

蛋白質構造データバンクのデータ駆動型研究基盤への拡張

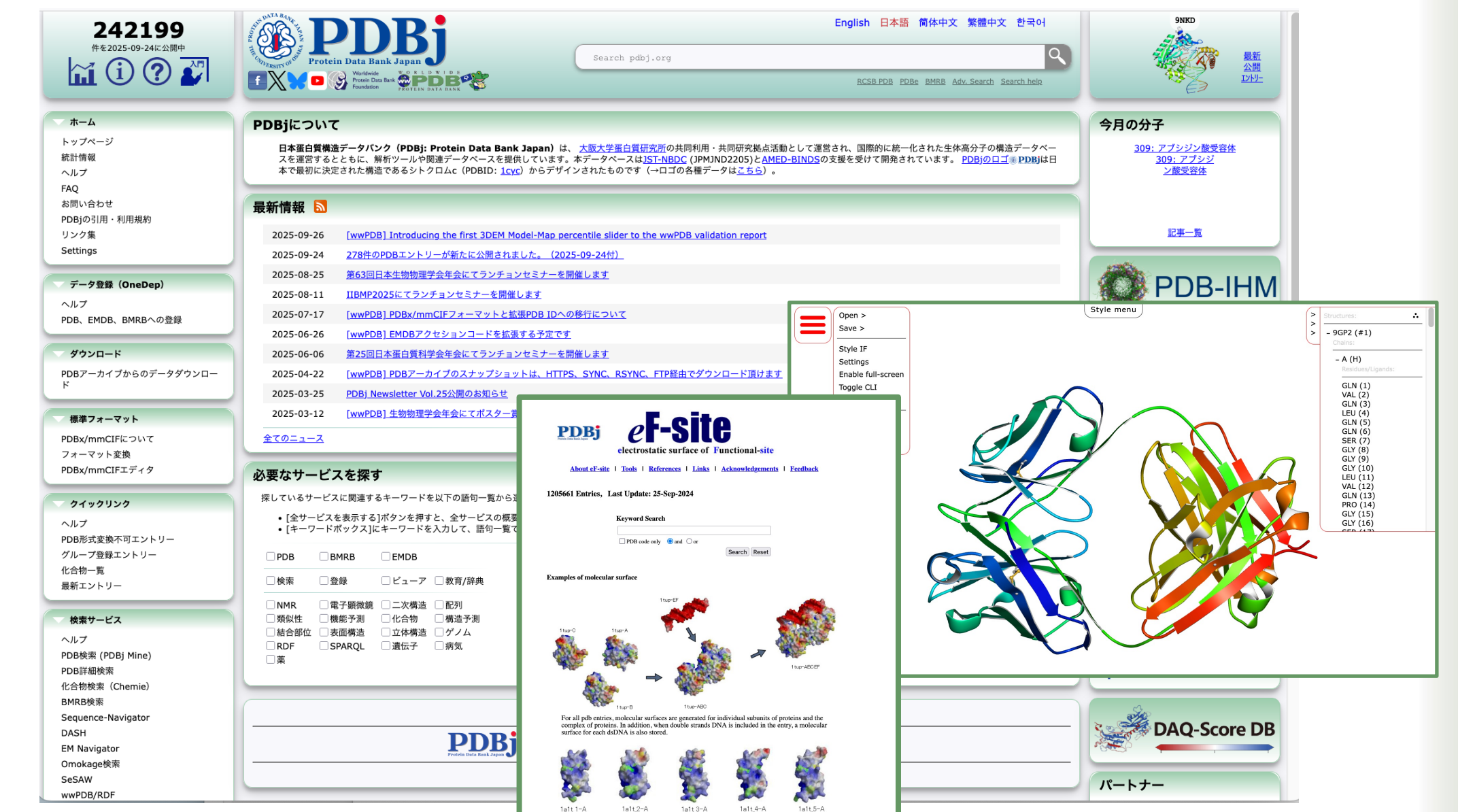
