

Title	タンパク質立体構造データベースにおける代表タンパク質チェーン決定システムに関する研究
Author(s)	野口, 保
Citation	大阪大学, 2001, 博士論文
Version Type	VoR
URL	<a href="https://doi.org/10.11501/3184211">https://doi.org/10.11501/3184211</a>
rights	
Note	

*Osaka University Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名	野口 保 <small>のぐち たもつ</small>
博士の専攻分野の名称	博士(工学)
学位記番号	第 16349 号
学位授与年月日	平成13年3月23日
学位授与の要件	学位規則第4条第1項該当 基礎工学研究科情報数理系専攻
学位論文名	タンパク質立体構造データベースにおける代表タンパク質チェーン決定システムに関する研究
論文審査委員	(主査) 教授 橋本 昭洋  (副査) 教授 萩原 兼一 教授 柏原 敏伸 教授 中村 春木 助教授 松田 秀雄

### 論文内容の要旨

タンパク質立体構造データベース (PDB) は、近年の X線結晶解析や NMR による構造解析技術の進歩により急激に増加し、その内容は2000年11月の時点で13,600エントリーを越えている。今後も、各種生物種のゲノムプロジェクトの後を受けて開始した“構造ゲノミクスプロジェクト (ゲノムの中に含まれるタンパク質の立体構造をすべて決める。)”によって、さらにその増加は加速すると予想されている。しかしながら、冗長性やデータの不完全性のために、PDBの全てのエントリーがタンパク質の立体構造の解析に適しているとは言えない場合 (例えば、タンパク質立体構造予測の研究) があり、何らかの基準でタンパク質立体構造を分類して、その中から代表タンパク質を決定する必要がある。代表の決め方としては、分類されたグループ内の中心を求めて、そのタンパク質を代表にする方法があるが、そうした場合、その代表タンパク質データの質が悪く (分解能が悪く、チェーンが途中で切れているなど)、タンパク質立体構造予測の基礎データとしては、相応しくないことがある。そのような観点では、各エントリーの内容を調べ、解析に適さない質の悪いデータを除去し、各エントリーに対して他のエントリーが配列および立体構造上、類似のタンパク質かどうかを調べあげ、分類した上で代表を決定する必要がある。PDBのエントリー内には、構造上、一つにつながったタンパク質チェーンが複数本含まれる場合があるため、一般に、そのチェーン同士を比較・分類し、その中から任意の優先度で選ばれたタンパク質チェーンを代表にしている。タンパク質チェーン同士の比較・分類は、立体構造の取扱いの困難さとそれに基づく分類に膨大な計算が必要なため、近似的に配列の類似性 (ID%) を指標にして行なわれてきた。

本研究では、まず従来のID%による分類に、タンパク質分子を重ね合わせた時の原子間距離の最大値 (Dmax) を分類の指標として加え、部分構造の違いも考慮した、より正確な立体構造分類を行って、非冗長な PDB 代表タンパク質チェーンデータベース (PDB-REPRDB) を作成することを可能にした、PDB 代表タンパク質チェーン決定システムの初期バージョンの作成を行った。次に、自動化が不十分であった初期バージョンの自動化を進め、高速に代表タンパク質チェーンを得られるように、PDB 代表タンパク質チェーン決定システムの改良を行った。新しい分類指標 (Dmax) の追加による計算量増加の問題は、MPI ライブラリを用いてプログラムを並列化することによって処理の高速化を行い解決した。また、あらかじめ決めた基準 (配列の相同性: ID% ≥ 25% ~ 95% で 10% 刻みの 8 通りと、構造の相同性: Dmax ≤ 10 Å ~ 50 Å まで 10 Å 刻みと ∞ Å を加えた 6 通りの基準を組み合わせ、合計 8 × 6 = 48 通り) の代表タンパク質チェーンを決定し、それらの結果を PDB-REPRDB として WWW 上に公開した。

しかしながら、研究の内容や研究者によって、代表を選ぶ基準は様々で、とてもすべての要求に応じきれない。そこで、最新バージョンでは、WWWによるインターフェースを作成し、研究者が得たい基準での代表セットを自身で指定することができるように改良し、代表タンパク質チェーンをオンデマンドで提供することを可能にした。

本論文では、PDB代表タンパク質チェーン決定システムの初期バージョン、並列版および最新の会話形式バージョンにおけるシステム構成及び特徴について述べる。

最後に、本システムで作成した代表タンパク質チェーンの利用例として、タンパク質二次構造予測の基礎データとなる構造ライブラリのセットや、並列タンパク質情報解 (PAPIA) システムの検索対象となる立体構造データベースを作成し、実際のタンパク質立体構造予測の研究に役立てているので、それについても述べる。

## 論文審査の結果の要旨

タンパク質立体構造予測問題は、現在実験的に明らかになっているアミノ酸配列に比べ、その立体構造が不明な場合が遥かに多く、DNA自動読み取り装置に相当するような大量解析技術が立体構造解析の技術にも登場しない限り、今後もこの状況が変わることはないと思われているため、タンパク質の構造と機能の関係を理解し、生物のメカニズムを理解する上で重要な研究課題となっている。タンパク質立体構造予測法には、物理化学的な力だけを考慮してタンパク質の折れたたみをシミュレーションすることによって予測する手法もあるが、多くは既知のタンパク質立体構造とそのアミノ酸配列の関係を様々な手法で表現し、それを基に未知の立体構造を予測する経験 (統計) 的手法である。

本論文では、そのようなタンパク質立体構造予測の研究の学習セットとなる非冗長な代表タンパク質チェーンをタンパク質立体構造データベース (PDB) から得る手法を提案し、かつ、そのPDB代表タンパク質チェーンをタンパク質立体構造予測の研究者の様々な要望に対応して選出するシステムを開発・公開することによって、その実用性が示されている。

従来のPDB代表タンパク質チェーン決定法の研究では、計算機で扱い易いアミノ酸配列の情報だけを用い、その相同性から近似的に立体構造の類似性を定義して、タンパク質立体構造の分類を行い、その中から代表を選んでいた。それに対して、本論文で提案されているPDB代表タンパク質チェーン決定法では、タンパク質分子を重ね合わせた時の原子間距離の最大値を分類の指標として加え、より正確な立体構造分類を行って、非冗長なPDB代表タンパク質チェーンデータベース (PDB-REPRDB) を作成することを可能にしている。また、PDBのデータ量増加に伴う大量データ処理に対応するため、処理の並列化を行い、処理の高速化を実現することによって、実用的な時間で処理が完了する改良が加えられている。さらに、利用者の要求に対応できるようなWWWを利用した会話形式のPDB代表タンパク質決定システムを開発し、様々な基準でPDB-REPRDBを作成できるようにしている。これにより部分構造の違いも見落とさず分類することが可能になり、タンパク質立体構造予測の基となる学習セットを厳密に選び出すことができるようになり、またタンパク質立体構造の部分的な構造変化を調べる際のデータベースを得ることもできるようになった。

さらに、実際の応用例として、二次構造予測の研究と立体構造検索システムで利用した研究成果を示し、提案するPDB代表タンパク質チェーン決定システムの有効性が示されている。本研究成果により、PDB代表タンパク質チェーン決定法の新たな機構が与えられており、より正確で様々な視点から代表を決定することが可能となり、タンパク質立体構造予測及びタンパク質部分構造の研究の効率化に寄与することによって、タンパク質立体構造研究の発展に大きく貢献している。よって、博士 (工学) の学位論文として価値のあるものと認める。