

学 位 論 文 題 名

Computer-aided molecular modeling
and sequence analyses of repeat-containing proteins

（繰り返しアミノ酸配列を持つ蛋白質に対する、
コンピュータ技術を用いた分子構造モデリングおよび配列解析）

学位論文内容の要旨

<序>

近年の遺伝子工学的技術や装置の発展により、11万を超える数の蛋白質のアミノ酸配列がこれまでに決定されているが、その配列中に、単一のアミノ酸から数十残基のオリゴペプチドを一単位とし、それが連続して繰り返す配列（タンデムリピート）を含む蛋白質が多く存在していることが知られている。これらのタンデムリピート領域は、その構造的・進化的意味についてはもちろん、基礎医学的な観点からも注目されているが、これまでにこの種の領域を網羅的に扱った研究はほとんど行なわれていなかった。その理由として、タンデムリピート領域に関するしっかりとした定義、情報の収集および整理がなされていなかった事、またタンデムリピート領域の原子レベルでの立体構造決定はあまり進んでおらず、研究の基礎にできる構造情報がほとんど無かった事などが挙げられる。

本研究ではこの非常に特徴的な配列領域に注目し、コンピュータ技術を積極的に利用して大量の情報を高速かつ網羅的に検索・処理することで、その構造的および進化的な意味を明らかにすることを目的として、代表的な以下の2種類の繰り返し配列について種々の解析を行なった。

<p26olfの立体構造予測とCa²⁺結合能力の測定>

p26olf は、カエル嗅上皮細胞質より単離された217残基からなる Ca²⁺結合蛋白質である。p26olf のアミノ酸配列は、その前半部分と後半部分が非常に高い相同性を示し、さらに各々が S100 蛋白質とよく類似している。

S100 蛋白質はその配列中に、EFハンドとして知られるヘリックスループ-ヘリックスの超二次構造を持った Ca²⁺結合モチーフを2つ（N 末端側に位置する S100 に特有な pseudo EF ハンドと、C 末端側に位置する典型的な EF ハンド）含んでいる。p26olf の前半および後半部の配列と S100 蛋白質の配列とのアラインメントの結果、p26olf が2つの S100-like ドメインと、それらのドメイン間をつなぐリンカーからなる事が示された。各ドメインの N 末端側に位置する配列領域は S100 蛋白質の pseudo EF ハンドとよく類似していたが、C 末端側の EF ハンド領域には典型的な EF ハンドモチーフと比較して、Ca²⁺結合ループの部分に4残基ずつの挿入が見られた。

我々はこのアラインメント結果と、apo-S100β 二量体の溶液構造を元にしてホモロジーモデリングを実行し、p26olf の立体構造モデルを構築した。得られた p26olf 一分子の予測構造は、この蛋白質が S100β 二量体と著しく類似したフォールディングを十分にとり得ることを明らかに示した。二つの S100-like ドメインを結ぶリンカーの構造および配置はドメイン本体の立体構造に全く障害を与えずに存在していた。また各 S100-like ドメインの C 末端側に位置する EF ハンド領域は、4残基の挿入にも関わらず、通常より長い Ca²⁺結合ループをもってヘリックスループ-ヘリックスの超二次構造を完全に保っており、構造的な矛盾は見られなかった。よって我々はこれを新規なパターンの EF ハンドであると提案した。

実際の p26olf 溶液に対する Ca²⁺-titration の結果から、p26olf は一分子あたり4個の Ca²⁺を結合し、円偏光二色性(CD)及び蛍光スペクトルの測定結果は、その Ca²⁺の結合と同時に大きな立体構造変化

が起きていることを示した。

以上の結果から我々は、p26olf が S100 蛋白質ファミリーの新たな一員であり、また p26olf が持つ新規なパターンの EF ハンドも Ca^{2+} を結合する能力があると結論づけた。

<ロイシン-リッチ-リピート蛋白質ファミリーの配列解析と分子進化>

ロイシン-リッチ-リピート(LRR)は、ロイシン残基を豊富に含んだ 20~30 残基が一単位として繰り返されるタンデムリピートであり、バクテリアからヒトに至るまでの多くの生物種で、100 種類以上の蛋白質中に見いだされている。しかし、一般に同一の蛋白質内では LRR の繰り返し単位同士が比較的によく類似しているが、異なる蛋白質間ではその繰り返し単位の長さや保存性の高い残基の種類などが著しく異なるため、これまで LRR 蛋白質の分子進化の解析はほとんど進展していなかった。

タンデムリピート配列は一般に、その元となるモチーフ配列が不等交差などの遺伝子重複 (gene duplication) によって、そのリピート数を増加させることで生成されていったと考えられている。我々は、個々の蛋白質の LRR 領域が祖先の LRR 単位からどのような進化過程を経て形成されていったのかを明らかにする為、LRR の各繰り返し単位をそれぞれ独立に扱い、分子進化の系統樹を作成するためのプログラムパッケージである MOLPHY を用いて、それら単位配列間の相関関係を総合的・系統的に評価した。

