

ゲノムワイドな構造・機能分類による膜蛋白質の機能理解 ：G蛋白質共役型受容体

●諏訪牧子 ◆広川貴次

産業技術総合研究所生命情報科学研究センター

＜研究の目的と進め方＞

膜蛋白質は、ゲノム全体で見ると、ORFの20%～30%を占め、細胞活動の主要な機能を担うことから、網羅的な機能解明が特に急がれている。機能及びその発現メカニズムを本質的に理解するためには立体構造情報が欠かせないが、結晶化等の実験的困難性などにより、実験的に立体構造が解明された膜蛋白質は極めて少なく（数十個）、本来上記のような要求に答えるには、未だ大きなギャップがある。

本提案では、ゲノムワイドに配列情報のみから、膜蛋白質の立体構造情報、機能情報を抽出して分類することで機能発現のメカニズムを理解することを目的とする。この目的で、具体的に行うことは（1）各生物種ゲノムにおいて膜蛋白質を判別し、比較ゲノム的な観点から機能を考察する。（2）配列情報から構造の違いを識別する認識法を確立し、ゲノム中での構造型の個数を明らかにする。（3）構造の観点からの分類情報と機能との相関性を体系化して整理し、機能のバリエーションも含めた分類にする。（4）分類後の代表配列に対し、膜蛋白質に特化した精度の良い立体構造モデリング法を確立する。（5）以上の情報を含む総合データベースを作成する。

もちろん本研究の最終的な目的は、全ての膜蛋白質に対して構造、機能情報を与えることにあるが、限られた3年間ではモデルケースとしてGタンパク質共役型受容体（GPCR）ファミリーに焦点を絞る。GPCRは免疫系、神経系、循環器系など多様な重篤な疾病原因となる膜タンパク質で、機能解明は極めて急がれている。蛋白質は、ゲノム全体で見ると、ORFの20%～30%を占め細胞活動の主要な機能を担うことから、網羅的な機能解明が特に急がれている。

＜研究開始時の研究計画＞

（A）構造、機能分類

①GPCR遺伝子の網羅的収集、比較ゲノム解析：GPCR遺伝子予測システムを使い、ゲノム配列が決定したすべての生物種に対して、GPCR候補を予測し、比較解析する。網羅的に収集した配列は、GPCRデータベース（SEVENS：<http://sevens.cbrc.jp>）に収める。

②構造、機能分類：極性表面による立体構造認識プログラム（13年度までに開発済み）を整備し、上記で予測した全てのGPCR遺伝子（機能未知の遺伝子まで含める）を構造情報（蛋白質内部の極性表面）で分類する。

次に構造分類した配列メンバーに対して機能分類を行う。具体的には、GPCRの機能（結合リガンド種、G蛋白質種）に物理量（膜貫通ヘリックス部分、ループ部分、リガンド、結合Gタンパク質等の物理化学的特徴量）を定量的に関係付け、その関係式から機能分類する。この機能分類法をより整備し、配列から機能を予測するプログラムを作成する。

（B）立体構造モデリング

まず、膜貫通ヘリックスに特化したアラインメント法、ループ部分の構造予測法の確立を目指す。具体的には、膜

内環境で、ヘリックス間の極性相互作用部分が保存されるよう制限がかかるようなアラインメント法を開発する。

このアラインメント法を基に、GPCRに特化した精度の良い立体構造モデリングを行う。（A）で機能分類した配列から、特にClass Aファミリーに属する配列を選択する。これらは、立体構造既知のロドプシンが含まれるため、ロドプシンの構造をテンプレートにして、比較モデリングを行う。また、立体構造既知配列のホモログで無い場合（Class Aファミリー以外）についても立体構造モデリングの方法を確立する。ヘリックス間相互作用と格子モデルを用いた方法を用いる。予測した膜貫通ヘリックス部分とループ部分を結合させ、分子動力学法により、全体構造を精密化する。また、膜タンパク質立体構造を用いて膜タンパク質専用の環境プロファイルを作成し、モデリング構造の評価に用いる。（A）、（B）の研究を相互にフィードバックし、機能のバリエーションを生ずるメカニズムを理解する。

