

新しい蛋白質ネットワークを求めて —完全長cDNAバンクからのアプローチ—

総括責任者 加藤誠志

1. はじめに

生命現象を分子レベルで解明しようという研究が急速に進展している。生物の特徴である代謝や自己複製といった機能も、分子レベルで記述できつつある。このような研究の根底には、細胞というものが分子同士の相互作用のネットワークからできており、このネットワークの構造がわかれば、生命現象に特有な機能も説明できるという仮定がある。これまでの膨大な生化学的・分子生物学的研究は、この仮定が正しいことを裏付けるものであった。その結果、代謝ネットワーク、蛋白質合成・分解ネットワーク、遺伝子複製・転写ネットワーク、情報伝達ネットワークなどが次々と明らかにされてきた。これらのネットワークの主要構成要素である蛋白質とその他の分子群をすべてまとめて我々は「たん白生態」系と呼んでいる。細胞の生命現象の全貌を分子レベルで理解するには、「たん白生態」系の全構成要素とそれらのつながりを明かにする必要がある。しかし、これらの蛋白質の中でこれまでその機能が明らかになったものは、ヒトを例にとると十万種類と見積られている蛋白質の数%にも満たない。ヒトゲノムがコードしている蛋白質の大半は機能未知ということになる。このような機能未知の蛋白質からなる未解明部分のネットワークの構造と機能を明らかにしていくことが、これからの生命科学研究の最重要課題の一つと言える。

この課題に対して、従来とられてきたアプローチは、細胞を構成している蛋白質の中から、ある生物活性を持った蛋白質を探し出すという生化学的アプローチ、すなわち、まず活性という機能が先にあり、この機能を担っている蛋白質を探索するというアプローチであった。このアプローチによって、酵素のように一つの蛋白質が一つの活性を有しているものが次々と発見されてきた。さらに、近年の遺伝子工学の進展に伴い、細胞を構成する蛋白質を遺伝子の形でクローン化し、これを用いて蛋白質を容易に調製できるようになってきたことにより、新しい蛋白質の探索が加速している。ただ、活性を指標として蛋白質を探すというこれまでのアプローチは、一つの機能を発現するのに複数の蛋白質が関与している場合には、まだ多くの困難を伴っているというのが実情である。

このような状況の中で、我々が提案したアプローチは、まず細胞を構成している蛋白質をすべて集めてしまい、これを出発材料にして、それぞれの蛋白質が関与している分子ネットワークを明らかにしていこうというアプローチ、すなわち「物質から機能へ」というアプローチである。このような発想の基になったのは、ヒト蛋白質を完全長cDNAの形で集めようというホモ・プロテインcDNAバンク構想である（加藤、蛋白質核酸酵素 **38**:458-467,1993; 加藤、*BIO INDUSTRY* **11**:760-770, 1994）。この構想を実現するために必須となる完全長cDNA合成技術は、我が国で世界に先駆けて開発され、ほぼ

完成の域に達している(Kato et al., *Gene* 150: 243-250,1994) (本報告書付録参照)。この技術を用いれば、キャップ部位からポリ(A)テールまでの全領域を有し、欠失や変異を含まない正真正銘の完全長cDNAを高収率で合成することができる。しかも、これらの完全長cDNAは発現ベクターの形でバンクに登録されているので、インビトロやインビボで直ちに蛋白質を作ることができる。

本プロジェクトが取り組んできたのは、約4,000種類のヒト完全長cDNAクローンバンク(本報告書付録のリスト参照)を出発材料にして、まだ未解明の蛋白質ネットワークの一旦を解明しようという試みであった。特定の生物現象の機構解明と違って、どのような結果が得られるのかを全く予測できない文字通りの探索研究となった。本報告書では、この5年間、様々な背景を持った研究者がプロジェクトに集い、試行錯誤しながら機能未知の完全長cDNAクローンと格闘して得られた研究成果を紹介し、我々のとった「物質から機能へ」というアプローチが、今後のポストゲノムシーケンス時代の分子生物学研究において有効な手段になりえることを示したい。

2. 研究体制

本プロジェクトは、ホモ・プロテインcDNAバンクの存在を前提として、細胞システムにアプローチするための道筋をつけることを目指した。この目的を達成するために、次の3つのグループを設けて研究をスタートした。一応グループ分けしているが、ほとんどの研究は一箇所で行われたので、研究者間のつながりは緊密であり、日常的にグループ間での議論がなされ、全員が一体となって問題に取り組んだというのが実態に近い。

