

機関番号：13901

研究種目：基盤研究(B)

研究期間：2008～2010

課題番号：20370058

研究課題名(和文) タンパク質のファミリー分布に注目したタンパク質ワールドの研究

研究課題名(英文) PROTEIN WORLD STUDIED BY PROTEIN DISTRIBUTION IN FAMILIES

研究代表者：

美宅 成樹 (MITAKU SHIGEKI)

名古屋大学・工学研究科・教授

研究者番号：10107542

研究成果の概要(和文)：物理的なパラメータによる高精度タンパク質分類法を用いて、ゲノムからの全タンパク質の高精度分類を行い、タンパク質ワールドというコンセプトが進化のプロセスで実際に成立しているかどうかを検討した。真核生物のゲノムでは、膜貫通領域を含むエキソンが体系的にシャプリングを受けていることが分かった。また原核生物では、突然変異だけで一定割合の膜タンパク質を形成できることが分かった。現実の生物ゲノムはエキソンシャプリングや突然変異などのランダムプロセスの平衡状態にあることが示された。

研究成果の概要(英文)：The validity of the concept of protein world was investigated by analyzing all amino acid sequences from the total genomes of many organisms, using a high performance membrane protein predictor. The analyses of the distribution of exons containing membrane spanning regions lead to the conclusion that the exon shuffling extensively occurred in eukaryotic genomes. The number distribution of transmembrane regions in membrane proteins was analyzed for real genomes and simulated genomes of prokaryotic genomes. The result indicated that the single exponential distribution is due to the equilibrium of the change in the transmembrane numbers. The results for both eukaryotes and prokaryotes suggested that the families of membrane proteins have already reached to the equilibrium of random processes in genomes.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2008年度	6,700,000	2,010,000	8,710,000
2009年度	4,600,000	1,380,000	5,980,000
2010年度	4,600,000	1,380,000	5,980,000
総計	15,900,000	4,770,000	20,670,000

研究分野：生物物理

科研費の分科・細目：生物科学・生物物理

キーワード：理論生物学、バイオインフォマティクス

## 1. 研究開始当初の背景

タンパク質ワールドというコンセプトがある。20世紀の末に大量のゲノム情報が報告されるようになり、それから得られるアミノ酸配列にアノテーションが付けられてきた。そこから興味深い結果が得られている。一つは、既知配列との類似性により未知配列のアノテーションを付けていくと、ほぼ半分のアミノ酸配列について意味を明らかにす

ることができるが、残りの半分は意味を推定することができない。もう一つは、意味が推定されているタンパク質のファミリー分類をしてみると、その数に奇妙な関係があるということである。ファミリーの間にタンパク質の変換可能性があると考えると理解しやすいような関係にあり、それに基づいてタンパク質は全てお互いに変換可能なタンパク質ワールドのメンバーとなっているという

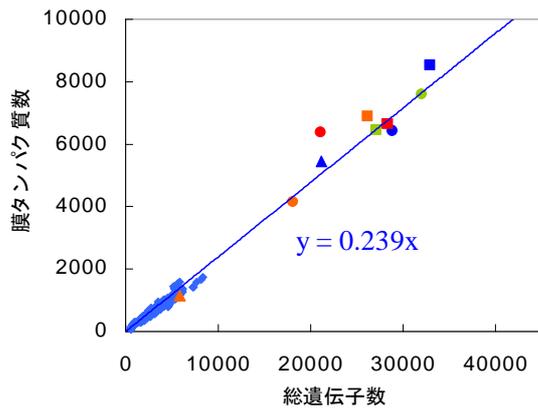


図 1. 各ゲノムについての総遺伝子数に対する膜タンパク質数の関係

コンセプトが提出されたのである。しかし、これまでの解析はすべて文字配列の類似性を基本としており、約半分のタンパク質について意味が分からない状態で、議論が進められてきたという弱点がある。この問題に対してははっきりした結論を出すには、全てのタンパク質（アミノ酸配列）に対して高精度の分類・予測できるシステムを開発し、全ゲノムからの解析を行う必要がある。研究を開始する時点で、世界的に以上のような状況であった。

