

令和元年6月20日現在

機関番号：12608

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K00388

研究課題名(和文) 網羅的ドッキングを用いた熱帯病関連タンパク質間相互作用解析

研究課題名(英文) Tropical diseases-related protein-protein interaction analysis by using exhaustive docking

研究代表者

松崎 由理 (Matsuzaki, Yuri)

東京工業大学・リーダーシップ教育院・特任准教授

研究者番号：30572888

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：本研究では、網羅的タンパク質ドッキングによって対象タンパク質群から相互作用し得るペアを予測する。まず、既存手法を改良して網羅的タンパク質ドッキングによる高精度な相互作用予測手法を開発した。次に、寄生原虫のタンパク質、ウイルスと宿主であるヒトのタンパク質を対象に本手法を適用してタンパク質間相互作用(Protein-protein interaction, PPI)探索を実施し、新規PPI候補を67件得た。本手法を用いて予測した2件の内1件について、免疫沈降法を用いて実際に結合を確認した。さらに、ウイルス-宿主タンパク質間相互作用の特徴について、ドッキング結果に基づく比較解析を行なった。

研究成果の学術的意義や社会的意義

熱帯病を引き起こす寄生生物やウイルスに関する基礎的な生物学研究は、経済的な問題からあまり進んでおらず、多くのPPI情報が既に知られているヒトとは異なり、新規PPI発見やPPIネットワークの理解が創薬におけるターゲットタンパク質の選択に有用な情報となる。PPI予測は困難な問題である一方、現時点で15万件をこえるタンパク質立体構造データが公開データベース(PDB)に蓄積されている。従来のPPIネットワーク予測研究では十分には活用されてこなかったこれらのデータを活用することで、未知PPIの発見や、ウイルス-宿主間PPIの特徴の発見などの結果を得ることが期待できる。

研究成果の概要(英文)：We have developed a computational method of protein-protein interaction (PPI) prediction by using interaction profile analysis of rigid-docking decoys. By combining originally developed interaction fingerprint representation of docking decoys and an extended machine learning (extra-trees) method, we have successfully improved prediction performance of conventional docking based prediction methods. We applied this PPI prediction method on predicting i) PPIs of *Trypanosoma cruzi* and ii) interactions of dengue virus and human proteins and obtained in total 67 novel PPI candidates. Two candidate pairs of high prediction score were tested by immunoprecipitation experiments and one was confirmed to have affinity. We have also conducted comparison analysis of virus-human and human-human protein interactions in terms of interaction fingerprints of docking decoy set.

研究分野：システム生物学

キーワード：タンパク質間相互作用 タンパク質ドッキング タンパク質間相互作用ネットワーク

### 1. 研究開始当初の背景

タンパク質間相互作用 (Protein-protein interaction, PPI) はシグナル伝達など細胞機能の制御における中心的現象である。ウイルスの感染、増殖過程の多くは、宿主の制御機構を担う PPI を攪乱する。様々な感染症の理解には、PPI の全体図である PPI ネットワークの正確な把握が必要である。PPI ネットワークの研究には、実験、計算の両面から多くのアプローチが適用されてきた。Yeast two-hybrid 法 (Y2H) などハイスループットな実験手法も利用されてきたが、高速かつ信頼性の高い方法はまだ開発されていない。計算機を用いるアプローチでは、既知 PPI との配列モチーフの類似性やタンパク質の階層的分類 (オントロジー)、配列情報やタンパク質の特徴量を入力として二つのタンパク質の親和性を予測する機械学習による予測などが報告されている。しかし、タンパク質の機能において本質的である立体構造を活用した研究は、分子動力学法 (MD) などの動的解析が一部行われているほか、既知の複合体構造を参照した予測が試みられはじめたのみであった。MD は詳細にタンパク質の運動をシミュレーションできる一方、膨大な計算時間を必要とするため、解析の規模は 1 対 1 の PPI のへの試用にとどまっている。

