

L0023-13	SVMを用いた β シート構造の解析と予測					
	氏名	小林 貴幸	主査	黒田	副査	朝倉 中村俊 山田和 中村徳

[背景・目的]

タンパク質の機能はその構造に密接に関係している。そのため、タンパク質を理解・利用するためにタンパク質の立体構造情報はきわめて重要である。実験的にタンパク質の立体構造を求めるのは高コストであるため、低コストな計算的手法でタンパク質の立体構造を求めるアプローチは大変有用である。近年のタンパク質立体構造情報の急増に伴い、種々の二次構造・タンパク質立体構造予測手法が開発されてきた。しかし、主流になっている手法はいずれも近傍な配列の特徴・相互作用を基に構造予測を行うため、アミノ酸配列上で離れている残基間の相互作用の影響を受ける構造についての予測を苦手としている。そこで本研究はそのような構造の一つである β シート構造、すなわち β ストランドの水素結合形成についての予測を行う。

[研究方法]

本研究では機械学習法の一つであるSVMを用いた β ストランド対の水素結合形成予測法を開発する。本研究は大きく分けて、以下の4段階から成る。

- ①. データセットの構築、解析
- ②. SVMを用いて β 残基対の水素結合形成の予測
- ③. β 残基対の水素結合形成度数を基に β ストランド対の水素結合度数スコアの算出
- ④. β ストランド対の水素結合度数スコアを用いた β シート構造の予測

[結果・考察]

PDBから配列・構造の冗長性、解析に不適な構造を除いた1754タンパク質からなるデータセットを構築した。立体構造から二次構造を計算するDSSPを用いて二次構造を定義し、解析を行った。その結果、アンチパラレルな水素結合を形成する β ストランド対、パラレルな水素結合を形成する β ストランド対においてアミノ酸組成、水素結合形成残基対の頻度(Figure.1)、残基間のアミノ酸配列上の距離による水素結合形成頻度の違いといった差異が存在することが判明した。

これらの特性を基にSVMに学習させるデータを作成、予測を行った。その結果、構築した予測機の β ストランド対の水素結合の検出感度は62.2%、正確性は53.6%であり、コントロールとして用いたナイーブアルゴリズムと比べて検出効率で11%、正確性で4%、他の既存の手法と比較して検出効率が3%の予測精度の向上を果たした (Table 1.)。学習・予測の際の特徴記述子を変化させた結果、比較的近距離のストランド対の予測にはアミノ酸の電荷と残基間のアミノ酸配列上の距離、中・遠距離の β ストランド対の予測にはPSSMが予測精度の向上に貢献することが判明した。

最後に、実際の運用に近い形式での予測機の性能を測るため、二次構造を既存の二次構造予測機であるPSIPREDを用いて定義し、アミノ酸配列より予測した二次構造を基に β ストランド対の水素結合予測を行った。結果は、ナイーブアルゴリズムに対して5%の検出効率の向上を果たした (Table 1.)。この結果は、本研究で開発した手法により、アミノ酸配列から β シート構造のトポロジーを予測できることを示している。

アミノ酸配列上で離れた残基間の相互作用を考慮した、 β シート構造の予測が可能となった本研究の成果を応用することで、フォールト探索コストの低減など、新規タンパク質の構造予測の精度向上へ貢献することが期待できる。今後はさらなる予測精度の向上を目指しパラメータ・アルゴリズムの最適化に取り組みたい。

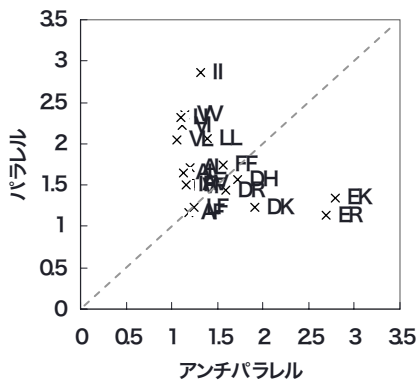


Figure 1.水素結合形成残基対の頻度の差 (一部を抜粋)

	検出効率(%)	正確性(%)
SVM	62.2	53.6
ナイーブアルゴリズム	51.2	49.1
既知の手法	59	53
SVM+PSIPRED	59.4	40.6
ナイーブ+PSIPRED	54.1	41.1

Table 1.各予測機・アルゴリズムにおける予測精度