

蛋白質構造データベースの国際的な構築と統合化

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

代表研究者: 中村春木(大阪大学蛋白質研究所・教授)
[代理: 金城玲(同・准教授)]

1. 本研究開発課題の目的・要旨

国際協力のもと、蛋白質構造データバンク(PDB)の構築・運営を実施するとともに、NMR実験情報のデータバンク(BMRB)、電子顕微鏡やX線小角散乱による構造情報データベース(DB)を構築し公開する。オントロジー、データ管理システム、新たなデータ書式、利用者インターフェースや二次的DBおよび種々のサービスも開発して公開し、PDBの高度化・統合化を図り、人材育成も実施して、他の生命DBとの統合化による高次生命機能の理解に資する研究開発を行う。

The screenshot shows the PDBj website homepage. At the top, there's a navigation bar with language options (English, Japanese, simplified Chinese, traditional Chinese, Korean) and utility links (統計情報, ヘルプ, FAQ, お問い合わせ). Below this is a main header with the PDBj logo and a brief description of the Protein Data Bank Japan. The left sidebar contains various navigation links such as 'データ登録', '検索', 'サービス&ソフトウェア', and 'ダウンロード'. The main content area features a search bar with options for 'PDB検索' and 'Mine', and a '最新情報' (Latest News) section with several entries dated from 2011/5/27 to 2011/12/6. The right sidebar includes a '73503 entries available on 1 Jun. 2011' counter, logos for 'eProtS', 'Protein Globe', 'DBCLS', 'Tanpaku.org', and 'NBDC'.

PDB (Protein Data Bank):

X線結晶解析、核磁気共鳴法 (NMR)、電子顕微鏡観測によって得られた73,600件以上の蛋白質・核酸・糖などの生体高分子の3次元立体構造の国際的データベース。原子種とその座標、アミノ酸残基、実験手法、実験時の情報、実験観測データを含む。

2. 背景とこれまでのPDBjの活動状況

(2-1)国際蛋白質構造データバンク(wwPDBとwwPDB Foundation)の創設と協力

(2-2)蛋白質立体構造データバンク登録作業

2006~2010年には登録数10,450件:
世界の26%の登録をPDBjにて実施。

(2-3)BMRB(NMRデータベース)登録作業

2006~2010年には登録数628件:
世界の18%の登録をPDBj-BMRBにて実施。

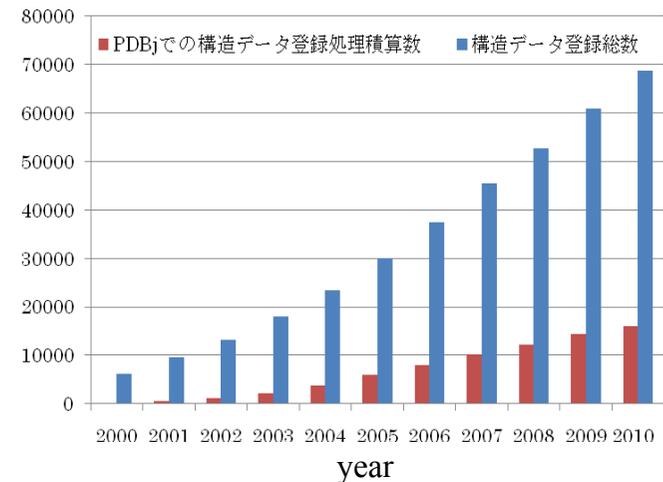
(2-4)蛋白質構造情報の標準XML記述(PDBML)の応用

PDBMLに基づく全データのvalidationを随時実施。PDBMLとその拡張情報に基づくRESTfulなRDB検索サービス(日本語キーワード検索を含む)を有するデータビューアPDBjMineの開発と稼働。

(2-5) 関連する新規二次データベースの構築と解析ツールの開発

jV3, *GIRAF*, *eF-site/eF-surf/eF-seek*, *ProMode*, *ASH*, *Seala*, *Sequence/Structure Navigator*, *MAFFTash*, *SeSAW*, *Spanner*, *EM Navi*, *Yorodumi*

