

Title	タンパク質配列における共通保存領域の探索手法に関する研究
Author(s)	川路, 英哉
Citation	大阪大学, 2003, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/44351">https://hdl.handle.net/11094/44351</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉</a> 大阪大学の博士論文について <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈/a〉</a> をご参照ください。

***Osaka University Knowledge Archive : OUKA***

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	川路英哉
博士の専攻分野の名称	博士(工学)
学位記番号	第 17918 号
学位授与年月日	平成 15 年 3 月 25 日
学位授与の要件	学位規則第 4 条第 1 項該当 基礎工学研究科情報数理系専攻
学位論文名	タンパク質配列における共通保存領域の探索手法に関する研究
論文審査委員	(主査) 教授 松田 秀雄 (副査) 教授 柏原 敏伸 教授 増澤 利光

#### 論文内容の要旨

生物を構成する最も重要な分子の一つであるタンパク質は、その機能に関して生物学や医学、創薬等の観点から様々な研究が行われている。タンパク質は 20 種類のアミノ酸が直鎖状に結合した分子であり、その結合順序を文字列で表現したものはタンパク質配列と呼ばれる。生物学的な実験を用いたタンパク質の機能解明は大きなコストが必要になることから、計算機を用いたタンパク質の機能解析に関する研究が行われてきた。特に現在では、配列解析に基づいたタンパク質の機能予測に関する研究は非常に重要なものである。様々な実験により多くのタンパク質の機能が解明されてきた結果、機能的に関連のあるいくつかのタンパク質の配列上に共通保存領域と呼ばれる局所的なアミノ酸の出現パターンが見られることが明らかになった。

これまで共通保存領域を求めめるために、閾値以上の類似を示す部分配列を計算する手法や、ある小さなタンパク質配列集合の中で、すべてのタンパク質配列に共通して存在する配列パターンを計算する手法、あるいは多くが互いに類似する部分配列集合を計算する手法等が提案されてきた。しかしこれらの手法では、a) タンパク質同士の機能的な関係は、予め与えられない、b) 様々な種類の共通保存領域を持つタンパク質を含む大規模なタンパク質配列集合が入力として与えられる、の両方の条件を満たす場合の共通保存領域探索は難しいものであった。

そこで本論文では、これらの探索が可能な共通保存領域探索手法を提案する。提案手法は、1) タンパク質配列集合をより小さな部分集合へ分類する、2) 各々の部分集合に対して共通保存領域を探索する、の二ステップから構成される。第一ステップでは p 準完全連鎖法を、第二ステップでは最大密度部分グラフ法を使用することで、実際に大規模なタンパク質配列集合から、これまでに見付かっていない共通保存領域が探索可能であることを示した。また、より大規模なタンパク質配列集合からの探索を可能にするために、p 準完全連鎖法よりも小さな計算量でタンパク質配列集合の分類を行う手法を提案し、これが実際に有効であることを示した。

## 論文審査の結果の要旨

本論文では、タンパク質を構成するアミノ酸配列（以下、タンパク質配列）の集合に対し、配列間に存在する共通保存領域を探索する手法を提案している。共通保存領域は機能の同じタンパク質の配列に見られる、互いに類似した部分配列集合であり、タンパク質の解析に広く用いられている。しかし、共通保存領域間の配列類似度の明確な基準を決めることが困難であるなどの理由から、従来手法では、数万個以上の規模のタンパク質配列集合から、体系的に共通保存領域を探索することは困難であった。

本論文で提案されている手法は  $p$  準完全連鎖法によるタンパク質配列集合のクラスタリングを行うステップと、クラスタリング結果の部分配列集合から最大密度部分グラフ法により共通保存領域を探索するステップで構成される。共通保存領域探索に、タンパク質配列集合のクラスタリングを導入することで、多数のタンパク質配列からの探索を可能にしている。

また、さらに多くのタンパク質配列からの探索を可能にするため、グラフ分割を用いたクラスタリング手法について提案している。 $p$  準完全連鎖法が互いに重複する部分集合へ分類を行うのに対し、分類結果間の重複を考えないことで、より計算量の小さいクラスタリングを可能にしている。

本論文で提案されている手法の有効性は、マウスの約1万5千個のタンパク質配列に対する実験結果により示されている。 $p$  準完全連鎖法を使用した場合の探索した結果、実際にこれまでに見付かかっていない共通保存領域が探索されていることが示されている。また、グラフ分割によるクラスタリング手法を用いた場合も、同様の結果がより小さな計算量で得られることが示されている。これにより、従来手法では困難であった多くのタンパク質配列からの共通保存領域探索が可能となり、計算機によるタンパク質配列解析の発展に大きく貢献している。よって博士（工学）の学位論文として価値あるものとして認める。