

[別紙 2]

審査の結果の要旨

氏名 尾山 大明

本研究は膨大なヒト完全長 cDNA 配列情報から示唆される未知低分子タンパク質の存在を明らかにするため、ヒト K562 細胞中に存在する低分子タンパク質群を対象として質量分析計(LC-ESI-MS/MS)を用いたタンパク質レベルの発現解析を行い、下記の結果を得ている。

1. 培養したヒト K562 細胞破碎液から、2通りの方法（電気泳動-ゲル内消化法、酸抽出-直接消化法）で低分子タンパク質由来のペプチド断片が濃縮された試料を作製し、LC-ESI-MS/MS を用いて発現解析を行ったところ、NCBI の Reference Sequence (RefSeq) タンパク質データベースに登録されている 671 個の低分子タンパク質(ORF : 100 アミノ酸残基以下)(2003 年 3 月現在)の中で、前者の調整法で作製した試料からは 37 個、後者の試料からは 24 個、計 52 個（9 個の重複を含む）のタンパク質が同定された。これらの中には、細胞内で重要な機能を担っている低分子タンパク質が多く含まれ、リボソームやスプライソソーム複合体に関してはデータベースに登録されている低分子構成タンパク質(ORF : 100 アミノ酸残基以下)の半数以上を当解析系で同定する事が出来た。

2. 上記の RefSeq タンパク質データベースに対する検索において、同定に至らなかった MS/MS スペクトルに関して、更に 2 種類 (FLJ 及び RefSeq) のヒト完全長 cDNA データセットに対して検索を行ったところ、前者から 4 つ、後者から 3 つ、計 7 つの新規コード領域(ORF : 100 アミノ酸残基以下)が同定された。

3. 当解析系で同定された7つの新規コード領域の中で、5つのコード領域は各完全長 cDNA 配列中で最も長い ORF の上流に位置している事が分かった。下流の長い ORF はタンパク質コード領域として既知あるいは強く推定されるものであり、これらの遺伝子は 2 つのコード領域を持ちうる事が示された。

4. 最長 ORF の上流に位置している 5つのコード領域の開始コドンは全て、各完全長 cDNA 配列中で最も上流に存在していた。また、RefSeq のタンパク質データベースから同定された 52 個のタンパク質に関して、対応する RefSeq の完全長 cDNA 配列の情報に基づいて同様に開始コドンの位置を調べたところ、44 個(85 %)のタンパク質が最も上流に位置する ATG コドンから翻訳を開始している事が分かった。この解析結果から、リボソーム前駆体が mRNA の 5' 端から 3' 端に向かってスキャンをし、最初に遭遇した開始コドンから翻訳を開始するという典型的な翻訳開始のメカニズムの普遍性が示唆された。

5. FLJ、RefSeq 双方の完全長 cDNA データセットの中には、配列中の最も上流に短い ORF を持つ cDNA 配列が非常に多く存在することが分かっており、本研究の結果からこれらの ORF がコードする低分子タンパク質が、実際に細胞中で多く存在することが示唆された。

以上、本論文はヒト K562 細胞中で発現している低分子タンパク質群を対象として、質量分析計を用いたタンパク質レベルでの発現解析から、完全長 cDNA 配列中においてタンパク質コード領域の上流に存在する短い ORF がコードするタンパク質が、実際に細胞中で発現していることを明らかにした。本研究は、ヒトの細胞中でこれらのタンパク質が発現していることを直接的に示した最初の研究であり、ヒトのプロテオームに関して、その真の全体像の解明に重要な貢献をなすと考えられ、学位の授与に値するものと考えられる。