Vol.2018-MPS-118 No.46
Vol.2018-BIO-54 No.46
2018/6/15

ミトコンドリアに関連したヒトタンパク質間相互作用予測データベースMEGADOCK-Web-Mitoの開発

渡辺 紘生[1,2]　林 孝紀[1]　大上 雅史[1]　秋山 泰[1,a)]

概要：ミトコンドリアに関するヒトタンパク質の機能解明の支援を目的として，ミトコンドリアに関するヒトタンパク質 654 種類，全 1,675 種の立体構造に対するタンパク質間相互作用を網羅的に予測し，全ての予測結果を参照可能なデータベース MEGADOCK-Web-Mito を構築した．また，これまでヒトのミトコンドリアに関係するタンパク質の注釈情報が様々なデータベースに散在していたため，これらの情報を統合化し，相互作用予測結果を効率良く詳細に調べることを可能にした．特に，遺伝子の情報をヒトのミトコンドリアに関係するタンパク質に対応づけることで，ヒトのミトコンドリアに関連するタンパク質の機能解明に特化した．ミトコンドリアに関係するタンパク質の網羅的な相互作用予測結果とヒトのミトコンドリアに特化した検索機能により，ミトコンドリア病等の疾病研究の一助となることが期待される．

キーワード：タンパク質間相互作用，PPI 予測，MEGADOCK，MEGADOCK-Web-Mito，ミトコンドリア，タンパク質ドッキング，PPI 予測データベース

Development of MEGADOCK-Web-Mito: human mitochondrial predicted protein-protein interaction database

Hiroki Watanabe[1,2]　Takanori Hayashi[1]　Masahito Ohue[1]　Yutaka Akiyama[1,a)]

Abstract: For the purpose of helping elucidate the functions of human mitochondrial proteins, we performed a comprehensive protein-protein interaction (PPI) prediction for 654 mitochondria-related human proteins the number of whose structures is 1,675, and constructed a database, MEGADOCK-Web-Mito, that enables users to obtain every prediction result. In addition, since annotations of human mitochondrial proteins are scattered around various public databases until recently, we integrated these information and developed a system in MEGADOCK-Web-Mito that enables users to search a PPI prediction result efficiently and in detail. Especially, we linked gene functional annotation to each human mitochondrial protein to make MEGADOCK-Web-Mito specialized in elucidation of human mitochondrial proteins' functions. It is expected that comprehensive human mitchondrial PPI prediction results and searching function of MEGADOCK-Web-Mito specialized in human mitochondria supports researches on mitchondrial diseases whose one of the most major cause is dysfunction of mitochondria.

Keywords: protein-protein interaction (PPI), protein-proten docking, MEGADOCK, MEGADOCK-Web-Mito, mitochondria

[1] 東京工業大学 情報理工学院 情報工学系，
Department of Computer Science, School of Computing, Tokyo Institute of Technology
[2] 産業技術総合研究所 産総研・東工大実社会ビッグデータ活用オープンイノベーションラボラトリ，
AIST-Tokyo Tech Real World Big-Data Computation Open Innovation Laboratory (RWBC-OIL), National Institute of Advanced Industrial Science and Technology (AIST)
[a)] akiyama@c.titech.ac.jp

1. 導入

ミトコンドリアは真核生物の細胞小器官の一つであり，エネルギー産生の場として知られている．また，ミトコンドリアはエネルギー産生に加えて，活性酸素産生，アポトーシス，カルシウムイオンの貯蔵，感染防御などにも関

© 2018 Information Processing Society of Japan

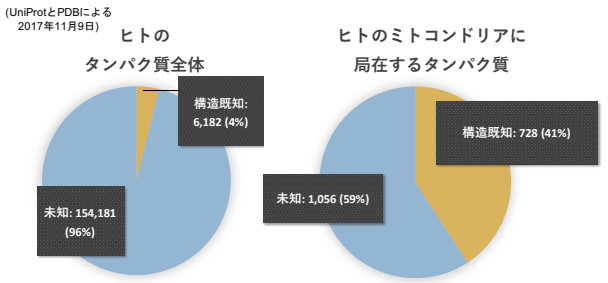


図 1 タンパク質における構造が既知のタンパク質の割合

わっている．さらに，これらのミトコンドリアの機能が障害されることが原因となって発症する，パーキンソン病や糖尿病などの病気を総称してミトコンドリア病 [1] という．これらのことから，近年ミトコンドリアの機能，とりわけミトコンドリア病の原因となるミトコンドリアの機能に対する注目が高まっている．ミトコンドリアに関連するタンパク質の機能の解明がミトコンドリアの機能の解明の一助となる．

