

A-017

多変量解析を用いた血友病 B の研究  
Study of Hemophilia B with Multivariate Analysis

宇都宮 真†      坂本 真人†      古谷 博史†  
Makoto Utsunomiya   Makoto Sakamoto   Hiroshi Furutani

## 1. はじめに

血友病は血液凝固因子欠乏症の1つで伴性劣性遺伝による先天性疾患であり、血友病 B は第 IX 因子の欠損あるいは活性低下によって発症する。

第 IX 因子遺伝子の突然変異の大部分は点突発変異である。点突発変異ではアミノ酸配列の置換が最も多く、活性度は変異の起きた部位、アミノ酸置換の種類に依存してさまざまな値をとる。一般に、重要な部位における変異や性質の異なるアミノ酸への置換は、因子活性の大幅な低下をもたらす、重要でない部位の変異や性質の似たアミノ酸への置換は影響が小さいと予想される[1]。

血友病 B 患者の第 IX 因子の変異に関しては、Haemophilia B Mutation Database (fixhome) にデータが蓄積されている[2]。このデータベースには 2925 人のデータが存在するが、今回は女性のデータなどを除いた 1494 人のデータを使用した[3]。表 1 はデータベースの一部を抜き出したものである。

表 1 血友病 B データベースの一部

活性度	アミノ酸置換		
	発生部位	前	後
4.7	182	V	L
0	182	V	F
23	182	V	A

本研究では、fixhome から得たデータとアミノ酸の種類による 4 つの物理・化学的パラメータを用い、第 IX 因子活性とアミノ酸置換の関連について回帰分析を用いて調べた。また、ロジスティック回帰分析を用いてアミノ酸置換による血友病 B 重症度の判別を行った。

## 2. 方法

本研究で用いた分析法について説明する。

## 2.1 アミノ酸間距離

アミノ酸のもつ物理・化学的パラメータに着目し、第 IX 因子活性との関連を調べた。本研究では以下の 4 つのパラメータを採用した。

- (1) 分子体積 (Molecular volume)
- (2) 疎水性 (Hydrophathy)
- (3) 極性要求 (Polar requirement)
- (4) 等電点 (Isoelectric point)

パラメータの各々について、アミノ酸  $i$  と  $j$  の間の距離  $D_{ij}$  を次式で定義する[4]。

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

ここで  $f_i, f_j$  はアミノ酸  $i$  および  $j$  の各パラメータの値である。アミノ酸間距離としては、この他様々な方法があるが、今回はとりあえずこの方式を採用した。

## 2.2 回帰分析

アミノ酸間距離を用いて第 IX 因子活性の予測を行った。

## 相関分析

いくつかの変量の間に関連性の強さの程度を解析する手法を相関分析という。2 変数  $x$  と  $y$  の値の関連の強さの尺度としては、相関係数  $r$  が用いられ、 $|r|$  が 1 に近ければ強い相関、0 に近ければ弱い相関がある。

## 重回帰分析

$$y_\alpha = \beta_0 + \beta_1 x_{\alpha 1} + \beta_2 x_{\alpha 2} + \cdots + \beta_i x_{\alpha i} + \cdots + \beta_p x_{\alpha p} + \varepsilon_\alpha$$

このモデルを重回帰モデルという。

ここで、 $y_\alpha$  として第 IX 因子活性、 $x_{\alpha i}$  としてアミノ酸間距離  $D_{ij}$  を用いた。 $\alpha$  は血友病患者の番号である。

## 2.3 ロジスティック回帰分析

ロジスティック回帰分析を血友病患者の重症度の判別分析のために用いた。

$$p(x) = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r)\}}$$

$p(x)$  は重症となる確率を表す。このモデルをロジスティック回帰モデルという。この式を変形すると、

$$\log \frac{p(x)}{1-p(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_r x_r$$

となり、重回帰モデルのような式が現れる。これがロジスティック回帰モデルと呼ばれる由来である[5]。

$p$  は 0~1 の値をとり、最尤法を用いることで  $\beta_r$  を求める。

† 宮崎大学, University of Miyazaki

3. 結果

3.1 重回帰分析(1)

y にデータベースから得た第 IX 因子活性度を当てはめて分析した結果を下に示す。

表 2 分析結果 (重回帰分析(1))

	係数	p 値
M	-0.028314	*
H	-0.36848	**
P	-0.145484	
I	-0.415529	**
相関係数	0.2510253	
有意		**

結果が 1%水準で有意なら p 値は\*\*、5%水準で有意なら\*、5%水準で有意でないなら空白で示してある。極性要求以外は 1%あるいは 5%水準で有意であるという結果になった。これより疎水性(H)と等電点(I)が第 IX 因子活性にとつて重要である、ということが示唆される。

次にこの結果を散布図に示した。

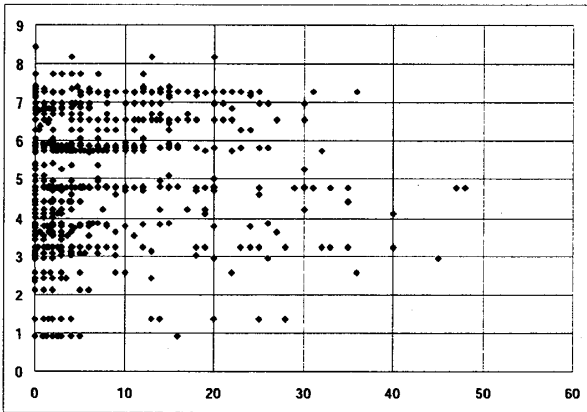


図 1 実測値と予測値の関係 (重回帰分析(1))

