

News Letter

Vol. 2
June 2003

蛋白質立体構造データベース高度化事業
ニュースレター

Protein Data Bank Japan
<http://www.pdbj.org/>

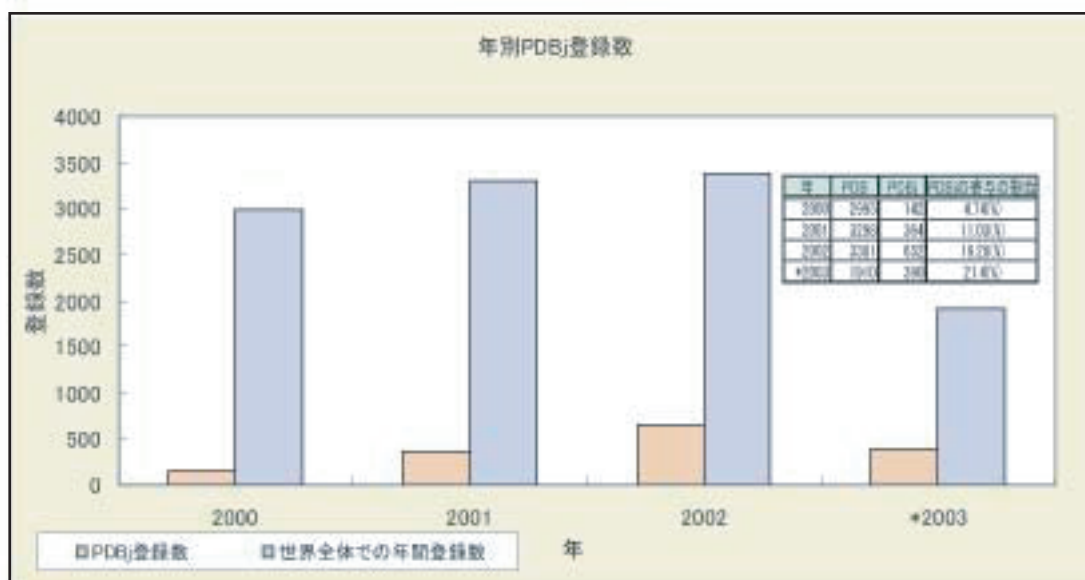
PDBj
Protein Data Bank Japan

現在、多くの遺伝子情報が解き明かされようとしています。遺伝情報を有効利用するためには、各遺伝子がコードする蛋白質分子を理解することが必要です。特に、各蛋白質の立体構造がその特異的な機能発現を与えているため、分子レベルにおいて遺伝子の進化と生物学的機能を把握するには蛋白質の構造は不可欠です。私達PDBjでは、科学技術振興事業団の支援を受けて、これら蛋白質立体構造のデータベースを運営しております。



このデータベースは、米国のRCSB (構造バイオインフォマティクス研究共同体)、欧州EBI (ヨーロッパ・バイオインフォマティクス研究所) のMSD (高分子構造データベース) グループと共同で運営しており、データの登録・編集・配布作業を行っております。私もPDBjの活動についての詳細は、<http://www.pdbj.org/>をご覧ください。

PDBj構造登録データ統計



*2003年5月31日現在

2003年5月31日現在、世界全体でのPDBデータベースの登録数は21,055となり、PDBjの寄与の割合も年々増加しています。また、Webページ上における登録案内ページ、一日平均のアクセス要求数は6,894件、また1日平均ページビュー数は717件とこちらも増加しています。