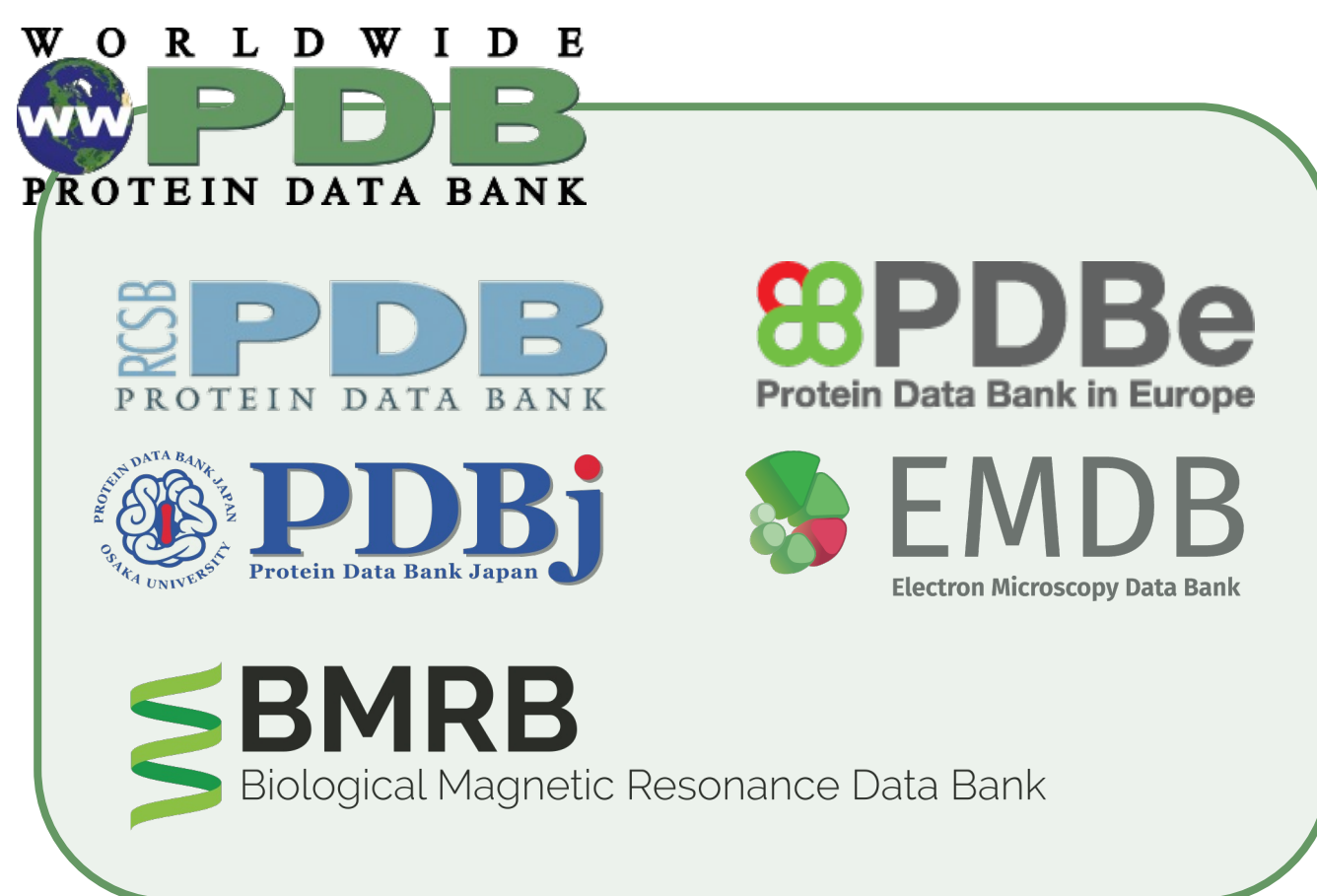
長尾知生子¹ 工藤高裕² Gert-Jan Bekker¹ 山下鈴子¹ 横地政志¹ 栗栖源嗣^{1,2}