生体内の多くのタンパク質は相互作用しながら機能を果たしている [2]．この相互作用はタンパク質間相互作用 (Protein-Protein Interaction, PPI) と呼ばれており生命現象の中核を担っている．また，近年では疾患の要因とされるタンパク質間相互作用を阻害することで効果を得る医薬品の開発も進んでいる [3, 4]．しかし，現在ヒトのタンパク質だけでも約 16 万種（2017 年 11 月 9 日時点，UniProt (Universal Protein Resource) [5] の AC エントリに基づく）のタンパク質が発見されており，それらのタンパク質に対して生化学的実験で網羅的にタンパク質間相互作用を決定することは，時間や費用の面で困難である．したがって，計算機を用いてタンパク質間相互作用を行い，相互作用する可能性の高いタンパク質をあらかじめ選別することが有用である．

計算機によってタンパク質間相互作用予測を行う手法には，タンパク質の立体構造を用いる手法 [6] のほか，アミノ酸配列 [7] や共進化情報 [8] に基づく手法がある．この中で，計算機による既知の相互作用情報を利用しない立体構造ベースのタンパク質間相互作用予測手法のひとつに，当研究室で開発された MEGADOCK [9] がある．MEGADOCK は 2 つのタンパク質構造情報を入力として，タンパク質ドッキング計算を行うソフトウェアであり，既知の相互作用情報に依らない，新規のタンパク質間相互作用を発見することができる．ヒトのタンパク質全体における立体構造が既知のタンパク質の割合が 4%であるのに対して，ヒトのミトコンドリアに局在するタンパク質における立体構造が既知のタンパク質の割合は 41%である（2017 年 11 月 9 日時点，UniProt，PDB (Protein Data Bank) [10] に基づく，**図 1**）ことから，ミトコンドリア関連タンパク質には立体構造を用いる手法が特に有用と思われる．

一方，MEGADOCK によるタンパク質間相互作用予測結果をより利活用しやすい方法で提供するために，我々はタンパク質間相互作用予測統合データベースである MEGADOCK-Web [11] を開発してきた．MEGADOCK で計算された予測タンパク質間相互作用に関する情報の検索・表示機能が実装されており，ヒトの網羅的なタンパク質間相互作用予測を予測構造を含めて閲覧することが可能である．しかし，MEGADOCK-Web はヒトのタンパク質一般に関するタンパク質間相互作用予測データを対象としたデータベースであり，ミトコンドリアに関連するタンパク質間相互作用予測に特化した機能をもたない．ミトコンドリア病の主要な原因の一つがミトコンドリアに関連するタンパク質の配列情報をもつ遺伝子の変異や異常であることなどから，ミトコンドリアのタンパク質の機能の解明には遺伝子の情報も重要である．よって，ミトコンドリアの生命機能やミトコンドリアに起因する疾病の原因の解明のためには，ミトコンドリアに関連するタンパク質の配列情報をもつ遺伝子の情報も対応づけるなど，ミトコンドリアに関連するタンパク質間相互作用予測データに特化したデータベースが必要である．

本研究では，核ゲノムにコードされるタンパク質を含む，ミトコンドリアに局在する，ヒトのタンパク質に関する立体構造を取得し，それらの情報を様々な公共データベースが提供する情報から抽出する．その後，取得したタンパク質の立体構造を用いてミトコンドリアに関連するタンパク質間相互作用予測を網羅的に行う．そして，予測結果をミトコンドリアに関連するタンパク質，タンパク質の配列情報をもつ遺伝子の情報とともに MEGADOCK-Web-Mito としてデータベース化することを目的とする．

2. 関連研究

2.1 MitoProteome

MitoProteome [12,13] は，UCSD Supercomputer Center の Cotter らによって開発された，ヒトの核ゲノムに主にコードされている，ミトコンドリアに局在する遺伝子・タンパク質配列のデータが収録されているオブジェクト関係データベースである．各エントリはミトコンドリアに局在する遺伝子に対応し，各遺伝子の基礎情報，アイソフォームやスプライスバリアント，対応するタンパク質の機能等について注釈付け（アノテーション）がされている．

2.2 MEGADOCK-Web

MEGADOCK-Web [11] は当研究室で開発した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で，MEGADOCK によるヒトの代表的なタンパク質に対する網羅的なタンパク質間相互作用予測の結果が収録されている．現行のバージョンである MEGADOCK-Web version 1.0 のページ構成を**図 2** に示す．

