Vol.2013-MPS-92 No.31
2013/2/28

タンパク質の立体構造を利用した遺伝病の解析

汪巍[†1]　末吉健二[†1]　青木謙二[†2]　坂本眞人[†3]　古谷博史[†3]

血友病 B は第 IX 因子遺伝子の突然変異によって引き起こされる．我々は，第 IX 因子遺伝子の突然変異のデータベースを用いて研究を行った．本研究では，タンパク質の立体構造を使用し，重回帰分析を用いて突然変異の効果を解析した．この結果，立体構造と第 IX 因子活性の関連が示唆された．

Analysis of Genetic disease using three-dimensional protein structure

GI OU[†1]　KENJI SUEYOSHI[†1]　KENJI AOKI[†2]
MAKOTO SAKAMOTO[†3]　HIROSHI FURUTANI[†3]

Hemophilia B is caused by mutation in factor IX gene. We study the factor IX missense mutation database. We analyze the mutations using three-dimensional protein structure, and apply the multiple regression method. The results show the importance of three-dimensional protein structure.

1. はじめに

最近，遺伝病の原因となる突然変異のデータが大量に蓄積され，遺伝病の病因解明や治療法の研究においてなくてはならないものとなりつつある．血友病 B もその病因となる遺伝子，血液凝固第 IX 因子，における患者の突然変異が蓄積されデータベースとして公開されている．我々はこれまで，患者突然変異のうちアミノ酸置換（ミスセンス変異）について，血液凝固能（活性度）と変異アミノ酸の種類のかんけいについてアミノ酸の物理・化学パラメータを用いて解析してきた[1][2].

本報告では，これに加え，第 IX 因子の立体構造から抽出したデータを用いて解析を行い，従来法より良い結果を得た．

2. 理論

ASA-View[3]を用いて,血液凝固第 IX 因子の立体構造から抽出したデータを求めた．

解析には、(1)分子体積，(2)疎水性，(3)極性要求，(4)等電点の 4 つのパラメータに立体構造のデータを加えた計 5 つのパラメータを用いた．(1)は大きさ，(2)と(3)は極性，(4)は電荷に関係した量である．各パラメータにおいて，アミノ酸 i と j の間の距離$\mathbf{D_{ij}}$を次の式で定義する．

$$\mathbf{D_{ij}} = |\mathbf{f_i} - \mathbf{f_j}|$$

ここで$\mathbf{f_i}, \mathbf{f_j}$は各パラメータのアミノ酸 i 及び j における値を表す.

5 つのパラメータを用いて，第 IX 因子活性の予測値を求め，重相関係数を求めた.予測値を求める式は次の式で表わされる．

$$\mathbf{E_{ij}} = \boldsymbol{\alpha}|\mathbf{a_i} - \mathbf{a_j}| + \boldsymbol{\beta}|\mathbf{b_i} - \mathbf{b_j}| + \boldsymbol{\gamma}|\mathbf{c_i} - \mathbf{c_j}| + \boldsymbol{\delta}|\mathbf{d_i} - \mathbf{d_j}| + \boldsymbol{\zeta}\mathbf{e}$$

ここで，i,j はアミノ酸の種類を表し，a は分子体積，b は疎水性, c は極性要求, d は等電点, e は類似スコアを表し，係数$\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\zeta}$は重回帰分析により求める．

最後に,サポートベクターマシンを使って,重症と軽症の分類を行った.

サポートベクターマシンは, ソフトマージンとカーネルトリックを用いた．このとき，サポートベクターマシンの特徴空間での双対問題は次式で表される．

$$\mathbf{max} \quad \mathbf{Q}(\boldsymbol{\alpha}) = \sum_{\mathbf{i=1}}^{\mathbf{n}} \boldsymbol{\alpha}_\mathbf{i} - \frac{\mathbf{1}}{\mathbf{2}} \sum_{\mathbf{i,j=1}}^{\mathbf{n}} \boldsymbol{\alpha}_\mathbf{i}\boldsymbol{\alpha}_\mathbf{j}\mathbf{y_i}\mathbf{y_j}\mathbf{K}(\mathbf{x_i}, \mathbf{x_j})$$

$$\mathbf{s.t} \quad \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{y_i}\boldsymbol{\alpha}_\mathbf{i} = \mathbf{0}, \quad \mathbf{0} \le \boldsymbol{\alpha}_\mathbf{i} \le \mathbf{C} \quad (\mathbf{i} = \mathbf{1} \ldots \mathbf{n})$$

ここで$\mathbf{y_i}, \mathbf{y_j}$は正解クラスラベルであり，活性度１％未満を重症とし，クラス-1 を重症，+1 を軽症とする．K はカーネルで，ここでは RBF カーネルを使用している．これを用いて$\boldsymbol{\lambda}$を求め，決定関数を次式で決定する．

$$\mathbf{D}(\mathbf{x}) = \sum_{\mathbf{i} \in \mathbf{S}} \boldsymbol{\alpha}_\mathbf{i}\mathbf{y_i}\mathbf{K}(\mathbf{x_i}, \mathbf{x_j}) + \mathbf{b}$$

活性度の実測値及びサポートベクターマシンで予測値を用いて感度と特異度を導いた．求めた感度及び特異度から ROC 曲線を使って最適なカットオフポイントを選んだ.