(1) 大阪大学蛋白質研究所 (2) 一般財団法人 蛋白質研究奨励会

蛋白質構造データバンク（Protein Data Bank：PDB）は蛋白質や核酸、糖鎖などの生体高分子の原子分解能の構造情報を集めた世界で唯一のデータアーカイブである。PDBjは、国際的組織であるwwPDBの一員として、共通した品質管理によるデータ登録とダウンロードサイトの運営を担うとともに、データベースへの登録と利用促進を目的とした独自のサービスやツールの開発・公開を行なっている。本発表では、利用者の動向調査に基づいて、現在、構築と改良を進めている、配列、化合物データベースとの統合利用のためのポータルサイトについて紹介する。

日本蛋白質構造データバンク（Protein Data Bank Japan, PDBj） <https://pdbj.org>

- 生体高分子の立体構造データベース
- 国際蛋白質構造データバンク（worldwide Protein Data Bank: wwPDB）メンバー
- アジア・中東からのデータ登録
- 独自のサービスとツール開発（ProMode, eF-site, wwPDB/RDF, Molmilなど）

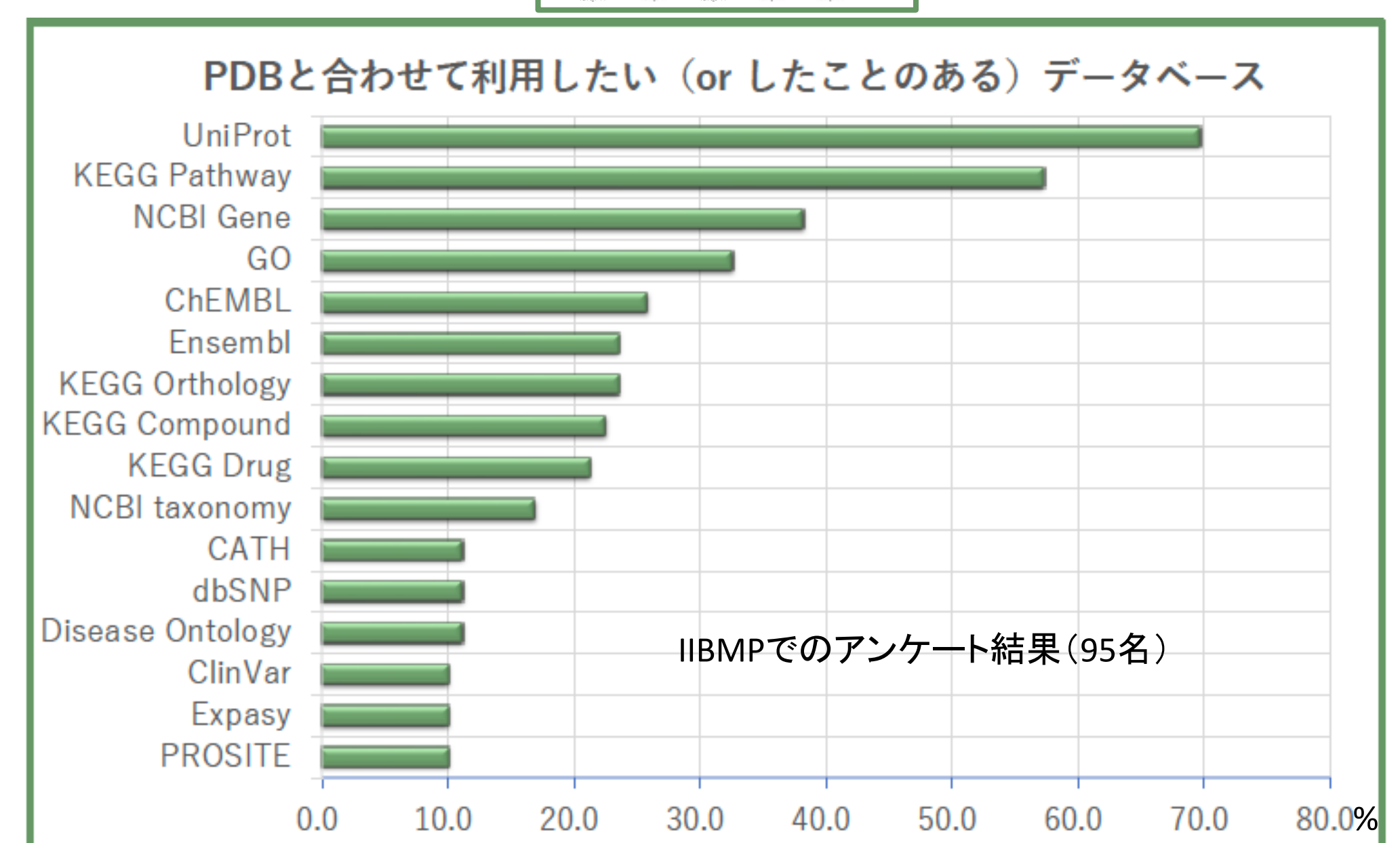


データ駆動型研究基盤への拡張

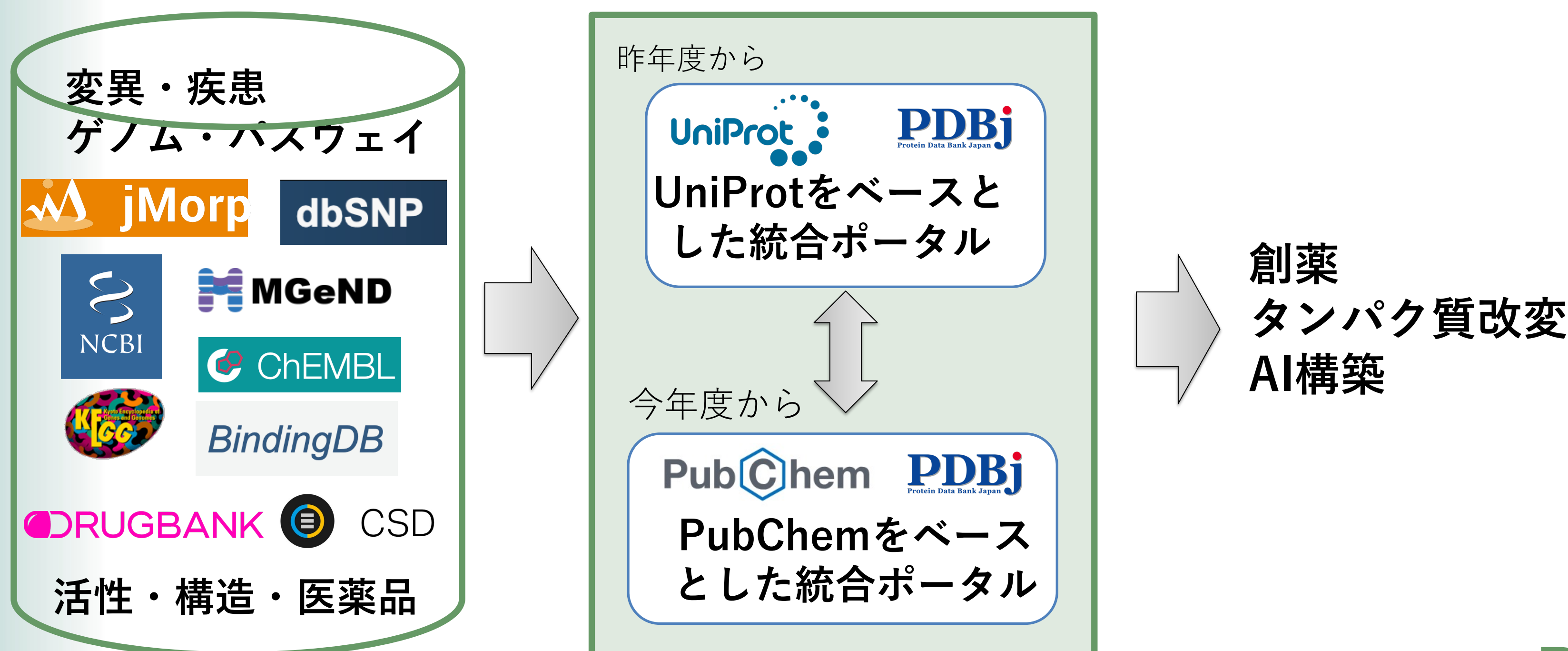
- さまざまな分野でのAI開発の加速
- 創薬、育種、発酵、食品、エネルギー生産などの社会の諸問題の解決支援
- 高度な検索および部分データ取得システム
- ー情報科学や構造生物学に詳しくないユーザーが生物学・化学的興味から適切なデータを発見
- ー情報科学研究者が機械学習などに必要なデータセットをプログラマ的に生成

ユーザへのアンケート実施

- 蛋白質科学会、結晶学会、日本バイオインフォマティクス学会年会・生命医薬情報学連合大会（IIBMP）、CBI学会などでアンケートを実施（2022年）
- 遺伝子、化合物、パスウェイデータベースとの接続への需要
- **配列・化合物情報とタンパク質の立体構造をつなぐ**