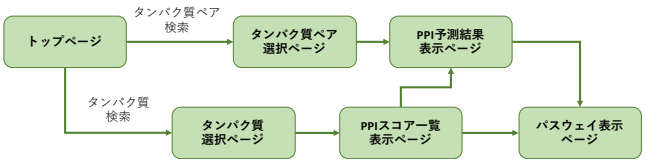


図 2 MEGADOCK-Web version 1.0 のページの遷移図

以下に各ページについて説明する．

- **トップページ**： MEGADOCK-Web のトップページで，このページから検索機能を使用する．
- **タンパク質選択ページ**： 単一のタンパク質の検索結果が表示される．
- **タンパク質ペア選択ページ**： タンパク質のペアの検索結果が表示される．
- **PPI スコア一覧表示ページ**： タンパク質選択ページで選択されたタンパク質とデータベース中のタンパク質との PPI スコアの一覧が表示される．
- **PPI 予測結果表示ページ**： タンパク質ペア選択ページまたは PPI スコア一覧表示ページで選択されたタンパク質のペアの予測立体構造が表示される．
- **パスウェイ表示ページ**： タンパク質との PPI スコアが一定以上のタンパク質が属する KEGG パスウェイへのリンクが表示される．

PPI スコアとは MEGADOCK によるタンパク質間相互作用予測計算の結果から算出される値で，2 つのタンパク質が相互作用するかを予測する指標である．通常は PPI スコアにある一定の閾値を設定し，閾値以上の PPI スコアをもつタンパク質のペアが相互作用すると予測する．また，KEGG (Kyoto Encyclopedia of Genes and Genomes) [14] は遺伝子の機能を体系的に分析し，ゲノム情報を高次機能情報とともに関連させたデータベースであり，生化学反応における経路（パスウェイ）が KEGG パスウェイとして収録されている．KEGG にデータベース化された生化学反応における経路はパスウェイと呼ばれる．KEGG パスウェイとは，代謝経路の情報を統合化し，パスウェイ図で可視化できるようにしたデータベースである．

MEGADOCK-Web の主な機能を以下に挙げる．

(1) MEGADOCK で計算された予測タンパク質間相互作用の検索機能
(2) 予測された相互作用相手の KEGG パスウェイマップ上での可視化機能
(3) 複合体モデルの可視化機能

3. MEGADOCK-Web-Mito

MEGADOCK-Web-Mito は本研究で開発した，ミトコンドリアに関連するタンパク質間相互作用の予測結果を収録したデータベースである．MEGADOCK-Web version 1.0 のフレームワークを元に，新規のデータベースとして開発を行った．

3.1 MEGADOCK-Web-Mito のタンパク質のデータリソース

MEGADOCK-Web-Mito で対象としたタンパク質は，MitoProteome に登録された全タンパク質のうち，UniProt を介して Protein Data Bank (PDB) に問い合わせを行い，一定の基準を満たしたタンパク質とした．以下に各データベースの詳細を述べる．

MitoProteome [12, 13]

本稿ではミトコンドリアに関連するタンパク質を MitoProteome に収録されているタンパク質とし，ミトコンドリアに関連するタンパク質を MitoProteome に収録されている遺伝子とする．MitoProteome において，ミトコンドリアの各遺伝子には “MT000001” のような “MT” と 6 桁の数字で表される `Mito ID` が付与されており，この `Mito ID` とタンパク質配列の ID である `UniProt AC` との対応関係が提供されている．

Universal Protein Resource (UniProt) [5]

Universal Protein Resource (UniProt) はタンパク質配列のデータと遺伝子などの機能に関わるアノテーション情報を収録したデータベースである．UniProt ではタンパク質の配列それぞれに UniProt Accession Number (`UniProt AC`) が割り当てられている．また，REST 形式の API を利用して UniProt 上のデータをテキストや XML などの形式で取得することができる．

Protein Data Bank (PDB) [10]

Protein Data Bank (PDB) [10] は，タンパク質と核酸の立体構造のデータがまとめられたデータベースである．PDB が提供する構造情報には，それぞれに 4 文字の `PDB ID` と呼ばれる識別子が割り当てられていて，タンパク質などのポリペプチドが複数個合わさった複合体の状態での構造情報として提供される場合がある．そして，`PDB ID` が示す構造に含まれるポリペプチド鎖それぞれに対して `chain ID` が決められている．例えば，1A4I という構造情報には 2 つのチェイン（チェイン A とチェイン B）が含まれている．ここで，本稿ではこのような状態の構造を PDB 構造，`PDB ID` が示す構造に含まれるポリペプチド鎖単体をチェイン構造と呼ぶこととする．また，それらを総称してタンパク質構造とする．また，PDB が提供する PDB 構造に含まれる各チェイン構造には対応する `UniProt AC` が accession（以降，チェイン accession と呼ぶ）として付されている．

