

氏名	はやし だ もり ひろ 林 田 守 広
学位(専攻分野)	博 士 (情 報 学)
学位記番号	情 博 第 150 号
学位授与の日付	平 成 17 年 3 月 23 日
学位授与の要件	学 位 規 則 第 4 条 第 1 項 該 当
研究科・専攻	情 報 学 研 究 科 知 能 情 報 学 専 攻
学位論文題目	Computational Analysis and Inference of Protein-Protein Interactions from Domain Information (ドメイン情報からのタンパク質間相互作用の解析と予測)
論文調査委員	(主 査) 教 授 阿 久 津 達 也      教 授 小 林 茂 夫      教 授 後 藤 修

### 論 文 内 容 の 要 旨

本論文は、ドメイン情報に基づくタンパク質間の相互作用の計算機による解析法および予測法について述べたもので、6章から構成されている。

第1章は序論で、本論文で扱うタンパク質間の相互作用の解析および予測問題について概観している。特に、実験技術による相互作用検出の問題点や、計算機による解析や予測手法の現状について述べるとともに、本論文の構成の概要を示している。

第2章では、まず、本論文の基礎となるドメイン情報に基づくタンパク質相互作用の既存の確率モデルについて説明した後、そのモデルに基づく既存の相互作用推定法である Association 法および EM 法について説明している。本論文では、確率モデルに基づき、既知のタンパク質間相互作用データからドメイン間の相互作用の確率を推定することにより、未知のタンパク質間相互作用を推定するという予測方式を採用しているが、第2章では、ドメイン間の相互作用確率推定問題を線形計画問題として定式化して解くという新しいアプローチによる予測方式である LPBN 法を提案している。さらに、この方式が既存の EM 法と組み合わせることも可能であることを示している。提案方式の有用性を検証するために、公開されているデータベース中のタンパク質相互作用データを用いて計算機実験を行い、少なくともテストデータに対しては既存手法と同程度以上の精度が得られることを確認している。

第3章では、確率モデルに基づき定式化したタンパク質間の相互作用問題の計算論的側面について理論的考察を行っている。具体的には、学習データとして与えられた既知相互作用データをできるだけ正しく分類するドメイン間の相互作用確率を計算する問題が、NP 困難さらには MAX SNP 困難であることを証明している。

第4章では、実際の実験によって得られるデータの性質を考慮し、相互作用の有無ではなく、何回実験を行って何回相互作用が観測されたかという相互作用の観測頻度をもとに、タンパク質間相互作用の確率(強度)を推定する問題を提案している。この問題を解くために、本論文では LPNM と ASNM という二種類の予測手法を提案している。LPNM 法は第2章で提案した線形計画法に基づく LPBN 法を相互作用の観測頻度データに対応できるように拡張した手法であり、ASNM 法は従来手法である Association 法を観測頻度データに対応できるように拡張した手法である。本章では、これらの手法の有用性を検証するために、公開されているデータベース中のタンパク質相互作用の頻度データを用いて計算機実験を行い、相互作用の強度推定問題に対しては、LPNM 法、および、ASNM 法は精度の面で従来手法である Association 法や EM 法をはるかに上回ることを確認している。また、LPNM 法と ASNM 法に関しては、予測精度の観点からは LPNM 法が ASNM 法より良いが、計算時間の観点からは ASNM 法がはるかに優れていることを確認している。

第5章では、ドメインの重複、突然変異、融合を反映した確率モデルに基づくタンパク質のドメイン構成の進化、および、その結果として得られるネットワークについて解析を行っている。理論解析のために、ドメインの重複および突然変異のみを反映した確率モデルを定義し、その漸近的な性質が既存の優先的選択法の解析手法を拡張することにより解析可能である

ことを示し、解析の結果として、このネットワークモデルが「スケールフリー性」、すなわち、頂点の次数分布がべき乗則に近似的に従うという性質を持つことを示している。また、提案する確率モデルのシミュレーションを行い、その結果得られるネットワークの次数分布が、実際のタンパク質のドメイン構成から得られるネットワークの次数分布と類似していること、および、両者ともにスケールフリー性が近似的に成立していることを確認している。

第6章は結論であり、本研究のまとめと今後の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、ドメイン情報に基づくタンパク質間の相互作用の計算機による解析法と予測法について述べたもので、得られた成果は以下の通りである。

- (1) タンパク質は一般にドメインと呼ばれる部分構造の集合により構成されると考えられ、そのドメイン情報に基づくタンパク質間相互作用の確率モデル、および、その確率モデルに基づく相互作用の予測法が従来から提案されてきた。本論文では、その予測問題を線形計画問題として定式化して解くという新たな予測手法を開発した。そして、公開データベースにおけるタンパク質相互作用データを用いて、既存手法との計算機実験による比較を行い、その有効性を確認した。
- (2) 上記で述べた問題に対して、計算論的側面から理論的な解析を行い、学習データに対する分類精度を最大化するという定式化のもとで、相互作用の予測問題がNP困難、さらには、MAX SNP困難であることを示した。
- (3) 実際の生物学実験により得られるデータの性質を考慮し、相互作用の有無ではなく、相互作用が観測される頻度情報をもとに、ドメイン間の相互作用確率を推定する問題を提案し、線形計画法を用いた予測法と、従来の association 法を拡張した予測法の二種類を開発した。そして、公開データを用いて計算機実験を行い、新たな問題設定のもとでは、いずれの提案手法を用いても従来手法を大幅に上回る精度が得られることを示した。
- (4) ドメインの重複、突然変異、融合を反映した確率モデルに基づくタンパク質のドメイン構成の進化、および、その結果として得られるネットワークについて理論および計算機実験による解析を行った。理論解析においては、重複および突然変異のみに基づく確率モデルについて解析し、このモデルにより得られるネットワークにおいて、頂点の次数分布がべき乗則に近似的に従う、すなわち、スケールフリー性を持つことを示した。一方、重複、突然変異、融合を考慮した確率モデルについては、計算機によるシミュレーション結果と、実際のタンパク質のドメイン構成から得られるネットワークを比較することにより、両者における次数分布が類似し、かつ、両者ともスケールフリー性を持つことが確認され、提案する確率モデルの妥当性が示唆された。

以上、本論文は生命情報学（バイオインフォマティクス）において重要な研究テーマであるタンパク質間相互作用の計算機による解析および予測のための独創的かつ有用な手法を示しており、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。

また、平成17年2月23日実施した論文内容とそれに関連した試問の結果合格と認めた。