

“生命医薬情報学連合大会 2014”

PDBj&創薬等支援技術基盤プラットフォームランチョンセミナー

「ビッグデータ時代における構造生命科学：ゲノム情報の蛋白質構造によるアノテーション」

H26 年 10 月 3 日（金）12:40-13:40 仙台国際センター 2 階・大会議室『萩』

蛋白質や核酸、糖鎖等の生体高分子の原子レベルの分解能での立体構造を決定することを主な目標としていた「構造生物学」研究は、構造を基にした生命科学という概念の「構造生命科学」に進化しており、生命科学、医薬学、農学等と強く結びつき、バイオ産業への応用により社会におけるイノベーションへも向かおうとしております。これら構造情報は、国際的なデータバンクである PDB (Protein Data Bank) に蓄積され世界中に無償で公開されていますが、日本では大阪大学蛋白質研究所が組織する PDBj (PDB Japan, <http://pdbj.org/>) が、国際組織である wwPDB (worldwide PDB, <http://wwpdb.org/>) のメンバーの一員として、高い品質管理を行いつつデータの登録・編集・公開作業を行っています。一方、2012 年度から開始された文部科学省創薬等支援技術基盤プラットフォーム事業においては、生命科学や医科学に重要な生体高分子および複合体構造の構造解析に加えて、構造情報の生命科学や医薬品開発への応用についても支援を行っています。本セミナーでは、構造情報を蓄積している PDB の活動と、その生命科学および創薬への応用について、具体的な利用法を交えて紹介します。

講演 1：PDBj と wwPDB の活動

中村春木（阪大蛋白研）

蛋白質を初めとする生体高分子の 3 次元（3D）立体構造情報は、1971 年以来、国際的な蛋白質構造データバンク PDB (Protein Data Bank) に蓄積され公開されています。このデータバンクは、設立以来米国によって運営がなされてきましたが、2001 年度から大阪大学蛋白質研究所においても日本蛋白質構造データバンク (PDBj: PDB Japan, <http://pdbj.org/>) なる組織を創設して構造データを受け付け、検証・編集・登録処理を開始しました。2003 年に米国ラトガース大学の Helen M. Berman 教授が率いる RCSB-PDB と、欧州 EBI-PDBe および日本の PDBj が協力して、国際蛋白質構造データバンク (wwPDB: worldwide PDB, <http://wwpdb.org/>) を設立し、さらに NMR の化学シフト等の実験データを含む米国の BMRB (BioMagResBank) を加え、日米欧の 3 極での生体高分子構造データバンク構築の体制を整えています。2014 年 5 月には 10 万件を超える構造が公開されており、日本の PDBj による登録処理は全世界からの約 1/4 ほどとなっています。セミナーでは、PDBj と wwPDB の活動を紹介し、特に新規の PDB のデータフォーマットと validation report について説明します。

講演 2：HOMCOS サーバを用いた複合体立体構造の検索とモデリング

川端 猛（阪大蛋白研）

一つの蛋白質だけの単体の立体構造を眺めていても、その機能についてわかることには限界があります。しかし、他の分子がその蛋白質に結合している状態の立体構造からは、相互作用部位やその結合機

構など、その分子の機能の理解につながる有用な生物学的な知見を得ることができます。我々が開発した HOMCOS (<http://homcos.pdbj.org>)は、こうした複合体立体構造を PDB から検索し、さらにそれを鋳型として、複合体のホモロジー・モデリングを行うサーバです。蛋白質—蛋白質複合体、化合物—蛋白質複体の両方を扱うことができます。検索も、蛋白質（アミノ酸配列、UniProt の ID）と化合物（SDF/MOL2 ファイル、SMILES）の両方から行うことができます。アミノ酸配列による検索には BLAST、化合物の検索には我々が開発した KCOMBU プログラムが使用されています。類似した複合体を発見できた場合、直ちに簡易モデリング構造が表示され、複合体の結合部位などを WEB 上で検討することができます。また、蛋白質のサイトごとの相互作用予測のサマリを作成することも可能で、アミノ酸変異の解析に貢献することも目指しています。

講演 3：ゲノム変異解釈への立体構造情報の活用

木下賢吾（東北大学）

大小様々なヒトゲノム配列解析プロジェクトにより、ヒト個人のゲノム配列情報が蓄積され、その結果として、少数のヒトしか持たない頻度の低い変異（レアバリエント）が多数存在することがわかってきました。頻度の高い変異では、その変異の意味づけに統計解析が非常に有効に機能し、実際、数多くのゲノムワイドな関連解析により変異の意味が明らかにされているケースも増えてきています。一方で、レアバリエントでは、そもそも変異を持っているヒトが少ないため、統計解析に持ち込むためにはきわめて多くのサンプルを集めゲノム解析を行う必要があり実行が困難です。そこで我々は、レアバリエントをタンパク質立体構造上にマップし、その変異の及ぼす影響を直接的に検討する仕組みの開発を行いました。具体的には、ゲノム上の座標を refseq を介して Uniprot, PDB へとマッピングすることで、機能上重要なアミノ酸への影響を見積もりますが、立体構造上へマッピングすることで、配列上は離れていても空間的に機能性残基に近い変異を洗い出すことができるようになりました。また、タンパク質を利用した多くの変異実験の情報を利用してゲノムの変異の解釈を行えるようにもなりました。本口演では、我々が開発を行ったマッピングツール SNVMap と、国際 6500 人エキソームプロジェクトへの適用した結果を報告します。