

学 位 論 文 題 名

Structural Principles of Proteins Revealed by Computer-aided Geometrical Analysis

(蛋白質の立体構造原理解明へのバイオインフォマテイクス研究)

学位論文内容の要旨

最近、コンピュータを駆使した大量のゲノム配列や立体構造の情報解析から、生物の基本的な原理を解き明かすバイオインフォマテイクス研究が注目を集めている。本論文においてコンピュータを利用した幾何学的解析から蛋白質の立体構造の原理を解明した。まず、最小二乗法により 3 次元のデータ点から最適な円を求める方法 (3D circle fitting 法) と最適なヘリックスを求める方法 (Helix fitting 法) を開発した。この 3D circle fitting 法をロイシンリッチリピート (LRR) 蛋白質の立体構造に適用し、一つの構造原理を見出した。さらに、Helix fitting 法を二次構造の一つである 3_{10} -helix に適用し、蛋白質中の 3_{10} -helix は本質的に不完全なヘリックスであることを見出した。

3D circle fitting 法と Helix fitting 法の開発： 最小二乗法を利用して、これらのフィッティング法を開発した。最小二乗法は、二次元のデータ点から最適な直線の方程式を導くためによく使われており、この方法を 3 次元のデータ点から最適な円を求める方法に拡張した。データ点 X_i から任意の円までの最短距離を $d(X_i)$ とする。そして、Iteration のテクニックによりこの $d(X_i)$ の二乗の和 (rmsd) を最小にした。この rmsd が最小の時、最適な円の半径 R と円の中心座標を与える。この方法により最小 3 個のデータ点から正確な半径 R を求めることができる。

ヘリックスはラセン構造で、ヘリックスピッチ (P) (1 回転当たりのヘリックス軸方向の距離) とヘリックスの半径 (r) を用いて表すことができる。データ点 X_i から任意のヘリックスまでの最短距離を $d(X_i)$ とする。この $d(X_i)$ の二乗の和 (rmsd) が最小になる時、最適な P と r 、さらにヘリックス軸を与える。この方法により、最少で 4 個のデータ点から正確なヘリックスパラメータを計算することができる。この計算のためのアルゴリズムは、世界ではじめて独自に開発された。

馬蹄形ロイシンリッチリピート蛋白質の幾何解析： LRR 蛋白質は、バクテリアからヒトまで 4 千種類以上の蛋白質中に見いだされており、その数は爆発的に増加している。植物 LRR 蛋白質においても 2 千個以上存在することが知られている。この LRR 蛋白質は、植物の病原体抵抗性、動物の自然免疫、アポトーシス、細胞接着、遺伝子の組み替え、DNA 修復、転写などの様々な機能発現に関与している。例えばリボヌクレアーゼインヒビター (RI) の一次構造は、28 残基あるいは 29 残基からなるリピートが連続して 17 回繰り返している。この LRR のコンセンサス配列は、 $Lx\boxed{Lx}Lxx(N/C)xLxxxxxxxLxxxLxxxxx$ で表わされる。ここで、L はロイシン、N はアスパラギンを示す。この RI の立体構造は、分子全体として馬蹄形をしていることが明らかにされた。また、それぞれの LRR モチーフは β -シート (コンセンサス配列の $\boxed{xLx}$ の部分)、ループ、 α -ヘリックス (コンセンサス配列の $xxxxxLxxxL$ の部分) からなる超

二次構造をとっている。ここで、この超二次構造を β - α モチーフと名づける。しかしながら、RI の LRR とは異なるさまざまなモチーフが、少なくとも 6 種類以上存在していることが知られている。我々は、このような著しい多様性が、蛋白質立体構造の原理を理解する上で格好の材料を提供していると考え、これまで立体構造が解かれた 15 個の LRR 蛋白質に 3D circle fitting 法を適用した。

すべての LRR 蛋白質は馬蹄形をしているので、円弧とみなすことができる。3D circle fitting 法により円弧の半径 (R) と回転角を求めた。実際のデータは β -シート部分に相当するコンセンサス配列の 4 番目のロイシンの α 炭素の座標を用いた。RI では LRR が 17 回繰り返しているので、データ点は 17 個となる。

RI は β - α モチーフ、internalin は β - 3_{10} モチーフ、yopM は β -ポリプロリン II (pll) モチーフ、Nogo receptor は β - β ターンモチーフをとる。これらのモチーフの違いに依存して、さまざまな値の半径をとることが明らかにし、 β - α 、 β - 3_{10} 、 β -pll、 β - β の順序で、円弧の半径 R が大きくなることをみいだした。逆に、繰り返し単位当たりの回転角 ϕ は、半径 R とは異なり、この順序で小さくなる。隣り合ったデータ点とデータ点の間の平均距離すなわち β -シート間の平均距離は、半径 R と回転角 ϕ を使って $2R\sin(\phi/2)$ と表わされる。計算結果は、この平均距離が立体構造が解かれたすべての LRR 蛋白質において一定であることを示した。