3.1.1 データ取得手順

MEGADOCK-Web-Mito に収録するタンパク質間相互作用予測結果は，タンパク質間相互作用予測にチェイン構造を用いる．タンパク質は配列が同一でも取り得る構造は

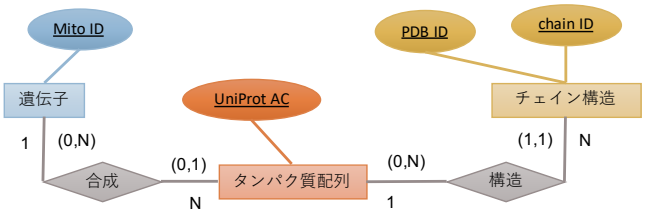


図 3　Mito ID，UniProt AC，チェイン構造の関係

複数存在することを考慮し，対象としたチェイン構造の取得のために，以下の 10 工程を行った．

(1) MitoProteome からミトコンドリアに関連するタンパク質の UniProt AC を取得する．具体的には，MitoProteome（2017 年 10 月時点）の各 Mito ID に対し，対応するミトコンドリアに関連するタンパク質の UniProt AC を以下の手順で取得した．
- 各 Mito ID を検索したときの検索結果として表示されるページのソースから，Python の正規表現ライブラリ re を用いて UniProt AC を検索する．
- 解析結果から，Mito ID と UniProt AC の 1 対多の対応関係を抽出する．

(2) UniProt では収録されているタンパク質配列を廃止することがあり [15]，本稿では廃止されたタンパク質配列に割り当てられた UniProt AC を obsolete UniProt AC と呼ぶ．(1) で得られた UniProt AC から obsolete UniProt AC を除外する．

(3) (2) で得られた各 UniProt AC に対し，対応する PDB ID を UniProt の API を用いて取得する．

(4) (3) で得られた各 PDB ID から，以下の 2 条件
  (a) 実験によって構造が決められた PDB 構造に割り当てられている PDB ID である．
  (b) 立体構造情報（PDB ファイル）が存在する．
  をどちらも満たすものを抽出する．

(5) (4) で得られた各 PDB ID に対応する PDB 構造をチェイン構造に分割し，チェイン accession が (1) で得られた UniProt AC のいずれかに一致するチェイン構造を，PDB ID と chain ID の 2 つ組（以下，⟨PDB ID, chain ID⟩ と表す）として得る．

(6) (5) で対応する ⟨PDB ID, chain ID⟩ をもたない UniProt AC に対応する Mito ID を除外する．

(7) (1) から (6) までで得られた
- Mito ID と UniProt AC との対応関係
- UniProt AC と ⟨PDB ID, chain ID⟩ との対応関係

から，Mito ID と ⟨PDB ID, chain ID⟩，すなわち Mito ID とチェイン構造との対応関係を得る（**図 3**）．

(8) (7) でいずれかの Mito ID と対応関係のあるチェイン構造が含まれる PDB 構造を，同一のタンパク質配列，つまり同一の UniProt AC の組み合わせによって構成される PDB 構造毎にグループ化する．

表 1　タンパク質間相互作用予測計算に用いるチェイン構造のタンパク質配列の種類，遺伝子の種類の数

遺伝子	タンパク質配列	チェイン構造
654	654	1,675

(9) (8) で得られた各グループから，以下の基準に従って 1 つ代表 PDB ID を選択する．
  (a) PDB 構造を構成する各チェインの残基数の合計が最大のものの集合を A_1 とする．$|A_1| = 1$ ならば A_1 の要素を代表 PDB ID とする．
  (b) $|A_1| \neq 1$ ならば A_1 の要素のうち X 線結晶構造解析によって構造が決定されたもの，かつ解像度の最小のものの集合を A_2 とする．$|A_2| = 1$ ならば A_2 の要素を代表 PDB ID とする．
  (c) $|A_2| > 1$ ならば A_2 の要素のうち適当に 1 つ選び代表 PDB ID とする．$|A_2| = 0$ ならば A_1 の要素のうち適当に 1 つ選び代表 PDB ID とする．

(10) (9) で選出された PDB 構造をそれぞれチェイン構造に分割し，得られたチェイン構造の中から (7) で得た Mito ID と対応関係のないチェイン構造を除外する．

以上の手順により，異なる 1,675 種類のチェイン構造が得られた．得られたチェイン構造 c は，以下の条件を満たす．
- チェイン構造 c はあるミトコンドリアに関連する遺伝子 m から翻訳されるタンパク質 u がとりうる構造である．
- チェイン構造 c を含む PDB 構造 p と UniProt AC が同一である PDB 構造 $p'(\neq p)$ が存在する場合，p の各チェイン構造の残基数の総和は p' の各チェイン構造の残基数の総和以上である．各チェイン構造の総和が p と p' で等しいときは，p' は X 線結晶構造解析以外のデータであるか，X 線結晶構造解析によるデータだが p より解像度が悪い．