X 軸は活性度の実測値、Y 軸は分析の結果から求めた予測値である。

3.2 重回帰分析(2)

表 3 のような、アミノ酸置換の前と後が同じものは活性度の平均値を求め、表 4 を求めた。例えば、表 3 では G→S の変異について活性度の平均値を求める。

表 4 の活性度 y について同様な重回帰分析を行った。

表 5 にその結果を示す。疎水性(H)と全体の結果は 5%水準で有意であったが、それ以外は 5%水準で有意でないという結果になった。

表 3 活性度平均化の例(1)

活性度	アミノ酸置換		
	発生部位	前	後
5	3	S	P
20	4	G	S
17	4	G	S
11	4	G	S
4	5	K	E

表 4 活性度平均化の例(2)

活性度	アミノ酸置換		
	発生部位	前	後
5	3	S	P
16	4	G	S
4	5	K	E

表 5 分析結果 (重回帰分析(2))

	係数	p 値
M	0.0006	
H	-0.5087	*
P	-0.0061	
I	-0.4324	
相関係数	0.3061	
有意		*

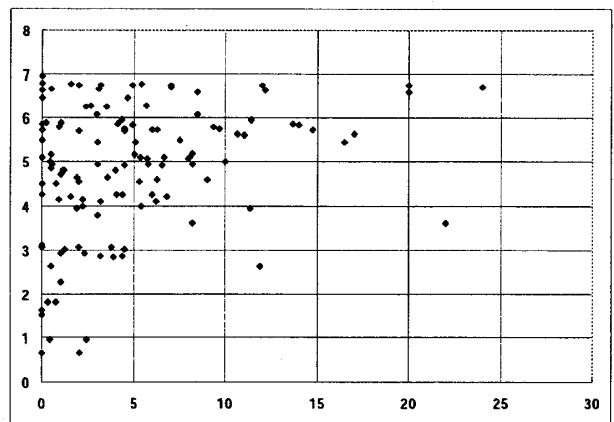


図 2 実測値と予測値の関係 (重回帰分析(2))

X 軸は平均化した活性度の実測値、Y 軸は分析の結果から求めた予測値である。

3.3 ロジスティック回帰分析(1)

次に重症度の判別を行うためロジスティック回帰分析を行った。第 IX 因子活性度を 100 で割ったものを  $p$  に当てはめて分析を行った。

$p$  の実測値、予測値に関して 0.05 以上のものは軽症、0.05 未満のものを重症として、それぞれのデータ数を表 6 にまとめた。

表 6 実測と予測における軽症、重症の度数 (ロジスティック回帰分析(1))

		予測		
		軽症	重症	計
実測	軽症	315	150	465
	重症	277	752	1029
	計	592	902	1494

3.4 ロジスティック回帰分析(2)

活性度が 5 以上のものは  $p=0$ 、5 未満のものは  $p=1$  において分析した結果を下に示す。

表 7 分析結果 (ロジスティック回帰分析(2))

	係数	p 値
M	0.0136	**
H	0.1844	**
P	0.0958	
I	0.2669	**

極性要求(P)以外は 1%水準で有意であるという結果になった。

$p$  の実測値が 0 であれば軽症、1 であれば重症とする。また、 $p$  の予測値が 0.5 以下ならば軽症、0.5 より大きければ重症として、それぞれのデータ数を表 8 にまとめた。

表 8 実測と予測における軽症、重症の度数 (ロジスティック回帰分析(2))

		予測		
		軽症	重症	計
実測	軽症	187	278	465
	重症	108	921	1029
	計	295	1199	1494

3.5 ロジスティック回帰分析(3)

活性度が 5 以上のものは  $p=0$ 、5 未満の部分については活性度を  $x$  において、 $p = -\frac{1}{5}x + 1$  の式で変換した。その分析結果を下に示す。

表 9 分析結果 (ロジスティック回帰分析(3))

	係数	p 値
M	0.0123	**
H	0.0786	*
P	0.0900	*
I	0.1591	**

分子体積(M)と等電点(I)は 1%水準、疎水性(H)と極性要求(P)は 5%水準で有意であるという結果となった。

$p$  の実測値、予測値に関して 0.4 以下のものは軽症、0.4 より大きいものを重症として、それぞれのデータ数を表 10 にまとめた。

表 10 実測と予測における軽症、重症の度数 (ロジスティック回帰分析(3))

		予測		
		軽症	重症	計
実測	軽症	192	273	465
	重症	117	912	1029
	計	309	1185	1494

4. まとめ

重回帰分析では、疎水性と等電点が最も重要なパラメータであることが分かった。ロジスティック回帰分析では、それらに加えて分子体積も重要であることが分かった。アミノ酸パラメータを用いた血友病 B 重症度の判別は、有意ではあるが実用レベルにはまだ達していない。これは突然変異の発生部位依存性が考慮されていないためと思われる。今後この点を取り入れた方法を開発していきたい。

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 (1993).
- [2]Makoto Utsunomiya, Makoto Sakamoto, Hiroshi Furutani, “Analysis of Amino Acid Substitutions and Factor 9 Activities in Hemophilia B”, APBC2008 (2008).
- [3]“Haemophilia B Mutation Database”, Ver.13 (2004). <http://www.kcl.ac.uk/ip/petergreen/haemBdatabase.html>
- [4]Makoto Utsunomiya, Makoto Sakamoto, Hiroshi Furutani, “Regression Analysis of Amino Acid Substitutions and Factor IX Activity in Hemophilia B”, AROB 13th '08 (2008).
- [5]丹後 俊郎, 山岡 和枝, 高木 晴良, “ロジスティック回帰分析 -SAS を利用した統計解析の実際-”, 朝倉書店 (1996).