

氏 名	木 原 大 亮
学位(専攻分野)	博 士 (理 学)
学 位 記 番 号	理 博 第 2107 号
学位授与の日付	平 成 11 年 5 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 生 物 科 学 専 攻
学位論文題目	膜タンパク質の予測法の開発とゲノムにおける分布の解析

論文調査委員	(主 査) 教 授 金 久 實	教 授 藤 吉 好 則	教 授 平 野 丈 夫
--------	--------------------	-------------	-------------

論 文 内 容 の 要 旨

近年決定されている数多くの生物種の完全ゲノム配列には、生命現象の原理を探る情報が含まれていると考えられる。これを明らかにするには、ゲノムに含まれる遺伝子領域を予測し、配列と並びから遺伝子の機能を予測する方法の開発が必須である。本論文では、輸送やシグナル伝達など生体内で必須の様々な機能を担う膜タンパク質に焦点を絞り、まずアミノ酸配列から膜タンパク質を予測する方法の開発を行った。さらにこれを用いてゲノムに含まれる膜タンパク質を同定し、その分布の解析を行っている。

論文の前半において、申請者が新規に構築した、タンパク質の膜貫通部位の予測法について報告している。膜タンパク質の膜貫通部位には、貫通本数や順番によって疎水性の強さや周期性に特徴があることが知られている。このため、まずデータベース中の膜タンパク質の膜貫通部位をいくつかのグループに分類し、各グループの特徴を複数の判別関数として考慮した膜貫通部位予測法を構築した。この予測法の大きな特徴は、ある膜貫通本数をもつ膜タンパク質を、異なる特徴をもった膜貫通部位グループの並びとしてパターン化したことであり、これを膜タンパク質のモデルと呼んでいる。膜タンパク質でないモデル、すなわち球状タンパク質のモデルも陽に考慮されている。これは球状タンパク質の立体構造予測法である 3D-1D 法の概念を取り入れたもので、ある配列に対して最も適合したモデルを探すことが予測である。さらにこの予測法のもうひとつの特徴は、スコアによって各モデルの結果をランクづけして、複数の可能性を示し得ることである。予測精度の評価は、トポロジーが実験的に確認されている膜タンパク質を文献から収集し、これをテストセットとして行った。用いたパラメータによる予測率の影響、判別関数の複数化の効果、スコアの信頼度などについても議論されている。予測率はタンパク質単位、膜貫通部位単位、残基単位の 3 つの正解率で評価されており、この方法はタンパク質単位の予測率が他の手法と比較し優れていることが示されている。膜貫通部位の本数はそのタンパク質の機能と関係がある場合が多いので、申請者が開発した膜タンパク質の予測法は機能の推定に役立つことが期待される。

論文の後半では、この予測法を用いて実際に完全ゲノム配列が決定されている古細菌、真正細菌、真核生物あわせて 16 種の生物種のゲノムに含まれる膜タンパク質の予測と分布の解析を行っている。その結果、まず全 ORF の数が多くなるほど、すなわちゲノムが大きくなるほど、膜タンパク質の割合が高くなることが示された。一般に環境適応能力が高い生物種は遺伝子の数が多くゲノムが大きい傾向があるが、ここでは相対的にさらに多くの膜タンパク質が細胞膜に存在することが示唆されたわけである。さらに最も興味深い結果として、各ゲノム中、約半数の膜タンパク質の遺伝子が隣接して存在すること、すなわちタンデムクラスターになっていることが示された。これは統計的に有意な数である。生物種間で保存されたクラスターのオーソログ、同一生物種内で重複したクラスターのパラログのうち、機能が明らかなものは全体の 15% に過ぎなかった。これらのクラスターの機能は、輸送系、電子伝達系、細胞の運動に関わるものが大半である。一方、個々のクラスターの内部での遺伝子重複をみると、各生物種とも 10～20% のクラスターには、ホモロジーの高い膜タンパク質のペアが含まれていた。これら膜タンパク質の膜貫通部位の本数とクラスターのサイズは、その機能と関連していることが明らかになり、これまで機能未知の 14 個の膜タンパク質の機能推定を行っている。さらに、膜タンパク質と ATP 結合タンパク質の

ペアのクラスターをゲノム中で探索し、生物種間で保存されるそのようなクラスターは輸送系、その中でも ABC トランスポーターが多いことが示されている。ABC トランスポーターの遺伝子はオペロンを形成している場合が多いが、孤立した遺伝子はどの生物種でも 10% 以上存在すること、*A. aeolicus* と *Synechocystis* では 40% も存在すること、膜タンパク質同士のペアの方が膜タンパク質と ATP 結合タンパク質のペアよりオペロン内に存在する傾向が強いこと、すなわちクラスターとしての結びつきが強いことが明らかになった。最後に膜タンパク質のタンデムクラスター形成のメカニズムが考察されている。

