

## ゲノム・プロテオームデータベースの連携による高次生物知識発見に関する研究

●高橋 勝利 ◆富永 大介

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

ゲノム塩基配列データベース、mRNA発現データベース及びプロテオームデータベースを相互に連携し、各データベースに含まれる情報の相互補完・対応づけを行ったうえで、高次の生物知識を発見することを目的とし、

- 1) ゲノム塩基配列データベースと質量分析技術を利用したタンパク質同定を含むプロテオーム解析を総合的に支援するコンピュータシステムの開発及びプロテオーム関連データの蓄積
- 2) mRNA発現パターンとタンパク質発現パターンとの相互比較による生物知識発見

を目標として研究を実施した。

### 2002年度の研究の当初計画

- 1) PiKA2システム（WEBインターフェイスを使ってタンパク質二次元電気泳動像を収集し、タンパク質スポットの自動認識や、複数の電気泳動パターンの自動比較を行うシステム）を改良し、様々なプロテオームデータベースや、質量分析によってタンパク質スポットの同定を支援する機能を追加する。
- 2) パブリックに公開されているプロテオームデータベース及びタンパク質二次元電気泳動画像の収集を行い、ゲル上のタンパク質スポットの同定結果、同定に用いた質量スペクトルデータなどを上記1)によって構築した統合プロテオーム解析支援システムに入力する。
- 3) 収集したmRNA発現データベース及びタンパク質発現データベースから、対象生物種、詳細な実験条件等を抽出し相互に関係付ける作業を行う。

### 2002年度の成果

- 1) 簡単にエンドユーザーが、自分の持つタンパク質二次元電気泳動像を登録し、電気泳動パターンの自動解析を行いことができるようにPiKA2システムのGUIを改良した。また、質量分析によってタンパク質スポットの同定を支援するシステムの構築を開始した。
- 2) PCクラスターシステムを構築し、大量の二次元電気泳動像データの解析を高速に行えるように、PiKA2システムの解析プログラムを並列化した。
- 3) パブリックに公開されているmRNA発現データベース、プロテオームデータベースの調査を行い、プロテオームデータベースに登録されているのと類似条件でのmRNA発現パターンがどの程度利用可能かを調査した。

### 国内外の成果の位置づけ

WEBベース・インターネット経由でタンパク質二次元電気泳動像のカスタムデータベースを作成し、その解析を自動的に行う解析システムは国内外を見ても存在しない。このようなインターフェイスによりタンパク質二次元電気泳動像データを大量に収集し、そこからプロテオ

ーム発見に関する情報を得ようとする試みは新規性が高い。

### 達成できなかったこと、予想外の困難、その理由

- 1) タンパク質二次元電気泳動像自動解析システムにタンパク質同定機能を組み込むことが出来なかった。これは様々な種類の質量分析計からの出力の違い（フォーマットの違い、質量分解能の違い、質量精度の違いなど）を一手に取り扱うことが非常に困難であったことに起因している。
- 2) 公開されているデータベースに限定して、相互比較可能な類似条件に関してのタンパク質発現量データとmRNA発現量データの探索を行ったが、残念ながら十分な統計解析に耐えうるような多量のエントリーを見出すことが出来なかった。

### 今後の課題

PiKA2システムと質量分析によるタンパク質同定支援システムとの相互連携を進める。このために、様々な質量分析計に固有の情報（質量スペクトルフォーマット、質量分解能、質量精度など）のデータベースを作成し、そのデータにもとづいて、質量スペクトルの品質の違いを吸収するための仕組みを確立しなければならない。

公開データベースだけでは、プロテオーム発現データとトランスクリプトーム発現データを相互比較して高次生物知識を発見するために十分な量のエントリーが抽出できなかったため、今後様々な研究機関との共同研究に基づいて非公開データベースを利用させていただく可能性や、独自に実験を行ってデータを収集する必然性について考慮する必要がある。