(2-7)教育用の蛋白質構造百科事典(eProtS/MoM)の開発



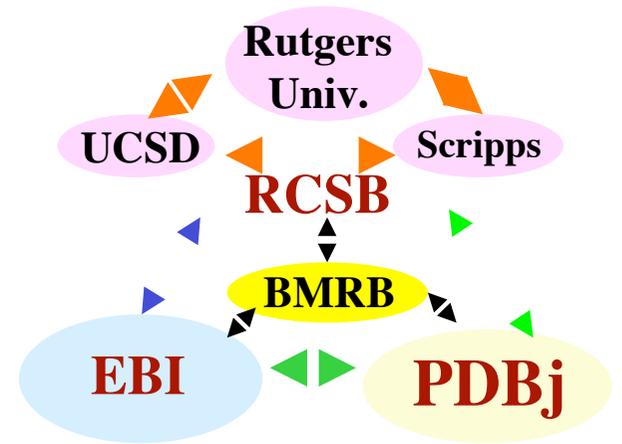
3. 国際協力・国際連携について

(3-1) 米国RCSB-PDB, BMRB, 欧州PDBe-EBIと共に、三極を結んだwwPDBを運営。

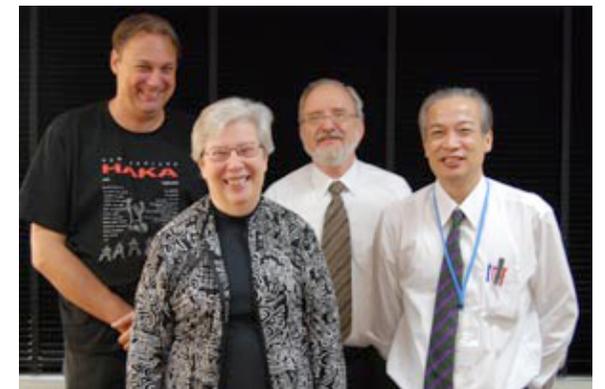
- ・シングル・データ・アーカイブを構築。(ユニークなPDBID)
- ・データ記述法を共同で修正・確立。
- ・データ内容の検証(validation)を実施。
- ・wwPDBAC会議(S. K. Burley議長)を開催。

2006年10月東京, 2007年9月Princeton, 2008年9月EBI, 2009年11月大阪, 2010年9月Rutgers Univ.

- ・wwPDB Foundation 活動(PDB 40th Ann. Symposium *etc*)
- ・登録業務の標準化・自動化システム開発を共同で実施。
- ・それぞれ独自のviewer, API, Web service を開発。



<http://www.wwpdb.org/>



Kleywegt, G, Markley, JL
Berman, HM, Nakamura, H
p. 3

(3-2) 米国Wisconsin大学とBioMagResBank (BMRB)とSMSDepを共同で運営。

(3-3) 中国、インド、韓国等のアジア諸国とも連携。

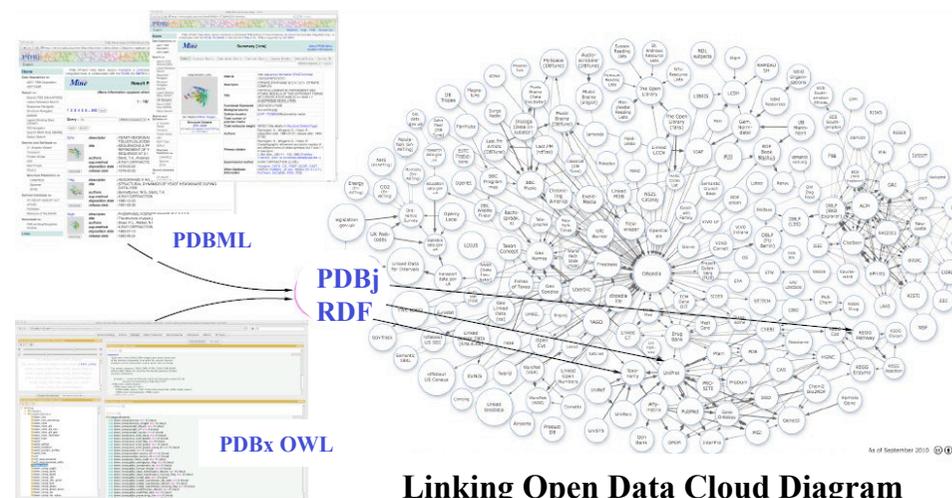
4. 本研究開発課題の詳細

(4-1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発

- Primary Annotatorsによる欧米との協力したデータ登録の継続的实施。
- 登録業務の標準化・自動化 (Common Annotation & Deposition) システム開発の国際協力 (データ検証部分を担当)。
- wwPDB活動を推進。
- 電子顕微鏡およびX線小角散乱の実験による構造情報データベースの構築と公開
- NBDCとの統合的データベース運営の仕組みを構築し実施。

