

Title	蛋白質機能部位データの比較ならびに抽出に関する研究
Author(s)	兼田, 佳和
Citation	大阪大学, 2005, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/45752">https://hdl.handle.net/11094/45752</a>
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉</a> 大阪大学の博士論文について <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈/a〉</a> をご参照ください。

***Osaka University Knowledge Archive : OUKA***

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

Osaka University

氏名	兼 田 佳 和
博士の専攻分野の名称	博士 (情報科学)
学位記番号	第 19621 号
学位授与年月日	平成 17 年 3 月 25 日
学位授与の要件	学位規則第 4 条第 1 項該当 情報科学研究科マルチメディア工学専攻
学位論文名	蛋白質機能部位データの比較ならびに抽出に関する研究
論文審査委員	(主査) 教授 薦田 憲久  (副査) 教授 藤原 融    教授 松田 秀雄    教授 西尾章治郎 教授 岸野 文郎    教授 下條 真司    助教授 大川 剛直

### 論文内容の要旨

蛋白質は、数十～数千個のアミノ酸分子が一本の鎖状に繋がった構造を成す。鎖状のアミノ酸は複雑に折り畳まって三次元立体構造を構成し、活性部位と呼ばれる局所的機能部位によって様々な生体機能を発現する。機能を理解するためには膨大な数の蛋白質を解析する必要があるが、その処理を生化学実験のみで行うのは困難であるため、蛋白質機能部位データを活用した網羅的比較による知識発見、大規模実験データからの機能部位データの獲得・整理といった情報処理技術の活用が必要不可欠である。機能部位データの活用においては、蛋白質の機能が分子表面上における物性に深く関わっていることから、機能部位と蛋白質の分子表面データを比較することで蛋白質の機能予測を実現する仕組みが求められている。一方、機能部位データの獲得においては、多数の文献に機能部位データ（活性部位情報）が内在しているため、そのデータベース化が望まれている。

そこで本研究では、蛋白質分子表面データと機能部位分子表面データとの比較と文献からの活性部位情報の抽出を実現する方式を提案する。分子表面比較では、蛋白質の機能が突起・窪みの部位に強く発現する性質があることに着目し、突起・窪み部分から属性付き法線ベクトルを導出し、これを用いて高精度かつ高速な比較を実現する。活性部位情報の抽出では、表現が平易でテンプレート化しやすい文と、表現が例外的でテンプレート化しづらい文が混在していることに着目し、テンプレートマッチングに基づく情報抽出手法と表層的特徴に依存しない文抽出手法とをハイブリッド化した情報抽出システム ExAct (EXtraction system of ACTive site information) を提案する。

本論文は、全 6 章から構成される。

第 1 章では、生体情報科学分野における蛋白質機能理解の背景を説明し、本研究の位置付けを示す。また、三次元形状比較と文献からの情報抽出の関連研究について検討し、本研究における課題とその解決方針を明確にする。

第 2 章では、分子表面を高速かつ厳密に比較する手法について論じる。まず、高曲率な突起・窪み部分の法線ベクトルに物性を付加した属性付き法線ベクトルを定義する。ベクトルの全組み合わせについての位置合わせは非効率であるため、パケット分割法を利用して類似属性を持つベクトル組のみを位置合わせ対象とする。また、より最適な位置合わせを実現するために反復改善法を適用する。

第 3 章では、提案する活性部位情報抽出システム ExAct の概要を示し、その第一段階の処理である固有表現同定手法について論じる。同定対象とする固有表現の一覧を示し、特に同定が困難な蛋白質名と化合物名に対処するため、

ゲノムネットの検索結果を利用する。検索結果から直接的に蛋白質名と化合物名を判別できないため、検索結果に含まれる特徴を表す特徴パラメータを定義する。

第4章では、ExActにおいて固有表現同定の後に続く処理であるテンプレートマッチングと照応解析の手法について論じる。テンプレートマッチングでは、複文の単文用テンプレートとの照合を可能にするため、構文解析を利用した照合方法について述べる。照応解析では、複数の先行詞を参照する照応関係に着目し、表層情報から先行詞候補をグループ化し、それと照応詞との数量一致性を考慮して先行詞を特定する。

第5章では、蛋白質立体構造データを利用することで相互作用記述文を抽出する手法について論じる。立体構造データにおける複合体と単体のそれぞれの場合に対する抽出方法を述べる。複合体の場合には、相互作用している対象間の距離を求めることで相互作用記述文かどうかを判断し、不要文の抽出を回避するために作用対象判定ルールを定義する。単体の場合には、立体構造データに作用対象が含まれていないため、単体と相同な複合体を利用する。

第6章では、本研究で得られた成果を要約し、今後に残された課題について述べる。

### 論文審査の結果の要旨

蛋白質はその構造に基づき生体内で様々な機能を発現することから、蛋白質の機能解析に関わる研究が活発化している。蛋白質の機能解析においては、生化学的な実験結果の収集のみならず、実験結果の網羅的かつ統計的な分析が重要である。そのための一つのアプローチとして、蛋白質の機能部位に対応する分子表面データの網羅的比較が有望視されている。一方、生化学実験の結果として得られた機能部位データは文献内に潜在していることから、これらを自動抽出し、データベース化することによって様々なアプリケーションでの活用が期待されている。

本論文は、蛋白質機能部位データの比較ならびに抽出に関する研究成果を纏めたものである。その主要な成果を要約すると次の通りである。

(1) 機能未知蛋白質と機能既知蛋白質の間の分子表面データの比較による蛋白質機能予測の実現を目的とし、機能部位となりやすい突起・窪みを重視した比較を実現するために属性付き法線ベクトルを導入したマッチング方法を提案し、検証実験により高精度かつ実時間内での計算が可能であることを示している。

(2) 文献から機能部位データを抽出するための第一の処理、すなわち専門用語とその種類の特定において、新語出現頻度の高い膨大な数の蛋白質名・化合物名に対応するために、WWW上で公開されているデータベースを利用した特定手法を提案している。また、蛋白質構造解析に関わる文献への適用実験により、その効果を示している。

(3) 機能部位データ抽出のための第二の処理として、専門用語特定済みの文献から必要な情報を抽出することを目的とし、文献内に存在する複文を考慮した構文解析によるテンプレートマッチング方法、及び、一对多の関係にある照応を特定するための数量表現に着目した照応解析手法を提案し、実問題へ適用した結果、従来手法と比較して精度が向上することを確認している。

(4) 機能部位データ抽出のための第三の処理として、テンプレートとの照合が困難な例外的記述形式の文へ対処するため、蛋白質立体構造データに着目し、その原子間距離を用いることで機能部位データに関わる文そのものを抽出する手法を提案している。また、この方式を用いた検証実験により、抽出漏れの減少を示している。

以上のように、本論文は蛋白質機能部位データの比較・抽出に関して成果を挙げた先駆的研究として、情報科学に寄与するところが大きい。よって本論文は博士（情報科学）の学位論文として価値あるものと認める。