我々の研究としては、膜タンパク質を中心にタンパク質ワールドのコンセプトにより立ち入った先行研究を行っていた。膜タンパク質予測システム SOSUI を既に開発しているので、これを用いて 300 近い生物ゲノムからの全アミノ酸配列に適用した。その結果、現実のゲノムの解析を行い、ゲノム中での膜タンパク質の数を各生物種で調べ、膜タンパク質の割合がほとんど一定だという結果が得られたのである（図 1）。これは進化の過程を考える上で、非常に強い制約となる。その制約がどのようなメカニズムで起こっているかを明らかにするために、進化に対して新しい考え方を持ち込むことが必要となった。そして、タンパク質ワールドというコンセプトが図 1 の結果を説明するのに有効であることに気が付いた。図 2 は大きなタンパク質の集団としての水溶性タンパク質と膜タンパク質の間で変換が可能であると考えた時の反応式である。簡単な考察から分かることであるが、もし反応が平衡状態に到達していたら、2つの状態（この場合水溶性タンパク質と膜タンパク質）の比は、反応の速度定数比で一定となる。これはほとんどすべての生物ゲノムの中で膜タンパク質の割合が一定であるということの説明となりえる。言い換えれば、タンパク質ワールドにおいて、水溶性タンパク質の集団と膜タンパク質の集団が突然変異を通して深く関係づけられることを示唆

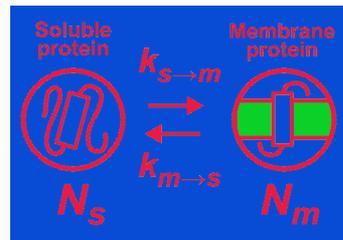


図 2. 配列の変異に対する膜タンパク質形成反応式

している。このような結果を得て、本研究を始めたのである。

## 2. 研究の目的

我々は、アミノ酸配列の持つ物理化学的なパラメータの分布を用いて、高精度のタンパク質分類法を開発してきた。このアプローチでは、未知配列も含めて高精度分類ができるので、全タンパク質のファミリー分類が可能となる。そこで本研究では、物理化学的なパラメータによる高精度タンパク質分類法を用いて、ゲノムからの全タンパク質の高精度分類を行い、タンパク質ワールドのメンバーであるファミリーの変化を調べることを目的とする。

実際に遺伝子からアミノ酸配列が作られるプロセスが真核生物と原核生物で違っている（真核生物ではスプライシングというプロセスがある）、それに対応して 2 つの研究を計画した。

真核生物ではスプライシングに対応して遺伝子領域にエキソン-イントロン構造が存在している。そして、膜貫通領域を含むエキソンが生成消滅することによって膜貫通領域の数が変わり、膜タンパク質のファミリー間の変換が起こる可能性がある。そこで遺伝子を構成するエキソンの数に対して遺伝子数がどのような分布を示すか、また膜貫通領域を含むエキソンの数に対して遺伝子数がどのような分布を示すかを調べる。

他方、原核生物ではスプライシングのプロセスがないので、配列の変化は主に突然変異によるものとなる。そこで DNA の塩基配列に対して突然変異を導入するシミュレーションを行い、膜タンパク質の割合がどのように変化するかを調べる。このように多くの生物種についてのゲノム比較を行うことで、タンパク質ワールドについてのより立ち入った議論を行うことができるようになる。

