

課題名「蛋白質立体構造データベースの高度化」

代表研究者 大阪大学蛋白質研究所 教授 中村 春木

1.代表研究者による成果概要報告

1-1.研究開発のねらい

欧米・日本を中心とする国際的な構造ゲノム/構造プロテオミクス・プロジェクトの進展によって、多くの蛋白質立体構造が従来に比べて迅速に決定される時代を迎えつつある。蛋白質立体構造データベース(PDB: Protein Data Bank)に登録されている立体構造データは 1990 年代から急増し、毎年世界で 5,000 件以上の実験によって決定された構造が登録される。PDB は地球規模で利用されてきたものの、構造生物学研究の成果をまとめあげただけのデータベースに留まってきた。そこで本研究開発では、PDB へのデータ登録作業に積極的に協力して蛋白質の立体構造とゲノム情報との結びつきを強める一方、XML などの最新情報技術を利用し、国際的な連携のもとに世界標準としての新しいデータ記述(PDBML)と解析ツールや二次データベースを開発して付加価値を付け加え、構造生物学者だけを対象とする専門的なデータベースから、広く生命科学の研究者、産業界、さらには一般の人にも役立つデータベースに高度化することを目的とする。

1-2. 研究開発の成果

1-2-1. PDB データベース業務の日本の分担作業の実施

PDBj では、専門のキュレータを育成し、wwPDB の一員として、日本国内はもとより、アジア・オセアニア地区からの登録を原則的に担当している。PDBj における登録処理件数は、2001 年 356 件、2002 年 648 件、2003 年 935 件、2004 年 1,586 件、2005 年 2,101 件と急増している。世界全体では 2005 年は 6,043 件であり、日本の寄与は 34.8%にのぼった。

PDBj において生体高分子の立体構造を登録する際には、インターネット上の登録サーバ ADIT を利用する。日本語による案内ページを作成・公開し電子メールによる PDB データ登録に関する質問を受付け、登録者に対する便宜を図っている。

一方、生体分子に対する核磁気共鳴(NMR)実験データのデータベースである BMRB(BioMagResBank : 米国 Wisconsin 大学が開発) のミラーサイトを共同して立ち上げ、日本での登録作業も 2005 年から開始した。これまでに、31 件の登録が PDBj においてなされている。

1-2-2.蛋白質立体構造データベースの標準化記述

PDB では、Brookhaven National Laboratory による伝統的な PDB フォーマットと呼ばれるフラットファイル・フォーマットが現在でも標準的に用いられているが、いろいろな歴史的経緯から、記述法が必ずしも統一的でなく、特に

古いエントリーに関しては正しくデータがフォーマットどおりに記述されているかどうかの検証(以下、このことを validation と記す)も十分とは言えない状態にあった。そのため、データベースにアクセスする際には、利用者はその多様な記述法と例外処理に悩まされてきた。PDB データそのものの品質管理を将来にわたり保つだけでなく、他のゲノム配列やプロテオームデータベースとのデータグリッドなどによる統合化のためには、validation のためのツールがそろっており、他のデータベースにおいても最近利用が始まった eXtensible Markup Language (XML) を用いた記述が強く望まれていた。我々 PDBj のグループは米国の RCSB-PDB と共同し、国際標準としての正規 XML 記述を PDBML という名称で 2004 年に確立した。XML による一般的な問題として増大するファイルサイズの問題を避けるため、データの 9 割以上を占める原子情報を別ファイルに格納する ext-atom 方式が、PDBj からの提案として採用された。この方式によって検索が高速となり、ファイルサイズも 2~3 倍程度に納めることが可能となっている。

さらに、拡張が容易に行える XML 書式の特長を活かして、オリジナルの PDB データに多く欠損している分子機能や実験条件等の情報を文献から抽出して追加する作業を行い、その数は 16,221 件(平成 18 年 2 月 20 日付け)に達している。同時に関連する 13,997 件の文献をスキャンし、PDF および WORD の書式による電子化を行って蓄積しアーカイブも作成している。これらの文献から抽出された情報は XML 化されオリジナルの PDBML データに加えられ、native XML-DB による蛋白質構造データ検索システムを構築して、一般に公開している。この検索システムは、native XML-DB の特徴を生かして、XPath による検索が行える Web サービスも実現しており、SOAP (Simple Object Access Protocol) サービスも運用している。

これらテキスト・ベースでの情報検索に加えて、2006 年 3 月には、蛋白質の類似構造(フォールド)の検索(Structure Navigator)と、eF-site を拡張した蛋白質類似分子表面の検索のサービスを開始し、蛋白質構造・形状というアナログデータが直接アナログ情報のままで検索できるようにした一方、PDB にこれまで蓄積されてきた実験データである構造因子から電子密度マップ(2Fo-Fc)を作成して DB 化し、その interactive 画像表示を行うサービスを、jV3 を利用して運用を開始した。