3.2　チェイン構造・タンパク質の分類

3.1 節で得られた 1,675 種類のチェイン構造をタンパク質の種類，すなわちタンパク質配列（UniProt AC ベース）で分類すると 654 種類に分類される．タンパク質間相互作用予測計算に用いるチェイン構造のタンパク質配列の種類，遺伝子の種類の数を**表 1** に示す．

さらに，1,675 種類のチェイン構造を解像度で分類する．解像度は X 線結晶構造解析によるタンパク質の立体構造決定において，決定された立体構造の正確さを表す指標である．タンパク質間相互作用予測にタンパク質の立体構造を用いる MEGADOCK では，立体構造の正確さは相互作用予測の精度に関係する．そこで，解像度 2.5 Å を基準値として，チェイン構造を
- 解像度 2.5 Å 以下

表 2　1,675 種類の立体構造の構造決定手法による分類

構造決定手法		数
X 線結晶構造解析	解像度 $\leqq$ 2.5 Å	886
	解像度 $>$ 2.5 Å	458
X 線結晶構造解析以外の手法		331

表 3　654 種類のタンパク質の内訳

酵素・非酵素		数
酵素	酸化還元酵素	97
	転移酵素	116
	加水分解酵素	94
	付加脱離酵素	17
	異性化酵素	12
	合成酵素	28
非酵素		290

- 解像度 2.5 Å より大きい
- X 線結晶構造解析以外の方法で決定された

の 3 種に分類すると**表 2** のようになる．

さらにここで，得られた 654 種類のタンパク質を EC 番号（酵素番号）に基づいて分類した．酵素とは生体で生じる化学反応に対して触媒として機能する分子である．EC 番号は酵素を分類するもので，“EC”に続く 4 組の数字（例：EC4.2.1.3）で表される．4 組の数字の最初の数字は 1 から 6 まであり，それぞれ

- EC1.$x.y.z$：酸化還元酵素（酸化還元を触媒する酵素）
- EC2.$x.y.z$：転移酵素（原子団をある分子から別の分子に移動する酵素）
- EC3.$x.y.z$：加水分解酵素（加水分解を触媒する酵素）
- EC4.$x.y.z$：付加脱離酵素（二重結合生成や結合解離を触媒する酵素）
- EC5.$x.y.z$：異性化酵素（分子の異性体を生成する酵素）
- EC6.$x.y.z$：合成酵素（高エネルギー化合物の加水分解エネルギーにより 2 分子を結合させる酵素）

を表す．ここでは，654 種類のタンパク質を

(1) 酵素か酵素でないか
(2) 酵素に対して上記の 6 分類

で分類した（**表 3**）．

3.3　タンパク質間相互作用予測計算

3.1 節で得られた 1,675 種類のチェイン構造をタンパク質間相互作用予測計算に用いる．タンパク質間相互作用予測計算には東京工業大学が所有するスーパーコンピュータ TSUBAME 3.0 上で MEGADOCK を用い，$1{,}675 \times 1{,}675 = 2{,}805{,}625$ 組のチェイン構造ペアに対して計算を行った．TSUBAME 3.0 の f_node 1 ノード（Intel Xeon E5-2680 V4 Processor (14 cores, 2.4GHz) × 2 基，GPU (NVIDIA TESLA P100 for NVlink-Optimized Servers) × 4 基）を使用した場合，1 チェイン構造ペアあたり約 10 秒計算に要するため，得られたチェイン構造全 1,675 種類を用いた網羅的なタンパク質間相互作用予測計算にはおよそ $1{,}675 \times 1{,}675 \times 10[\mathrm{node \cdot s}] \approx 3.2 \times 10^2[\mathrm{node \cdot day}]$ の時間を要する．そこで，TSUBAME 3.0 のノードを複数同時に使用するスクリプトを作成し，用いることで計算時間 4.3 日で網羅的なタンパク質間相互作用予測計算を実現した．

3.4　MEGADOCK-Web-Mito 独自の機能

本研究で実装した，既存の MEGADOCK-Web とは異なる MEGADOCK-Web-Mito 独自の機能は以下の通りである．

(1) タンパク質への
- MitoProteome の各レコードの識別子である `Mito ID`
- Entrez Gene の各レコードの識別子である GeneID（他のデータベースの遺伝子の ID と区別するため `Entrez Gene ID` と呼ぶ．）

