

## 研究開発課題別事後評価結果

### 1. 研究開発課題名

メタボローム・データベースの開発

### 2. 代表研究者名

奈良先端科学技術大学院大学 情報科学研究科 教授 金谷重彦

### 3. 研究実施概要

メタボローム研究においては、質量分析装置により出力されるピークデータをもとに代謝物の推定を行い、機能を注釈づけることが必要とされる。本課題は、質量分析装置から得られるピークデータ、ピークデータと代謝物の関係づけ、さらには、代謝物の機能注釈についてのデータベース構築ならびにツール群の研究開発および、低コストで継続維持可能なデータベースシステムの研究開発を実施した。

#### 1) 質量スペクトルデータベース「MassBank」の拡充

「MassBank」を一括してスペクトル検索を行うための SOAP-API インターフェイスを開発提供した。また、「MassBank」データにクリエイティブ・コモンズ (CC) ・ライセンスの表示を行った。さらに、ユーザーの利便性を図る目的で、MassBank.jp と対等な機能を有する MassBank.eu を EU に設置し、安定的なサービスの提供を実現した。また、メタボローム・マススペクトル・データベース「Bio-MassBank」に設計し、マススペクトルの類似性を比較する検索機能を追加した。

#### 2) 代謝物情報データベースの構築

学術論文のデータを蓄積し、生物種と代謝物の関係データベース「KNAPSAcK Core DB」を開発した。また、二次代謝物における機能アノテーションを可能にするため、文献調査により集めたデータをもとに、KNAPSAcK Family 代謝物活性データベース「Metabolite Activity DB」の開発を進めた。さらに、未同定代謝物の化学構造推定を行う「Metabolite Prediction Tool」を開発し「MassBank」に実装した。

#### 3) メタボローム統合データベース

「Bio-MassBank」へのデータの預け入れ、研究者間のデータ共有を目指してメタボロームデータの標準化フォーマット TogoMD (The Togo Metabolome Data Format) の開発を行った。TogoMD フォーマットは、現在でも統一化されていないメタボロミクスデータ記述方法の世界標準となりつつある。データ処理パイプラインの改良を行い、「Bio-MassBank」からのスペクトルデータ公開は加速度的に進んだ。

さらに、データ公開における最大の律速要因である実験手法の詳細情報（メタデータ）の管理を効率化し、データ公開を加速する目的で、メタデータの管理と共有を専門としたデータベースシステム「Metabolonote」の開発等を行い、メタボローム実験データの標準的記載、生物サンプルにおける質量分析データベースとそのケミカルアノテーションならびに機能アノテーションまでの一連の作業を容易にするためのデータベースならびにツール群を開発した。

## 4. 事後評価結果

### 4-1. 当初計画の達成度

本課題は、1) 質量スペクトルデータベース「MassBank」の拡充、2) 代謝物質情報データベースの構築、3) メタボローム統合データベースの構築・公開や、各種ツールの開発、提供という当初の計画を達成した。加えて、「MassBank」システムを研究分野に応じてカスタマイズしたいという希望に応えるために、全てのソースコードを Java だけで記述した MassBank Developer's Kit を開発し公開した。代謝関連情報の統合化が進み、データの RDF 化も実施した。NBDC の生命科学系アーカイブサービスへの預け入れと公開も実現しており、統合化は進んだといえる。

### 4-2. 研究開発成果の公開および利用の状況等

「MassBank」、「Bio-MassBank」などのデータベースが公開されている。各データベースのアクセス数は様々であるが、「MassBank」においては、月間のユニーク IP アクセス数は 1.5 万件程度と多く、広く活用されている。

### 4-3. 研究開発成果によるライフサイエンス分野のデータ活用への波及効果

今後メタデータの更なる充実や他のライフサイエンスデータベースとの統合を進めることで、代謝物とそれらの生物活性との相関を研究する上で有用なデータベースとなることが予想される。

### 4-4. 広報・アウトリーチ活動等

論文発表、学会講演・発表などで研究成果を周知するとともに、展示会においても成果を紹介し、ユーザーからの要望収集などを積極的に行っており評価できる。

## 5. 総合評価

本課題で新規に構築、維持されたメタボローム・データベース及び関連データベースは、化合物のマススペクトルデータベースと代謝物との関係や、代謝物と生理活性の関係を整理したデータベースなど多岐にわたっており、メタボロミクス研究や、医薬品、生態学の

様々な研究分野に影響を及ぼす非常に重要なデータベース群である。また、マススペクトルでの標準データベースを通じて、生理活性情報等の収集・統合が順調に進捗しており、さらに、新規に開発されたデータフォーマット **TogoMD** の国際標準化をはじめ、メタボローム・データベースの国際標準化にも積極的に取り組んでいる。今後、メタデータを中心としたデータのデータベース化、検索システムの充実、高性能マススペクトルデータへの対応等により、この分野を代表する国際的な基盤データベースに成長していくことが期待される。