(1) 蛋白解析グループ

本グループは、ターゲット蛋白質と相互作用する分子を探索することによって、ターゲット蛋白質を中心にしたネットワークの解明を試みる。具体的には、手持ちのバンクに含まれている新規ヒト完全長cDNAクローンの中から、アミノ酸配列や発現産物の挙動などの情報を手掛かりにしてターゲット蛋白質を選択し、これと相互作用する分子を各種手法を駆使して探索し、新しい経路・複合体・修飾などの発見を目指す。

(2) 局在解析グループ

本グループは、細胞内における蛋白質の位置情報を得ることによって、蛋白質の機能を推定するための手掛かりを得ることを目指す。具体的には、ヒト完全長cDNAを用いて、蛋白質の局在を効率的に決めるのに必要となる手法を開発する。とくに、免疫染色による局在解析に必要な抗体を、遺伝子免疫法で作製することを試みる。また、新規cDNAクローンを用いて網羅的局在解析を行う。

(3) 動態解析グループ

本グループは、機能未知の完全長cDNAを細胞内で強制発現させた時に引き起こされる蛋白質ネットワークの変化を観察することによって、その機能に関わる手掛かりを得ることを目指す。具体的には、二次元ゲル電気泳動法による完全長cDNA-細胞内蛋白質マッピング、細胞内で発現させた完全長cDNA翻訳産物の動きやその結果生じる表現系の変化の解析などを行う。

3. 研究成果

研究成果は大きく二つのカテゴリーに分けられる。一つは、ヒト完全長cDNAバンクを出発材料にして、新しい蛋白質ネットワークの構成要素となるような蛋白質複合体、蛋白質修飾の発見であり、もう一つは、このようやアプローチに適した新しい手法の開発である。それぞれについて、トピックスを以下に列挙する。詳細については、原著（引用文献番号は本報告書の外部発表リストに対応）ならびに付属の講演要旨集や研究成果集を参照していただきたい。

A. 新しい蛋白質複合体・蛋白質修飾の発見

(1) ヒト完全長cDNAがコードする新規蛋白質群の解析

ヒト完全長cDNAバンクに含まれている新規クローンについて、全長塩基配列決定インビトロ翻訳、緑色蛍光蛋白質(GFP)との融合蛋白質を発現による細胞内局在部位決定を行い、ターゲット蛋白質を選択するための基礎データを蓄積した（本報告書研究成果集と付録参照）。

(2) NEDD8修飾経路の発見[3,8]

cDNAが本来コードしている蛋白質の分子量と大きく異なるインビトロ翻訳産物を生成するクローンを出発材料として、その生成機構を検討した結果、共有結合による新しい蛋白質修飾システムNEDD8経路を発見した。

(3) NEDD8化による細胞周期制御[9]

分裂酵母を用いた実験により、NEDD8化がユビキチンリガーゼ系であるSCF (Skp1-cullin-F-box)と呼ばれる蛋白質複合体の機能制御を介して、細胞周期の調節に関与していることを証明した。

(4) インビトロ翻訳産物のマルチユビキチン化[6,15]

インビトロ翻訳によって生成した新生小胞体膜蛋白質は、マルチユビキチン化を受けられることを見だし、この系を用いてマルチユビキチン化反応を追跡できることを示した。

(5) 新規核蛋白質複合体Npw38-NpwBPの発見[5,7]

蛋白質-蛋白質相互作用に関与する配列モチーフWWドメインを有する蛋白質を出発材料として結合蛋白質の探索を行い、転写機構に関与する新しい核蛋白質複合体Npw38-NpwBPを発見した。

(6) 新規スプライセオソーム構成成分の発見[14]

新規蛋白質の網羅的局在解析によって核内に斑点状に発現するクローンを見だし、この蛋白質Nps20はスプライシング因子SF3a60/SAP61と結合し、スプライシングに関与している新規のスプライセオソーム構成成分であることを示した。

(7) 不死化細胞で発現昂進する核蛋白質の同定[11]

cDNA全長配列の情報に基づいて発現プロフィール解析を行い、SV40によって不死化した細胞で発現が昂進する2種類の核蛋白質IMUP-1、IMUP-2を同定した。

(8) 細胞内蛋白質のO-グルコシル化の発見[12]

