバストなモデルを構築した上で次元数を選択することが適切であると考えられる。

5.3 残基を分割する手法

本研究では、残基を取り出す際に特徴量の余計な増加を防ぐために、Nメチル化と水素付加を行った。残基の情報をを用いた予測では環状ペプチド全体の情報をを用いる予測に対して若干の精度向上が確認できた。そのため、環状ペプチドの膜透過性予測において環状ペプチドの残基に着目したアプローチを行った提案手法は精度向上に有効である可能性を示した。また、環状ペプチドの膜透過性は分子内の水素結合と関係があるとされている [10] ため、隣接する残基や向かい合う残基の情報が重要であると思われる。ゆえに、1残基の特徴量に加えて、2-mer, 3-mer といった複数残基の特徴量を組み込むことで、さらなる精度の向上が期待できる。

5.4 提案手法の問題点

本研究では、提案手法において各残基の特徴量の平均を環状ペプチドの特徴量として扱っている。これは、今回のデータセットのように同一残基数の環状ペプチドのみを用いる場合には大きな問題とならないが、異なる残基数の環状ペプチドがデータセット中に含まれる場合、それらを区別することが困難である。今後、7残基、8残基と残基数の異なる環状ペプチドに対しても同様に予測を行うには、環状ペプチド全体の情報も反映できる特徴量（分子量、残基数など）も組み入れる必要がある。

また、提案手法では2D記述子による膜透過性予測モデルの提案を行ったため、光学異性体を区別して扱うことができないモデルとなっている。しかし、膜透過性の研究においてアミノ酸のD体とL体による膜透過率の大きな変化も報告されており [18]、光学異性体を区別した予測は実現すべき課題である。光学異性体を扱うためには、3D記述子の利用や立体配座の検討が必須である。最近の研究では極性表面積 (polar surface area, PSA) が膜透過性の高さに大きく関与すると報告されている [10] が、この PSA の

推定にも立体配座の情報が必要である。配座探索技術の導入や、適切な配座情報から求める3D記述子の検討を進める必要があると考えられる。

6. 結論

本研究では、6残基の環状ペプチドの公開データを用いて、特徴選択と機械学習手法による膜透過性予測を行った。環状ペプチドの各残基に着目した予測モデル構築のため、a. 環状ペプチド全体から計算した2D記述子と、b. 環状ペプチドを分割した各残基より計算された2D記述子の残基平均を用いた特徴ベクトルを比較した。外れ値特徴の除去と、Random Forest の feature importances に基づく特徴選択を訓練データに基づき行い、選択された最良特徴次元数によって検証データに対する予測を行った結果、全体の情報をを用いたモデル a. と比較して残基に着目した特徴量を計算した手法 b. の方が実験値と関連のあるモデルの生成に成功した。

本研究では2D記述子のみを用いて特徴ベクトルの生成を行ったが、本研究では扱えなかった光学異性体への適用も含め、さらなる予測性能の向上のためには環状ペプチドの立体的な情報を取り込む3D記述子の利用が重要であると考えられる。また、分子内水素結合を構成する残基の情報 [10] なども加味するために残基分割を2-mer, 3-mer と増やしたり、異なる残基数の環状ペプチドを含むデータセットへの適用のために環状ペプチド全体の情報も含めることを検討する必要がある。これらは本研究の今後の課題である。

謝辞 本研究の一部は、JSPS 科研費 (17H01814)、JST CREST 「EBD: 次世代の年ヨッタバイト処理に向けたエクストリームビッグデータの基盤技術」(JPMJCR1303)、JST リサーチコンプレックス推進プログラム、文部科学省地域イノベーション・エコシステム形成プログラムの支援を受けて行われた。

参考文献

- [1] Yang NJ, Hinner MJ: Getting across the cell membrane: an overview for small molecules, peptides, and proteins. *Methods in molecular biology*, *Methods Mol Biol*, 1266:29–53, 2015.
- [2] Dreyfuss M, Härril E, Hofmann H, et al.: Cyclosporin A and C, *Eur J Appl Microbiol Biotech*, 3(2):125–133, 1976.
- [3] Buckton LK, McAlpine SR: Improving the Cell Permeability of Polar Cyclic Peptides by Replacing Residues with Alkylated Amino Acids, Asparagines, and d-Amino Acids, *Org Lett*, 20(3):506–509, 2018.
- [4] Orsi M, Sanderson WE, Essex JW: Permeability of small molecules through a lipid bilayer: a multiscale simulation study, *J Phys Chem B*, 113(35): 12019–12029, 2009.
- [5] Sun R, Han Y, Swanson JMJ, et al.: Molecular transport through membranes: Accurate permeability coefficients

- from multidimensional potentials of mean force and local diffusion constants, *J Chem Phys*, 149(7):072310, 2018.
- [6] Lee CT, Comer J, Herndon C, *et al.*: Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds, *J Chem Inf Model*, 56(4):721–733, 2016.
- [7] Leung SS, Sindhikara D, Jacobson MP: Simple Predictive Models of Passive Membrane Permeability Incorporating Size-Dependent Membrane-Water Partition, *J Chem Inf Model*, 56(5):924–929, 2016.
- [8] Kansy M, Senner F, Gubernator K: Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes, *J Med Chem*, 41(7):1007–1010, 1998.
- [9] Avdeef A, Strafford M, Block E: Drug absorption in vitro model: filter-immobilized artificial membranes: 2. Studies of the permeability properties of lactones in Piper methysticum Forst, *Eur J Pharm Sci*, 14(4):271–280, 2001.
- [10] Rossi Sebastiano M, Doak BC, Backlund M, *et al.*: Impact of Dynamically Exposed Polarity on Permeability and Solubility of Chameleonic Drugs Beyond the Rule of 5, *J Med Chem*, 61(9):4189–4202, 2018.
- [11] Artursson P, Karlsson J: Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells, *Biochem Biophys Res Commun*, 175(3):880–885, 1991.
- [12] Yee S: In vitro permeability across Caco-2 cells (colonic) can predict in vivo (small intestinal) absorption in man—fact or myth, *Pharm Res*, 14(6):763–766, 1997.
- [13] 橋田 充: 薬物動態特性の *In Silico* 予測に関する研究, 薬理学雑誌, 125(11):853–861, 2005.
- [14] 舛屋 圭一: 特殊環状ペプチドがもたらす創薬研究開発の新潮流, 日本薬理学雑誌, 148(2):322–328, 2016.
- [15] MOE. Chemical Computing Group: Montreal, 2003.
- [16] Breiman L: Random forests, *Machine Learning*, 45(1):5–32, 2001.
- [17] Thissen U, Pepers M, Üstün B, *et al.*: Comparing support vector machines to PLS for spectral regression applications, *Chemometrics Intelligent Lab Syst*, 73(2):169–179, 2004.
- [18] Furukawa A, Townsend CE, Schwochert J, *et al.*: Passive Membrane Permeability in Cyclic Peptomer Scaffolds Is Robust to Extensive Variation in Side Chain Functionality and Backbone Geometry, *J Med Chem*, 59(20):9503–9512, 2016.
- [19] Wang CK, Northfield SE, Swedberg JE, *et al.*: Exploring experimental and computational markers of cyclic peptides: Charting islands of permeability, *Eur J Med Chem* 97:202–213, 2015.
- [20] Pedregosa F, Varoquaux G, Gramfort A, *et al.*: Scikit-learn: Machine Learning in Python, *J Machine Learning Res*, 12:2825–2830, 2011.
- [21] Ertl P, Rohde B, Selzer P: Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties, *J Med Chem*, 43(20):3714–3717, 2000.