1-2-3. 蛋白質構造を基にした解析システムの開発と二次データベースの構築

スタンドアローンとしてもアプレットとしても利用できる Java ベースの新しい 3D ビューア(PDBjViewer (jV))を JOGL を採用して実行速度を確保できるものを開発し、パブリックドメインとしてソースおよびバイナリプログラムの公開を行った。また、このアプレットを用いて、xPSSS、eProtS、eF-site などの PDBj が開発した様々のサービスのビューアに応用している。

一方、蛋白質表面形状と物性に関するデータベース(eF-site)、蛋白質ダイナミクス・データベース(ProMode)、マルチプル・アラインメントによる蛋白質立体

構造の比較サービス(ASH/GASH)の開発を行い、いずれも Web ページ上で公開している。

高度化された PDB データベースを基にして、立体構造解析を専門としない一般の生物学・生化学の研究者や大学生および高校生を対象とした、教育用の蛋白質構造データベース(eProtS: encyclopedia of Protein Structures)の英語版と日本語版の公開を継続的に進めた。これまでに 236 件の蛋白質についての解説を実施している。

1-2-4. wwPDB の活動と諮問委員会の開催、その他の広報・学会活動

国際蛋白質構造データバンク(wwPDB)の第 1 回諮問委員会(wwPDBAC: world wide Protein Data Bank Advisory Committee)を、米国ワシントンDCにおいて 2004 年 11 月 21 日に開催した。Structural GenomiX社のStephen K. Burley 博士が議長となり、その他にWayne A. Hendrickson 教授(RCSB, コロンビア大学)を始めとするPDB 各サイトの代表諮問者、国際学会の代表者、wwPDB 各サイトのリーダー、運営開発基金の配分機関代表者ら、総勢 16 名が参加した。平成 17 年 8 月 30 日には、イタリア・フィレンツェにおいて、wwPDB の第 2 回助言者委員会が開催された。会議では、PDBデータ記述の誤記述の修正と統合化、wwPDBとBMRBとの関係の強化、電子顕微鏡データベースとの協力、理論モデルの取り扱い、登録数の急増への対応、wwPDB諮問委員会メンバーによる運営開発基金の配分機関への働きかけの必要性等が議論され、wwPDB のメンバーに具体的な課題が課された。PDBデータ記述の誤記述の修正と統合化における日本の分担は、Primary Citationに対しての誤記述や記述の欠落を修正・追加することであり、平成 17 年 2 月から開始し、人手によって、多くの誤記述の修正と欠落している文献情報を追加した。

その他の広報・教育関連の活動として、ニュースレターを年間 2 回発行する他、国際学会(2004 年 4 月第 1 回環太平洋蛋白質科学国際会議(横浜)、2005 年 8 月第 20 回国際結晶学会(フィレンツェ))・国内学会(2004 年 12 月日本生物物理学会第 42 回年会(京都))へのブース出展やシンポジウムの共催による開催を行った。また、国立遺伝学研究所のDDBJ と協力し、大阪大学中之島センター・キャンパス・イノベーションセンターを利用して、DDBJing & PDBjing –講習会 in 大阪–を、2004 年 3 月 2 日、2006 年 2 月 2-3 日の 2 回にわたって開催し、一般社会人や関西、中国地方の大学生・大学院生を対象として、データベース利用法についての講習会を無料で行った。2 回とも約 25 名程度の参加者があり、講演と、自らの PC を持ち込んで行う実習が実施され、好評であった。

2. 事後評価結果

2-1. 当初計画の達成度

国内では、「日本蛋白質構造データバンク(PDBj:Protein Data Bank japan)」を設立し、複数の国内の大学の研究室をサテライト・サイトとする研究開発体

制を構築する一方、国際的には、米国 RCSB や欧州 EBI と共同で「国際蛋白質構造データバンク(worldwide Protein Data Bank:wwPDB)」を設立して国際的な連携を強化させ、研究開発を推進した。

当初計画の、1)新規蛋白質立体構造データベース構築、2)蛋白質構造を基にした解析システムの開発と二次データベースの構築、3) PDB データベース業務の日本の分担作業の実施、4) BioMagResBank(BMRB)の日本の分担作業の実施、5)日本国内での教育用データベースの作成と公開、はどの項目も実施され、その成果が Web を通して公開されている。それぞれの内容が充実しており、想定以上の成果が得られ、望ましい展開となった。

2-2.知的財産権、外部発表(論文等)等研究開発成果の状況

PDBjホームページ：<http://www.pdbj.org/>

RCSB-PDBデータ閲覧サイト：<http://pdb.protein.osaka-u.ac.jp/pdb/>

xPSSSブラウザサイト：<http://www.pdbj.org/xpsss/>

PDBj登録サーバ：<http://www.pdbj.org/deposit.html>

BMRBミラーサーバ：<http://bmrbl.protein.osaka-u.ac.jp>

BMRBデータ登録日本語ウェブマニュアル：

http://bmrbldep.protein.osaka-u.ac.jp/manual_top.html

蛋白質分子表面形状と物性・機能データベースeF-site：

<http://ef-site.hgc.jp/eF-site/>

蛋白質分子表面計算サーバeF-surf：<http://ef-site.hgc.jp/eF-surf/>

蛋白質類似分子表面検索サーバeF-seek：<http://ef-site.hgc.jp/eF-seek>

分子グラフィックス・プログラム jV (PDBjViewer)：

<http://pdbj.protein.osaka-u.ac.jp/PDBjViewer/>

蛋白質ダイナミクス・データベースProMode：

<http://promode.socs.waseda.ac.jp/>

蛋白質立体構造比較ツールASH：<http://timpani.genome.ad.jp/~ash/>

教育用蛋白質構造百科事典eProtS：

http://eprots.protein.osaka-u.ac.jp/eProtS/Top_e.jsp(英語版)