細胞内レクチンを探索する過程で、クレアチンキナーゼが、O-グルコシル化されていることを見いだした。さらに細胞内蛋白質の多くが新しいO-グリコシド型の糖鎖修飾を受けていることを見だし、これらの修飾による細胞内蛋白質の新しい制御機構の存在の可能性を示した。

(9) 二次元電気泳動におけるマルチスポットの成因説明[13]

インビトロ翻訳産物の二次元電気泳動の結果、等電点の異なるマルチスポットが生成することを見いだした。これらのマルチスポットは、等電点電気泳動中に高濃度の尿素の存在によって、カルバミル化された産物に因ることを明らかにした。

B. 新しい手法の開発

(1) 高感度レクチン検出法[1]

細胞ライセートなどに含まれている微量レクチンを、高感度で検出する方法を開発した。

(2) 膜蛋白質のトポロジー解析[4]

膜貫通ドメインと思われる配列の下流にウロキナーゼのプロテアーゼドメインを融合させたものを培養細胞内で発現させ、細胞表面のウロキナーゼ活性を指標として、膜蛋白質のトポロジーを決定する方法を開発した。

(3) cDNA免疫による抗体作製[10,16,17]

完全長cDNAの発現ベクターを注射や遺伝子銃でマウスやラットに投与することにより、cDNAがコードしている蛋白質に対する抗体を作製できることを示し、さらに抗体産生効率を上げるための手法を開発した。

(4) 2ハイブリッド局在化法による蛋白質-蛋白質相互作用の検出

局在化シグナルを有する蛋白質と、レポーター機能を付与した蛋白質を細胞内で共発現させ、レポーター機能を付与した蛋白質の局在変化を指標として、両者の相互作用を細胞内環境下で迅速にかつ視覚的に検出する方法を開発した。

4. 総括

本プロジェクトは、細胞の中の新しい蛋白質ネットワークを見つけるのに、完全長cDNAバンクを出発材料とする「物質から機能へ」というアプローチがどの程度有効なのかどうかを検証する実験の場となった。ここで我々のとったアプローチをもう一度振

り返ってみよう。出発材料は、一次構造が決定され、インビトロや細胞内で発現を確認できた完全長cDNAクローンからなるバンクである。基本戦略は、このバンクの中から機能未知のターゲット蛋白質を選び、この蛋白質と相互作用するものを探索することによって、未知の蛋白質ネットワークを見つけようというものである。ターゲット蛋白質は、配列情報や発現した蛋白質の情報に基づいて、他の蛋白質や生体分子と相互作用している可能性の高いものを選択する。例えば、結合モチーフを有しているもの、インビトロ翻訳産物の分子量が予想値と著しく異なるもの、特定の局在パターンを示すものなどである。蛋白質の相互作用をたどることによって、もし機能既知の蛋白質にたどり着けば、得られたネットワークは、その機能の一端を担っている可能性が高いということになる。

当初、このような趣旨のプロジェクトに果たして研究者が集まってくれるだろうか、このようなアプローチで新しい研究の種を本当に作りだせるのだろうかという危惧も抱いていた。幸い、チャレンジ精神旺盛な若手研究者が集まり、新しい試みに挑戦してくれた結果、当初の目標であった新しい蛋白質ネットワークの発見、具体的には、新しい経路、新しい蛋白質複合体、新しい蛋白質修飾の発見、および新しい手法の開発において、上記の成果をあげることができた。ここにあげた成果は、ある程度芽が出始め、今後大きく花開く可能性を秘めているものである。これ以外にも、まだ種の段階であるが、今後展開の仕方によっては面白い芽を出す可能性のあるクローンを数多く見いだすことができた。したがって、我々のとったアプローチは、これまで見逃されていた新しい蛋白質ネットワークを見つけるのにたいへん有効であるという結論を得た。

