

ゲノムにコードされるアミノ酸配列におけるジスルフィド結合パターンの推定

綾田 瑛美(指導教員: 由良 敬)

1. はじめに

タンパク質は、アミノ酸が一次的に連結して構成される直鎖状の生体高分子である。生物が用いているタンパク質は、20種類のアミノ酸で構成されている。アミノ酸配列は、DNA に与えられた情報を「転写」と「翻訳」することによって得ることができ、アミノ酸配列が決定すると、自発的に立体構造を形成する。

ゲノムプロジェクトによりアミノ酸配列の情報は簡単に得られるようになった。しかし、アミノ酸配列の情報から、立体構造を導出するには、X線結晶解析などの実験で測定するしか方法がない。立体構造の測定には、時間と費用が膨大にかかる。そこで、計算機を用いてタンパク質のアミノ酸配列から立体構造を推定する研究が、世界中で進められている。しかし現在までに、成功例は数えるほどしかない。様々な種類のタンパク質すべてに適用可能な立体構造予測方法を作ることには難しい。そのため、特殊なタンパク質の立体構造を予測できる方法を開発することは価値がある。

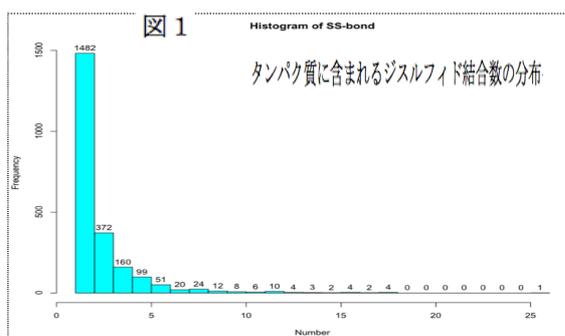
上記の通り、タンパク質の直鎖状の高分子である。実際のタンパク質は、折れたたまれた状態で存在する。折れたたまれ方には膨大な数の可能性があるが、実際はアミノ酸配列によって一意的に決まる。そして様々な結合が、タンパク質の立体構造を支えている。

そのなかの一つにジスルフィド結合がある。ジスルフィド結合とは、2つのシステインというアミノ酸に含まれる硫黄同士が共有結合したもので、これらのシステインは空間的に接近する。ただしすべてのシステインがジスルフィド結合に関与するわけではない。アミノ酸配列情報のみから、どのシステインがジスルフィド結合に関与し、どのシステインのペアがジスルフィド結合するかを知ることで、タンパク質の立体構造がわかるのではと考えた。

2. 手法

2.1. 対象データ

タンパク質立体構造データベース Protein Data Bank(PDB)に含まれている、既知の相同性のないタンパク質の立体構造16753個から



少なくとも1つのジスルフィド結合を持つタンパク質2264個を取り出した。

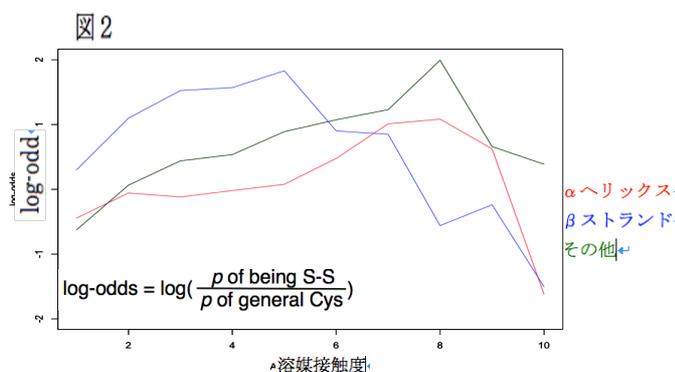
ジスルフィド結合を持つタンパク質は PDB に含まれているタンパク質の約14%であった。また、ジスルフィド結合を持つタンパク質のうち約半数は、ジスルフィド結合を1つしか持たなかった。図1のグラフはジスルフィド結合数の分布をあらわしたものである。

本研究では、異なるタンパク質間のジスルフィド結合は考えずに、一つのタンパク質内にあるジスルフィド結合のみを対象としている。

2.2 パラメーター

ジスルフィド結合に関与するシステインを特定するためのパラメーターを決定するため、システインとシスチン(システインが酸化して、ジスルフィド結合をしない状態のアミノ酸)が、タンパク質の二次構造であるαヘリックス、βストランド、その他の領域(コイル)、のどこに頻出するかを調べた。また、二次構造の表面と内部のどちらに位置しているのかも考慮した。

次にジスルフィド結合をするペアのシステインが一次的にどのくらい離れているかに着目し、2つのシステイン間にあるアミノ酸の残基数を距離とした。



上記図は、横軸を溶媒接触度、縦軸を二次構造のパラメーターとしたグラフで、αヘリックス(赤)、βストランド(青)、その他の領域(緑)が、タンパク質の内部に位置しているのか、表面に位置しているのかを表している。ジスルフィド結合するシステインは、βストランドでは内部に、αヘリックスとその他の領域においては表面に多く表れていることを示している。

ここでパラメーターは対象二次構造にジスルフィド結合が存在する確率に依存する。

図3は、横軸をジスルフィド結合する2つのシステイン間の距離(アミノ酸残基数)、縦軸を距離の確率に依存するパラメ