

審査の結果の要旨

氏名 中村 司

本論文は三章からなる。第 1 章では、タンパク質のアミノ酸配列から三次元構造を予測する新たなシステムに関して述べている。本システムでは、複数のタンパク質からなる複合体の構造を予測することもできる。Template-based Modeling (TBM) では、データベースに存在する三次元構造既知のタンパク質の配列から、ターゲットとしているタンパク質に対応する部分を発見することが重要である。そのためには、単なるアミノ酸行列同士のアラインメントではなく、profile-profile アラインメントを用いる必要がある。Profile-profile アラインメントでは、あるタンパク質と類似する配列をデータベースより取得し、マルチプルアラインメントを行なった上で、確率的なプロファイルとして扱う。そして、プロファイル同士のアラインメントを行うことで、元の配列では明らかでない関連性を発見することができる。本研究では、profile-profile アラインメントにアルゴリズム上での工夫を施すことで性能を向上させた。その結果、2016 年に行われた CASP 競技会において、優秀な成績を収めることができた。

第二章では、三次元構造予測でも重要な役割を果たす、マルチプルシークエンスアラインメントの並列化法について記述している。MAFFT は、世界でも広く使われているタンパク質配列のマルチプルシークエンスアラインメントソフトウェアであるが、全配列対全配列のペアワイズアラインメントを行なっているため、多数の配列を扱うと計算時間が過大になるという問題があった。本研究では、5000 本以上の配列を扱える G-large-INS-1 という手法を提案している。この手法では、MPI (Message Passing Interface) を用いて、多数の計算機上で同時に実行可能である。QuanTest (Le et al., 2017)、HomFam (Sievers et al., 2011)、OXFam (Raghava et al., 2003)、ContTest (Fox et al., 2016) という 4 種類のベンチマークデータで他の手法に比べて精度が優れていることが確認された。

第三章では、タンパク質の表面に存在しているポケットに関して新たな特徴量を提案している。タンパク質が薬剤などのリガンドと結合する上で、表面の凹んだ部分であるポケットは重要な役割を果たす。リガンドが結合することが確定しているポケットに対し、タンパク質三次元構造データベース上で類似したポケットを探索することで、当該リガンドが結合する他のタンパク質を発見することができる。従来は、アミノ酸が構成する

三角形を抽出し、その幾何的な情報を表現することで、特徴量を計算していたが、この方法は、関連しているポケットの発見に失敗する場合があることが知られている。本研究では、三角形に関する類似度行列を作成し、その固有値分解を行うことで特徴量を計算する方法を提案した。大規模な計算実験の結果、従来法より精度が優れていることが確認された。

全ての章の研究は他の研究者との共同研究であるが、論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。したがって、博士(科学)の学位を授与できると認める。

以上、1367字