このアプローチがうまく機能した最大の要因は、インビトロやインビボで直ちに発現可能な完全長cDNAバンクを出発材料として用いたということにある。完全長cDNA発現ベクターの集団をすでに手にしているので、一次構造の決定された蛋白質を容易に作るができる。さらにこれらのcDNAを使って変異体や融合蛋白質などを自由自在に作製し、それらの発現実験を行うことにより、機能を推定するために有用な情報を数多く得ることができた。従来法のように機能から出発して、それを担う物質を探索する場合、ほとんどの時間が物質のスクリーニングにとられる。最後にものが得られれば良いが、徒労に終わることも少なくない。このような徒労をいかにしたら無くせるかという問題意識が、ホモ・プロテインcDNAバンク構想の発想の原点にあった。

もちろん、我々のアプローチにも弱点がないわけではない。蛋白質によっては、様々な手法を用いても、相互作用する蛋白質を見いだせないものもあるだろうと思われる。したがって、我々のアプローチの成否は、ネットワークの核となるようなターゲット蛋白質をいかにうまくバンクの中から見つけられるかにかかっている。成功例となったNEDD8修飾経路やNpw38-NpwBP核蛋白質複合体は、完全長cDNAの網羅的解析によって得られたアミノ酸配列、インビトロ翻訳、細胞内局在などの実験データから、相互作用する蛋白質の存在が前もって予測できたクローンを出発材料としたものである。今後、バンクが充実し、実験データの蓄積が進めば、ターゲット蛋白質の選択はより容易になり、より多くの未知ネットワークがあきらかにされるものと期待される。

今後このアプローチをより有効なものとするためには、相互作用する蛋白質を効率良

く探索する手法の開発が必須の課題である。NEDD8経路やNpw38複合体の場合は、アフィニティー精製や酵母2ハイブリッドシステムなどの従来法を用いて探索を行ったが、このプロセスに最も労力と時間をとられた。この問題の一つの解決法として考案したが、2ハイブリッド局在化法である。この方法を用いれば、完全長cDNA発現システムだけを使って蛋白質間相互作用を網羅的に調べることが可能となる。ただし、この方法を適用するには、完全長cDNAバンクが完成しているということが大前提になる。完全長cDNAすなわち蛋白質ピースを全部そろえて、本プロジェクトに最初につけた副題「蛋白質ジグソーパズルを解く」が早く現実となることを願っている。

5. おわりに

過去5年間のゲノムプロジェクトの進展は著しく、酵母、線虫、ショウジョウバエなどの全ゲノム解読は終了し、ヒトゲノムの完全解読もこれから1~2年以内にはほぼ終了する勢いである。その結果、世の趨勢は、遺伝子の機能解析に向かっており、いかにしてゲノムの配列情報から遺伝子・蛋白質機能へアプローチするかということが、ポストシーケンス時代のメインテーマになってきた。これは、まさに本プロジェクトが掲げてきたテーマそのものである。現在、ゲノムプロジェクトでは、配列情報、発現プロフィール、遺伝子ノックアウトなどのデータに基づく機能予測が大規模に行われているが、いずれも予測でしかない。細胞内での機能を知るためには、蛋白質ネットワークへの位置づけが不可欠であろう。そのためには遺伝子を実際に発現させて蛋白質にしなければ、それらの相互作用や、その結果として出てくる機能を実証できない。今後、完全長cDNAバンクの必要性が益々高まると思われる所以である。

完全長cDNAバンクの中から、鍵となる蛋白質をコードするクローンを実験によって見つけだし、このターゲット蛋白質と相互作用する蛋白質の連鎖をたどることによって新しい蛋白質ネットワークを解明していこうという本プロジェクトのアプローチが、これからのポストシーケンシング時代の一つの柱となることを期待したい。また、本プロジェクトでターゲット蛋白質の個別的な解析によって得られた成果についても、これらの蛋白質が関与するネットワークの全貌を知るためにはまだ多くの実験的検討が必要である。今後、担当した研究者や共同研究者によって、さらに研究が続けられ、それぞれについて新しい研究分野が築き上げられることを願っている。

最後に、この5年間の研究活動に対して、ご支援をいただいたすべての関係者の皆様方に心から感謝の意を表したい。未踏領域に新しい道をつけてくださった研究者・共同研究者の皆さん、いつも有益なご助言を与えてくださった研究顧問・研究推進委員の先生方、裏方としていつも万全の体制で研究を支えてくれた技術・事務両参事と事務スタッフの皆さん、道に迷いそうになってもいつも温かく見守って下さった理事長・理事をはじめとする科学技術振興事業団の皆様、そして研究場所の設置において多大なるご配慮をいただいた(財)相模中央化学研究所と北里大学の関係者の皆様方に心から御礼を申し上げます。