（C）総合的データベース作成：以上の結果をデータベースとしてまとめる。データベースは、分類、予測した構造・機能情報に加え、既知の立体構造、機能情報にもリンクした総合的なものにする。

＜研究期間の成果＞

構造機能分類（発表論文1,2）膜貫通ヘリックス部分の極性領域の擬似座標、ループ部分の長さ平均疎水性値、既知の低分子リガンド、ペプチド性リガンドの分子量、疎水性値、電荷、などのGPCRの各ドメインの物理化学的パラメータと、結合Gタンパク質の種類（Gi/o, Gq/11, G12/13, Gs, Gt等）とは、有意な相関性があり、これを用いて結合G蛋白質のタイプを判別が可能であると示唆された。これらのパラメータは、Gタンパク質と直接には接触していない細胞外ループの部分の特徴量も含まれており興味深い1）。パラメータは多次元的にG蛋白質種と関連していることから、これらを特徴ベクトルとして、機械学習ツール（SVM）により、結合Gタンパク質を予測するプログラムを作成した。結合G蛋白質既知のGPCRに対して、予測精度は85%以上になっている2）。これを用いてヒトの機能未知GPCR、数百配列に対して機能予測を行ったところ、約3分の2がGs結合タイプであることが判った。

GPCRの比較ゲノム解析（発表論文3）：これまで開発済みのGPCR遺伝子予測自動化システム（発表論文3,5）を用いて、223種の細菌と、7種の真核生物のGPCR遺伝子を同定した結果、細菌類ではGPCRは見出せなかったが、多細胞生物からGPCR数が増加し、線虫、ショウジョウバエ、高等生物あたりから急増することが判った。また、これらGPCRをゲノム上に貼り付け、解析したところ、特定の染色体で遺伝子重複により、増大していることが示唆された（ヒト11番染色体、線虫V番染色体など）。特にヒトゲノムの嗅覚・味覚受容体など感覚器の受容体は、ゲノム配列上で、狭い領域に高密度に集積したクラスター領域を形成することが判った。一方、Science誌で、

遺伝子クラスターの上流の制御領域が、クラスター内の1つの遺伝子を1種類の嗅覚細胞に発現させるメカニズムが報告されていることから、ゲノム上の遺伝子分布構造からも機能が予測できる可能性を示された。上記のデータは<http://sevens.cbrc.jp>により公開している。

構造モデリング手法確率のための基礎研究(発表論文4)：クラスAファミリーに属するGPCRに関して、比較モデリングで立体構造を予測し、結合リガンド構造を最適化する方法を開発した。比較モデリングの際には、膜タンパク質に特化したアラインメント法を開発した。具体的には、膜内環境で、ヘリックス間の水素結合など、極性の相互作用部分が保存されるよう制限がかかるようなアラインメント法である。また、結合リガンド構造最適化の際には、他の既知リガンド配座との類似度と、結合エネルギーのバランスから評価関数を作成した。

この手法をヒスタミン受容体や、嗅覚受容体に適用し、変異体実験による機能変化をよく説明できる結果を得た(4)。この結果は、上記で示した手法の妥当性を示している

総合データベース作成

以上の結果を基にしてゲノム配列が決定している全ての生物種に対して、GPCR配列候補を収め、これに対して機能分類データを付加した総合データベースを構築した。この中の一部の生物種に対してはSEVENSデータベースの中に組み込んだ。

また、上記で開発した機能予測プログラム(結合Gタンパク質予測プログラム)をWEBサーバーとして組み込み、公開準備中である。

〈国内外での成果の位置づけ〉

構造、機能分類の観点では、配列情報から膜貫通ヘリックスの本数や、トポロジーで分類を行う研究が国内外で数グループある。(Von Heijne et al. Jones, D.T et al. Stevens T. J et al. Mitaku. S et al.など) これらの方法では、より詳細な構造型の違いを識別できないが、当該研究は膜貫通ヘリックスの極性表面を比較することでこれを可能にする。GPCRに限れば国外では隠れマルコフモデルを使った構造認識法(Baldi, et al.)やニューラルネットを使った認識法(Arrigo, et al.)、サポートベクターマシンを使った方法が報告されている。一方、当該研究は、極性表面の差、あるいはGPCR上の物理化学的特徴量を指標としており、全く新しい方法である。