新しい統合ポータル像



独自のスコアによる構造の選定

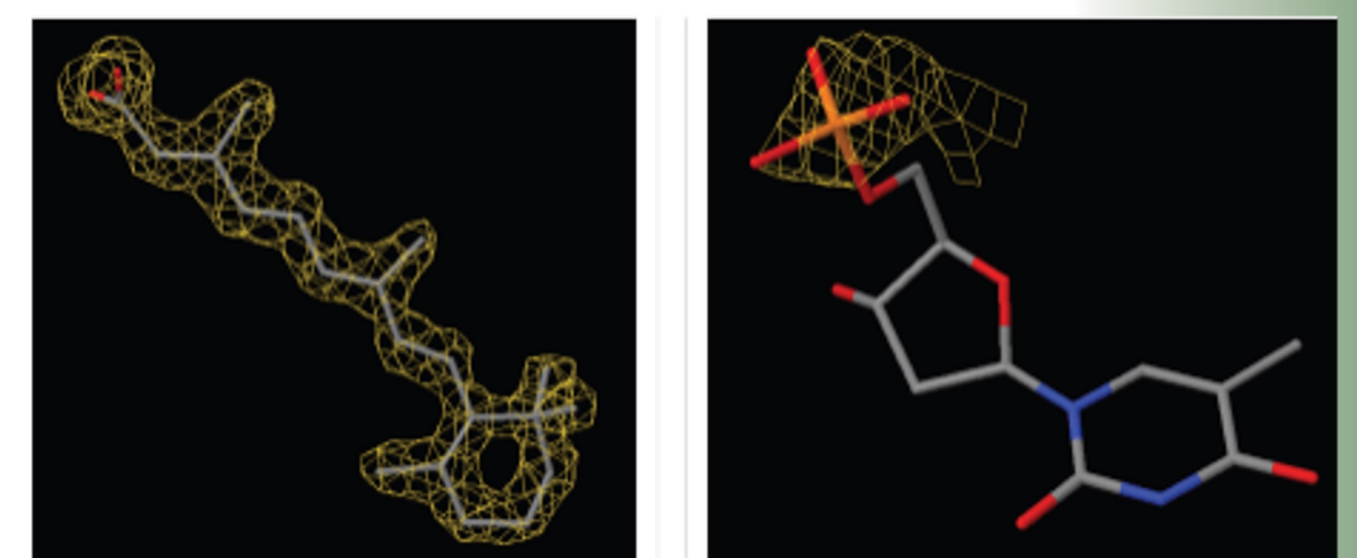
分解能だけでは、解析に適した構造を選択するのは難しい

タンパク質
・ 分解能
・ coverage

化合物
・ 実空間信頼度因子（Real Space Reliability Factor: RSR）
・ マップファイルとの相関係数（RSCC）
・ 化学結合距離や結合角度の標準的な値

