

Title	Clique-based algorithms for protein structure prediction(Abstract_要旨)
Author(s)	Dukka Bahadur K.C.
Citation	Kyoto University (京都大学)
Issue Date	2006-03-23
URL	http://hdl.handle.net/2433/143887
Right	
Type	Thesis or Dissertation
Textversion	none

氏名	ドゥッカ バハドゥル ケーシー DUKKA BAHADUR K. C.
学位(専攻分野)	博士(情報学)
学位記番号	情博第193号
学位授与の日付	平成18年3月23日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	情報学研究科知能情報学専攻
学位論文題目	Clique-based Algorithms for Protein Structure Prediction (クリークアルゴリズムに基づくタンパク質構造予測アルゴリズム)

論文調査委員 (主査) 教授 阿久津達也 教授 小林茂夫 教授 後藤 修

論文内容の要旨

本論文は、アミノ酸配列データからのタンパク質立体構造予測に関する二種類の問題である「タンパク質スレッディング」および「タンパク質側鎖パッキング」に対する最大クリークに基づくアルゴリズムについて述べたもので、7章から構成されている。

第1章は序論で、タンパク質立体構造予測問題の背景について述べ、本論文で提案するアルゴリズムの簡単な説明を行い、本論文の構成を示している。

第2章では、タンパク質立体構造の特徴および立体構造予測法について概観し、特に本論文で対象とするタンパク質スレッディング問題とタンパク質側鎖パッキング問題に対する既存手法やNP困難性に関する既知の結果について説明するとともに、本論文で利用する最大クリークアルゴリズムおよび最大重みクリークアルゴリズムについても説明している。

第3章では、構造を予測したいタンパク質のアミノ酸配列と構造既知のタンパク質立体構造間の対応関係(アライメント)を計算することによりタンパク質立体構造の概要構造を予測しようというタンパク質スレッディング問題について論じている。特に、予測したいアミノ酸配列に関して一部分のアミノ酸間の距離に関する制約情報がわかっている場合に、その距離制約を有効に活用して予測精度を向上させる試みについて論じており、具体的には、既存の最大重みクリークに基づくアルゴリズムを改良した二種類のアルゴリズムを提案している。一つは精度を保ちながら速度向上を図ったもので、もう一つは制約に関する条件を緩和することにより予測精度の改善を図ったものである。両方のアルゴリズムについて計算機実験を行い、もとのアルゴリズムと比較して意図した改良がなされていることを確認している。また、距離制約を用いないタンパク質スレッディング法との比較も行い、距離制約を用いることの有効性を確認している。

第4章から第6章では、側鎖パッキング問題と呼ばれる、タンパク質の概要構造が既知の場合に詳細構造を予測する問題に取り組み、三種類のアルゴリズムを提案している。

第4章では、幾何情報のみを考慮した側鎖パッキング問題を取り扱い、その問題を最大クリーク問題に帰着し、既存の最大クリークアルゴリズムを適用することにより最適解を計算し、予測を行うというアルゴリズムを提案している。そして、計算機実験により、比較的小さなサイズのタンパク質に対し、最適解が現実的な時間で計算でき、良好な精度での予測が行えることを示している。

第5章では、より高精度の予測を行うために、擬似エネルギーを用いた側鎖パッキング問題を取り扱い、その問題を最大重みクリーク問題に帰着し、それに最大重みクリークアルゴリズムを適用し、最適解もしくは準最適解を計算することにより予測を行うアルゴリズムを提案している。なお、擬似エネルギー関数に関してはアミノ酸間の距離の分布に基づく既存の関数を用いている。そして、計算機実験により、第4章で述べた手法、および、他の既存手法との比較を行い、提案するアルゴリズムの有効性を確認している。

第6章では、大規模なタンパク質の側鎖パッキングを行うために、分割統治法に基づくアルゴリズムを提案している。よ

り具体的には、タンパク質をいくつかの部分に分割し、不整合が起きないように工夫を凝らしながら分割された領域ごとに第5章で提案したアルゴリズムに基づく手法を適用し、最後にそれぞれの解を組み合わせるにより全体構造を予測するというアルゴリズムを提案している。そして、計算機実験により、800個程度のアミノ酸からなるタンパク質に対しても現実的な時間で、かつ、良好な精度で予測が行えることを確認している。

第7章は結論であり、本研究のまとめと今後の課題について述べている。

論文審査の結果の要旨

本論文は、タンパク質立体構造予測に関する二種類の問題に対するアルゴリズムについて述べたもので、得られた成果は以下の通りである。

(1) タンパク質スレッディングでは構造を予測したいタンパク質のアミノ酸配列と構造既知のタンパク質立体構造間の対応関係（アライメント）を計算する。一方、近年の実験技術の進歩により、一部分のアミノ酸間の近似的な距離が計測できるようになってきた。そこで、この距離に関する制約を用いることによりスレッディングの精度を向上させることが期待できる。本論文では、この距離制約を用いた場合のスレッディング問題に対する最大重みクリークに基づく既存のアルゴリズムに、二種類の改良を加えた手法を開発した。一つは精度を保ちながら速度向上を図ったもので、もう一つは制約に関する条件を緩和することにより予測精度の改善を図ったものである。

(2) タンパク質側鎖パッキングはタンパク質の概要構造がわかっている場合に詳細な構造を予測する問題である。側鎖パッキング問題の幾何学的な定式化に対し、その問題を最大クリーク問題に帰着させることにより、予測を行うというアルゴリズムを提案した。

(3) 側鎖パッキングの予測精度の向上を図るために、擬似エネルギーを考慮した定式化を行い、その問題を最大重みクリーク問題に帰着させることにより、予測を行うというアルゴリズムを提案した。

(4) 大規模なタンパク質に対して側鎖パッキングを計算するために、分割統治法に基づき、タンパク質をいくつかの部分に分割し、不整合が起きないように工夫を凝らしながら分割された領域ごとに(3)で述べたアルゴリズムに基づく手法を適用し、最後にそれぞれの解を組み合わせるにより全体構造を予測するというアルゴリズムを提案した。

さらに本論文では、上記(1)―(4)のいずれのアルゴリズムに関しても、公開されているタンパク質立体構造データを用いた計算機実験を行い、その有効性を確認した。これらの問題はNP困難な問題であり、それに対して実際のタンパク質に適用可能で、かつ、現実的な時間で最適解もしくは準最適解を計算することを示したことは高く評価できる。

以上、本論文はバイオインフォマティクス（生命情報学）において重要な研究テーマであるタンパク質構造予測に関連して、有用かつ独創的な手法を示しており、当該分野の発展のために十分な寄与をしている。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成18年1月26日実施した論文内容とそれに関連した試問の結果合格と認めた。