### 2. 研究の目的

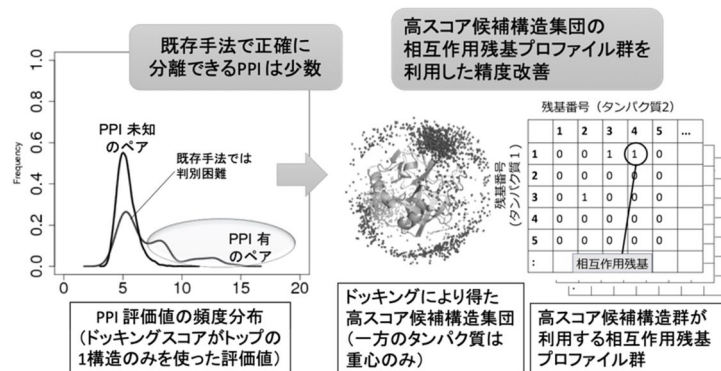
我々は、高速実行可能なタンパク質ドッキングソフトウェア「MEGADOCK」を並列計算機上に実装し、候補タンパク質構造間の網羅的ドッキング結果を利用して PPI 予測を行なう手法を開発した。(ここでは、物理的な結合が起き得るタンパク質ペアの予測を「PPI 予測」と呼ぶこととする。)本研究では、タンパク質立体構造データを活用し PPI 予測技術を改良し、大規模解析に応用することを目的とする。応用として、熱帯地方に見られるシャーガス病の原因となるトリパノソーマ原虫や、デング熱を引き起こすデングウイルスとヒトのタンパク質を対象として PPI 予測を実施し、ヒト体内での原虫の生存や、ウイルスの感染、増殖に重要な PPI を探索する。

タンパク質ドッキングは、もともと二つのタンパク質の結合ポーズを予測するための手法であり、典型的なドッキングソフトウェアでは一つのペアにつき数千の候補ポーズをスコア付けの上出力する。我々は既に、一般的なベンチマークデータおよび *E. coli*, *T. maritima* の走化性系タンパク質群に本手法を適用してランダムより有意に高い精度で予測に成功している。本研究ではこの手法をさらに高度化し、医学的に重要な実問題に適用することを目指す。

### 3. 研究の方法

#### 3-1. 網羅的タンパク質ドッキングによる高精度な相互作用予測手法の開発

ドッキング計算を用いた PPI 予測について、既存手法では主に候補構造の中で最も高いドッキングスコアのみ依存して PPI 予測を行ってきた (Matsuzaki, *et al.*, *Protein Pept Lett*, 2007)。一方我々は、実際に二つのタンパク質が近くに位置してから正しい場所で結合するまでに、いわゆる「正解」結合構造以外の場所でも弱いコンタクトを繰り返しながら結合状態に至っているのではないかという見地から、最良スコアの構造のみでなく、高スコア構造群の性質にも注目することにより PPI 予測の精度を上げることが



図：本研究で実施する PPI 予測

できる可能性があると考えている。同様の仮定から、高スコア構造群のスコア平均値を用いた予測方法が既に提案されている (Wass, *et al.*, *Mol Syst Biol*, 2011)。しかし、ドッキングスコアのみを用いるこの手法では、得られた高スコア構造群の構造情報が全く生かされていない。そこで本研究では、ドッキングの結果得られた高スコア構造群の相互作用プロフィールを利用した予測手法の改良を試みた (図)。

相互作用プロフィールは、研究分担者の内古閑が開発したタンパク質複合体の相互作用残基のコンパクトな表現手法である (Uchikoga, *et al.*, *BMC Bioinform*, 2010)。本研究では、相互作用プロフィールとドッキングスコアの組み合わせを入力とし、あるタンパク質に複数の相互作用候補タンパク質をドッキングした結果から、相互作用するタンパク質を予測する手法を新たに提案した。まず高スコアドッキングモデル群の相互作用残基プロフィール表現方法を開発した (Uchikoga, *et al.*, *Biophys Physicobiol*, 2016)。具体的には、ドッキングの結果得られた数千の高スコア構造群で相互作用に利用されている残基の集合を、ドッキングスコアと組み合わせたプロフィールとして表現した。このデータを入力とし、SVM, Extra-trees などの機械

学習手法を用いる複数の PPI 予測手法を実装し、一般的なベンチマークデータおよびヒトの既知 PPI 情報に適用して評価した。

### 3-2. 熱帯病関連タンパク質間相互作用の探索

シャーガス病を引き起こす寄生原虫 *Trypanosoma cruzi* 体内の PPI 予測、また、その分泌するタンパク質とヒト体内のタンパク質との PPI 予測を、3-1. で開発した手法および従来のドッキングベースの予測手法を用いて実施した。

*T. cruzi* については、517 の立体構造データを収集し、網羅的 PPI 予測に必要なドッキング計算を行った。 Dengue ウイルスについては 83 構造、ヒトについては 30,544 構造を収集した。ヒトのタンパク質構造については、過度に冗長なデータを除去するなどして 10,000 構造程度にしぼり、830,000 件程度のドッキング計算を実施した。

得られた予測結果を既知データと照合して未知の PPI を抽出し、そのうち特に評価値の高い 2 件の PPI 候補について、免疫沈降法を用いて相互作用の有無を検証した。

### 3-3. ドッキング結果に基づくウイルス-宿主タンパク質間相互作用の解析

3-2 でドッキング計算を実施済の宿主-ヒト PPI の相互作用プロファイル群と 3-2. で得られたウイルス-宿主タンパク質のドッキング結果を比較し、ヒトタンパク質をウイルスのタンパク質とドッキングした場合に、本来の結合相手とのドッキング結果と比べ相互作用残基の分布に特徴が見られるか調べた。

