

氏 名	こ いけ りょう た ろう 小 池 亮 太 郎
学 位 の 種 類	博 士 (理 学)
学 位 記 番 号	理 博 第 2707 号
学位授与の日付	平成 15 年 9 月 24 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	理 学 研 究 科 化 学 専 攻
学位論文題目	Comparison of Protein Sequences and Structures based on the Partition Function Formulation: Probabilistic Alignment (分配関数を用いたタンパク質配列・構造の比較法：確率的アラインメント法)
論文調査委員	(主 査) 教 授 加 藤 重 樹 教 授 三 木 邦 夫 教 授 寺 嶋 正 秀

論 文 内 容 の 要 旨

核酸・アミノ酸配列の比較は、進化的類縁関係を同定することによって、分子機能、分子進化を研究する基盤となる情報を与えてきた。比較の方法では、ある最適なアライメントを構築する。アライメントとは2つの配列の残基の対応関係を記述したものである。配列の類似性は、類似残基ができる限り対応付けられるように並べた最適アライメントに対して定義され、そこから進化的類縁関係の有無が統計的に判定される。しかし、このようにして得られる最適アライメントは、評価関数や最適化法に大きく依存する。とりわけ、類似性の低い配列の比較においては、最適アライメントとはほぼ同程度の類似性を与える準最適アライメントが多数存在することが知られており、ただ一つの解で類似性を定義することには大きな問題がある。この問題を解決するための方法として、準最適アライメントを逐次的に検出するが開発されてきた。ここで、申請者はこの問題に対する別の解決法として、全てのアライメントを分配関数として記述することで、準最適アライメントを網羅的に確率を付して記述「確率的アライメント法」と採用した。

確率的アライメント法は、基本的に1次元のイジング模型に基づいている。残基間の対応関係（アライメント）がスピン状態にあたり、ギャップペナルティーが近接スピン間の相互作用として記述される。また、これは標準的な配列比較法である動的計画法を有限温度に拡張したものと理解することもできる。この方法では、全てのアライメントは温度 T での分配関数として扱われる。このようにアライメントを統計力学的枠組みで表現することで、比較の対象となる2つの配列のそれぞれの残基 p_i と q_k の類似度を、従来の方法が用いてきたスコア行列ではなく、分配関数から導出される確率行列によって評価することが可能となる。 $T \rightarrow 0$ の極限では、確率行列によるアライメントは動的計画法により構築されるアライメントと一致する。

しかし、有限温度への拡張は致命的な問題を引き起こす。最も単純な例として、各残基間の類似度が一定となるスコア行列が与えられた場合を考える。このとき、各残基間の対応確率も同様に一定となることが期待される。しかし、実際には確率行列上の対角線に近い対応、すなわちギャップの少ないより多くの残基を対応付けるアライメントに包まれる対応ほど高い確率をもつこととなる。この対応確率のバイアスは有限温度に拡張したことで生じるエントロピーに起因する。申請者はこの問題を解決するために、分配関数に周期境界条件を導入し、エントロピーのバイアスを完全に解消した。

申請者はこの確率的アライメント法による配列比較を行い、進化的類縁関係の検出を行った。また、その検出能を評価するために、進化的類縁関係の検索に広く使われる PSI-BLAST でも同様のテストを行った。PSI-BLAST はデータベース中の進化的類縁関係にある配列を効率よく利用することで検出感度を高めているが、確率的アライメント法を用いることで、データベース中の配列を利用することなく PSI-BLAST に匹敵する高い検出能が得られることが示された。この結果は、アライメントを最適解だけではなく、すべての準最適解を網羅的に確率として評価することの妥当性を示しているといえる。

さらに、申請者はもう1つの重要な情報である立体構造の比較の問題に関しても考察している。進化的類縁関係の検出において構造情報の最大の利点はその保存性の高さである。実際に配列の相同性が20%に満たないようなタンパク質間でも、

その構造は高い類似性を示す場合があり、配列からのアプローチが困難な際に、進化的類縁関係を示唆する傍証として有益な情報源となっている。立体構造の比較も、配列のときと同様にアライメントを構築するのだが、アライメントのエネルギー関数に困難さを引き起こす問題がある。立体構造における二つの残基間の類似性は、ある残基と他のすべての残基との距離の集合の間の類似性によって表現される。この場合エネルギー関数は近接相互作用に止まらず、遠距離相互作用（鎖の上で離れた残基間の相互作用）を含めなければならず、容易に分配関数を書き下すことができない。申請者はこの問題を回避するために、平均場近似を用いて定式化を行った。この近似は、立体構造比較で広く用いられているダブルダイナミックプログラミング法の有限温度への一種の拡張と見ることもできる。

