

論文内容の要旨

博士論文題目 複合体立体構造モデルを用いたタンパク質間相互作用予測

氏名 福原直志

タンパク質間相互作用は多くの生体内プロセスを支えており、特に相同タンパク質の中から特異的に相互作用する相手タンパク質を見つけることは、シグナル伝達などの細胞機能の解明に極めて重要である。これらの相互作用を決定するために様々なハイスループットの実験手法が開発され、多量のデータが得られてきた。しかしながら、これらの手法が適用された生物種の数はいくつか、その実験精度にも限界があるため、計算機を用いてアミノ酸配列や立体構造情報からタンパク質間相互作用を予測する手法の開発が望まれている。

本研究では、ホモロジーモデリングされた複合体立体構造に基づいて、特異的に相互作用するタンパク質を予測する手法を開発した。これは、本質的には相同なタンパク質ペアの中から特異的に相互作用するタンパク質ペアを予測することを意味する。先行研究で用いられてきたコンタクトエネルギーに加え、新たにシンプルな静電エネルギーとテンプレート構造との配列類似度の二つのスコアを導入し、これらの単独および結合スコアを計算することで、各複合体立体構造モデルの妥当性を推定した。予測結果については、DIP データベースに登録されているタンパク質ペアを相互作用するペア、登録されていないペアを相互作用しないペアとする新たな基準を設け、両者の識別力によって評価を行った。

本手法を酵母のヘテロの全タンパク質ペアに適用した。10,325 個のタンパク質の二量体モデル構造が作成され、そのうち 417 個が相互作用するタンパク質ペアとして DIP に登録されていた。Recall-Precision プロットと F-measure の最大値による評価の結果、配列類似度スコアは立体構造に基づくスコアよりも高い識別力を有していることが分かった。また、配列類似度にコンタクトエネルギーを結合したスコアを用いることで識別力を有意に改善することができた。

これらの研究結果を踏まえ、いくつかの事項に関して検討を行った。まず、複合体立体構造モデルに基づく先行研究(Davis *et al.*, *Nucleic Acids Res.*, **34**, 2943-2952 (2006))の手法と本手法の性能比較を行った。先行研究はコンタクトエネルギーのみならず細胞内局在情報も特徴量として用いているため、本研究のコンタクトエネルギーよりも優れていたが、本研究の配列類似度の性能には及ばないことが分かった。また、これまで評価データとして用いてきた相互作用するタンパク質ペアと相互作用しないペアの中から、細胞内局在情報を導入し新たに信頼性の高い評価データを選択した。再度、評価を行ったところ特徴量の識別力の優劣に変化は見られず、配列類似度とコンタクトエネルギーを結合した場合に最良の性能が得られるという本研究の結論を確認することができた。

最後に、本研究手法を他の研究者が自由に利用できるように、ウェブサーバを開発した。これは、二つの標的配列を入力するとテンプレート構造の候補を列挙し、ユーザーが指定したテンプレートに基づいて作成した複合体立体構造モデルをエネルギー値とともに表示するものである。酵母以外のタンパク質配列に対しても適用可能であるため、有用性の高いツールであると考えている。

論文審査結果の要旨

平成18年12月28日に行われた公聴会の結果を参考に、平成19年6月5日に本博士論文の審査を実施した。本論文は、二つのアミノ酸配列から、それらのタンパク質が相互作用するかどうかを計算機科学的に予測する方法の開発に関する研究である。そのために、ホモロジーモデリングによって予測された二量体の立体構造を生成し、予測された二量体構造群を、「相互作用するペア」と「相互作用しないペア」に識別するための、立体構造及び配列の特徴を用いたスコア関数を開発した。本論文の主たる成果は以下に要約される。

1. 相互作用するペアとしないペアを識別するためのスコア関数として、先行研究で用いられていたコンタクトエネルギーに加え、本論文では、静電エネルギーと、さらにテンプレート構造との配列類似度のスコアを導入した。それぞれのスコアの予測性能を評価データセットを用いて調べたところ、配列類似度は、他に比べて格段に性能が高いこと、しかしながら、コンタクトエネルギーと配列類似度を組み合わせたときに、最良の予測性能が発揮されることがわかった。従来、コンタクトエネルギーと配列類似度は、それぞれ別の研究グループによって相互作用予測に使われていたが、その二つを組み合わせ使用し、単独で用いた場合より改善したという報告は初めてである。配列と構造を組み合わせたスコアを設計するための最初の提案として、この研究分野への貢献は高いと考えられる。
2. 立体構造を用いた相互作用予測の先行研究では、相互作用データベースに登録されているタンパク質ペアを「相互作用する」ペアの正解とみなして、どれだけ予測相互作用と重なるかで評価をしていたが、「相互作用しない」と予測されたペアが真に相互作用しないかどうかの検討は十分になされていなかった。本論文では対象を酵母の全タンパク質に限定し、相互作用データベースに含まれていないペアを全て「相互作用しない」ペアの正解とみなしたデータセット、さらに細胞内局在データを用いて信頼性の高いデータのみを抽出したデータセットの2種を用いて、これまでより精密な予測精度の評価を行った。前項に挙げた配列のスコアは構造のスコアより識別能力が高いという事実、配列と構造のスコアを組み合わせたときに最良であるという結果を明示できたのも、この「相互作用しないペア」の導入によるものであり、後続の研究の予測性能の評価法の指針を示すうえで、この研究分野への貢献は小さくないと考えられる。

以上のように、本論文は、配列からのタンパク質間相互作用予測という問題に対して、スコア関数の改良及び評価用データセットの改善において、十分の新規性があり、当該研究分野への寄与は少なくない。また、開発された予測手法、および予測WEBサーバは、タンパク質間相互作用に関心のある広い層の研究者にとって有用であると思われる。したがって、本論文は博士（理学）の学位論文として価値あるものと認める。