

## アミノ酸の使われ方の非対称性に注目したゲノム情報工学

ゲノム非対称性グループ 平木秀明

近年、ゲノムプロジェクトの進行に伴い生物のゲノムがコードしている全蛋白質（プロテオーム）を全体として捉えることが可能になった。我々はプロテオームにおけるオリゴペプチドの頻度解析によりアミノ酸の使われ方に非対称性があることを見出し、これに基づいてオリゴペプチドの“珍しさ”に対するアミノ酸残基の寄与度を表す新たな指標“珍しさ度数”を考案した。この指標は既知の蛋白質間の相同性に基づく従来の機能部位予測法とは異なる概念に基づくものであり、ゲノム情報の新しい利用法である。

### 1. はじめに

生物は各細胞にゲノム DNA を持っている。生物の多種多様な構造・機能はゲノム DNA に記された情報をもとに形作られ、発揮される。いわばゲノム DNA は一個の生物を形作るのに必要な設計図にあたる。いくつかの生物のゲノムプロジェクトによりゲノム DNA の全塩基配列が解明され、ゲノム情報を全体的に扱うことが可能になってきた。このため旧来の生命活動の個々の部品にかかわる研究手法だけでなく、一個の生物の設計図全体を見渡した新たな研究手法が重要になってきている。

ゲノム情報にはその生物に存在するすべての蛋白質のアミノ酸配列がコードされている。よく似たアミノ酸配列を持つ蛋白質はその構造・機能が似ているが、それらのアミノ酸配列を比較すると配列中に特によく保存されている部分があり、そういった部位はその蛋白質が機能するために重要な機能部位に対応することが知られている。

我々は、ゲノムプロジェクトによって得られたゲノム中の全蛋白質（プロテオーム）のアミノ酸配列の集合を解析し、オリゴペプチド（アミノ酸数残基の部分配列）の出現頻度がオリゴペプチドごとに異なることを見いだした。すなわちオリゴペプチドの使われ方は非対称であり、多くの蛋白質に使われプロテオーム中に大量に存在するオリゴペプチド、特定の蛋白質にしか使われておらずプロテオーム中では珍しいオリゴペプチド、プロテオーム中には存在しないオリゴペプチドがある。このことから各蛋白質の特異的な機能を担っている機能部位は、プロテオーム中に遍在しているオリゴペプチドではなく、その蛋白質に特異的な“珍しい”オリゴペプチドの部分に存在する可能性が高いと予想される。この考えに基づき、プロテオーム中のオリゴペプチドの頻度から計算されるアミノ酸残基の“珍しさ度数”という指標を考案した。この指標が高い残基はその蛋白質の特異的な機能に関わっている可能性が示唆される。

### 2. 研究の内容

#### 2. 1 *M. jannaschii* のオリゴペプチド頻度解析

ゲノム DNA が解明された生物のうち、本稿では主にメタン古細菌 *Methanococcus jannaschii* の解析を紹介する。*M. jannaschii* は原核細胞生物、真核細胞生物と並ぶ第三のグループ・古細菌に属する生物の中で最初に全塩基配列の解明されたものである。*M. jannaschii* のプロテオームに存在する各オリゴペプチドのコピー数と、同じコピー数を持つオリゴペプチドの種類数の関係を見ると、各オリゴペプチドは均等に使われておらず、コピー数の多いものと少ないものがあることが分かる。例えばテトラペプチド（長さ4残基のオリゴペプチド）の16000種類（アミノ酸20種類の四乗）の可能性のうち、プロテオームで実際に使われているのは104420種類（65%）で32210種類（20%）は1回しか使われていないが、10回以上使われているものが10891種類（8%）ある。我々はこのようなプロテオームにおけるオリゴペプチドの使われ方の非対称性に注目した。

#### 2. 2 *M. jannaschii* と *P. furiosus* に共通する蛋白質の比較

超好熱性古細菌 *Pyrococcus furiosus* は *M. jannaschii* と同じく古細菌に属する生物である。公開データベース上に登録されている *P. furiosus* の蛋白質との相同性により対応付けられる22種類の蛋白質を *M. jannaschii* のプロテオームから取り出し、この22種類の蛋白質に含まれるオリゴペプチドについて相同蛋白質間の保存性とプロテオームにおける出現頻度の関係を調べた。解析の結果、プロテオーム中における出現頻度の低い“珍しい”テトラペプチドほど相同蛋白質間で一致残基数が多くなる、すなわち保存されている傾向が見られた。蛋白質の機能にとって重要な場所は突然変異によって機能低下を招く可能性があるため相同蛋白質間で保存されやすい。したがって“珍しい”オリゴペプチドが機能部位に対応する傾向がある

