

# タンパク質間相互作用予測結果 データベース及び表示系の構築

長澤 一輝<sup>1</sup> 松崎 由理<sup>2</sup> 大上 雅史<sup>3</sup> 秋山 泰<sup>2,3,a)</sup>

**概要:** タンパク質間相互作用 (protein-protein interaction, PPI) の網羅的な理解に向けて、我々は立体構造情報から高速に PPI を予測するソフトウェア MEGADOCK を開発してきた。この PPI 予測結果の利活用には、結果をデータベースとして集約し、俯瞰的に閲覧できる表示系を構築することが必要不可欠である。そこで、本研究では MEGADOCK による予測 PPI のデータベースとその表示系である MEGADOCK-WEB の開発を行った。開発したデータベース及び表示系は、あらかじめ計算された PPI 予測結果の検索に加え、構築された複合体モデルの可視化や相互作用相手のパスウェイ上での可視化が可能で、PPI 予測結果や未知 PPI の可能性の検討に役立てられる。

**キーワード:** タンパク質間相互作用 (PPI), MEGADOCK-WEB, 予測結果データベース

## 1. 導入

生体内のタンパク質のおよそ 3 分の 2 はタンパク質間相互作用 (protein-protein interaction, PPI) によってその機能を発揮していると言われ、PPI は生命現象の中核を担っている。近年では PPI 阻害薬も注目され、PPI の理解は生命現象の解明だけでなく新規薬剤の開発を行う上でも重要となる。しかし、特定のタンパク質群に対して実験で網羅的に PPI を決定することは時間や費用の面で困難である。したがって、計算機を用いて PPI 予測を行い、相互作用の可能性が高いタンパク質のペアをあらかじめ選出することが有用である。

計算機によって PPI の予測を行う手法には、立体構造情報を用いる手法の他、タンパク質のアミノ酸配列情報や共進化情報に基づく手法などが挙げられる。このうち立体構造情報を用いる手法のひとつに、本研究室で開発された PPI 予測ソフトウェア MEGADOCK [1,2] がある。MEGADOCK は 2 つのタンパク質の立体構造を入力として、高速フーリエ変換を用いた剛体グリッドモデルによるタンパク質のドッキング計算を行い、その結果に基づいて PPI を予測するソフトウェアである。評価関数設計や GPU 並列化などの工夫により、大量のタンパク質の中から高速に PPI の可能性の高いタンパク質ペアの候補を予測

することができるようになってきている [2]。

一方、計算機による PPI 予測結果の有効な利活用には、予測データを広く公開して多くの生物学者による検証を必要とする。近年では、PrePPI [3] など、インターネット上で利用できる予測 PPI 情報を含んだデータベースも複数存在する。このように、PPI の情報を公開する方法として、ウェブ上にデータベースと表示系を構築する方法が一般的にとられている。本研究では、立体構造情報を利用したドッキング計算に基づいて予測を行う MEGADOCK によって予測された PPI 情報のデータベースの構築および、それに関わる表示系である MEGADOCK-WEB の開発を行った。

## 2. 関連研究

予測 PPI 情報を閲覧できる既存データベースに PrePPI [3,4] がある。PrePPI は、既の実験的に決定されている PPI の複合体の構造情報を複数組み合わせることでテンプレートを作り、これを用いて予測された PPI の情報をデータベースとしてまとめたものである。このような既知の複合体構造を用いる予測では、既知の PPI に対応する複合体の立体構造と大きく異なる複合体構造を形成するような予測結果は得られにくく、そのような新規の PPI の発見が難しいという問題点が存在する。

<sup>1</sup> 東京工業大学 工学部 情報工学科  
<sup>2</sup> 東京工業大学 情報生命博士教育院  
<sup>3</sup> 東京工業大学 大学院情報理工学研究科 計算工学専攻  
a) akiyama@cs.titech.ac.jp

### 3. MEGADOCK-WEB の実装

#### 3.1 収録データ

タンパク質の構造情報は Protein Data Bank (PDB) [5,6] から入手した 3,780 種類のヒトのタンパク質の立体構造情報を使用した。このとき、タンパク質の配列の類似度が 90% 以上の構造データは 1 つの PDB ID のみ収録することで、データの冗長性を抑えた。PPI 予測情報は MEGADOCK 4.0 [2] による全対全計算の予測結果  $3,780C_2 + 3,780 = 7,146,090$  件を収録した。

