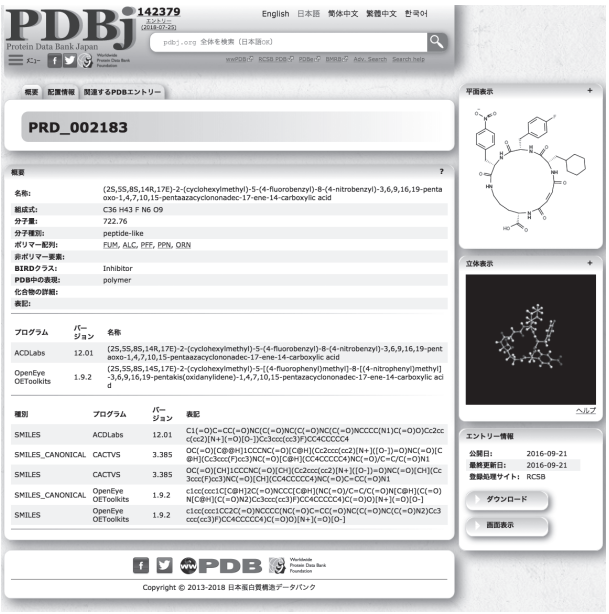
開発者から

蛋白質構造データバンクの データ検証高度化・統合化と創薬研究

栗栖 源嗣
大阪大学蛋白質研究所 教授

蛋白質構造データバンク（Protein Data Bank：PDB）は、蛋白質、核酸、糖鎖など生体高分子の立体構造情報を集めたデータベースである。放射光施設の高度化、超高磁場NMR装置の普及や革新的な電子顕微鏡装置の開発により、PDBとBMRB（NMR実験情報データバンク）、そしてEMDB（電子顕微鏡マップのデータバンク）の登録件数は増加の一途をたどっている。2018年7月末時点でのPDB登録件数は14万件を優に超えている。PDBj（PDB Japan: https://pdbj.org/）は、国際組織であるworldwide PDB（wwPDB：https://wwpdb.org）の一員として、共通した品質管理によりPDB、BMRB、EMDBの各データ登録を行ってきた。また、ダウンロード・サイトの運営や、横断的な検索サービス・ツール類を独自に開発するとともに、ライフサイエンス・データベースの統合的な利用を可能にするため、PDB関連データのセマンティック化を実施している。

今期プロジェクトでは、データベース利用者にデータの品質を保証するとともに、エントリーごとの品質を識別しながら利用することを促す目的で、PDBjではPDBデータの検証と検証レポートの利用拡大に力を入れている。具体的には、これまでPDFフォーマットで提供してきた検証レポートをRDF化し、データの質を機械的に評価できる仕組みを構築中である。

PDB利用研究のうち、機能分子である蛋白質の立体構造に基づいて創薬を加速させようとする創薬研究の割合が、年々大きくなっている。創薬プロジェクトの現場では、蛋白質に結合した生理活性ペプチド様化合物や薬剤など低分子化合物を、高効率・高精度で解析するため、効率よく信頼性の高い構造情報にアクセスすることが重要となっている。これらユーザーからの声に対応して、wwPDBの枠組みの下でChemical Component Dictionary（CCD）やBiologically Interesting Molecule Reference Dictionary（BIRD）の整備が進められてきた。PDBjでは、PDBに含まれる低分子化合物情報の検索機能の高度化を進め（**図**）、化学分野のデータベースであるケンブリッジ結晶構造データベース（CSD）との統合化も進めている。