の対応付け

(2) 個々のタンパク質への KEGG パスウェイの対応付け

3.4.1　タンパク質への `Mito ID`・`Entrez Gene ID` の対応付け

MEGADOCK-Web にはタンパク質，PPI 予測結果，KEGG パスウェイの情報が収録されているが，MEGADOCK-Web-Mito ではこれらのタンパク質の情報に

- `Mito ID`
- `Entrez Gene ID`

を追加し，それぞれ MitoProteome，Entrez Gene のエントリへのリンクを付加することで，タンパク質の詳細な情報へのアクセスを容易にするとともに，`Mito ID` や `Entrez Gene ID` による検索機能も追加した．

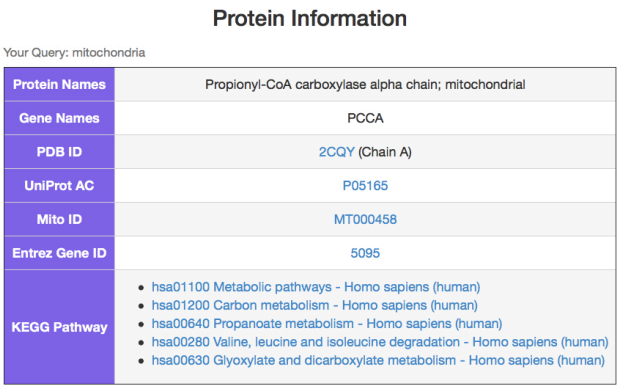
MEGADOCK-Web には検索したタンパク質と相互作用すると予測されたタンパク質が含まれる KEGG パスウェイの一覧を表示する機能はあるが，検索したタンパク質自身が含まれる KEGG パスウェイの一覧を表示する機能は実装されていない．そこで，MEGADOCK-Web-Mito では検索したタンパク質自身が含まれる KEGG パスウェイの一覧を表示する機能を実装した（**図 4**）．各 KEGG パスウェイをクリックすることで KEGG パスウェイが可視化される（**図 5**）．可視化の際，検索したタンパク質が橙色に着色される．

4.　MEGADOCK-Web-Mito の構成と使用例

4.1　MEGADOCK-Web-Mito の構成

MEGADOCK-Web-Mito のページ構成を**図 6** に示す．

- **トップページ**：MEGADOCK-Web-Mito のトップページは，MEGADOCK-Web と同様，4 つの検索窓としてのテキストフィールドをもつが，それぞれのテキストフィールドの左側に配置されたプルダウンメニューに



Protein Information

Your Query: mitochondria

Protein Names	Propionyl-CoA carboxylase alpha chain; mitochondrial
Gene Names	PCCA
PDB ID	2CQY (Chain A)
UniProt AC	P05165
Mito ID	MT000458
Entrez Gene ID	5095
KEGG Pathway	• hsa01100 Metabolic pathways - Homo sapiens (human) • hsa01200 Carbon metabolism - Homo sapiens (human) • hsa00640 Propanoate metabolism - Homo sapiens (human) • hsa00280 Valine, leucine and isoleucine degradation - Homo sapiens (human) • hsa00630 Glyoxylate and dicarboxylate metabolism - Homo sapiens (human)

図 4　MEGADOCK-Web-Mito の PPI スコア一覧表示ページのタンパク質情報のテーブル（例）の拡大図

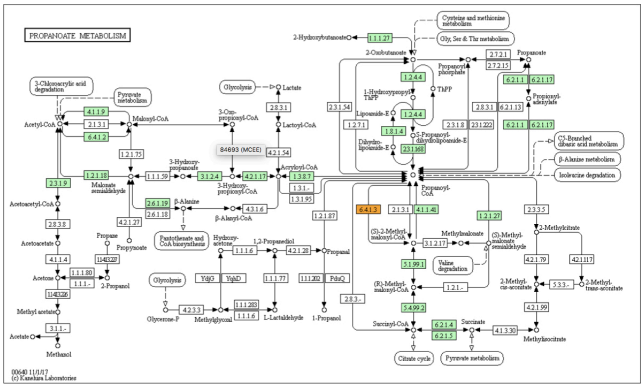


図 5　KEGG パスウェイ図可視化の例（プロパン酸代謝 hsa00640，検索したタンパク質が橙色に着色される）

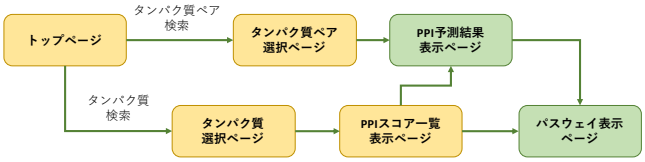


図 6　MEGADOCK-Web-Mito のページの遷移図（黄色で示されたページは MEGADOCK-Web に対して変更・拡張を行ったページである）

含まれるメニュー項目に Mito ID，Entrez Gene ID を追加し，Mito ID によるタンパク質の検索，Entrez Gene ID によるタンパク質の検索を可能にした（図 7）．

