

論文審査の結果の要旨

申請者氏名 角田 将典

本論文は、カーネル法を用いたタンパク質の構造および機能の予測についての研究であり、サポートベクターマシン(SVM)を用いたタンパク質間相互作用予測と、Isomapを用いた非線形次元削減によるポリペプチドの自由エネルギー解析について報告している。

第一章の序論に続き、第二章では、タンパク質ペアに対するカーネル関数を新たに設計し、これを SVM で用いたタンパク質間相互作用予測法の開発について報告している。タンパク質間相互作用は、多くの生物学的機能に関与している。そのため、細胞内でどのようなタンパク質間相互作用が存在するかを知ることは重要である。しかし、取りうるタンパク質の組合せは膨大であるため、全ての組合せを実験によって確かめるには多大な労力を必要とする。これを解決するため、本論文では、機械学習アルゴリズムのひとつである SVM を用い、与えられたタンパク質ペアの配列をもとに、そのペアが相互作用するか否かを予測する手法を開発した。SVM はデータ間の関係（類似度）をカーネル関数によって計算する。SVM はこのカーネル関数によって計算された値を用いて識別関数を構築するため、カーネル関数は予測精度に大きく影響し、精度の良い予測器を得るためには、適切に類似度を計算できるカーネル関数を用いる必要がある。そのため、タンパク質配列のペアに対するカーネル関数の設計を行った。カーネル関数設計は、構造を有するデータに対するカーネル関数の設計法として知られる Haussler の畳み込みカーネルに基づいて行われている。タンパク質ペアのカーネルを、ペアの構成要素であるアミノ酸配列のカーネルの組み合わせで定義し、このアミノ酸配列のカーネルには文字列に対する畳み込みカーネルである spectrum カーネルを用い、配列ペアに対するカーネルとして、アミノ酸配列のカーネルの組み合わせ方の異なる 10 種のカーネル関数を作成した。作成したカーネル関数をヒト及び *H. pylori* のタンパク質間相互作用データを用い、5 分割交差確認法で評価を行い、その結果、もっとも良い性能のカーネル関数では、予測の評価指標の一つである AUC (area under the ROC curve) の値として、それぞれ、0.88、0.93 を報告している。また、既存のタンパク質ペアを入力とする予測手法と比較し、予測精度が向上していることを確認し、ペア（非順序対）を適切に扱えるように設計したカーネル関数の予測精度がもっとも高いことを明らかにした。また、配列に対するカーネルとして、タンパク質ドメインを考慮した新たなカーネル関数の設計についても報告されている。

第三章では、分子動力学シミュレーションで生成されたポリペプチドの構造群に、非線形次元削減法の Isomap を適用することで、自由エネルギー解析を行う手法について報告している。分子動力学シミュレーションやモンテカルロシミュレーションなどで生成された分子構造群から安定状態や準安定状態、遷移状態の解析を行う際に、しばしば、次元削減が行われる。これは、高次元のデータのままで解析が可能なほどの大量の構造を、シミュ

レーションでサンプリングするのが困難なためであるのと、データの解析を容易にするためである。サンプリングした構造群の分布に基づいて系統的に次元を削減する手法として、しばしば、主成分分析や多次元尺度構成法などが用いられる。これらは、構造間の距離関係をできるだけ保つような直交座標に次元削減を行う。ここで問題となるのは、主成分分析や多次元尺度構成法で保存するのは構造間の直線距離である点である。分子は結合長、結合角、二面角、衝突、静電相互作用などから制約を受け、どのような構造でもとれるわけではないので、サンプリングされた構造の周辺を分子が取り得る構造と仮定し、取り得る構造のみを経る道なりの距離関係をできるだけ保つような次元削減が適当な場合もあると考えられる。この様な非線形の次元削減を行うのが **Isomap** である。**Isomap** は距離の近いデータ間を辺でつないだ重み付きの無向グラフを作成し、データ間の距離をこの無向グラフ上での最短距離で定義し、最後に、得られた距離行列に多次元尺度構成法を適用して次元削減を行う手法である。**Isomap** のシミュレーションで生成された分子構造への適用は報告されているが、主成分分析や多次元尺度構成法との比較は、どちらが無向グラフ上での最短距離と相関が強いのか、という観点からのみ行われている。しかし、自由エネルギー解析では、安定状態や準安定状態の数、その相対配置なども重要である。本論文では、**Isomap** と多次元尺度構成法を、ペントアラニンとデカアラニンの 100ns (10 万構造) の定温分子動力学シミュレーションの結果に適用し、両手法から求まる空間での自由エネルギー曲面を比較した。その結果、多次元尺度構成法による次元削減では、構造の異なる他の安定状態と重なってしまっている準安定状態が、**Isomap** による次元削減では分離されていることを確認し、**Isomap** による非線形次元削減が自由エネルギー解析に対し有効であることを示した。

以上、本研究は、カーネル法を用いた、アミノ酸配列ペアからのタンパク質間相互作用予測法および分子動力学シミュレーションで得られた構造群からの自由エネルギー解析法の開発について報告するものであり、学術的・応用的意義は大きい。よって、審査委員一同は、本論文が博士（農学）の学位論文として価値あるものと認めた。