#### 3.2 予測 PPI データベースの実装

データベースおよびその表示系の実装にはフレームワークとして Play Framework 2.2 [7] を使用した。データベースの管理は Java で実装されている SQL データベースである H2 Database Engine [8] を使用した。本データベースではクエリとして PDB ID, UniProt AC, タンパク質名, 遺伝子名を用いて PPI の検索を行うことを可能にするため、PDB の REST サービスを利用して PDB ID と鎖名から UniProt AC を取得し、続いてその UniProt AC を使用して UniProt [9] からタンパク質名, 遺伝子名, 生物種, KEGG ID などの情報を取得し、データに関連付けを行った。

#### 3.3 予測された相互作用相手のパスウェイ上へのマッピング

本システムは、KEGG [10] に登録されたパスウェイ情報を利用して、クエリタンパク質との相互作用が予測されるタンパク質がパスウェイ上のどの場所に存在しているかを表示することで、予測 PPI を生化学的な側面から検証するための手段を提供する。このパスウェイ上での相互作用相手の可視化は、KEGG が提供する経路上の指定したタンパク質を着色して表示する REST 形式の API を使用して、予測 PPI の相手のタンパク質が属する KEGG パスウェイ ID を集約して着色することで実現した。

#### 3.4 JSmol による複合体モデルの可視化

本システムでは、分子ビューワ JSmol [11] による予測複合体の 3D モデルの表示を可能とした。複合体構造をインタラクティブに表示することで、ユーザーは MEGADOCK によって予測された PPI に対応する予測複合体構造を実際に見て視覚的に評価を行うことができる。

#### 3.5 既知 PPI 情報の表示

本システムでは、予測 PPI が、既に実験的にも知られているものである際に、実験的に決定された PPI のデータベースへの横断検索を行い、それらのデータをユーザー

が参照できるようにしている。参照するデータベースは BioGRID [12], DIP [13], HPIDB [14], IntAct [15], Mentha [16], MINT [17], VirHostNet [18] の 7 データベースである。各データベースの PPI の情報はデータベースへの横断検索を提供する PSICQUIC [19] を利用し、行っている PSICQUIC へのアクセスは Python ライブラリである Bioservices [20] を使用した。

### 4. MEGADOCK-WEB の利用例

MEGADOCK-WEB の全体図を図 1 に示す、また、ページ遷移の流れを図 2 に示す。本システムではまずトップ画面でクエリを入力し、検索結果から興味のあるタンパク質とその相互作用相手のタンパク質を選択することで PPI 予測の結果を表示する。予測結果のテーブルから複合体を選択することで、複合体の分子ビューワによる表示を行う (図 3)。また、相互作用相手の選択画面から予測 PPI の相手のタンパク質が属するパスウェイの一覧に移動し、パスウェイを選択して相互作用相手のパスウェイ上での着色を行うことができる (図 4)。

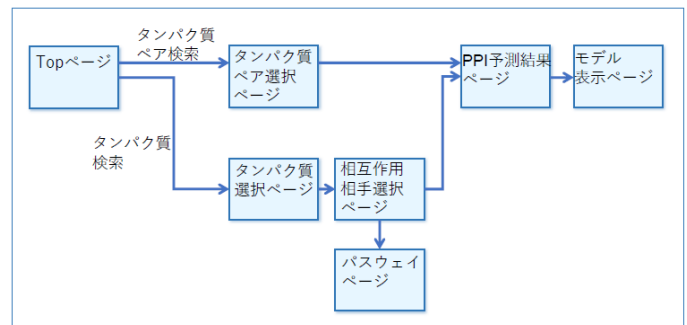


図 1 MEGADOCK-WEB の全体図

### 5. まとめ

本研究では、MEGADOCK によって予測された PPI の情報を集約したデータベースとして MEGADOCK-WEB の開発を行った。また、予測された PPI の評価のために有用な機能として、予測複合体構造の可視化に加え、予測された相互作用相手の生化学経路上へのマッピングといった従来の予測 PPI データベースでは提供されていなかった機能を実装した。

なお、現時点では MEGADOCK-WEB の使用についてユーザーからの意見を得るまでには至っていないため、生物学の研究者に使用してもらい、意見に基づいて MEGADOCK-WEB の改良を行うことが課題である。