- **タンパク質選択ページ・PPI スコア一覧表示ページ・タンパク質ペア選択ページ**：MEGADOCK-Web-Mito のタンパク質選択ページでは，検索にヒットしたタンパク質一覧のテーブルに “Mito ID” 属性と “Entrez Gene ID” 属性の列を追加した．さらに Mito ID，Entrez Gene ID をそれぞれ MitoProteome，Entrez Gene へリンク付けすることによりタンパク質の配列情報をもつ遺伝子に関する情報へ容易なアクセスを可能にした（図 8）．
- **PPI スコア一覧表示ページ**：MEGADOCK-Web-Mito の PPI スコア一覧表示ページでは，検索対象のタンパク

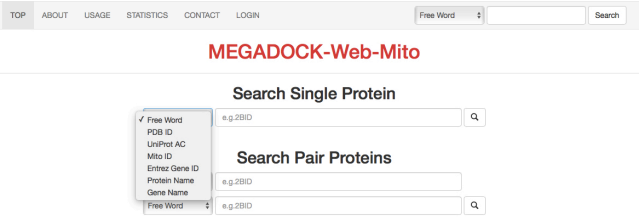
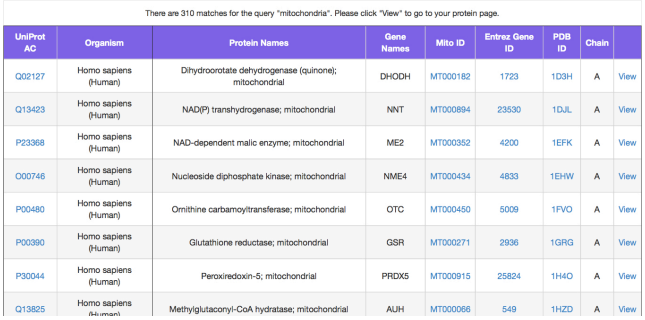


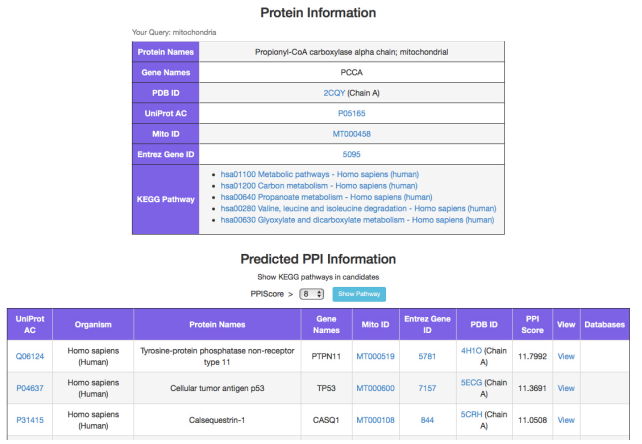
図 7　MEGADOCK-Web-Mito のトップページのプルダウンメニュー



There are 310 matches for the query "mitochondria". Please click "View" to go to your protein page.

UniProt AC	Organism	Protein Names	Gene Names	Mito ID	Entrez Gene ID	PDB ID	Chain	
Q02127	Homo sapiens (Human)	Dihydroorotate dehydrogenase (quinone); mitochondrial	DHODH	MT000182	1723	1D3H	A	View
Q13423	Homo sapiens (Human)	NAD(P) transhydrogenase; mitochondrial	NNT	MT000894	23530	1DJL	A	View
P23368	Homo sapiens (Human)	NAD-dependent malic enzyme; mitochondrial	ME2	MT000352	4200	1EFK	A	View
O00746	Homo sapiens (Human)	Nucleoside diphosphate kinase; mitochondrial	NME4	MT000434	4833	1EHW	A	View
P00480	Homo sapiens (Human)	Ornithine carbamoyltransferase; mitochondrial	OTC	MT000450	5009	1FVO	A	View
P00390	Homo sapiens (Human)	Glutathione reductase; mitochondrial	GSR	MT000271	2936	1GRG	A	View
P30044	Homo sapiens (Human)	Peroxiredoxin-5; mitochondrial	PRDX5	MT000915	25824	1H4O	A	View
Q13825	Homo sapiens (Human)	Methylglutaconyl-CoA hydratase; mitochondrial	AUH	MT000066	549	1HZD	A	View