(4-2) 統合化に向けたデータベースの高度化

- 蛋白質構造データの管理・配布システムの高度化: 他のDBとの関連性メタデータとRDF化したPDBデータを提供し、DB統合化環境を整備する。
- 構造データのオントロジーの改良・開発: mmCIFの記述に基づき、OWLに準拠したRDF/XMLの確立と、他のRDF化がなされたDBとの統合的利用法の確立。



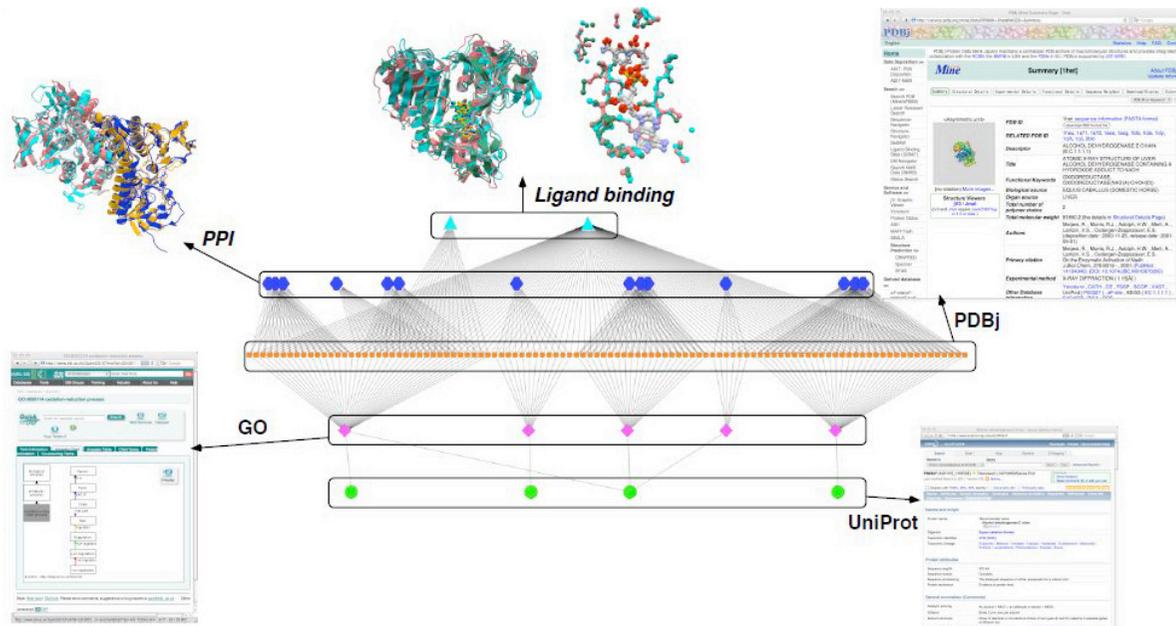
Linking Open Data Cloud Diagram
by Cyganiak & Jentzsch

(4-2) 統合化に向けたデータベースの高度化(つづき)

- ・データ記載のフォーマットの改良・開発: 新たなPDBフォーマットの開発と公開、およびマスターフォーマットの確立。NBDCとの連携による構造データのフォーマットの標準化。
- ・ユーザインターフェースの改良: 多言語でのweb pageや、jVを用いたGUIの高度化
- ・NMR実験データの管理・配布システムの高度化: オントロジー工学に基づく標準的なXML記述法を開発し、外部DBとの統合的利用を可能とし、新規実験データを収集する。