配列比較と同様に、構造比較においてもエントロピーによるバイアスの問題が起こる。そこで、構造比較においても周期境界条件が導入された。これにより、エントロピーのバイアスの無い比較法が構造においても確立された。申請者はこの分配関数による配列・構造比較法を内部対称性のある構造であるTIM-バレル構造と β -trefoil構造を持つタンパク質の比較の問題に適用した。その結果から、最も高い確率をもつ対応は配列と構造で一致し、配列・構造の比較における確率的アライメント法の妥当性とこれらのタンパク質間で円順列変異が起こったことを示した。

この分配関数を用いた枠組みは配列・構造比較の問題に止まらず、各タンパク質のとりうる立体構造の特徴抽出の問題にも拡張することができる。申請者は同様の方法を「All- β 蛋白質」として分類されている蛋白質立体構造の特徴抽出に適用した。ここでは、スコア行列として分子内のC α 原子間の距離を記述する距離行列を用いている。その解析の結果、確率行列上に特徴的なパターンが繰り返し現れることを申請者は発見した。このパターンは、All- β 蛋白質における「リング構造」や「ジッパー構造」という、All- β 蛋白質の立体構造の最も明瞭な特徴である。申請者はこの特徴に基づいて、これまでになく進化的類縁関係を越えた、見通しのよいAll- β 蛋白質の分類を行った。このように進化的類縁関係に関わらずに一般的に見いだされるリング・ジッパー構造は、物理化学的制約、即ちフォールディング過程における制約として現れてきたものと考察される。

論文審査の結果の要旨

申請論文は分配関数によるタンパク質の配列と構造の比較法に関する研究成果であり、主論文は3つの部分に分かれている。第一部はタンパク質配列・構造比較におけるアライメントを確率的に記述するための理論の構築について、第二部はこの理論の微弱な配列類似性を持つタンパク質配列の比較問題への応用に関して、第三部は第一部で構築した理論のタンパク質立体構造の特徴抽出問題への応用に関する研究である。

タンパク質の配列・構造比較法は生物学の主要な課題であるタンパク質の機能や進化の解明において基礎的な役割を担っており、配列や構造の比較法のもつ特徴や問題をより深く理解することは、タンパク質の諸性質を正しく記述するための重要な基礎となる。申請者の研究は配列・構造の比較法において不可欠なアライメントを確率的に正しく記述するための理論的な枠組みを初めて構築したものであり、配列・構造の比較法を理解する上で大きく貢献することが期待される。それ故、この研究は十分に評価に値すると思われる。

第一部では、申請者はまず、配列比較におけるアライメントが長距離相互作用のない1次元スピン系のスピン状態に一意に対応することを示し、分配関数を与えている。これにより、確率的アライメントにおける問題点である「ターミナルの問題」がエントロピーのバイアスに起因することを明らかにした。また、このエントロピーのバイアスを除くために周期境界条件を導入するのであるが、このとき最大の問題点となるのが周期境界条件下の分配関数である。申請者は集合論を用いることでこの分配関数を計算可能な形に展開した。これはエントロピーバイアスのない正確な確率アライメントの記述を可能にする初めての理論であり、この研究は高く評価することができる。また、申請者は配列比較にとどまらず構造の比較においても確率的アライメントの構築に必要な理論を確立した。構造比較におけるアライメントは長距離相互作用を含む1次元スピン系のスピン状態に対応する。この長距離相互作用のため分配関数が分割できず、分配関数を求めることができない。申請者はこの長距離相互作用を自己無撞着な平均場に近似することで分配関数を計算可能なものとした。構造比較において確率的にアライメントを記述する方法は申請者のものが初めてでありこの点も高く評価できる。

第二部では、申請者は第一部で構築された確率的アライメント法を用い、配列比較の重要な課題である微弱な配列類似性

の検出問題に適用した。その結果、最も広く用いられている標的と成る配列の中間配列を利用する方法と同等の検出感度を確率的アライメント法は示した。確率的アライメント法がいかなる中間配列の情報を使うことなしに標的となる配列間の類似性を感度良く検出できることを明らかにした点は評価に値すると思われる。

第三部では、申請者は第一部で構築された理論を応用し、タンパク質立体構造の特徴を抽出するための方法を開発した。この方法を all- β タンパク質に適用することで多数のタンパク質が「リング構造」「ジッパー構造」をとることを見出した。このような広範なタンパク質で保存される共通構造を見出すことは進化の過程における物理化学的な制約の考察につながり、この点においても申請者の研究は十分評価できる。

本論文は、博士（理学）の学位論文として価値のあるものと認める。

論文内容とそれに関連した口頭試問の結果合格と認めた。