<http://eprots.protein.osaka-u.ac.jp/eProtS/Top.jsp> (日本語版)

蛋白質機能情報抽出支援システムPROFESS：未公開

原著論文発表、学会発表ともに海外におけるものが多く、国際的に評価されている。各データベース・ツールはほぼ公開され、アクセス数は順調に増加している。また、ニュースの定期発行、講習会開催、学会展示などを積極的に行い、研究開発成果の普及に努めている。

2-3. 研究開発成果の公開による波及効果

native XML-DB のデータベース xPSSS では、単なる蛋白質の原子座標だけが入った従来の PDB データベースを、ゲノムやネットワーク等の他の様々なバイオインフォマティクス関連データベースと統合して理解できる仕組みへと高度化している。また、利用者への便を図るためにビューアや各種ツールを開発し、講習会や学会発表を通じて、蛋白質立体構造に対する理解を浸透させた。XML-DB や SOAP の利用は、実施当初は情報科学分野で研究的に行われていたものを先取りする形で運用を始めており、ライフサイエンスだけでなく、国内の情報科学全体へも実施例としての波及効果を及ぼした。また、PDBML の開発は、ライフサイエンスにおけるデータグリッドの意味付けを高め、推進する意義を与えた。PDBML 開発等のデータ記述での標準化については、米国 RCSB-PDB との密な協力関係によって達成され、30 年以上に渡る懸案であった PDB 記述法における誤記の問題点をクリアーし、全くエラーのないデータベースが構築された。本研究開発の高度化・標準化は非常に意義があり、成果は多方面に大きな影響を与えている。

2-4. 成果の実用化の可能性及び成果から予想される波及効果

当研究開発で構築してきた蛋白質立体構造データベースは、医・薬学を含む生物系研究者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おり、今後はますます意義が高まると予想される。さらに、BMRB のデータベースは、PDB に加えて、NMR 実験からさらに多くの情報を生命情報科学者へ与え、構造・機能についてのより詳細な情報抽出に寄与し、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(eProtS)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が期待できる。

2-5. 総合評価

当初の想定を大きく上まわる、研究開発成果を上げた。世界に発信する我が国の蛋白質構造データベースの中核として、PDBj の設立と発展は大きな意義がある。特に当初想定になかった wwPDB の一員として国際的に大きな役割と責任をになうようになったことは、注目に値する。これらの活動を実現した代表研究者のマネジメント努力は高く評価できる。

生体高分子の立体構造データは今後さらに増大すると予想される。データベースについて精度の高い品質管理を行うとともに、高度なデータ検索サービスや利用しやすいツールの開発を進め、利用しやすく信頼性の高い蛋白質の構造と機能についての情報源として国内外へ発信し、人材育成、普及・教育活動等も積極的に行って、さらなる社会への貢献を期待するものである。国内唯一の国

際拠点である PDBj は国家戦略的に大変重要であり、さらに体制を充実させて他プロジェクトとの連携を深めながら継続して行うべき事業である。

3.主な論文発表

- 1)Standley DM, Toh H, Nakamura H “Detecting Local Structural Similarity in Proteins by Mixing the Number of Equivalent Residues”, *Proteins: Structure, Function and Genetics*, Vol. 57, pp. 381-391 (2004)
- 2)Kinoshtia K, Nakamura H “eF-site and PDBjViewer: database and viewer for protein functional sites.” *Bioinformatics* Vol. 20, Iss. 8; pp. 1329-1330 (2004)
- 3)Wako H, Kato M, Endo S “ProMode: a database of normal mode analysis on protein molecules with a full-atom model”, *Bioinformatics*, Vol. 20, Iss. 13; pp. 2035-2043(2004)
- 4)Kinoshita K, Nakamura H “Identification of the ligand binding sites on the molecular surface of proteins” *Protein Science*, Vol. 14, pp. 711-718 (2005)
- 5)Westbrook J., Ito, N., Nakamura, H., Henrick K., Berman H.M. “PDBML: The representation of archival macromolecular structure data in XML.” *Bioinformatics* Vol. 21, Iss. 7; pp. 988-992 (2005)

参考

論文発表 国内 4 件、海外 26 件
口頭発表 国内 50 件、海外 15 件
ポスター発表 国内 36 件、海外 12 件
ニュースレター発行 7 号