謝辞 本研究の一部は JSPS 科研費 (19300102, 11J08750, 14J30002, 15K16081), JST CREST 「EBD : 次世代の年ヨッタバイト処理に向けたエクストリームビッ

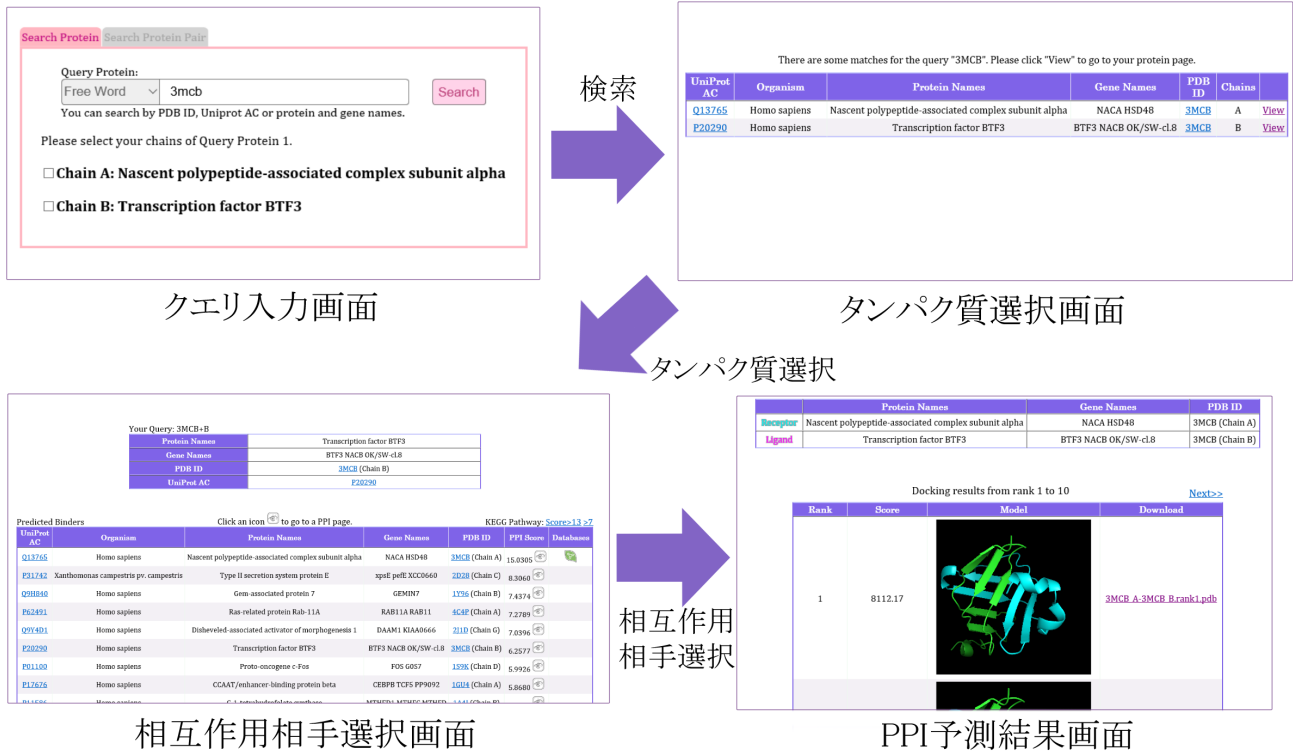


図 2 ページ遷移の流れ

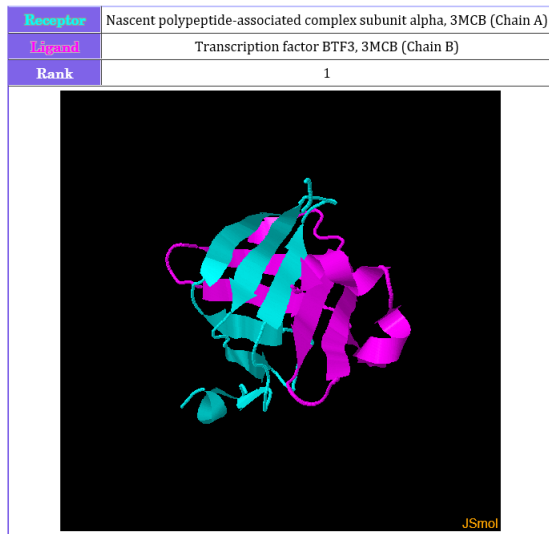


図 3 JSmol による複合体モデルの表示

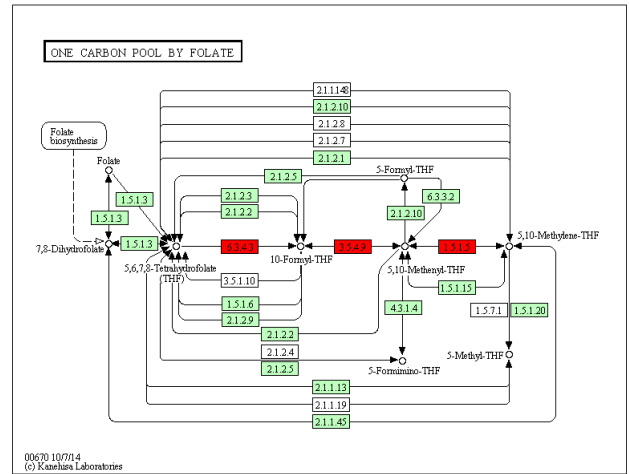


図 4 パスウェイ上での着色

データの基盤技術」, 文部科学省最先端・高性能汎用スーパーコンピュータの開発利用「次世代生命体統合シミュレーションソフトウェアの研究開発」の支援によって行われた。

参考文献

[1] Ohue M, *et al.*, MEGADOCK: An All-to-All Protein-Protein Interaction Prediction System Using Tertiary Structure Data., *Protein Pept Lett*, 21: 766-778, 2014.

[2] Ohue M, *et al.*, MEGADOCK 4.0: an ultra-high-performance protein-protein docking software for heterogeneous supercomputers, *Bioinformatics*, 30: 3281-3283, 2014.

[3] Zhang QC, *et al.*, PrePPI: a structure-informed database of protein-protein interactions., *Nucleic Acids Res*, 41: D828-833, 2013.

[4] <http://bhapp.c2b2.columbia.edu/PrePPI>

[5] Berman HM, *et al.*, The Protein Data Bank., *Nucleic Acids Res*, 28: 235-242, 2000.

[6] <http://www.rcsb.org>

[7] Play Framework <http://www.playframework.com>

- [8] H2 Database Engine <http://www.h2database.com>
- [9] UniProt Consortium, UniProt: a hub for protein information., *Nucleic Acids Res*, 43: D204–212, 2015.
- [10] Kanehisa M and Goto S., KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Res*, 28: 27–30, 2000.