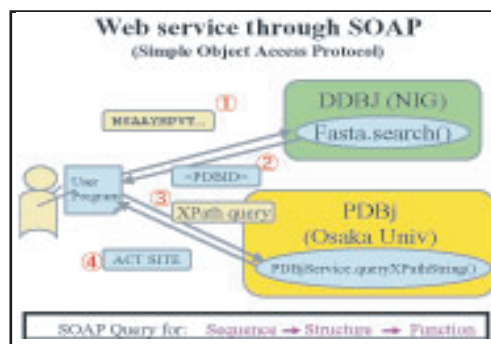


PDBのXML化と高度検索システム

昨年度から私どもは、mmCIFで記述されたPDBデータ内容が損なわれないよう注意を払いつつ、PDBデータをXML(eXtensible Markup Language)記述データへ変換する試みをしてまいりました。この準備段階としてのXMLデータベース開発を活用し、RCSB、EBI、PDBjの3者が協力して、PDBデータの正規XML記述法を、2003年6月までに確立いたしました。含まれるデータ全てにタグがふられたもののスキーマは<http://deposit.pdb.org/pdbML/pdbx-v0.904.xsd>に、また原子座標部分を他のXMLファイルに分離したもののスキーマは<http://deposit.pdb.org/pdbML/pdbx-v0.904-alt.xsd>に書かれています。

XML記述法の特徴である拡張性を利用して、さらに私たちは、個々の蛋白質に関連するより多くの情報を追加する努力を行っております。PDBで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し、その数は現在2000を超えています。上記PDBの正規XML表現スキーマを用い、私どもが追加した情報も加えて試験的なPDBデータのXMLデータベース(XML-DB_PDBj)を準備しております。

XMLデータベースの検索のために、X-Pathを用いた多様な検索手法を提供しています。XML化されたデータベースの高度な利用としてWebサービスの提供が挙げられます。これはSOAP(Single Object Access Protocol)と呼ばれるプロトコルによって実現しています。私どものこれら試験的なWebサービスは、http://www.pdbj.org/XML-DB_PDBj/で公開しております。



eF-site データベース

現在、蛋白質の構造と機能の関係を特定することが重要視されています。私たちは、蛋白質機能のうちで最も重要な分子認識において、その形状と物性が深く関わっていると考えられる分子表面の物性(静電ポテンシャル)と形状に着目しています。

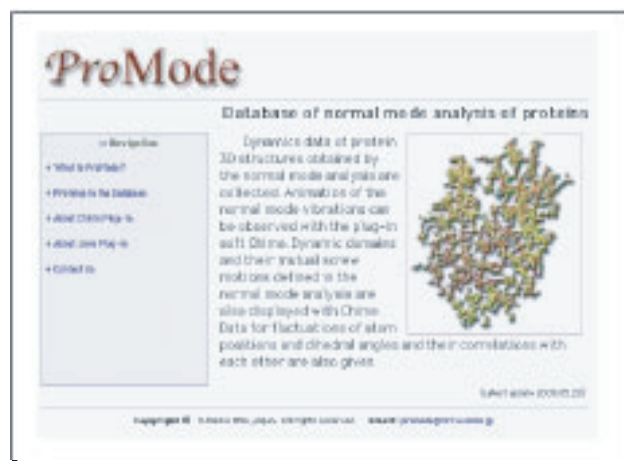
eF-Site(蛋白質機能部位データベース)では、これら分子表面の形状と物性をグラフィック表示しており、またこれらの情報に基づく検索も行えるようになっています。その機能部位情報として、これまでの抗体、prosite、活性部位、細胞膜、結合部位に、今年はモノヌクレオチド、DNAの情報を加え拡充しました。また、より利便性を高くするために表面形状と物性を同時に見ることが出来るように新たな分子表示プログラム(PDBビューア)を従来の分子画像機能により開発し、Webに公開しました。