クライオ電子顕微鏡
・ DAQ-score

深層学習を用いて、クライオEMマップから生体分子構造の特徴を計算。その局所の特徴と構造モデルの一致度で構造モデルの精度を評価（Purdue大学 木原教授）

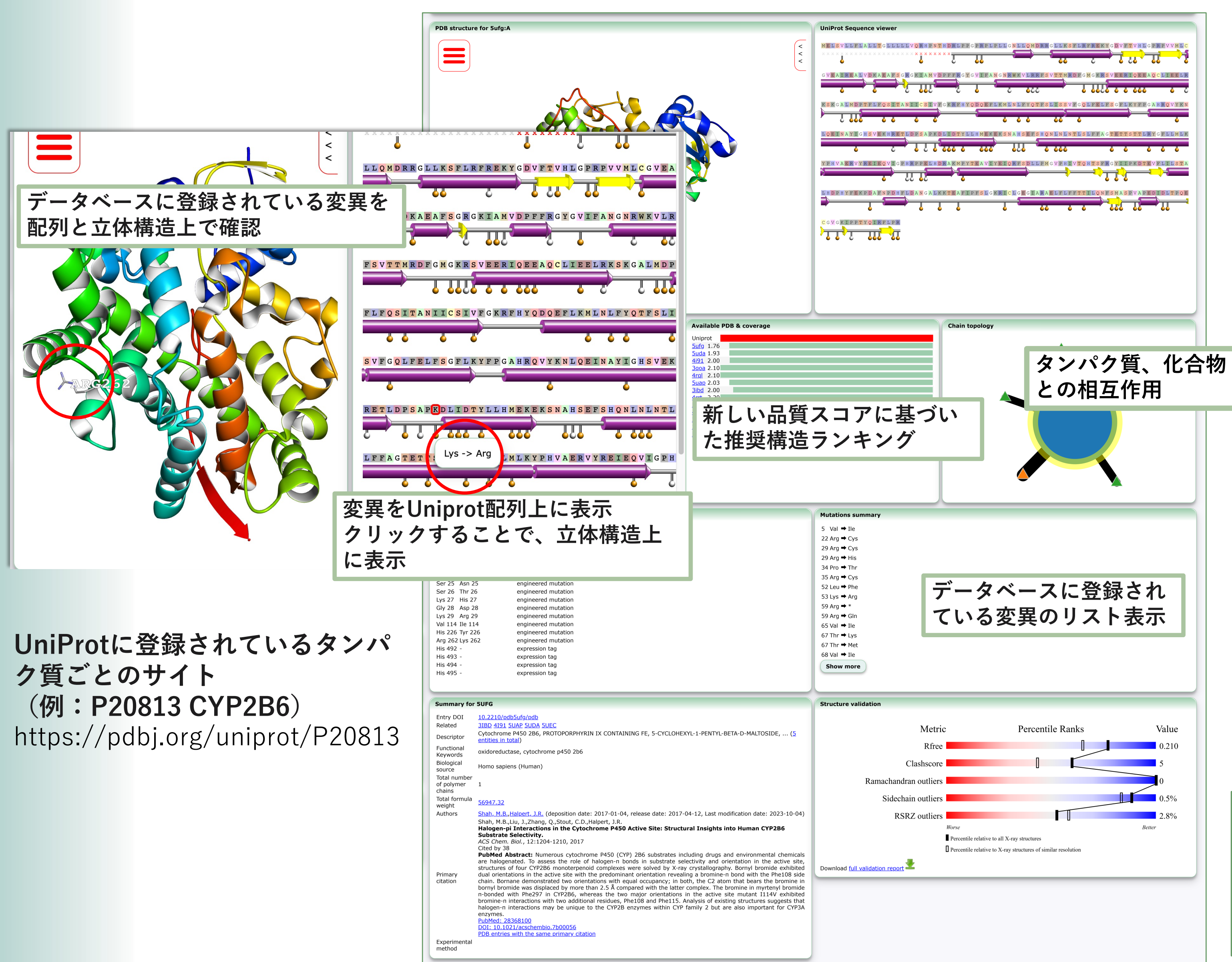


同分解能でも異なる精度が異なる構造の例
左：RSR=0.10, RSCC=0.95, 右：RSR=0.41, RSCC=0.70

PubChemをベースとした統合ポータル

UniProtをベースとした統合ポータル

<https://pdbj.org/uniprot>



データベースに登録されている変異を配列と立体構造上で確認

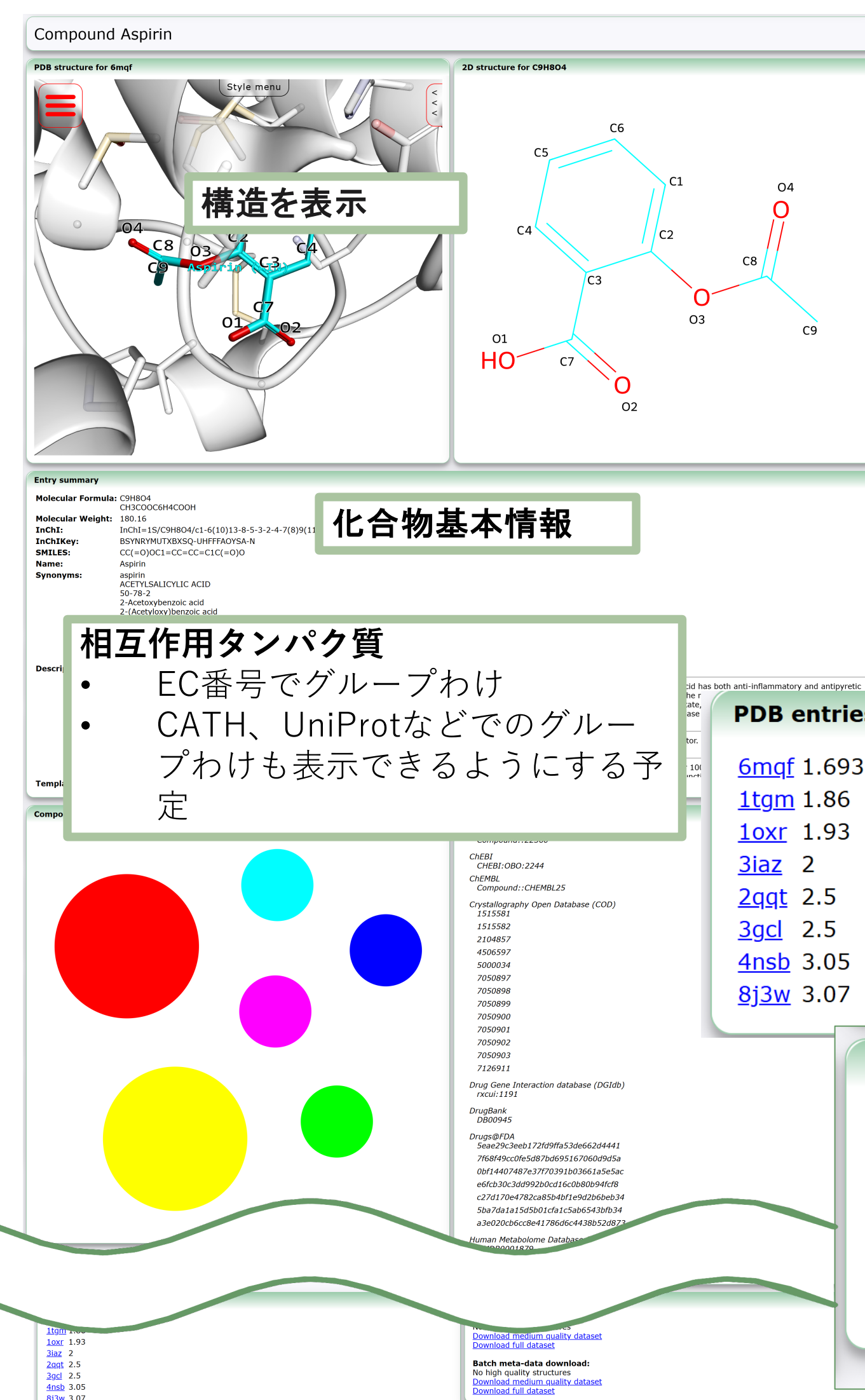
タンパク質、化合物との相互作用

新しい品質スコアに基づいた推奨構造ランキング

変異をUniProt配列上に表示
クリックすることで、立体構造上に表示

データベースに登録されている変異のリスト表示

UniProtに登録されているタンパク質ごとのサイト
(例：P20813 CYP2B6)
<https://pdbj.org/uniprot/P20813>



構造を表示

化合物基本情報

相互作用タンパク質
・ EC番号でグループわけ
・ CATH、UniProtなどでのグループわけも表示できるようにする予定

PDB entries

他データベースとの関連
BindingDB, ChEMBL, ChEMBL, Crystallography Open Database (COD), DrugBank, Drugs@FDA, Human Metabolome Database (HMDB), KEGG, The Cambridge Structural Database (CSD) など

PDB構造のリスト
・ 独自のスコアでランクづけ予定
・ 分解能だけでなく、RSRやRSCCを用いたスコアを用いる

Batch download
Batch file download:
No high quality structures
[Download medium quality dataset](#)
[Download full dataset](#)

Batch meta-data download:
No high quality structures
[Download medium quality dataset](#)
[Download full dataset](#)

バッチダウンロード
データセットの質を選択できる

・ 日本蛋白質構造データバンク(PDBj) は、JST-NBDC (科学技術振興機構NBDC事業推進部「統合化推進プログラム」JPMJND2205)、生命科学・創薬研究支援基盤事業 (BINDS)、「大阪大学蛋白質研究所に措置された共同利用・共同研究拠点経費 (文部科学省)」の支援を受けて開発されています。

日本蛋白質構造データバンク (PDBj)

大阪大学蛋白質研究所 蛋白質構造データバンク構築研究室 TEL：06-6879-4311 (全般) 06-6879-8634 (PDB登録)

(一財)蛋白質研究奨励会・データベース研究支援部門 Web問い合わせ: <https://pdbj.org/contact>

