

タンパク質立体構造および相互作用の情報解析手法

Computational methods for analyzing protein tertiary structures and protein interactions

京都大学化学研究所 バイオインフォマティクスセンター 数理生物情報 阮佩穎

背景と目的

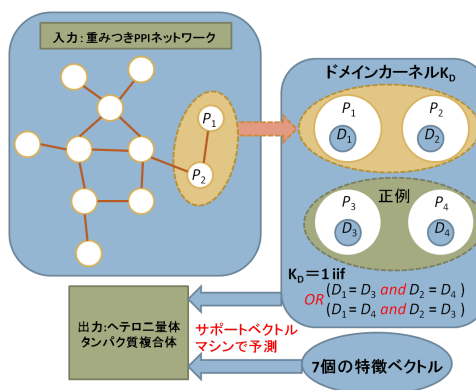
生物に様々な機能を賦与している遺伝子にとって、主要な機能発現の本体であるタンパク質を解析することは重要であり、タンパク質の機能と立体構造の解明が創薬などさまざまな分野への応用につながる。近年、種々の生物種のゲノム配列解析が急速な発展を見せる反面、タンパク質の立体構造の解明は実験によって決定するには時間がかかると同時に費用を要するため、立体構造データの増加はそれほど高くないのが現状である。そのため、予測によって決定する方法が重要になっている。また、より高次の構造として複数のタンパク質が結合した複合体と呼ばれる構造がある。複合体は転写や代謝などの多くの生命現象に関与しているので、複合体を予測することができれば、遺伝子制御ネットワークや代謝ネットワークを理解する手掛かりとなる。これまで、タンパク質立体構造および複合体予測の情報解析手法はいくつか研究されてきたが、現在では十分な精度や効率を持つアルゴリズムは提案されておらず、新しい手法の登場が期待される。それぞれのタンパク質が、どのような構造を持てばその機能を発現できるのか、を解明することが重要である。本研究では、タンパク質立体構造の解析、高精度・高速モデリング技術の開発、高精度の複合体予測手法の開発を目指す。

検討内容

複合体によっては酵素として化合物を反応させたり、転写因子として遺伝子を発現させたりすることが知られている。複合体を高精度に予測できれば、代謝ネットワークと遺伝子制御ネットワークの解明に貢献できる。しかし小さいサイズの複合体は、酵母の複合体データベース CYC2008[1]においては二量体が 42%、三量体が 21%のように多数を占めるにも関わらず、構造がより大きなサイズの複合体よりだいぶ単純であるため、これまで高精度に二量体を予測する手法は存在せず、ほとんどが少なくともサイズ 4 以上の複合体を対象にしていた[2,3]。したがってサイズ 2、3 の小さいサイズの複合体についての精度の高い予測手法の開発が期待される。

結果

本研究は右の図のように、タンパク質を頂点、相互作用を辺とした重みつきタンパク質相互作用 (PPI) ネットワークを入力として、ヘテロ二量体タンパク質複合体を予測する。PPI の重みはその相互作用の信頼度を表す。PPI の重み、隣接する辺、ドメインの数などに基づいた 7 つの特徴ベクトルを設計した。サポートベクトルマシン (SVM) を用いた計算機実験結果は従



来手法を予測精度において上回った。さらに、タンパク質のドメイン構成が既知の複合体と同じならば、それらのタンパク質が複合体になる可能性が高いというアイデアに基づき、ドメインカーネルというオリジナルなカーネルを提案した。既存手法で F-measure が 0.31 であったものがドメインカーネルにより 0.63 に大幅に改善した。

参考論文

[1]S. Pu et al, Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Research* 37:825-831, 2009.

[2]Y. Qi et al, Protein complex identification by supervised graph local clustering. *Bioinformatics* 24:i250-i258, 2008.

[3]O. Maruyama. Heterodimeric protein complex identification, *ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, 499-501, 2011.