現在、PDBに登録されている全低分子結合蛋白質を対象とした、より網羅的な結合部位データベースを近くWebにて公開する予定です。



ProMode データベース

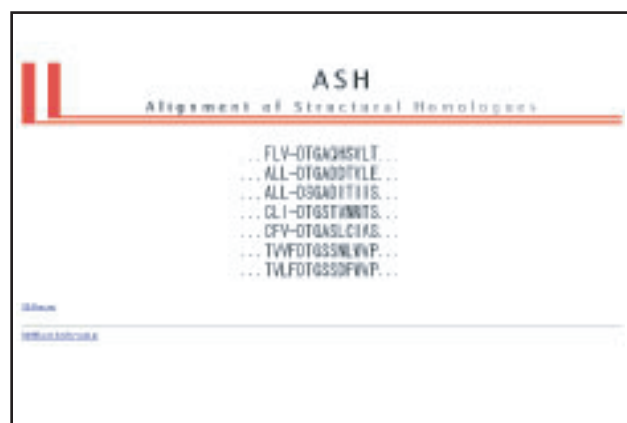
*ProMode*は早稲田大学、輪湖教授チームによって開発中の蛋白質の基準振動解析(NMA)のデータベースです。蛋白質の動的特徴(ダイナミクス)を理解することは、蛋白質の構造と機能の構築原理を知るためには大変重要です。NMAは調和振動近似に基づいており、その結果は多くの場合、合理的であるとされています。NMAは構造エネルギーの極小化に最も多くの計算時間を費やしますが、ダイナミクスのシミュレーションに通常使用されているモンテカルロ・シミュレーションや分子動力学計算に比べると、かかる時間や手順を大幅に軽減することができます。*ProMode*は迅速で効率的な計算法として、輪湖教授と日本原子力研究所・京都大学郷信広名誉教授らのグループが開発したプログラムFEDER(Noguti and N. Go (1983) *J. Phys. Soc. Jpn* **52**, 3685; H. Wako and N. Go (1987) *J. Comp. Chem.* **8**, 625 参照)を導入しています。現時点で、80の蛋白質が利用でき、近いうちには、より多くの蛋白質が解析され、登録される予定です。



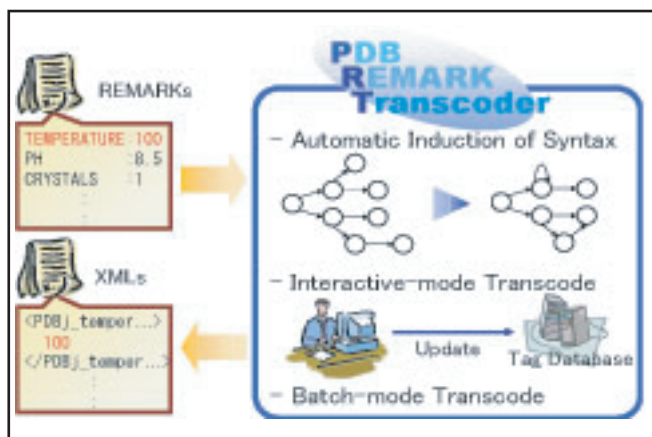
立体構造比較ツールの開発

蛋白質の立体構造の情報を利用して、その残基間の対応を行うことを構造アラインメントとよびます。より多くの配列を比較(マルチプル・アラインメント)することによって、抽出される情報の信頼性が向上します。1989年にTaylorとOrengoによって提案されたdouble dynamic programming アルゴリズムに基づき、京都大学、藤教授チームは蛋白質立体構造のマルチプル・アラインメントを構築するシステム-ASH-を構築しました。このシステムには、1997年に藤教授が導入した距離カットオフ近似と2-ステップアラインメント方式の採用、及びアラインメントへの二次構造情報の導入によって、処理速度を向上させると同時にアラインメントの精度を維持しています。

現在このシステムは<http://timpani.genome.ad.jp/~ash/>とPDBjのホームページにて公開しています。



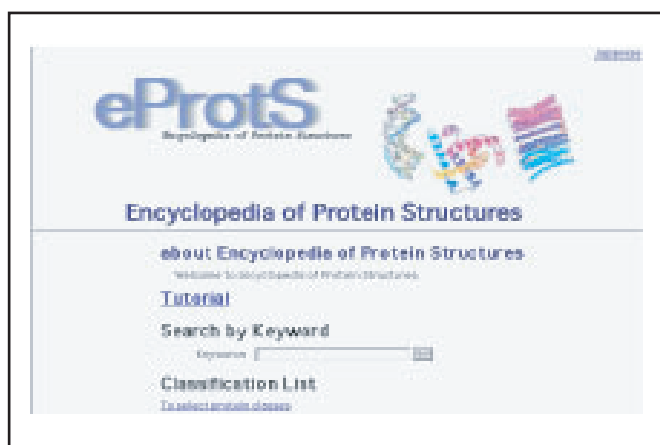
PDB REMARK transcoder



PDBのREMARK行に内包される各種情報を計算機により自動抽出し、これを元にXMLフォーマットに変換する効率的ツールを大阪大学、大川助教授チームが開発しました。まず、PDB REMARK transcoderにPDB内の100エントリーのREMARKデータをもとに、記述パターンを自動学習させ、これを用いて8906エントリー(約520万語)を対象としたXML変換を実施しました。その中から任意

