

(続紙 1)

京都大学	博士 (情報 学)	氏名	鎌田 真由美
論文題目	Analysis and Prediction Methods for Protein Structure and Function タンパク質立体構造と機能に対する解析及び予測手法の開発		
(論文内容の要旨)			
本論文はタンパク質のダイナミクスの情報解析手法と相互作用部位の予測手法、および、ペプチド配列の特性の予測手法について述べられており、5章から構成されている。 第1章では、研究の動機、タンパク質分子の特徴、実験による立体構造決定手法、構造ダイナミクスの特徴、タンパク質に対する従来のバイオインフォマティクス研究など、研究の背景について述べている。 第2章では、タンパク質の立体構造の時間変化データの解析手法を提案している。具体的には、分子動力学シミュレーションにより得られる立体構造の時系列データに対し、ウェーブレット変換を行い、さらに、それに特異値分解を用いて特徴抽出を行うことにより、ノイズを除去した低次元データを得る手法を提案している。さらに、そのようにして得た低次元データを用いて、各アミノ酸の近傍の集団運動の強さを表す指標を定義している。これらの指標を用いて、エステル結合を加水分解する酵素群であるリパーゼおよび酸素分子貯蔵、運搬を行うミオグロビン・タンパク質の時系列データに対してクラスタリングおよび正準相関分析を行うことにより、連動して運動するアミノ酸残基群を同定し、それらが実験などにより既に得られていた生物学的知見と合致していることを確認している。 第3章では、タンパク質配列データが与えられた時、タンパク質間および一分子タンパク質内で相互作用するアミノ酸残基を予測する手法を提案している。提案手法では、まず、複数の類似配列を多重整列することにより得られたアミノ酸間の相互情報量を計算する。次に、この一連の相互情報量データを画像データとして捉え、相互作用するアミノ酸対を予測する問題を画像のラベリング問題として定式化する。そして、画像のラベリング問題に対する有効な統計的モデルであるDiscriminative Random Fieldモデル、および、二次元のConditional Random Fieldモデルを用いた機械学習・予測手法を開発している。これらの手法を実際の相互作用タンパク質データをベンチマークとして用いて評価し、その有用性を確認している。 第4章では、少数のアミノ酸からなるペプチド配列に対する線形回帰を用いた特性の予測手法を提案している。ペプチド配列から既存の情報解析手法により様々な特徴量を得ることができ、それを用いた線形回帰により、ペプチドの性質を予測することができる。一方、線形回帰手法として、従来から2次の正則化項を持つRidge回帰と、1次の正則化項を持つLasso回帰が広く用いられてきた。提案手法では、Lasso回帰を最初に適用することにより特徴選択を行い、次に選択された特徴量のみを用いてRidge回帰を行うことにより最終的な予測を行う。ただし、損失関数としては単なる二乗誤差ではなく対数関数を用いた関数を用いている。そして、この手法をペプチド配列の抗体との結合親和性のベンチマーク・データを用いて評価し、その有効性を確認している。 第5章は結論であり、本研究をまとめるとともに、今後の研究の方向性や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、重要な生体内高分子であるタンパク質に関して、ウェーブレット変換と特異値分解を用いた構造変化の時系列データ解析方式、相互情報量とマルコフモデルを用いた相互作用部位の予測方式、Lasso回帰とRidge回帰を用いたペプチド分子の抗原に対する結合親和性の予測方式について述べたものであり、得られた成果は以下のとおりである。

(1) タンパク質立体構造のダイナミクスを理解することは、タンパク質の機能の予測やより深い理解のために重要である。これまでも様々な研究が行われてきたが、分子動力学シミュレーションにより得られた時系列データを解析するための手法は十分には開発されてこなかった。そこで、本論文では、ウェーブレット変換と特異値分解を組み合わせることにより、時系列データからノイズを除去し、かつ、低次元のデータに変換し、それを用いて時刻間、および、アミノ酸残基間の相関を測る手法を開発し、さらに、それらを用いて連動して集団運動を行うアミノ酸残基群を検出するための方法を開発した。これらの手法をTLLタンパク質およびミオグロビンの時系列データに適用した結果、既存の生物学的知見と合致する結果が得られ、提案手法の有用性を確認することができた。

(2) タンパク質の多くは他のタンパク質もしくは分子と相互作用することによりその機能を発揮するため、タンパク質配列中の相互作用部位予測は機能の予測や解明のために重要である。これまでに相互作用の有無を予測する手法は数多く提案されてきたが、相手側まで含めた相互作用部位の予測は十分に研究されてこなかった。そこで本論文では、配列進化に関係する相互情報量と画像ラベリングに用いられてきた統計的モデルであるDiscriminative Random Fieldモデル、および、Conditional Random Fieldモデルを組み合わせることにより、学習データからパラメータを推定し、それを用いて新規データに対する相互作用部位予測を行う方式を開発した。そして、実際のタンパク質結合構造データを用いて評価した結果、提案手法の有効性を確認することができた。

(3) 少数のアミノ酸からなるペプチドの抗原との結合親和性を予測することは免疫システムの理解や創薬のために有用であり、これまでも多くの研究が行われてきた。しかしながら、未だに十分な予測精度を得るに至っておらず、現在でも多くの研究が行われている。本論文では、1次の正則化項を持つLasso線形回帰により特徴量を選択した後、選択された特徴量のみを用いて2次の正則化項を持つRidge線形回帰により予測を行う二段階法を開発した。この手法に対してベンチマーク・データを用いて評価したところ、既存手法と同等以上の予測精度を持つことが示された。

以上、本論文ではバイオインフォマティクスにおける重要な研究テーマである、タンパク質立体構造のダイナミクス解析、2個のタンパク質間、そして一分子のタンパク質内におけるアミノ酸残基間の相互作用部位予測、ペプチドの特性予測に対して新規な手法を提案するとともに、シミュレーションデータやベンチマーク・データを用いた計算機実験により、それらの有効性を示している。提案手法のいずれもが新規性、有用性、拡張性の高いものであり、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。