

研究開発課題別中間評価結果

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高度化（PDBj）

2. 代表研究者名

中村 春木（国立大学法人大阪大学蛋白質研究所 教授）

3. 研究開発概要

2001 年度から日本蛋白質構造データバンク (PDBj) を組織し、米国 RCSB-PDB・欧州 MSD-EBI (PDBe-EBI) と協力して wwPDB (world wide PDB) を設立し、国際協力による PDB データベースの維持・管理とサービス・システムの開発・高度化を進めてきた。本研究開発課題においては、データ品質を保ちつつ、さらに大量のデータの登録・編集とその維持・管理・高度化を進めている。また、生体高分子の NMR データ、電子顕微鏡画像、理論モデル等の新たなデータベース構築を進め、グリッド技術も活用して蛋白質の骨格や分子表面の形状などのアナログ情報検索も高速に稼働させる。

4. 中間評価

4-1. 研究開発の進捗状況と今後の見込み

国際連携によるデータベースの開発・維持は順調に進んでおり、毎年行われている wwPDBAC の会合に参加する一方、wwPDB としての開発にも積極的に加わっている。PDBj データベースとサービスの高度化を継続的に実施している。データ登録・品質管理に関しては、2007 年 8 月より Data Uniformity project の成果を反映したファイルフォーマットでのデータ提供を開始している。また、2008 年 2 月より、データの登録に際して実験データの提供を必須とし、データベースの信頼性向上および格納データの充実に資することとしている。生体分子に対する NMR 実験のデータベースである BMRB (BioMagResBank) の構築と高度化に関しては、BMRB のミラーサイトを運営するとともに、NMR 実験データの登録作業も開始している。登録を支援するための「BMRB データ登録ウェブサイト」を開発し、専任アナレータを増員してこの登録・編纂事業をさらに継続するとともに、固体 NMR に対するデータ収集を特に推進している。さらに、タンパク 3000 プロジェクトにおいて多くの構造決定を行った大量の化学シフトデータを BMRB に登録するための方法を開発し、実施している。

研究開発は、事業の規模に合った人数を、開発内容にあった専門分野から確保して着実に進められており、増加するデータ量にもよく対応している。また、グリッド技術など最新のインフォマティクス技術を利用したデータベースサービスが進展している。当初の研究開発計画を若干前倒しにする形で進められている。

4-2. 研究開発成果の現状と今後の見込み

web 公開状況 PDBj トップページ <http://www.pdbj.org/>

国際協力の下生体高分子構造データベースの維持・管理を継続して進めている。PDBj は日本、アジア、オセアニア地区におけるデータ登録を担当し、常に世界全体の 27~28 % の登録をしている。

XML データベースを利用した xPSSS 閲覧システムを XQuery および XPath で利用する支援サービスとして、XQuad (PDBML XQuery Advisor)を開発・公開し、PDBML の知識のないユーザでも XQuad や XPath の利用ができる仕組みを整えている。一方、類似フォールドやファミリーの検索サービスをさらに高度化し、蛋白質フォールド全体を俯瞰的に理解できる Protein Globe や SeSAW のサービスを開発・公開している。さらに、PDBj の Web page を使い易くするためデザインも含めて修正し、中国語およびハングルによる新しいページも作成・公開している。EBI で収集している 3 次元電子顕微鏡構造データベース (EMDB) のデータ閲覧サービスである EM Navigator を開発し、データ公開を開始している。EMDB の全データ 587 件を登録し、ムービーを作成し公開している。

生体高分子のアナログデータ検索の高速な実施について継続的に高度化を進め、「蛋白質分子表面・機能データベース (*dF-site*) と検索システム (*dF-seeK*) の構築」、「蛋白質の動的性質のデータベース (*PrdMode*) と計算サービス」を、一部グリッド技術も用いながら高速に稼働させる仕組みを開発し、運営している。教育的なサイトとしての英語と日本語による蛋白質構造事典: *dProtS* (encyclopedia of Protein Structures)について、Wiki の仕組みを使って大幅に修正し、内容の拡充とともに、蛋白質および蛋白質構造についての情報が手軽に入手できるようにしている。さらに RCSB-PDB が作成している教育サイトである MOM (Molecule of the Month)について、日本語訳を作成し各月の初めに同時公開を行う仕組みとして稼働を始めている。

BMRB データベースのミラーサイトの維持・管理を行い、NMR 実験データの登録および研究開発を行っているが、2006, 2007, 2008 年 (10 月まで) には、それぞれ 98, 130, 44 件の登録を実施しており、世界全体の 13~20%に対応している。蛋白質フォールド・機能・進化のデータベースと検索システムの構築では、(a) フォールドの分類とそれを利用した高速検索システム (ASH) の構築と公開、(b) 構造・配列統合アライメントシステム (MAFFTash) の構築と公開、(c) 進化トレース解析の自動化とデータベース化 (β 版) の公開を行っている。

上記のとおり国際標準データベースの一部として十分な実績を上げており、データベース開発そのもの、データベースを利用した研究、どちらからも十分な成果が出されている。データベース高度化・標準化の研究成果が着実に出てきている。

4-3. 今後の研究開発にむけて

wwPDB メンバーの一員として、特に日本国内およびアジア・オセアニア地域の研究者との連携をはかりつつ、生体高分子の構造データベースを維持・発展させていき、日本国内のライフサイエンス統合データベースセンターに対しても、PDBj は基盤データベースとして、そのセンターの活動に積極的に協力していくこととしている。データベースの高度化については、データ閲覧システムを web サービス化し、開発を進めている種々のサービスについても、さらに利用しやすくするような工夫をする予定としている。

wwPDB Common Deposition and Annotation Project について、開発項目の精査および実装を進めていき、wwPDB が提供する chemical component library については、利用者が構造解析に使用するプログラムとの相互運用性を改善するため、フォーマットの文法的妥当性確認や情報の一部欠損等について検証および必要に応じて修正を行う計画である。

