課題紹介④　10月5日　10:00 - 10:15

蛋白質構造データバンクのデータ検証高度化と統合化

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）では、蛋白質、核酸、糖鎖など生体高分子の立体構造情報を集積し、2017年7月には13万件を超える構造データを公開している。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）の登録件数は増加の一途をたどっている。生体高分子の立体構造に基づいて高次の生命機序までを合理的に説明しようとする構造生物学が、医薬学、農学等とも強く結びつき、創薬等のバイオ産業への応用によりイノベーションにも大きく貢献しつつある。しかし実体のところ、PDBの登録データには品質にバラツキがあるため、PDBj（PDB Japan: https://pdbj.org/）ではデータ検証に力を入れている。利用者にデータの品質を保証し、専門家によるタスクフォースを組織してデータ精度を規格化・データ化する努力を続けている。具体的には、国際的組織であるworldwide PDB（wwPDB：https://wwpdb.org）の一員として、共通した品質管理によるデータ登録作業とダウンロード・サイトの運営、およびデータのセマンティック化を実施・計画している。また、構造情報を裏付ける実験データの同時登録や登録者を特定するためのORCID（http://orcid.org/）の必須化も計画している。新たな動向としては、X線小角散乱や低分解能電子顕微鏡データを併用したHybrid methods（ハイブリッド法）による生体超分子複合体の構造解析が進展している。創薬研究の現場では、薬剤など低分子化合物が生体高分子と複合体を形成する際の結合情報を、より高解像度・高精度で解析するために信頼性確保も重要な課題となっている。これらユーザーからの声に対して、国際組織であるwwPDBの枠組みの下、PDBjではハイブリッド法に拡張したデータ登録システムの構築や、化学分野のデータベースであるCSD（ケンブリッジ結晶構造データベース）との統合化も計画している。さらに、

蛋白質の種々の機能情報が利用できるように開発ツールを高度化するとともに、データベースの利用を促進するための講習会等の活動を幅広く行い、科学の進展に併せてデータベースを深化させていきたいと考えている。