## 3. 研究の方法

本研究では、大量に蓄積されているゲノム情報に対して、物理化学的なパラメータによる高精度タンパク質分類システムを用意し、生物種による違いを比較する。そして、進化の

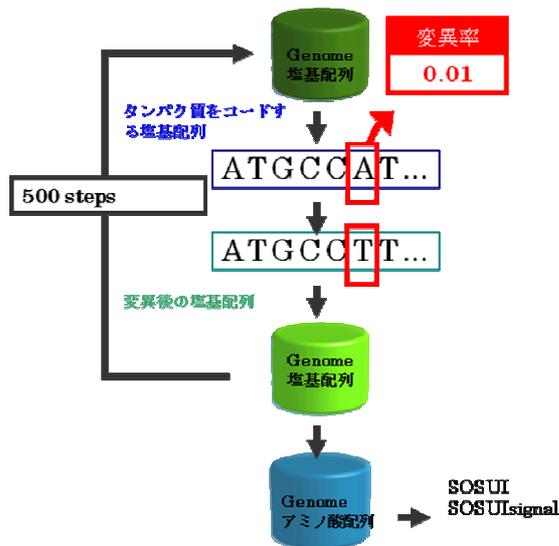


図3. 配列シミュレーションの流れ図

問題について、タンパク質ワールドという新しい視点から検討する。高精度予測システムは、すでにいくつか開発しているが、実際にタンパク質ワールドの研究に用いたシステムは、膜タンパク質予測システムである。このシステムは、アミノ酸の疎水性と両親媒性のインデックスを用いて膜貫通領域を予測するもので、任意のアミノ酸配列に対して適用可能である。図3は現実のゲノムに対して、DNA塩基配列に突然変異を導入する配列シミュレーションの流れ図を示したものである。シミュレーションは完全にランダムになるまで行うが、突然変異の条件を色々と変更を行ってみた。GC含量だけを固定した場合、コドンの位置によるバイアスを考慮した場合などである。

#### 4. 研究成果

多細胞真核生物の遺伝子は、イントロンによってコーディング領域が分断されていて、スプライシングによって最終的なRNA配列が作られる。その時にエクソンの中に一つでも膜貫通領域を含むものがあれば、最終的にその遺伝子産物は膜タンパク質になる。そこで膜貫通領域をマーカーとしてタンパク質のファミリーを容易に分類することができるのである。スプライシングは、一般に配列の中にイントロンが挿入されたもので、エクソンが位置を変えたり、生成、消滅をあまりしないと考えられてきている。しかし、タンパク質ワールドのコンセプトが正しいとすれば、スプライシングのプロセスでエクソンのシャフリングが起こり、アミノ酸配列が形成されているという可能性もある。さらにエクソンシャフリングの平衡状態がゲノムを比較することで見えるかもしれない。そこで、

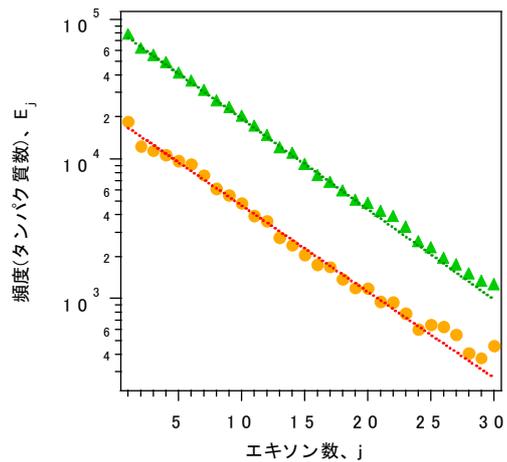


図4. 真核生物ゲノムの遺伝子における遺伝子あたりのエクソン数の分布 (▲、全遺伝子；●、膜タンパク質)

我々は膜貫通領域を遺伝子中のマーカーとみなして、分布を解析してみた。図4は、多細胞真核生物のゲノムから遺伝子領域を取り、各遺伝子が何個のエクソンからなっているかを数え、そのエクソン数に対する分布をプロットしたのが図4である。非常にきれいな指数分布となっている。一般に指数分布は素過程がランダムプロセスの場合に出てくる分布で、この場合もエクソンの数が変化するプロセスにランダム性があるということが示唆される。つまり、エクソンの生成消滅が進化的時間の中でランダムに起こっていると考えると、この非常にきれいな指数分布は説明できる。しかし、エクソンの位置の変化を含むエクソンシャフリングが起こっているかどうかをこのグラフから結論することはできない。そこで、膜貫通領域がタンパク質のどの位置に出現しているかということ、原核生物、初期の真核生物、高等真核生物で比較してみた(図5)。膜貫通領域が顕著に違うことが分かる。つまり、エクソンシャフリングが起こっていると考えないと理解できない結果が得られた。結論として、進化的時間スケールでは、真核生物はエクソンシャフリングによる遺伝子の変化を受けており、すでに平衡状態に達していると考えられる。