NMR によって解析された生体高分子構造データの登録では、化学シフト情報の添付を必須項目とすべきであるという議論があり、今後、BMRB データ登録数が急増する可能性が高いが、PDBj-BMRB 登録では、化学シフトも立体構造も同時に登録する仕組みが開発されて

おり、この ADIT-NMR を利用した PDB 登録手順書も公開する予定である。

フォールドの分類とそれを利用した高速検索システムの構築については、さらなる高速高感度検索システムの開発と、それに並行してこれまでに開発してきた検索プログラムの応用展開と、そのフィードバックによる改良を行って行くことが予定されている。

PDBj の維持発展にとって妥当な方向性を持った研究開発計画となっている。しかし、データ量の増加と高度化の進展が極めて早いので、後継者を含め本事業を維持するための方策の検討を始めてほしい。

4-4. 国内外のデータベース高度化・標準化の動向・状況と本課題の位置づけ

本研究開発課題の代表研究者は、2001 年度から JST のバイオインフォマティクス推進事業 (BIRD) における「蛋白質立体構造データベースの高度化」として PDBj を組織し、米国 RCSB-PDB、欧州 MSD-EBI と協力して、2003 年に wwPDB を設立し、国際協力による PDB データベースの維持・管理とサービス・システムの開発を進めている。さらに 2006 年からは米国 BMRB が NMR の実験データベースの構築を実施する機関として wwPDB に加わり、世界の 4 極で wwPDB の維持・運営が行われている。

この蛋白質を中心とする生体高分子立体構造のデータベースは、日本はもとより、アジア・オセアニア地域においては本 PDBj が唯一の機関として、他 3 組織との共同作業によって品質を保持しつつ、単一のデータアーカイブを管理・運営し、さらにデータベースの改善と新たなサービスの拡充に努めている。日本の PDBj のデータ登録への寄与は、2006-2008 年度において常に世界の 27~28% となっており、本研究開発の世界の生命科学への寄与は極めて大きいと言える。

登録データの品質管理は極めて重要であり、wwPDB 国際諮問委員会による討議と構造生物学の国際的な Society のバックアップによって、wwPDB のポリシーとして、登録時にデータ品質について精査を実施するとともに、実験データも原子座標データと同時に登録すべき必須項目となっている。構造生物学の進展とともに、従来のフラットな PDB フォーマットだけでなく、mmCIF, PDBML (XML 形式) などの情報学的に進んだフォーマットも同時に修正し、version 3.2 を確立中である。

PDBj-BMRB は国際的 NMR データの供給を目的に、BMRB との共同開発を行ってきた。BMRB ウェブサイトのミラーを PDBj-BMRB で公開し、BMRB 登録処理も開始している。国際的 NMR データベースの供給を BMRB と PDBj-BMRB の 2 極で担うまでに成長したと言える。

このように PDBj は当初から国際協調の枠組みの中にあり、高度化、標準化も協調して進められている。国際連携によるデータベースの開発・維持が順調に進んでおり、国際的な貢献をしている。

4-5. 研究開発成果の社会への貢献

本研究開発で構築する蛋白質立体構造データベースは、Structure-Based Drug Design (蛋白質立体構造に基づく医薬品設計) あるいは Structure-Guided Drug development (構造に指南される医薬品開発) と呼ばれる手法に応用されたり、NMR の化学シフト値の同定や高次構造の決定の迅速化、または人工蛋白質を設計・創製する蛋白質工学など、医・薬学を含む生物系研究者や製薬メーカーに蛋白質の立体構造データをフリーで使いやすい形で一元的に提供することで、産学を問わず、社会に大きく寄与している。

PDB データベースに対するアナログ情報検索手法は、データベースの高度利用のために重要である。さらに、BMRB と新たに構築する生体超分子の電子顕微鏡画像データベースは、

多くの情報を生命情報科学者へ与え、基礎研究に役立つだけでなく、医療や薬物開発に関連する標的蛋白質の知見が迅速に得られることが期待される。分子グラフィックス・ビューア (jV) の開発や教育用データベースである蛋白質構造百科 (dProtS、MOM) の開発と普及は、生体高分子の立体構造に関する高校生や一般の人々の啓蒙を目指しており、社会への直接的な貢献である。

PDBj はライフサイエンス分野の基盤データベースであり、現在も基礎応用両面で大きな貢献をしている。創薬などいくつかのバイオ産業で、蛋白質立体構造データベースは必須の情報基盤となっており、サービス高度化の研究開発は意義が大きい。将来の重要性はさらに高まると考えられ、その高度化・標準化がもたらす貢献は大きい。

5. 総合的評価

PDBj は日本、アジア、オセアニア地区におけるデータ登録を担当し、常に世界全体の 27 ~ 28 % の登録をしており、国際協力体制の一極を担う揺るぎない位置において、PDB データベースの維持・管理とサービス・システムの開発を行っている。wwPDB のデータ更新を国際協調体制で進めるなど国際的な貢献が大きい。国際スタンダードを備えた重要なデータベースとして長期的に維持、管理されるべきであり、データベース運営における持続可能なビジネスモデルを今後日本において提示しうる、数少ないデータベースの 1 つであると思われる。増加するデータへの対応やグリッド技術など最新のインフォマティクス技術を利用したデータベースサービスを運用している点が評価できる。機関としての蛋白研のサポートが積極的でないことがやや懸念されるが、今後の研究計画に沿ってさらなる高度化を期待する。