次に、LRR リピートのそれぞれの β -シートを 3 次元空間中に一つのベクトルで表した。このベクトル表示は、LRR において一つの面を形成し、この面の傾きは次第に変化することを明らかにした。この面はトポロジーの分野で有名なメビウスの帯で近似できることを示唆した。

二次構造 3_{10} -ヘリックスの構造解析： 蛋白質には α -helix、 3_{10} -helix、 π -helix の 3 種類のヘリックス構造が知られている。 3_{10} -helix は 3 残基で 1 回転するヘリックス構造である。この 3_{10} -helix は i 番目の残基のカルボニル酸素と $i + 3$ 番目の残基のアミドプロトンとの間に水素結合が繰り返して形成される。Ideal な 3_{10} -helix は、 -49° と -26° の 2 面体角 (ϕ , ψ) で表される。我々は立体構造が解明されているすべての蛋白質の 3_{10} -helix 部分を同定し、Helix fitting 法を適用した。1.6Å 以下の X 線構造解析 PDB ファイルから抽出した 689 個の蛋白質から検出した 3_{10} -helix 部分の α 炭素の原子座標を用いて、ヘリックスパラメーター(ヘリックスピッチ P 、1 回転当たりの平均残基数 n 、ヘリックス半径 r) を求めた。

まず、 3_{10} -helix の 2 面体角 (ϕ , ψ) を Ramachandran プロットした。5-7 残基からなる 3_{10} -helix において、各位置が C 末端側に近づくにつれ ϕ は増加し、逆に ψ は減少した。また、8 残基および 9 残基からなる 3_{10} -helix の場合、このようなシフトが 2 回繰り返されていた。

次に、我々は rmsd に基づいて、 3_{10} -helix を regular と irregular の 2 つのタイプに分類した。その結果、regular な 3_{10} -helix において、 P , n , r の全てが、ヘリックスの長さにかかわらず ideal な値 ($P=5.80\text{\AA}$, $n=3.0$, $r=1.89\text{\AA}$) よりも大きくなっていた。また、ヘリックスが 5 残基から 8 残基と長くなるにつれ、 P は増加した。逆に、 n と r は減少した。このことはヘリックスが長くなるにつれ、 3_{10} -helix は、細くて長くなることを示している。さらに、regular な 3_{10} -helix が存在する割合が、ヘリックスが長くなるにつれ減少した。そして、9 残基より長い 3_{10} -helix はすべて irregular を示した。平均構造を用いたエネルギー計算の結果から、regular な 3_{10} -helix がヘリックスが長くなるにつれ、不安定になることが示唆された。

以上の結果から、蛋白質中に存在する 3_{10} -helix は、無限に長い regular な helix は inherent な性質として不可能であることを示している。このような特徴をもつ 3_{10} -helix に対して、高分子結晶学の分野で知られているパラクリスタルとの類似性から、パラヘリックスと名づけた。

本研究において、3D circle fitting 法と Helix fitting 法を開発することを世界ではじめて成功した。こ

れら 2 つの方法は、蛋白質の立体構造の特徴およびその原理の解明に有効であることを明らかにした。これらの知見は蛋白質の立体構造予測法の開発やドラッグデザインの開発にも役立つものと期待している。

学位論文審査の要旨

主 査 教 授 大 崎 満

副 査 教 授 松 井 博 和

副 査 教 授 松 嶋 範 男 (札幌医科大学)

副 査 助 教 授 信 濃 卓 郎

学 位 論 文 題 名

Structural Principles of Proteins Revealed by Computer-aided Geometrical Analysis

(蛋白質の立体構造原理解明へのバイオインフォマテイクス研究)

本論文は6章からなり図20, 表6、総頁95からなる英文論文で、別に参考論文1編が付されている。

コンピュータを駆使した大量のゲノム配列や立体構造の情報解析から、生物の基本的な原理を解き明かすバイオインフォマテイクス研究が注目を集めている。本論文は蛋白質の立体構造の定量的幾何学的解析のための新しいプログラムを開発し、そのプログラムをロイシンリッチリピート(LRR)蛋白質と蛋白質中の 3_{10} -helixに適用したものである。