- [11] JSmol: an open-source HTML5 viewer for chemical structures in 3D.  
<http://wiki.jmol.org/index.php/JSmol>
- [12] Chatr-Aryamontri A, *et al.*, The BioGRID interaction database: 2015 update., *Nucleic Acids Res*, 43: D470–478, 2015.
- [13] Salwinski L, *et al.*, The Database of Interacting Proteins: 2004 update., *Nucleic Acids Res*, 32: D449–451, 2014.
- [14] Kumar R and Nanduri B., HPIDB - a unified resource for host-pathogen interactions, *BMC Bioinform*, 11: S16, 2010.
- [15] Orchard S, *et al.*, The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases., *Nucleic Acids Res*, 42: D358–363, 2014.
- [16] Calderone A, *et al.*, mentha: a resource for browsing integrated protein-interaction networks, *Nat Methods*, 10: 690–691, 2013.
- [17] Licata L, *et al.*, MINT, the molecular interaction database: 2012 update., *Nucleic Acids Res*, 40: D857–861, 2012.
- [18] Guirimand T, *et al.*, VirHostNet 2.0: surfing on the web of virus/host molecular interactions data, *Nucleic Acids Res*, 43: D583–587, 2015.
- [19] del-Toro N, *et al.*, PSICQUIC and PSISCORE: accessing and scoring molecular interactions, *Nucleic Acids Res*, 41: W601–606, 2013.
- [20] Cokelaer T, *et al.*, BioServices: a common Python package to access biological Web Services programmatically., *Bioinformatics*, 29: 3241–3242, 2013.