Search for novel protein networks: Approach from a full-length cDNA bank

Project Director Seishi Kato

Background

The cell is composed of tens of thousands of proteins which function by forming a complex network composed of equally complex subnetworks. To understand the whole picture of life phenomena at the molecular level, it is necessary to identify all components involved in the cytoprotein network and to uncover all interactions among them. However, only several percent of human proteins has been identified and analyzed. Most of proteins encoded by the human genome remain unidentified. One of the most important subjects of life science would be to uncover the structure and function of unidentified networks constructed by proteins with unknown functions.

The main purpose of the Kato cytoprotein network project was to find novel protein networks that can act as seeds capable of growing to break new ground in molecular cellular biology. The conventional approach to achieve this aim is to search for proteins possessing some target activity, such as enzymic. This approach, from function to protein, has been successfully carried out by many researchers to elucidate various protein network systems. In contrast, the approach taken by this project was opposite, that is, from protein to function. The starting material was a human cDNA bank comprising ca. 4,000 kinds of full-length cDNA clones. This idea is based on the concept of the Homo-Protein cDNA Bank whose purpose is to collect an entire set of human proteins as a form of a full-length cDNA clone. By using cDNA clones encoding novel proteins with unknown functions, this project undertook the task of uncovering their functions through characterizing their interacting proteins.

Research Results

Discover of a NEDD8 modification system regulating the cell cycle: During *in vitro* translation analysis of full-length cDNA clones, a ubiquitin-like protein, NEDD8, was found to covalently modify cullin-family proteins, and three proteins composed of this pathway were identified. Using a fission yeast system, NEDD8 was shown to regulate the cell cycle via the modification of an SCF (Skp1/cullin-1/F-box protein) complex. These results indicate that NEDD8 modification plays an essential role in various regulation processes involving cullins.

Discovery of a novel nuclear protein complex with interaction via the WW domain: A novel nuclear protein containing the WW domain, Npw38, was identified in the full-length cDNA bank. A search of Npw38-interacting molecules revealed that Npw38 can interact with NpwBP, DNA helicase, RNA helicase, RNA, and single-stranded DNA to form a nuclear protein complex which may be involved in transcriptional events as a member of an mRNA factory.

Discovery of a novel O-glycosylation of intracellular proteins: Rabbit muscle cytosolic creatine kinase was found to be O-glycosylated via Ser and Thr residues. Other intracellular proteins, such as metabolic enzymes, were shown to contain sugar molecules. These results imply the existence of a novel glycosylation system in the cytosol, by which some cellular function might be regulated.

Localization to function: A comprehensive localization analysis of novel full-length cDNA clones was carried out using fusion genes with urokinase cDNA for membrane proteins or green fluorescent protein cDNA for intracellular proteins, while identifying many novel proteins including type II membrane proteins, a spliceosome component, as well as RNA helicase, and immortalization-upregulated nuclear proteins.

Antibody production using cDNA immunization: When an expression vector carrying human full-length cDNA was inoculated into a mouse or rat using a syringe or gene gun, antibody production against the protein encoded by each cDNA was observed in the serum. Furthermore, a method to improve the efficiency of antibody production was developed.

Two-hybrid localization method for protein-protein interaction analysis: A novel method to rapidly and visually detect the protein-protein interaction occurred under physiological conditions in the cell was developed, based on the observation of localization change of reporter-fused protein induced by interaction with a protein possessing a localization signal. This method enables us to perform comprehensive analysis of protein-protein interactions among the proteins encoded by the full-length cDNA clones.

Perspectives

This five-year project has demonstrated that the approach from protein to function is a very effective methodology to find novel protein networks. This success was due to the possession of a set of full-length cDNA clones by which it was possible to obtain proteins when necessary. This approach, however, has one problem at present: the limited size of the bank. If the bank could be completed and comprehensive data accumulated, this would greatly accelerate the ability to select a target protein and to find its interacting proteins from the bank. Thus, the top priority is now to complete the construction of a human full-length cDNA bank. By doing this, it is expected that the present approach will be used as one of the major strategies at the stage of functional genomics in the post-genomic era.