立体構造モデリングにおいては、極めて多くの国内外のグループが行っている。しかしながら、構造テンプレートが無い場合も含め、膜環境に特化した方法を取り入れた精度の良いシステムに仕上げることで見据えているのは、当該研究のみである。

データベースの観点から言えば、国内では清水、中井らのTMPDBや、美宅らのデータベースがある。これらは、SWISS-PROT中の膜タンパク質を展開したものである。一方、国外ではAlliance for Cellular Signaling (<http://www.cellularsignaling.org/>) による膜蛋白質データベースや、GPCRに限ればGPCRDB (Horn. H et al <http://www.gpcr.org/7tm/>)が有名である。GPCRDBでは、既知の配列に関してあらゆる解析がなされていて非常に整理されたデータベースである。一方私たちは、遺伝子予測からの新規なGPCR候補も含めて、各生物種ゲノムにおいてファミリー全体の概観できる点で、メリットがある。

GPCRの総合的案機能分類データベースは従来のSEVENS (<http://sevens.cbrc.jp>)の中に組み込み、更新を

繰り返している。これは、所属機関のトップページからもリンクを張られ、現在、月に1000件程度の利用がある。また、このデータベースや機能分類ツールに関し得て企業が興味を示し、共同研究が2件スタートした。さらに、GPCRの生理活性に関連する学会から興味を持たれ、シンポジウム講演を数回依頼された。

〈達成できなかったこと、予想外の困難、その理由〉

研究目的のうち、各種ゲノムからGPCRを判別、構造、機能の観点から分類しゲノム中の構造型の数を解明し、データベース化するという部分に関しては、ほぼ達成できた。また、分類後の代表配列に対する詳細な立体構造モデリング法を確立する部分に関しては、立体構造モデリングの計算条件の設定など基盤整備に思いのほか時間がかかり、大量配列に応用することは出来なかった。しかし精度の良い方法論として確立できた意義は大きい。

〈今後の課題〉

本研究により、ゲノムが明らかになった生物種に関してはGPCRの機能分類まで収めた総合データベースができたので、これを基に今後は、機能の観点で、比較ゲノム解析を本格的に行う必要がある。

また、GPCRに特化した立体構造モデリングに関しては、今後方法を整備し、自動化、ハイスループット化することにより、GPCR 総合データベースの配列に全てに対して、立体構造構造予測を行う必要がある。

〈研究期間の全成果公表リスト〉

- 1) Statistical Analysis and prediction of functional residue sites which are effective to GPCR-Gprotein coupling selectivity. Takahiko Muramatsu and Makiko Suwa Protein Engineering, inpress.
- 2) Yukimitsu Yabuki, Takahiko Muramatsu, Takatsugu Hirokawa Hidehito Mukai and Makiko Suwa, GRIFFIN: A System for Predicting GPCR-G-protein Coupling Selectivity using Support Vector Machine based on physicochemical parameters, Nucleic Acid Research, 33, 148-153, 2005.
- 3) Yukiteru Ono, Wataru Fujibuchi, Makiko Suwa, Genome Scale Overview of GPCR Genes detected from Bacteria and Archer and Eukaryote. GENE, 30;364:63-73, 2005.
- 4) Sayoko Katada, Yuki Oka, Takatsugu Hirokawa, Makiko Suwa, Structural Basis for Broad but Selective Ligand Spectrum of a Mouse Olfactory Receptor: Mapping the Odorant Binding Site. The Journal of Neuroscience, 25, 1806-1815, 2005.
- 5) Suwa, M., Sato, T., Okouchi, I., Kumagai, T., Arita, M., Asai, K, Akiyama, Y., Matsumoto, S., Tsutsumi, S., Aburatani, H. "SEVENS" Nucleic Acids Research. 31, 1 Online summary paper (<http://www3.oup.co.uk/nar/database/summary/373>) 2003.

〈公開ホームページ〉

- 1) SEVENS (<http://sevens.cbrc.jp>) 各生物種ゲノム配列からGPCR遺伝子を網羅的に発見し、機能分類などの解析情報を付加したデータベース。
- 2) GRIFFIN (<http://griffin.cbrc.jp>) GPCRとGタンパク質との結合選択性を予測するツール。