と考えられる。

## 2. 3 アミノ酸残基の“珍しさ度数”

前述したように各蛋白質の特異的な機能を担っている機能部位はプロテオーム中に遍在しているオリゴペプチドではなくその蛋白質に特異的な“珍しい”オリゴペプチドの部分に存在する可能性が高いと考えられる。一方、アミノ酸配列の情報から蛋白質の機能部位を予測するには、従来相同蛋白質間の保存性に基づいた手法が用いられている。そこで我々は、与えられたアミノ酸配列からオリゴペプチドを“珍しく”しているような残基を見つけるため、アミノ酸残基の“珍しさ度数”という指標を考案した。

“珍しさ度数”は、与えられたアミノ酸配列の各残基について、その残基を含んだオリゴペプチドがその残基によってどれくらい“珍しい”ものになったかを計算して求めたものである。“珍しさ度数”が高い残基はオリゴペプチドの特異性に大きく寄与しているため、その蛋白質の特異的な機能に関わっている可能性が高いと推定される。*M. jannaschii* のプロテオームにおいて“珍しさ度数”は一峰性に分布しており、平均は3.62、標準偏差は0.75である。

## 2. 4 *Pfu* DNA 合成酵素の改変実験

古細菌 *P. furiosus* 由来の *Pfu* DNA 合成酵素は2.2節で解析した22種類の相同蛋白質の一つであり、*M. jannaschii* の DNA 合成酵素と約46%の相同性がある。*P. furiosus* と同属の *Pyrococcus* sp. KOD 由来の DNA 合成酵素と比べると、アミノ酸配列では約80%の相同性があるが DNA 鎖伸張効率は KOD DNA 合成酵素の方が約6倍良い。この *Pfu* DNA 合成酵素と KOD DNA 合成酵素の酵素活性の違いは両者の間で保存されていない20%の残基のいずれかに起因するはずである。そこで両者間であまり保存されていない5つの領域 I~V に注目し、各領域の“珍しさ度数”を計算するとともに、*Pfu* DNA 合成酵素の領域 I~V を遺伝子工学的に KOD DNA 合成酵素の配列に改変したキメラ DNA 合成酵素を作成した。領域 III は従来の相同蛋白質間の保存性にに基づく手法により予測される DNA 合成酵素の機能部位（モチーフ C）と重なっているが、他の4つは重ならない領域である。

領域 I~V における“珍しさ度数”を見ると、領域 III、II、V における“珍しさ度数”の平均（それぞれ4.07、3.91、3.74）が *M. jannaschii* のプロテオームにおける平均値より高かった。この差は統計的には有意ではないが、これらの領域がこの蛋白質の機能に関わっている部位である可能性を示唆している。

作成した変異体酵素を用いた DNA 合成実験

の結果、野生型酵素(WT)や変異体 I、II、IV などでは、使用したいずれの酵素量においても1000 base 位から様々な位置にバンドが確認でき、伸長反応の鋳型特異的な停止が起こっていることが分かる。一方、変異体 IIIb においては、300-600 base にかけてのバンドが薄くなっている。さらに変異体 III においては、ほぼ1000 base 程度のバンドが主になり、これら領域 III の変異体では伸長停止反応を解除する活性があることが分かる。そこで、IIIb のアミノ酸置換に対応する変化を詳細に検討した結果、この中の I540S の1アミノ酸の変化が DNA 鎖伸長停止を解除する活性に重要であることが判明した。また変異体 Vb' は伸長停止反応を解除する活性に関しては野生型酵素(WT)と変わらないが、DNA 合成量の増大に伴う高分子の産物が検出される。したがって、KOD DNA 合成酵素が *Pfu* DNA 合成酵素に比べより良い伸長活性を示すことに対応する領域は III と V であることが示唆される。

領域 III に関する実験結果から、DNA 鎖伸長停止を解除する活性は I540S のアミノ酸置換を軸として L545F 及び Y546F が補助的に働いていると考えられる。一方、これらのアミノ酸に対応する“珍しさ度数”を見ると *Pfu* に比べ KOD DNA 合成酵素のほうが高くなっている。すなわち本例では、“珍しさ度数”の高くなる傾向が活性を担うアミノ酸残基に対応している。

## 3. 終わりに

前節で述べた例のように、“珍しさ度数”による機能部位予測は近縁種の相同蛋白質間で保存されていない部分にも適用できる。このような手法は系統の離れた種間の相同蛋白質で保存されている部分に基づいて機能部位を推定する従来の手法と対照的であり、様々なゲノムプロジェクトにより大量に蓄積されつつある既知のものと同様性がない新規蛋白質の機能を同定するのに有用であると期待される。本法はいまだ発展段階であり、さらなる条件決めや実験的な証明が要求される。しかしながら、本研究のように全体的な眼をもってゲノム情報を解析していくことが、今後ますます重要になってくると思われる。