

論文審査の結果の要旨

氏名 伊東 純一

本論文は5章からなる。1章はイントロダクション、2章は類似したタンパク質とリガンドの結合部位を高速かつ正確に数え上げる計算方法を提案し、3章では提案手法を約120万箇所 of 結合部位へと適用し、4章では異なるタンパク質のファミリーへと適用し、5章では提案手法が予測したタンパク質リガンド結合部位のデータベースについて述べている。

1章では問題の背景について説明している。タンパク質とリガンドの結合部位を実験的に検証したデータは蓄積されているものの、タンパク質およびリガンドの数が多いため、結合部位についての我々の知識は乏しい状態にある。そのため、既存の知識から結合部位を精度良く予測できる方法への期待は大きい。

2章では、予測を実現するための基礎的な情報として、タンパク質とリガンドの結合部位の3次元構造情報から特徴量を抽出する方法を提案している。アプローチは、3次元構造を頂点間の辺から構成される三角形へとまず分解し、各三角形の辺の長さを5つのグループへと分類することで属性を抽出し、さらに各頂点のアミノ酸残基の電荷チャージも属性として利用している。ここまでは既存の研究が採用している典型的アプローチであるが、従来の方法が属性の存在数により特徴量を記述していたのに対して、本研究では Gaussian kernel を使うことを提案している点があたらしい。結合部位の3次元構造の類似性は、この特徴量ベクトルの正規化した内積として定義している点は従来手法を踏襲しているものの、Gaussian kernel による特徴量は従来の定義に比べて、類似構造の予測に威力を発揮することが ROC curve を使った分析から明らかになっている。

このように提案された新しい特徴量は予測に有効であることが明らかになった。しかしながら構造の類似性を比較する際にかかる計算時間は大きい。類似度の高い点の組を数え上げるステップが要となるが、対象とする点の数を n とするとき、すべての点間の類似度を計算する単純な網羅的な探索を実行すると n^2 に比例する時間がかかるため、現実的な時間で計算することは困難になる。そこで3章では、田部井らによって開発されたランダムアルゴリズムに基づく計算方法を採用した場合の効果を検討している。このアルゴリズムを用いると、ある類似度以下の類似構造を若干のエラーが含まれるものの n に線形的に比例する時間で高速に計算できることを示している。エラー率の期待値を小さくできるようにパラメータをチューニングできるため、実用上は極めて低いエラー率で高速に類似構造を網羅できることが検証されている。

この高速手法を約120万箇所の結合部位へと適用し、4章では異なるタンパク質のファミリーへと適用している。これほど多くの結合部位の類似度を計算した前例研究はない。さらに5章では提案手法が予測したタンパク質リガンド結合部位のデータベースを公開している。このデータベースは、特定のタンパク質に結合することを想定して設計されたある化合物が、他のタンパク質に結合するかどうか

かを判定するのに使うことができるので、たとえば薬の副作用の可能性を検証するにはとても有用である。

以上のように提案方法は、正確性、高速性において従来手法を凌駕しており、公開されたデータベースは薬の安全性を評価する際に幅広く利用される可能性がある。なお、本論文は、田部井靖夫、清水佳奈、富井健太郎、津田宏治との共同研究であるが、論文提出者が主体となって開発、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。