に抽出した10エントリーを対象として正確にタグ付けがなされているかを調査しました。記述パターン抽出時におけるクラス属性値とトークン属性値の距離を元に自動的に5段階のレベルにわけ、そのうちもっとも信頼性度の高いレベルに於いては99%以上正確にタグ付けが付与されていました。また、8906エントリー全てを対象としたログ集計の結果、もっとも信頼性が高いレベルが85%を占めていました。これらのことから、PDB REMARK transcoderは、人為的最終段階調整の前段階においては極めて有効であると言えます。私たちは既知のデータでのトレーニングによりtranscoderの信頼性を高めています。

eProtS(教育用データベース)



PDBデータベースは、構造生物学の専門家のためのデータベースであり、またすべて英語で表記されています。そのため構造生物学とは異なる分野の生物学者や学生、また、興味を持つ一般の人々にとっては、使いやすいデータベースとは必ずしもいえません。そこで、データベースをXML化した利点を活かし、一般的に必要な、また分かりや

すい蛋白質の情報を掲載したeProtS: encyclopedia of Protein Structures(蛋白質構造百科)を開発中です。現段階では100個程度の蛋白質を英語版、日本語版で公開しています。

統括責任者

中村春木(大阪大学蛋白質研究所・教授)

PDBjデータベース管理運営グループ:

楠木正巳(大阪大学蛋白質研究所・助教授)
小佐田高史(大阪大学蛋白質研究所・技官)
五十嵐令子(科学技術振興事業団・研究補助員)
見学有美子(科学技術振興事業団・研究補助員)
佐伯忍(科学技術振興事業団・研究補助員)
森田恭代(科学技術振興事業団・研究補助員)

新規蛋白質立体構造データベース構築グループ:

伊藤暢穂(東京医科歯科大学大学院・教授)
Standley, M. Daron(科学技術振興事業団・研究員)
小林香織(科学技術振興事業団・技術員)
坂本久(科学技術振興事業団・技術員)
清水有希子(科学技術振興事業団・研究補助員)
高橋亜紀(科学技術振興事業団・研究補助員)

eF-Siteデータベース開発グループ:

木下賢吾(横浜市立大学大学院総合理科学研究科・助手)

ProModeデータベース開発グループ:

輪湖博(早稲田大学社会科学部・教授)

ASH開発グループ:

藤博幸(京都大学化学研究所バイオインフォマティクスセンター・教授)

PDB REMARK transcoder開発グループ:

大川剛直(大阪大学大学院情報科学研究科・助教授)
兼田佳和(科学技術振興事業団・研究補助員)

BMRBデータベース管理運営グループ:

阿久津秀雄(大阪大学蛋白質研究所・教授)
松本陽(科学技術振興事業団・研究員)
中谷英一(科学技術振興事業団・技術員)

研究チーム事務員

鎌田知佐(科学技術振興事業団)



(後列左から)

中村春木、鎌田知佐、坂本久、Standley, M. Daron、小佐田高史、中谷英一、
佐伯忍、高橋亜紀、清水有希子、小林香織、
森田恭代、五十嵐令子、見学有美子、楠木正巳(敬称略)

連絡先

〒565-0871
大阪府吹田市山田丘3-2
大阪大学蛋白質研究所・附属プロテオミクス総合研究センター内
PDBj事務局: TEL(06)6879-4311、FAX(06)6879-8636
PDBjデータベース登録業務: TEL(06)6879-8638、FAX(06)6879-8636
E-mail: pdbj@protein.osaka-u.ac.jp