(4-3) 機能情報と構造情報との統合化システムの開発

- ・蛋白質相互作用部位データベースの改良とPDBデータとの統合化: 原子レベルでの蛋白質相互作用を記述するオントロジーを定義し、相互作用部位をRDFで記述する。これにより、原子レベル→残基レベル→分子レベル→分子複合体レベルへつなげて、高次生命機能情報へ、原子レベルの情報を統合化できる環境を整える。



原子レベルでの構造モチーフとして、低分子リガンド結合モチーフ (Ligand binding) と蛋白質間相互作用モチーフ(PPI)があり、それら複数のモチーフの固有の組み合わせを持った蛋白質および蛋白質複合体(PDBj)が、GOに定義される固有の高次生命機能を持つ蛋白質(複合体)として整理される。

(4-3) 機能情報と構造情報との統合化システムの開発(つづき)

・配列情報を入力とした、構造情報に基づく機能の解析・推定パイプラインの構築:
PDBjおよびその他の研究者が開発した、リモートホモログの推定や局所的蛋白質相互作用部位や機能的分子表面部位の推定のツールを基に、ゲノム解析等によって得られた配列情報から、構造情報あるいは複合体を含むモデル構造情報を一度経由し、高次生命機能情報を推定できる高速なパイプラインを開発し、提供する。

(*PreMAFFTash, MAFFTash, MSThread, Spanner, SeSAW, surFit* 等を改良・成熟させそれらをつないだ検索システムとする)

(manualでの実施例 1: TLR-inducible gene product の機能推定から同定へ: *Nature* (2009) 458, 1185-190)



Vol 458 | 30 April 2009 | doi:10.1038/nature07924

nature

LETTERS

Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay

Kazufumi Matsushita^{1,3*}, Osamu Takeuchi^{1,3*}, Daron M. Standley², Yutaro Kumagai^{1,3}, Tatsukata Kawagoe^{1,3}, Tohru Miyake^{1,3}, Takashi Satoh^{1,3}, Hiroki Kato^{1,3}, Tohru Tsujimura⁴, Haruki Nakamura⁵ & Shizuo Akira^{1,3}

(manualでの実施例 2: ROP16 (parasite-derived kinase)の機能推定から同定へ: *J. Exp. Med.* (2009) 206, 2747-2760)

A single polymorphic amino acid on *Taxoplasma gondii* kinase ROP16 determines the direct and strain-specific activation of Stat3

SeSAW correctly distinguished between dual specific (Ser/Thr or Tyr) and Ser/Thr-specific kinases based on our 20% ID homology model. In contrast, sequence alignment alone indicated only Ser/Thr kinase hits. We also modeled the polymorphic position 503 which has been shown to affect the virulence of *T. gondii* and showed that it sits beneath the predicted substrate binding cavity, leading to characteristic changes in shape and electrostatic potential.

(4-4) 人材養成の実施

- ・データ寄託・登録者に対する高精度のデータ整理と登録についての教育: 構造生物学の研究者・学生に対し、データの品質管理について講習会やセミナーで紹介する。
- ・データ利用者に対する初歩的および高度な利用法についての教育: 初歩的な利用法と、高度なweb serviceの利用法の双方について、講習会やセミナーで紹介する。
- ・キュレータ、アノテータの人材養成: 国際的なon-the-job trainingを実施し、国際感覚を身に付けたキュレータ、アノテータを育成する。

(4-5) 研究開発のスケジュール

	H23 年度	H24 年度	H25年度
(1) PDBデータおよびNMR実験情報の登録・編纂の国際連携による実施とデータ検証法の開発	<p>蛋白質構造情報とNMR実験情報の登録</p> <p>登録業務の標準化・自動化の開発</p> <p>wwPDB活動の推進</p> <p>電子顕微鏡およびX線小角散乱の構造情報DBの構築と公開</p> <p>NBDCとの統合的データベース運営の仕組みを構築し実施</p>		
(2) 統合化に向けたデータベースの高度化	<p>構造情報のRDF形式等による高度化と配布</p> <p>構造データのオントロジー化</p> <p>構造データの新フォーマットの開発・改良</p> <p>ユーザインターフェースの改良と高度化</p> <p>NMR化学シフトデータの管理・配布システムの高度化 (STAR形式への変換) (XML形式への変換)</p> <p>新たなNMR実験データのDB化と登録システムの開発</p>		
(3) 機能情報と構造情報との統合化システムの開発	<p>蛋白質相互作用部位データベースの改良とPDBデータとの統合化</p> <p>配列情報を入力とした構造情報に基づく機能の解析・推定パイプラインの構築 (パイプライン化と公開) (GIRAF, eF-seekの組込) (システムの確立と公開)</p>		
(4) 人材養成の実施	<p>データ寄託者・登録者に対する教育・講習会の実施</p> <p>データ利用者に対する教育・講習会の実施</p> <p>キュレーター・アナレータの人材育成</p>		

