

アミノ酸配列から膜タンパク質の判別と膜タンパク質データベースの開発

東京農工大学工学部 美宅 成樹

現在では微生物のみならず、多細胞生物である線虫のゲノム解析が完了し、2、3年の間にはヒトゲノムも解析されるという状況にある。そして、得られた大量の配列情報の意味（タンパク質の機能）を明らかにする技術が強く求められている。本計画では、全プロテオームの20～35%を占める、機能的にも重要な膜タンパク質に焦点を当て、その構造・機能の情報を抽出する技術を開発することである。このために私たちはプロテオーム情報解析システムを開発した（図1）。本研究の大きな目標は、全てのアミノ酸配列に対して、タンパク質の完全な立体構造と機能を与えるようなシステムを開発することであるが、限られた研究期間内に次の事を行った。

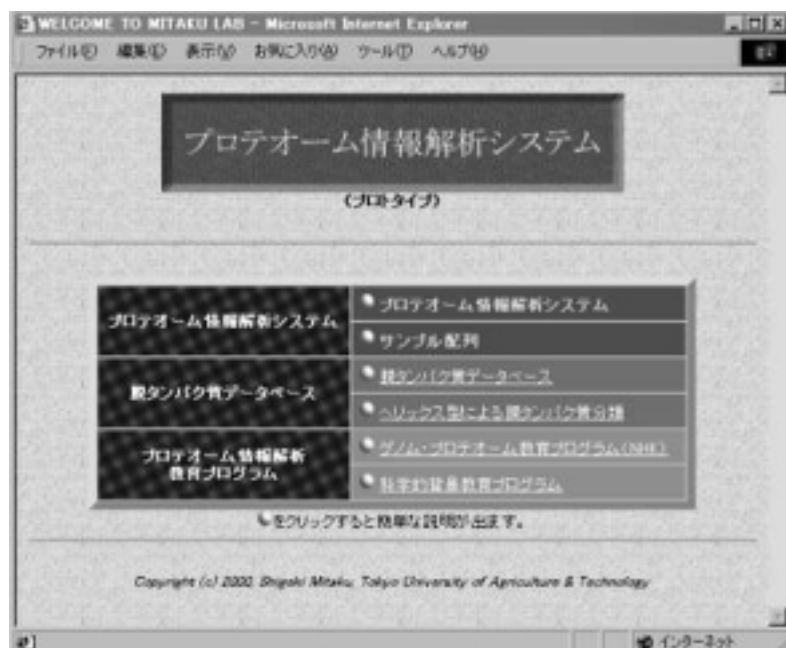


図1 プロテオーム情報解析システムのトップページ

- (1) アミノ酸配列から膜タンパク質かどうかを判別するシステムを作り、多数のアミノ酸配列に対して同時に解析できるようにした。これによって、例えばある生物種の持つ全てのアミノ酸配列の解析を一挙に行うことができるようになった。さらに、生物種の持つ性質として膜タンパク質の割合を提示することができる。タンパク質の判別には、アミノ酸配列からセグメントの平均疎水性インデックス、その両端に当たる領域の平均界面活性インデックス、およびタンパク質全体の残基数を計算する。それらのパラメータによって、もっとも疎水性の高いセグメントの性質から、それが膜に分配されるか、あるいは水溶性タンパク質の疎水性コアに分配されるかを予測し、膜タンパク質の判別を行なう。600個のタンパク質（そのうち約20%は膜タンパク質）を用いて評価した結果、膜タンパク質判別の精度は99%であった。
- (2) 膜タンパク質だけを扱うデータベースはこれまでもあったが、ヘリックスのタイプを分類した膜タンパク質のデータベースはなかった。これを可能にするために、SWISSPROTのデータを解析し、膜タンパク質の分類を行ない、データベース化した。膜タンパク質は一般に膜を貫通するヘリックスを持っており、膜タンパク質の性質や機能は膜貫通ヘリックスの本数によって特徴付けられる。そこでまず膜タンパク質を膜貫通ヘリックスで分類し、データの検索を行えるようにした。また、本研究の特徴としては、膜貫通ヘリックスを非常に疎水性の高い一次膜貫通ヘリックスとあまり疎水性が高くない二次膜貫通ヘリックスに分類できる。そこで、それぞれの本数によってさらに分類し、データの検索を行えるようにした。これによって従来にない新しい膜タンパク質データベースとなった。

- (3) アミノ酸配列の解析は様々な側面があり、ホモロジーの解析は良く知られているが、本研究のような物理化学的考えに基づくアプローチは少ない。したがって、単にソフトウェアを提供するだけでは本当の背景までは理解しにくい。そこで、教育プログラムが重要である。ここでは、ゲノム・プロテオーム教育プログラムと科学的背景教育プログラムとをシステムの中を含めた。前者は1999年にNHKが製作したプログラムの中で、FAQの部分を本研究代表者がまとめたが、これをリンクした。また後者は、やはり代表者が中心として、まとめた科学背景プログラムものがあるのでこれを含めた。
- (4) ソフトウェアシステムもユーザーがあって始めて意味がある。私たちのソフトウェアシステムは、分子生物学者が実際にアミノ酸配列を解析して、その結果を自分の仕事に反映して始めて意味があるわけである。そこで、このシステムを世界に情報発信できるようにするためにホームページ化した。そのトップページが図1である。ただこのホームページはまだ十分な機能を盛り込んでないので、実際の運用にはもう少し研究開発を進めたものを用いる予定である。

さて、本研究はプロテオームから得られるアミノ酸配列を情報処理するためのシステムを開発することであるが、このシステムを用いてどのような意味のある結果が得られるかということについても簡単に触れておきたいと思う。すでに多くの生物種について全ゲノム解析が行われているので、私たちは16種類の生物種について解析を行った。11種類の真性細菌、3種類の古細菌、2種類の真核生物（単細胞の酵母と多細胞の線虫）である。その結果、単細胞生物は大きな生物の分類に関わらずおよそ25%の膜タンパク質を持っていることが分かった。また、多細胞生物の例は少ないが、単細胞生物よりもはるかに膜タンパク質の割合が多く、約35%であった。その意味については、まだ確証的なことは言えないが、一つの可能性としては細胞間コミュニケーションのために膜タンパク質が必要とされているということが考えられる。このことは本研究のようなかなり単機能のシステムであっても高精度のシステムを作れば、分子生物学的に意味のある情報を得られる可能性が高いことを示している。

本研究には、共同研究者の園山正史以外にも、広川貴次、謝文清、赤沢史嗣、高江州宏智、小野満、業務委託の菱化システムの中川敬一、谷川恒夫の寄与が大きいことを指摘して、深く感謝したいと思う。

これは平成12年3月9日に開催した
計算科学技術活用型特定研究開発推進事業
研究報告会（主催 科学技術振興事業団）
の予稿集から抜粋したものです。