## 4. 研究成果

### 4-1. 網羅的タンパク質ドッキングによる高精度な相互作用予測手法の開発

3-1 で述べた複数の予測手法を、一般的なベンチマークデータ (Vreven, *et al.*, *J Mol Biol*, 2015) に適用して評価した。予備研究では、11 のターゲットタンパク質に対し 120 のタンパク質の中から相互作用する相手を識別する問題セットで予備解析を実施し、有意な精度向上が見られることを確認した。このうち、Extra-trees 法を拡張した手法を用いて、最新のベンチマークデータ(230 のターゲットタンパク質)に適用したところ Area Under the Receiver Operation Curve (0-1 の値、ランダム予測で 0.5、完全な予測で 1.0) の値を、レセプターとするタンパク質によって異なるが、平均 0.79 まで向上した。

そこで、対象をさらに広げてヒトのタンパク質間相互作用作用を対象に同手法を適用し、より広い対象での試験を行った。その結果、予測の有効性が対象とするタンパク質によって大きく異なることがわかった。本手法では、ドッキングの結果得た相互作用プロファイルを入力とし、既知のタンパク質間相互作用情報を学習して予測モデルを構築する。そのため、既知の情報が少ないタンパク質については、相互作用相手の予測に失敗しやすくなっていると考えた。

そこで、対象とするタンパク質ごとに予測モデルを立てるのではなく、より一般化したモデルの構築を行うことで、予測精度の低い対象について精度を向上することを試みた。入力情報として、高スコア構造群の相互作用面の原子レベルのグラフ表現を用いることとし、入力された候補結合構造の結合の強さを予測する手法を開発し、小規模なテストデータにて良好な結果を得た。

### 4-2. 熱帯病関連タンパク質間相互作用の探索

*T. cruzi* の PPI と、 Dengue ウイルス-ヒト PPI の予測を行った結果、検討すべき新規 PPI 候補を *T. cruzi* について 39 件、 Dengue ウイルス-ヒト間について 28 件得た。さらに、これまでに実施したヒトのタンパク質間相互作用予測について、得られた新規相互作用候補のうち 2 組について、免疫沈降法を用いた検証を行い、1 組について実際に結合がみられることを確認した。

### 4-3. ドッキング結果に基づくウイルス-宿主タンパク質間相互作用の解析

4-2. で予測されたウイルス-ヒト PPI と、既知のウイルス-ヒト PPI について、ヒト側のタンパク質の既知相互作用相手のドッキング結果と比較する。この比較から、ウイルスが特徴的に利用する相互作用残基パターンがあるのか、それとも宿主の PPI の特徴に似せた相互作用パターンを示すのかを見出すことを試みた。相互作用プロファイルの単純な比較においては、両者に違いを見出すことができなかった。

一方、ドッキングの高いスコア構造群の 3 次元空間分布について、既知の結合ポーズとその周辺の結合モデルの間に、本来の PPI ペアと外来のタンパク質のペアとの間で差がある可能性を示唆する結果を得た。PPI 予測手法の改善と合わせて、今後の課題とする。

## 5 . 主な発表論文等

[雑誌論文](計 3件)

1. Takanori Hayashi, Yuri Matsuzaki, Keisuke Yanagisawa, Masahito Ohue, Yutaka Akiyama, MEGADOCK-Web: and integrated database of high-throughput structure-based protein-protein interaction predictions., *BMC Bioinformatics*, **19**:62-72, 2018. (査読あり)
2. Yuri Matsuzaki, Nobuyuki Uchikoga, Masahito Ohue, Yutaka Akiyama, Rigid-docking approaches to explore protein-protein interaction space., *Adv Biochem Eng Biotechnol*, **160**:33-55, 2017. (査読あり)
3. Nobuyuki Uchikoga, Yuri Matsuzaki, Masahito Ohue, Yutaka Akiyama, Specificity of broad protein interaction surfaces for proteins with multiple binding partners., *Biophys Physicobiol*, **13**:105-115, 2016. (査読あり)

[学会発表](計 14件)