この解析手法を適用した結果、small proteoglycan ファミリーの蛋白質は、基本的に異なった 2 種類の LRR 単位配列 S と T が (STT) のように 3 つ一組となって高次のリピート単位を形成している事が示された。また Slit 蛋白質の場合は、(XABCABC) のように、より複雑な二次・三次の high-order なリピートが存在し、その上数カ所で LRR 単位での deletion が起きていることが示された。これらの結果を基にして構築された、各ファミリー内における LRR 領域の分子進化モデルは、LRR のリピート数増加の過程が不規則・無秩序に起きているのではない事を強く示唆するものであった。

<まとめ>

コンピュータの大量・高速な情報処理能力を利用して、実験的に明らかにすることが困難な蛋白質分子の立体構造のモデルを提案し、また、繰り返し配列領域の進化過程に関するモデルを構築した。一方で、これらは全てがコンピュータの能力によるものではなく、その出力を人間が生物学的・物理学的知識を以ってデータを解析し、またコンピュータにフィードバックすることなしには成しえなかった。

配列情報・立体構造情報の増加に伴い、本分野におけるコンピュータの有用性はますます高くなる事は明白である。そこにどれだけ専門の研究者の持つ生物学的・物理学的知見を積極的に取りこめるかが重要であることを、本研究では示すことが出来たと考える。

学位論文審査の要旨

主 査 教 授 新 田 勝 利

副 査 教 授 山 岸 皓 彦

副 査 助教授 渡 邊 信 久

副 査 助教授 出 村 誠

副 査 教 授 松 島 範 男 (札幌医科大学保健医療学部)

学 位 論 文 題 名

Computer-aided molecular modeling and sequence analyses of repeat-containing proteins

(繰り返しアミノ酸配列を持つ蛋白質に対する,
コンピュータ技術を用いた分子構造モデリングおよび配列解析)

本研究は以下の2つの部分からなる。

p26olf の立体構造予測と Ca^{2+} 結合能力の測定 p26olf は、カエル嗅上皮細胞質より単離された217残基からなる Ca^{2+} 結合蛋白質である。p26olf のアミノ酸配列は、その前半部分と後半部分が非常に高い相同性を示し、さらに各々が S100 蛋白質とよく類似している。p26olf の前半および後半部の配列と S100 蛋白質の配列とのアラインメントの結果、p26olf が2つの S100-like ドメインと、それらのドメイン間をつなぐリンカーからなる事が示された。各ドメインの N 末端側に位置する配列領域は S100 蛋白質の pseudo EF ハンドとよく類似していたが、C 末端側の EF ハンド領域には典型的な EF ハンドモチーフと比較して、 Ca^{2+} 結合ループの部分に4残基ずつの挿入が見られた。実際の p26olf 溶液に対する実験結果から、p26olf は一分子あたり4個の Ca^{2+} を結合し、その Ca^{2+} の結合と同時に大きな立体構造変化が起きていることを示した。

以上の結果から申請者は、p26olf が S100 蛋白質ファミリーの新たな一員であり、また p26olf が持つ新規なパターンの EF ハンドも Ca^{2+} を結合する能力があると結論づけた

ロイシン-リッチ-リピート蛋白質ファミリーの配列解析と分子進化 ロイシン-リッチ-リピート (LRR) は、ロイシン残基を豊富に含んだ20~30残基が一単位として繰り返されるタンデムリピートであり、バクテリアからヒトに至るまでの多くの生物種で、100種類以上の蛋白質中に見いだされている。申請者は個々の蛋白質

の LRR 領域が祖先の LRR 単位からどのような進化過程を経て形成されたのかを明らかにする為、LRR の各繰り返し単位をそれぞれ独立に扱い、分子進化の系統樹を作成するためのプログラムパッケージである MOLPHY を用いて、それら単位配列間の相関関係を総合的・系統的に評価した。

この解析手法を適用した結果、small proteoglycan ファミリーの蛋白質は、基本的に異なった 2 種類の LRR 単位配列 S と T が (STT) のように 3 つ一組となって高次のリピート単位を形成している事が示された。また Slit 蛋白質の場合は、(XABCABC) のように、より複雑な二次・三次の high-order なリピートが存在し、その上数カ所で LRR 単位での deletion が起きていることが示された。これらの結果を基にして構築された、各ファミリー内における LRR 領域の分子進化モデルは、LRR のリピート数増加の過程が不規則・無秩序に起きているのではない事を強く示唆するものであった。

学位論文の公開発表の質疑応答では、申請者は自ら行った実験の結果や過去の論文の内容を引用し、豊富な知識に基づいて質問に明快に回答した。

以上のように申請者は、EF-hand 型カルシウムタンパク質の新しいメンバーを発見し、タンデムリピート配列についていくつかの重要な知見を示した。審査員一同はこれらの成果を高く評価し、申請者が北海道大学博士（理学）の学位を授与される資格が充分にあるものと認定した。