次に、原核生物における膜タンパク質のファミリー分布の変化を調べてみた。ここでは膜貫通領域の本数分布がどのように変化するか注目した。原核生物の場合、スプライシングがないので、主な配列変化の原因は突然変異である。まず現実のゲノムからの全アミノ酸配列を膜タンパク質予測システムSOSUIで解析し、膜貫通領域本数分布を調べると、1本型が最も多く、本数が増えるにつれてタンパク質の頻度が減少する。しかし、4本あたりからプラトーとなり、12本付近

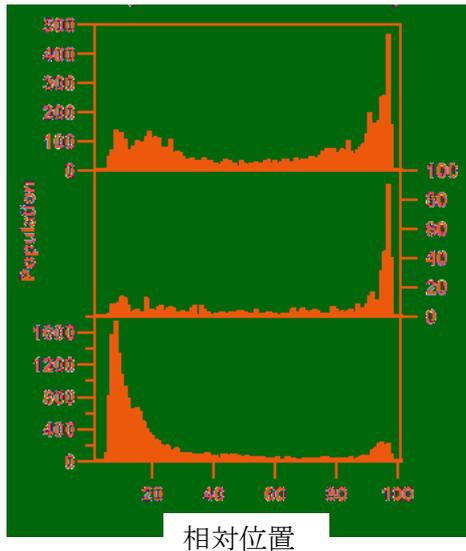


図 5. 高等真核生物（上）、初期の真核生物（中）、原核生物（下）における一本型膜タンパク質の膜貫通領域の位置分布（大きさが200~300残基のタンパク質）

から再び急激に減少する。この分布は詳細に調べれば生物種によって少し異なるが、平均的には同じ傾向を示している。図6にある本数分布の最も肩が突き出したグラフが、ここで述べた実際のゲノムに対する解析結果である。

これには2つの因子が関与しているはずである。一つは、突然変異によってランダム化した配列でもある割合で膜タンパク質ができるという因子である。また、何らかの自然選択の影響があつてそれが加わって本数分布が決まっているはずである。図6は、アミノ酸組成を実際の値に固定して、配列をランダム化した時に得られる膜貫通領域本数分布の変化である。最終的に、本数分布が指数分布となっていることが分かる。現実の膜貫通領域の本数分布は、ランダムプロセス（突然変異）による指数分布であり、12本

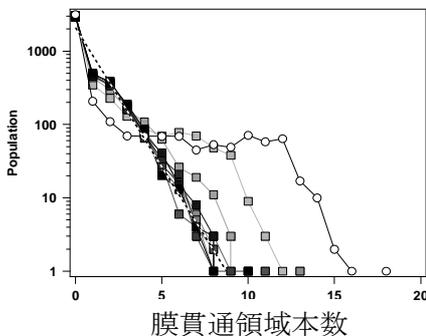


図 6. 膜貫通領域の本数分布。配列にランダムな突然変異を導入したシミュレーションで、指数分布に近づいた

型まで伸びている分布の肩は自然選択によるものと考えられる。

本研究によって、タンパク質ワールドの中で少なくとも膜タンパク質集団は素過程としてのエキソンシャプニングや突然変異などによる変化の平衡状態となっていることが強く示唆された。