3D circle fitting 法と Helix fitting 法の開発 (第2章および第3章): まず、最小二乗法により3次元のデータ点から最適な円を求める方法(3D circle fitting 法)と最適なヘリックスを求める方法(Helix fitting 法)を開発している。3D circle fitting 法から、最少3個のデータ点から最適な円の半径 R と円の中心座標を得られる。また、Helix fitting 法からは、最少4個のデータ点から最適なヘリックスピッチ P とヘリックス半径を同時に精度よく得られる。この Helix fitting 法のためのアルゴリズムは、本研究において世界ではじめて独自に開発されたものである。

馬蹄形ロイシンリッチリピート(LRR)蛋白質の定量的幾何解析(第4章): LRR 蛋白質は、バクテリアからヒトまで4,000 以上の蛋白質中に見いだされており、その数は爆発的に増加している。植物LRR蛋白質においても2,000 個以上存在することが知られている。ほとんどの LRR 蛋白質は、蛋白質—リガンド相互作用に関与することが知られている。この相互作用をどうして植物の病原体抵抗性、動物の自然免疫、アポトーシス、細胞接着、遺伝子の組み替え、DNA 修復、転写などの様々な機能発現に関与している。

これまでLLRDメインを含む立体構造は、15 種類の LRR 蛋白質について原子座標レベルで解かれている。それらの LRR 蛋白質はすべて、馬蹄形をしていることが明らかにされた。Enkhbayar 氏は、馬蹄形が円の一部(円弧)とみなせることに気づき、3D circle fitting

法から円弧の半径 R と回転角 ψ を求めた。この回転角 ψ を使って、リピート単位あたりの平均の回転角 (ω) を計算した。

3 Dcircle fitting の解析結果から、 R において 16Å から 46Å に値を得た。一方、 ω は 6° から 17° の値を得た。RI は β - α モチーフ、internalin は β - 3_{10} モチーフ、yopM は β -ポリプロリン II (pII) モチーフ、Nogo receptor は β - β ターンモチーフをとる。したがって、これらのさまざまな R と ω の値は、それぞれの LRR 蛋白質の構造を特徴づけている。さらに R と ω の値から、 $2 R \sin(\phi/2)$ の量を計算し、この量がすべての LRR 蛋白質において一定であることをみいだした。

次に、LRR リピートのそれぞれの β -シートを 3 次元空間中に一つのベクトルで表す方法を開発している。このベクトル表示から、LRR において一つの面を形成し、この面の傾きは次第に変化することを明らかにした。この面はトポロジーの分野で有名なメビウスの帯で近似できることを示唆した。

以上の、本研究において行われた定量的幾何解析は、他に類をみないまったく独自のアプローチである。

二次構造 3_{10} -ヘリックスの構造解析 (第 5 章): 3_{10} -helix は 3 残基で 1 回転するヘリックス構造である。Ideal な 3_{10} -helix は、 $n=3.0$ 、 $P=5.8\text{\AA}$ $r=1.89\text{\AA}$ のヘリックススパラメータをとる。ところが、蛋白質中の 3_{10} -helix のコンフォメーションは、Ideal な 3_{10} -helix と異なることが報告されている。そこで、Enkhbayar 氏は、蛋白質中の 3_{10} -helix の精度のよいヘリックススパラメータを、本研究において開発した Helix fitting 法を用いても求めた。同時に、蛋白質中の 3_{10} -helix を regular と irregular の 2 つのタイプに分類する評価法も開発している。その結果、regular な 3_{10} -helix において、 P 、 n 、 r の全てが、ヘリックスの長さにかかわらず ideal な値よりも大きくなることを示した。さらに、ヘリックスが長くなるにつれ、 3_{10} -helix は、細くて長くなり、また、regular な 3_{10} -helix が存在する割合が、ヘリックスが長くなるにつれ減少することをみいだした。さらに、エネルギー計算から、regular な 3_{10} -helix がヘリックスが長くなるにつれ、不安定になることを示した。

以上の結果から、蛋白質中に存在する 3_{10} -helix は、無限に長い regular な helix は inherent な性質として不可能であることを明らかにした。蛋白質中の 3_{10} -helix を、高分子結晶学の分野で知られているパラクリスタルとの類似性から、パラヘリックスと名づけることを提案している。

本研究で開発した 3 Dcircle fitting 法、Helix fitting 法は、蛋白質の立体構造解析のための基本的なツールになることが期待される。また、この立体構造解析手法を LRR 蛋白質とリガンドとの相互作用のメカニズムの解明に適応することにより、植物の抵抗性遺伝子の機能発現の解明にも役立つものと期待される。よって審査員一同は、Purevjav Enkhbayar が博士 (農学) の学位を受けるに十分な資格を有するものと認めた。