3. 数値計算

解析には fixhome[4]の 836 人のデータを用いて解析を行った．まず単回帰分析を行った．パラメータ毎に解析を行った結果，相関係数は 0.1721 で，4 つのパラメータの中で良かった分子体積の相関係数 0.1602 より良い結果が得られた．

†1 宮崎大学工学研究科
Graduate School of Engineering, University of Miyazaki
†2 宮崎大学情報基盤センター
Information Technology Center, University of Miyazaki
†3 宮崎大学工学部
Faculty of Engineering, University of Miyazaki

次に，4 つのパラメータと 5 つのパラメータを使って重回帰分析を行った．4 つのパラメータと立体構造のデータを加えた 5 つのパラメータを用いた重回帰分析では，重相関係数が 0.2824 で 4 パラメータの 0.1910 より高い結果が得られた．

次に，5 つのパラメータを用いて，サポートベクターマシンを使った重症･軽症の分類の結果を表 1 に示す．また，図 1 に活性度 1%未満を重症とした場合，図 2 に活性度 5%未満を重症とした場合の第IX因子全体におけるサポートベクターマシンの ROC 曲線の結果を示す．縦軸が真陽性率，横軸が偽陽性率の図である．

表 1：活性度 1%未満

		予測		
		軽症	重症	計
実測	軽症	489	79	568
	重症	51	217	268
	計	540	296	836

表 2：活性度 5%未満

		予測		
		軽症	重症	計
実測	軽症	128	142	270
	重症	9	557	566
	計	137	699	836

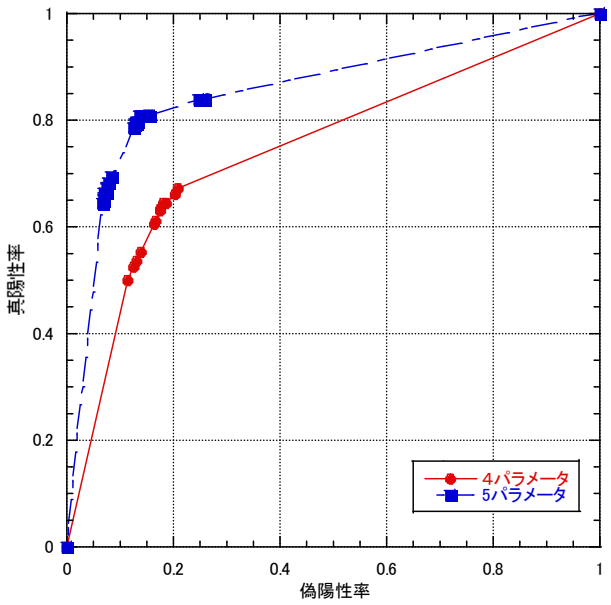

図 1：活性度 1%未満

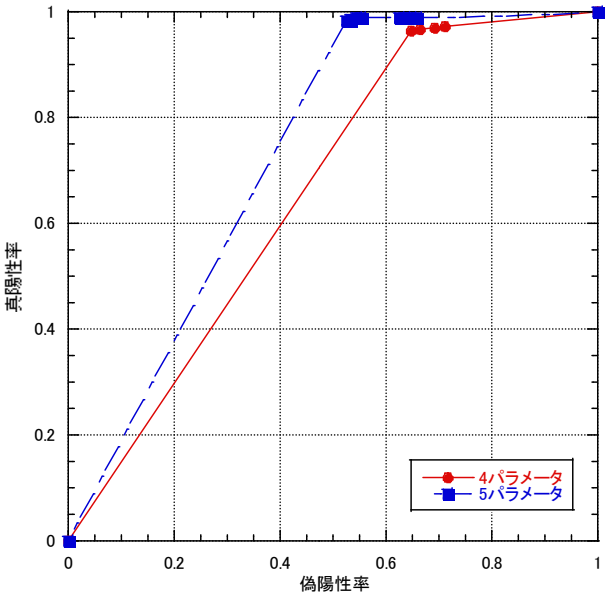

図 2：活性度 5%未満

このとき，活性度 1%未満のカットオフポイントは 0.70 で，活性度 5%未満のカットオフポイントは 0.01 である．表 1 から活性度 1%未満の感度は 87.97%，特異度は 86.09%であった．4 つのパラメータを使ったとき，感度は 67.16%，特異度は 79.04%であり，立体構造のデータを加えたほうが感度と特異度が高くなっていることがわかる．表 2 から活性度 5%未満の感度は 98.41%，特異度は 47.40%であった．4 つのパラメータを使ったとき，感度は 96.29%，特異度は 35.19%であり，立体構造のデータを加えたほうが感度と特異度が高くなっていることがわかる．

4. おわりに

アミノ酸の配列類似性を用いた重回帰，サポートベクターマシンによる解析を行った．その結果，立体構造のデータを加えたほうが良くなることがわかった．今後は立体構造とパラメータの組み合わせを変えて詳しい検証を重ねていきたい．

参考文献

1) 古谷博史,“血友病 B における第IX因子アミノ酸置換と活性の相関分析” 医療情報学 Vol.13 ,No.4 ,pp.211-220(1994)
2) 宇都宮 真，古谷 博史，片山晋，“第 IX 因子タンパク質のアミノ酸置換と血友病 B の重症度の相関分析”, 宮崎大学 情報システム工学科 卒業論文（2006）
3) ASA-View, http://gibk26.bio.kyutech.ac.jp/jouhou/shandar/netasa/asaview/
4) fixfome, http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html