## 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計10件）

- ① R. Sawada & S. Mitaku, How are exons encoding transmembrane sequences distributed in the exon-intron structure of genes?, *Genes to Cells*, 査読有, **16(1)**, 2011, 151-121.
- ② Y. Yokoyama, L. Negishi, T. Kitoh, M. Sonoyama, Y. Asami & S. Mitaku, Effect of lipid phase transition on molecular assembly and structural stability of bacteriorhodopsin reconstituted into phosphatidylcholine liposomes with different acyl-chain lengths, *J. Phys. Chem. B*, 査読有, **114(47)**, 2010, 15706-15711.
- ③ T. Tsuji, F. Akazawa, R. Sawada & S. Mitaku, SOSUImp1: high performance prediction system for single-spanning membrane proteins, *Chem-Bio Informatics J.*, 査読有, **10**, 2010, 61-73.
- ④ Y. Yokoyama, M. Sonoyama & S. Mitaku, Structural changes in bacteriorhodopsin in purple membranes induced by irreversible photobleaching with heterogeneous and homogeneous stability, *Photochem. Photobiol.*, 査読有, **86(2)**, 2010, 297-301.
- ⑤ N. Asakawa, N. Sakiyama, R. Teshima & S. Mitaku, Characteristic amino acid distribution around segments unique to allergens, *J. Biochem.*, 査読有, **147**, 2010, 127-133.
- ⑥ M. Sonoyama, T. Kikukawa, Y. Yokoyama, M. Demura, N. Kamo & S. Mitaku, Effect of molecular assembly on photocycle of reconstituted bacteriorhodopsin: significant blue shift of the late M intermediate in the liquid crystalline phase, *Chem. Lett.*, 査読有, **38**, 2009, 1134-1135.
- ⑦ H. Tanizawa, M. Taniguchi, G.D. Ghimire & S. Mitaku, Prediction of fragile points of coiled coils, *Chem-Bio Informatics J.*, 査読有, **38**, 2009, 12-29.

- ⑧ K. Imai, N. Asakawa, T. Tsuji, F. Akazawa, A. Ino, M. Sonoyama & S. Mitaku, SOSUI-GramN: high performance prediction for sub-cellular localization of proteins in Gram-negative bacteria, *Bioinformatics*, 査読有, **2(9)**, 2008, 417-421.
- ⑨ H. Tanizawa, G.D. Ghimire & S. Mitaku, A high performance prediction system of coiled coil domains containing heptad breaks: SOSUIcoil, *Chem-Bio Informatics J.*, 査読有, **8**, 2008, 96-111.
- ⑩ G.D. Ghimire, H. Tanizawa, M. Sonoyama & S. Mitaku, Physico-chemical properties of GPCR amino acid sequences for understanding GPCR-G-protein coupling, *Chem-Bio Informatics J.*, 査読有, **8**, 2008, 49-57.

[学会発表] (計 37 件)

- ① 美宅成樹、BMB2010 ワークショップ「膜タンパク質の構造・機能から見るオルガネラの進化—膜貫通分布領域からみたオルガネラの進化—、第 33 回日本分子生物学会年会・第 83 回日本生化学会大会、2010.12.7, 神戸ポートアイランド 他国内学会 24 件
- ② R. Sawada, R. Ke, T. Tsuji, M. Sonoyama & S. Mitaku, Numbers of membrane proteins coded in 248 prokaryote genomes, 53<sup>th</sup> Annual meeting of Biophysical Society, 2009.2.28, Boston Convention and Exhibition Center, USA 他国際会議 11 件

[図書] (計 1 件)

- ① R.A. ガイル・B. ガステル 著 美宅成樹 訳、丸善株式会社、世界に通じる科学英語論文の書き方 執筆・投稿・査読・発表、2010, 322.

[その他]

ホームページ等

[http://bp.nuap.nagoya-u.ac.jp/sosui/sosuiogramn/sosuiogramn\\_submit.html](http://bp.nuap.nagoya-u.ac.jp/sosui/sosuiogramn/sosuiogramn_submit.html)

## 6. 研究組織

### (1) 研究代表者

美宅 成樹 (MITAKU SHIGEKI)  
名古屋大学・大学院工学研究科・教授  
研究者番号：10107542

### (2) 研究分担者 なし

### (3) 連携研究者

横山 泰範 (YOKOYAMA YASUNORI)  
名古屋大学・大学院工学研究科・助教  
研究者番号：80402486