

研究開発課題別事後評価結果

1. 研究開発課題名

蛋白質構造データバンクの国際的な構築と統合化

2. 代表研究者名

大阪大学蛋白質研究所 教授 中村春木

3. 研究実施概要

本課題では、大阪大学蛋白質研究所内に構築された日本蛋白質構造データバンク PDB Japan (PDBj) が、米国 (RCSB-PDB)、欧州(EBI-PDBe)および NMR 実験情報のデータバンクである BMRB(BioMagResBank)と協力して国際蛋白質構造データバンク (wwPDB) を運営するとともに、新規に電子顕微鏡および X 線小角散乱の構造情報データベース Electron Microscopy Data Bank(EMDB)の構築を行った。また、高品質データ登録のための、データ管理・配布システム、構造データのオントロジー、データ記載のフォーマット、ユーザインターフェース等の開発・改良を実施した。さらに、「PDBj」および「PDBj-BMRB」のミラーサイトを構築して冗長化も行い、分散・統合的な環境で構造データベースを公開する体制を構築した。

1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発

平成23年度から登録業務の標準化・自動化の開発のため、wwPDBとして大規模で緻密な国際的なシステム (Common Deposition and Annotation System) を開発して、平成25年度 1月から新規の登録システムの一般利用を開始した。PDBjではデータ検証部分を担当・実施して、新フォーマットであるPDBx/mmCIFの策定を行った。さらに、BMRBと外部データベースとのインターフェース・システムの開発を行い、BMRBデータのXML等への変換やオントロジーにより記述された高度な検索操作や他のオーミクスとの連携を実現した。NMR実験データを統合管理し、データ解析とデータベース登録の両面を支援するプログラムであるMagROを開発し、NMRToolBoxにおいて公開し、NMR立体構造アンサンブルから収束領域を自動検出して分離するプログラムFitRobotも開発・公開した。

2) 統合化に向けたデータベースの高度化

wwPDBで従来から用いられている PDBx/mmCIF辞書に基づいて、それと互換な OWLオントロジーを構築した。さらに、このオントロジーに基づいてPDBの各エントリーPDBMLファイルから対応するResource Description Framework (RDF) ファイルを自動生成する仕組みを開発し、新規エントリーの更新に合わせて自動的に対応するようにした。各エン

トリーの各カテゴリーのRDFデータをユニークなURLで参照して、ウェブ上で取得する仕組みを、簡易検索機能とともに実装し、公開した。このSemantic Webのために開発されたフォーマットはPDB/RDFとしてwwPDBの正規フォーマットとして採用された。

3) 機能情報と構造情報との統合化システムの開発

網羅的構造比較プログラムGIRAFの改良、蛋白質構造や機能を推測、解析するツール、パイプラインのCompotif, eF-patch, SFAS を開発・公開し、これらのサービスの結果を機械可読化のためにセマンティックウェブ化した。さらに、蛋白質機能情報および高次生命機能情報と蛋白質構造情報との統合化システムを確立するため、「UniProt」等の他のデータベースとも統合することによって、配列情報から構造情報を基に蛋白質機能を同定・予測し、高次生命機能に対する原子レベルでの合理的な解釈を可能とする仕組みを開発した。

4) 人材養成の実施

wwPDB の講習会、テレビ会議等を通じ、On the job training 手法を中心に、データ管理技術者、アノテータの人材育成を実施した。また、一般社会人や高校生を含む学生向けの講演会等の実施も行った。

4. 事後評価結果

4-1. 当初計画の達成度

本課題は、1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発、2) 統合化に向けたデータベースの高度化、3) 機能情報と構造情報との統合化システムの開発、4) 人材養成の実施、という当初の研究計画を達成した。加えて、NPO法人を設立して社会に対する国際的アウトリーチ活動を積極的に行うことや、「PDBj」および「PDBj-BMRB」のWebポータルをNBDCのサーバにも設置し分散・統合的な環境で構造データベースを公開する体制も構築した。データがすべてRDF化されダウンロード可能となっており、サイトのミラーサイトも構築されていることから、データベースの統合化が非常に高いレベルに達している。

4-2. 研究開発成果の公開および利用の状況等

本課題で構築されている「PDBj」、「PDBj-BMRB」のポータルサイトと各種ツール等の成果はすべて公開され、月間ユニーク IP アクセス数は、それぞれ 2 万件および、1.5 千件程度となっており、多くの利用者に活用されている。

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

タンパク質構造データベースは元々ライフサイエンス分野の基幹的データベースのひとつ

であり、基礎生物学のみならず、医薬品創薬研究などの応用研究にとっても重要なデータを提供している。

本事業の成果により PDB の質・量ともに順調に充実化し、さらに統合化の推進によりタンパク質構造と機能の統合的解析が容易になったことで、ライフサイエンス研究者にとっての有用性が一段と向上した。基盤整備として当該データベースの貢献度は極めて高いといえる。

4-4. 広報・アウトリーチ活動等

論文発表、学会講演・発表などにおいて、研究成果を積極的に発表されており、展示会等でも成果の紹介を実施している。なお、九州、北海道、東京、大阪等、日本各地で「PDBj」講習会を実施しており、さらに、小学生から一般社会人を対象とした講演会への参加も積極的に行われた。

5. 総合評価

本課題が運営している蛋白質データベース「PDBj」は QC の行き届いた質の高いものであり、国際的にもライフサイエンス研究活動に不可欠なデータベース及び解析ツール群を提供しており、長期かつ安定的な運用が今後の最も重要な課題の一つである。また、X 線小角散乱等これまでに蓄積できていないデータも収集し、データベース化することで、より網羅性の高いデータベースを目指していただきたい。さらに、今後統合化推進の他の課題との連携により、生命科学情報の統合化を高め、様々なライフサイエンス分野の研究開発に応用・活用されることにより、ライフサイエンス分野における知識融合と発見に貢献することが大きく期待される。