論文審査の結果の要旨

ゲノム解析計画の進展に伴い、微生物から高等生物まで各生物種の全ゲノム配列データが蓄積されつつあるが、それらにコードされている 1/3 から半分の遺伝子の機能は未知のままである。機能推定のための最も有効な方法は、既存のデータベースに対するホモロジー検索である。しかし、ホモロジーが見つからなければ、この方法では機能の手がかりは得られない。一方、膜タンパク質は膜貫通部位の疎水性の強さなどの特徴によって、配列データだけからかなりよい精度で予測することができる。また、膜貫通部位の本数はその膜タンパク質の機能と関係しているため、予測によって機能推定の手がかりとなることが期待できる。本論文はこのような背景から、膜タンパク質の予測法の開発とこれを用いてゲノム中に含まれる膜タンパク質の分布の特徴を解析したものである。

タンパク質の膜貫通部位予測法は、膜タンパク質の立体構造のデータ数が少ないこともあって、球状タンパク質の構造予測法と比較して、Kyte-Doolittle によるハイドロパシープロット法以来、概念的に新しい方法論の発展はあまりなかったと言ってよい。これに対し申請者が開発した方法は、球状タンパク質の立体構造予測において近年成功を修めている 3D-1D 法 の概念を取り入れていることが、独創的で最も評価できる点である。すなわち、貫通本数が異なる膜タンパク質を 1 つのモデルと考え、与えられた配列に対して最も適合性の高いモデルを、予測された膜タンパク質と考える点である。別の観点からこれをみてみると、膜タンパク質の貫通部位の性質は一樣ではなく、膜タンパク質が全部でいくつの貫通部位をもつか、その中の何番目の部位であるかによって、疎水性やその周期性に違いがある。このような生物学的知見を予測法に積極的に取り入れている点が新しい。具体的には貫通部位を 5 つのタイプ分け、その組み合わせで、膜タンパク質のモデルを定義している。本方法の開発により、貫通本数を含めて膜タンパク質の貫通部位を予測することが可能となった。これまでの方法は貫通部位を独立の単位として予測精度の向上が行われており、タンパク質全体の予測という観点は必ずしも明確ではなかった。実際、申請者が開発した方法の予測結果は従来の予測法と比較して、タンパク質単位の予測率がよいことが示されている。よって、本方法は遺伝子の機能推定への手がかりを与えるという、当初の目的を達成していると考えられる。

この方法を用いた解析の部分では、古細菌から 4 種、真正細菌から 11 種、真核生物から 1 種 (*S. cerevisiae*) の全ゲノム配列を用いている。その結果、各ゲノムに含まれる膜タンパク質の約半数が相互に隣接して、つまりタンデムクラスターとして分布しているという、大変興味深い特徴が見いだされた。クラスター形成の理由としてオペロンなどの機能的な共役、膜輸送における有利さなどが挙げられているが、膜輸送における有利性は真核生物の場合に理由としてあてはまらない。

今後、他の真核生物のゲノムを対象とすることなどによって更なる解析が期待される部分である。膜タンパク質同定の際にはシグナル配列の除去、アミノ酸組成に偏りがある領域をマスキングしてのホモロジー検索、手作業によるチェックなど注意深い手順がとられており、申請者の研究態度がうかがえる。機能予測に関しては、予測される膜貫通部位の本数とクラスターのサイズがその機能と関連している例を示し、これに基づいた膜タンパク質の機能予測を行っている。これは、ゲノムに含まれる遺伝子の機能の推定においてホモロジー以外に、構造とゲノム上の配置の情報を取り入れたことであり、ゲノム配列の機能解析に、一つの有効な手段を提供したと言えるであろう。更に本論文で行われている、膜タンパク質と ATP 結合タンパク質のペアのクラスターの探索も、遺伝子間の機能上の関連を解析する一手段として今後様々な応用が期待できる。

以上申請者が行った膜タンパク質の予測法の開発とゲノムにおける分布の解析は、基礎的な研究成果としても実用的な応用価値としても高く評価できる。よって、本論文は博士(理学)の学位論文として価値あるものと認める。なお、主論文および参考論文に報告されている研究業績に関連した試問の結果、合格と認めた。