

論文 / 著書情報  
Article / Book Information

題目(和文)	立体構造情報に基づくタンパク質間相互作用ネットワーク予測
Title(English)	Protein-Protein Interaction Network Prediction Based on Tertiary Structure Data
著者(和文)	大上雅史
Author(English)	Masahito Ohue
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9553号, 授与年月日:2014年3月26日, 学位の種別:課程博士, 審査員:秋山 泰,佐藤 泰介,関嶋 政和,瀬々 潤,杉山 将
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9553号, Conferred date:2014/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

専攻： Department of	計算工学	専攻	申請学位 (専攻分野)： 博士 (工学)
学籍番号： Student ID Number			指導教員 (主)： 秋山 泰
学生氏名： Student's Name	大上 雅史		指導教員 (副)： 瀬々 潤

### 要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

近年、疾病の原因の特定や薬剤標的の決定を目的として、タンパク質間相互作用 (Protein-Protein Interaction, PPI) と呼ばれる細胞内生命現象の網羅的な解明を目指した研究が活発に行われている。PPI とは、生体内の複数のタンパク質が互いに結合するなどして、機能の促進・抑制や、新たな機能の獲得が行われる現象である。自己免疫疾患のように PPI の変動が原因となる疾病も存在し、生体内のタンパク質の制御関係を大規模かつ網羅的に理解することは、生命科学における重要課題とされている。PPI は生化学実験によって決定することが通常であるが、日々新たに発見され続けるタンパク質に対して実験コストは増加する一方であり、計算機による PPI 予測手法、特に、大量のタンパク質を扱うことのできる予測手法への期待が高まっている。

本論文は、立体構造情報に基づく網羅的な PPI 予測を実現させることを目的として、高速なタンパク質ドッキング計算を用いて予測を行うシステム「MEGADOCK」を開発し、またその応用を行った結果を報告する。計算機による PPI 予測の従来手法には配列情報や共進化情報などに基づく手法が存在するが、これらの手法は物理化学的側面に基づいた予測ができない、類似していない配列における PPI の検出が難しいといった問題があった。本研究ではこの問題を解決するため、立体構造情報を用いることを提案した。立体構造情報を扱う場合、原子レベルの時間発展をシミュレーションする計算を用いると 1 組のタンパク質ペアでも数日レベルの計算を要するが、本論文では剛体グリッドモデルによるタンパク質ドッキング計算を利用し、さらに新しい評価関数を提案することで、現実的な時間内での網羅的計算を可能とした。

MEGADOCK は大きく 2 つの部分によって構成されており、(1)タンパク質の結合を予測するタンパク質ドッキング計算を行う部分と、(2)ドッキング計算の結果に基づいた PPI の有無を判定する部分とに分かれている。(1)のタンパク質ドッキング計算では、形状相補性・静電相互作用・脱溶媒和自由エネルギーの 3 つの物理化学的要素を、新しい形状相補性モデル real Pairwise Shape Complementarity (rPSC) と簡易型脱溶媒和自由エネルギーモデル Receptor Desolvation Free Energy (RDE) の提案によって、1 回の畳込み演算で同時に計算可能とした。この提案手法を広く用いられているベンチマークデータセットである Protein-Protein Docking Benchmark 4.0 (Hwang, *et al.* 2010) によって評価した結果、現在一般的に用いられているタンパク質ドッキング計算の従来手法 ZDOCK (Mintseris, *et al.* 2007) と比較して、同等精度の計算を約 9.8 倍高速に行えることを示した。また、(2)の PPI 判定では、原子レベルのエネルギー計算による予測複合体の再評価を予測に組み入れることで精度を向上させられることを示した。14,400 組のタンパク質ペアに対する PPI 予測では F 値 0.231 の精度を示し、ネットワークレベルの予測が可能であることを確認した。

さらに、MEGADOCK を用いた網羅的 PPI 予測の実問題への応用として、細菌走化性シグナル伝達系 (約 1 万組)、ヒトアポトーシスシグナル伝達系 (約 4 万組) への応用を行った。細菌走化性シグナル伝達系の応用では、既知の PPI を F 値 0.464 の精度で再現することに成功し、従来まで発見されていなかった相互作用の可能性を持つタンパク質ペアを提案した。ヒトアポトーシスシグナル伝達系では、既知の PPI を F 値 0.220 の精度で再現することに成功し、酵素タンパク質群の新たな相互作用関係を示唆した。加えて、タンパク質と並んで重要である RNA 分子を扱うための力場パラメータの拡張を行い、RNA 結合タンパク質に対する応用も行った。約 6 千組のタンパク質と RNA のペアに提案手法を適用し、F 値 0.465 の精度でタンパク質と RNA の相互作用を予測できることを実験的に示した。

MEGADOCK は大規模並列計算機上で実行するための並列実装や GPU 実装が行われている。TSUBAME や「京」によるスケーラブルな計算が可能となっており、本研究によって数百万件規模の PPI 予測を立体構造情報に基づいて行うことが初めて可能となった。本研究の成果は、たとえばがん細胞や微生物の細胞などを対象としたシステム生物学の研究において、システム内のタンパク質の制御関係を理解・予測するための新たな解析支援ツールとして広く利用されることが期待される。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 2 部提出してください。

Note: Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 2 copies of 800 Words (English).

(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

専攻 : Department of	計算工学	専攻	申請学位 (専攻分野) : Academic Degree Requested	博士 Doctor of	( 工学 )
学籍番号 : Student ID Number			指導教員 (主) : Academic Advisor(main)		秋山 泰
学生氏名 : Student's Name	大上 雅史		指導教員 (副) : Academic Advisor(sub)		瀬々 潤

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words )

Protein-protein interactions (PPIs) are fundamental in the majority of cellular processes and their study is of enormous biotechnological and therapeutic interest. The computational prediction for elucidation of PPI networks is crucial in biological fields. However, the development of an effective method to conduct exhaustive PPI screening represents a computational challenge.

In this dissertation, we proposed a novel PPI network prediction system called MEGADOCK based on protein-protein docking calculation with protein tertiary structure information. MEGADOCK reduced the calculation time required for docking by using new score functions, rPSC and RDE, and was implemented on recent parallel high-performance computing environments by employing a hybrid parallelization with MPI and OpenMP and general-purpose graphics processing unit technique.

We showed that MEGADOCK is capable of exhaustive PPI screening and completed docking calculations 9.8 times faster than the conventional method (Mintseris, *et al.* 2007) while maintaining an acceptable level of accuracy. When MEGADOCK was applied to a subset of a general benchmark dataset to predict 120 relevant interacting pairs from 14,400 protein combinations, an F-measure value of 0.231 was obtained. Moreover, the system was scalable as shown by measurements carried out on two supercomputing environments, TSUBAME 2.0 and K computer.

It is now feasible to search and analyze PPIs while taking into account three-dimensional structures at the interactome scale. We demonstrated the applications to pathway analyses, bacterial chemotaxis, human apoptosis, and RNA binding proteins by using our system. As an example of the results, when analyzing the positive predictions of bacterial chemotaxis pathway from MEGADOCK, all the core signaling interactions were correctly predicted with the exception of interactions activated by protein phosphorylation.

Large-scale PPI prediction using tertiary structures is an effective approach that has a wide range of potential applications. This method is especially useful for identifying novel PPIs of new pathways that control cellular behavior.

備考 : 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 2 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 2 copies of 800 Words (English).