5. 終了時の達成目標・将来展望

(5-1) 達成目標

- ・立体構造データとNMR実験データを、wwPDBとBMRBの一翼を担って収集し、厳しい品質管理を行いつつコスト・パフォーマンスに優れたデータベース構築と公開を継続して行う。
- ・データ管理・配布システム、構造データのオントロジー、データ記載のフォーマット、ユーザインターフェース等の開発・改良を実施し、DB統合化環境を提供する。
- ・上記環境を用い配列から構造に基づき高次生命機能を推定する仕組みを開発・公開する
- ・データ寄託者、利用者、アノテータの3種の人材育成を行う。
- ・上記活動を通して、登録者・利用者から支持される恒久的に持続可能な統合的バイオサイエンス・データベースのありかたを、本研究開発終了時までNBDCと協力して確立する。

(5-2) 終了後の構想

- ・生命科学の基盤的データとして、構造生物学の専門家を擁する組織が、NBDCの組織の一部としてDistributed & Integrated の形態で構造データバンクを日本国内で維持すべき。
- ・「共同利用・共同研究拠点」の役割の一つとしてデータベース構築と運営があり、附置研究所の業務として、これらに対する予算面での配慮を国に求めていきたい。

(5-3) 将来展望と社会への寄与

- ・蛋白質機能が発揮される部位の立体構造とダイミクスをもとに、その部位をブロックする低分子化合物を探索するStructure-Based Drug Design(構造に基づく医薬品設計)への応用。
- ・天然蛋白質を改変し新たな機能を発揮する人工蛋白質の設計・創製(蛋白質工学)。
- ・教育用データベースである蛋白質構造百科事典の開発と普及、講習会等による高校生や一般の人々への啓蒙。

6. 2011年度からの実施体制(計画)

NBDC

wwPDB

代表研究者

中村春木
(阪大蛋白研)

研究チーム事務員

阪大蛋白研 共同利用共同研究委員会・
蛋白質立体構造データベース専門部会
(国内諮問委員会)

阪大蛋白研より2名
X線結晶学の専門家
NMR構造解析学の専門家
構造バイオインフォマティクスの専門家
PhotonFactoryの専門家
SPring-8の専門家

PDBj DB登録・管理・運営・開発・統合化

統括: 中村春木 (阪大蛋白研)

X線結晶構造DB登録・管理: 中川敦史(教授)

アノテータ(5名)、プログラマ(2名)、研究者(2名)

電子顕微鏡DB登録・管理・開発: 岩崎憲治(准教授)

研究者(1名)

統合化に向けたDB高度化: 金城玲(准教授)

プログラマ(2名)、研究者(2名)

機能情報と構造情報との統合化システムの開発: 金城玲(准教授)

プログラマ(2名)、研究者(1名)

配列からの機能推定パイプラインの開発: Standley, DM(iFREC准教授)

プログラマ(2名)、研究者(1名)

人材養成: 中村春木(教授)

中川敦史、岩崎憲治、金城玲、研究者(3名)

PDBj-BMRBDB登録・管理・運営・開発・統合化

代表: 藤原敏道(阪大蛋白研)

BMRB と外部DBとの統合化: 藤原敏道(教授)

児嶋長次郎(准教授)、アノテータ(1名)、

プログラマ(1名)、研究者(1名)

BMRB DB登録・管理・開発: 藤原敏道(教授)

児嶋長次郎(准教授)、アノテータ(1名)、

プログラマ(1名)、研究者(1名)

生体系NMRの国際的な高度化: 藤原敏道(教授)

児嶋長次郎(准教授)、アノテータ(1名)、

プログラマ(1名)、研究者(1名)