1. Serika Taga, Nobuyuki Uchikoga, Takanori Sasaki, Pattern of codon usage for chemotaxis related protein genes in *E. coli* and *T. maritima*., The 22nd International Conference on Research in Computational Molecular Biology (RECOMB2019), France, 2019.
2. Masahito Ohue, Takanori Hayashi, Yuri Matsuzaki, Keisuke Yanagisawa, Yutaka Akiyama, Megadock-Web: an integrated database of high-throughput structure-based protein-protein interaction predictions., *Biophysical Society 63rd Annual Meeting*, USA, 2019.
3. Masahito Ohue, Takanori Hayashi, Hiroki Watanabe, Yuri Matsuzaki, Nobuyuki Uchikoga, Yutaka Akiyama, Supercomputing-based exhaustive protein-protein interaction prediction and its open database., *The 56th Annual Meeting of the Biophysical Society of Japan*, Japan, 2018.
4. Nobuyuki Uchikoga, Yuri Matsuzaki, Analysis of amino acid sequences of protein interaction surfaces by rigid-body docking for known and unknown protein complex pairs., *The 56th Annual Meeting of the Biophysical Society of Japan*, Japan, 2018.
5. Takanori Hayashi, Masahito Ohue, Juliette Martin, Guillaume Launay, Yuri Matsuzaki, Nobuyuki Uchikoga, Yutaka Akiyama, A large decoy dataset for protein-protein docking model quality assessment., *The 56th Annual Meeting of the Biophysical Society of Japan*, Japan, 2018.
6. Yuri Matsuzaki, Jaak Simm, Nobuyuki Uchikoga, A docking based approach to analyze interaction surfaces of protein-protein Interactions., *The 25th Conference on Intelligent Systems for Molecular Biology (ISMB2017)*, Czech Republic, 2017.
7. Nobuyuki Uchikoga, Yuri Matsuzaki, Masahito Ohue, Yutaka Akiyama, Analysis of protein interaction surfaces using a profile method with rigid-body docking decoys., *The 25th Conference on Intelligent Systems for Molecular Biology (ISMB2017)*, Czech Republic, 2017.
8. Nobuyuki Uchikoga, Yuri Matsuzaki, Masahito Ohue, Yutaka Akiyama, Protein interaction surfaces of protein-protein interaction networks in bacterial chemotaxis networks using profile methods., *The 55th Annual Meeting of the Biophysical Society of Japan*, Japan, 2017.
9. Nobuyuki Uchikoga, Yuri Matsuzaki, Masahito Ohue, Yutaka Akiyama, Interaction surfaces of proteins involved in bacterial chemotaxis with rigid-body docking decoys., *Biophysical Society 61st Annual Meeting*, USA, 2017.
10. Masahito Ohue, Yuki Yamamoto, Takanori Hayashi, Yuri Matsuzaki, Yutaka Akiyama, Cloud computing for all-to-all protein-protein docking on Azure HPC., *Biophysical Society 61st Annual Meeting*, USA, 2017.
11. Yuri Matsuzaki, Nobuyuki Uchikoga, Jaak Simm, Yutaka Akiyama, A docking based approach to analyze interaction surfaces of virus-host protein-protein interactions., *Biophysical Society 61st Annual Meeting*, USA, 2017.
12. Yuri Matsuzaki, Rigid docking based protein-protein interaction prediction by using high scoring docking models., *The 54th Annual Meeting of the Biophysical Society of Japan (Symposium 2SEA: Applications of protein structure data for understanding biological phenomenon)*, Japan., 2016.
13. Nobuyuki Uchikoga, Profile analysis of protein interaction surface with rigid-body docking decoys., *The 54th Annual Meeting of the Biophysical Society of Japan (Symposium 2SEA: Applications of protein structure data for understanding biological phenomenon)*, Japan, 2016.
14. Yuri Matsuzaki, Nobuyuki Uchikoga, Jaak Simm, Protein-protein interaction network prediction by using rigid-body docking., *17th International Conference on Systems*

Biology (ICSB2017), Spain, 2016.

〔図書〕(計 0件)

〔産業財産権〕  
出願状況(計 0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年：  
国内外の別：

取得状況(計 0件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年：  
国内外の別：

〔その他〕  
ホームページ等  
なし

## 6. 研究組織

### (1)研究分担者

研究分担者氏名：内古閑 伸之  
ローマ字氏名：Uchikoga Nobuyuki  
所属研究機関名：明治大学  
部局名：総合数理学部  
職名：助教  
研究者番号(8桁)：20397483

研究分担者氏名：黒川 裕美子  
ローマ字氏名：Kurokawa Yumiko  
所属研究機関名：国立遺伝学研究所  
部局名：新分野創造センター  
職名：特任研究員  
研究者番号(8桁)：10381633

### (2)研究協力者

研究協力者氏名：Jaak Simm  
ローマ字氏名：Yaku Shimu  
所属研究機関名：KU Leuven

研究協力者氏名：Juliette Martin  
ローマ字氏名：Jurietto Marutan  
所属研究機関名：Lyon University 1

研究協力者氏名 : Guillaume Launey

ローマ字氏名 : Giyomu Renei

所属研究機関名 : Lyon University 1

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。