図 8　MEGADOCK-Web-Mito のタンパク質選択ページ（例）



Protein Information

Your Query: mitochondria

Protein Names	Propionyl-CoA carboxylase alpha chain; mitochondrial
Gene Names	PCCA
PDB ID	2CQY (Chain A)
UniProt AC	P05165
Mito ID	MT000458
Entrez Gene ID	5095
KEGG Pathway	• hsa01100 Metabolic pathways - Homo sapiens (human) • hsa01200 Carbon metabolism - Homo sapiens (human) • hsa00640 Propanoate metabolism - Homo sapiens (human) • hsa00280 Valine, leucine and isoleucine degradation - Homo sapiens (human) • hsa00630 Glyoxylate and dicarboxylate metabolism - Homo sapiens (human)

Predicted PPI Information

Show KEGG pathways in candidates

PPIScore > 8 Show Pathway

UniProt AC	Organism	Protein Names	Gene Names	Mito ID	Entrez Gene ID	PDB ID	PPI Score	View	Databases
Q06124	Homo sapiens (Human)	Tyrosine-protein phosphatase non-receptor type 11	PTPN11	MT000519	5781	4H1O (Chain A)	11.7992	View	
P04637	Homo sapiens (Human)	Cellular tumor antigen p53	TP53	MT000600	7157	5ECG (Chain A)	11.3691	View	
P31415	Homo sapiens (Human)	Calsequestrin-1	CASQ1	MT000108	844	5CRH (Chain A)	11.0508	View	

図 9　MEGADOCK-Web-Mito の PPI スコア一覧表示ページ（例）

質の情報のテーブルに “Mito ID” 属性 “Entrez Gene ID” 属性の行を追加し，それぞれ MitoProteome，Entrez Gene へのリンク付けを行った．さらに，検索対象のタンパク質が含まれる KEGG パスウェイの一覧を表示する行も追加し，個々のタンパク質に対する KEGG パスウェイの可視化も可能にした．また，PPI スコア一覧のテーブルにも “Mito ID” 属性と “Entrez Gene ID” 属性の列を追加し，それぞれ MitoProteome，Entrez Gene へのリンク付けを行った（図 9）．

- **タンパク質ペア選択ページ**：MEGADOCK-Web-Mito のタンパク質ペア選択ページでは，タンパク質の情報を表示する 2 つのテーブルに対して “Mito ID” 属性 “Entrez Gene ID” 属性の行を追加し，それぞれ MitoProteome，Entrez Gene へのリンク付けを行った（図 10）．

© 2018 Information Processing Society of Japan

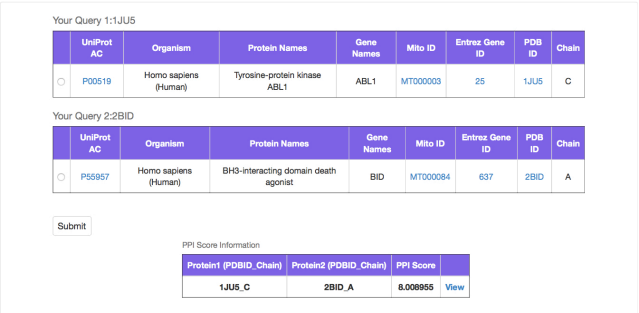


図 10　MEGADOCK-Web-Mito のタンパク質ペア表示ページ（例）

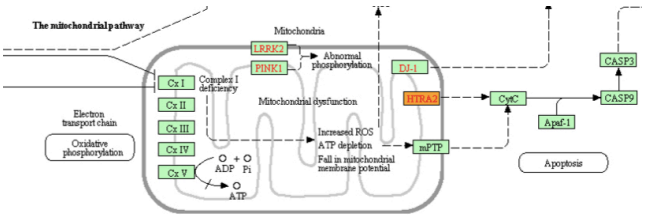


図 11　パーキンソン病の KEGG パスウェイ

4.2　MEGADOCK-Web-Mito の具体的な使用例

MEGADOCK-Web-Mito の具体的な使用例について説明する．MitoProteome において `Mito ID` が MT000972 の遺伝子はパーキンソン病に関連のある遺伝子である．以下では `Mito ID` が MT000972 の遺伝子から転写・翻訳を経て合成されるタンパク質に関するタンパク質間相互作用予測情報を検索する使用例を説明する．

(1) MEGADOCK-Web-Mito で `Mito ID` が “MT000972” である遺伝子から合成されるタンパク質を検索すると，以下のことが分かる．
- `Mito ID` が “MT000850” であり，パーキンソン病に関連している．
- `PDB ID` が “5FHT”，`chain ID` が “A” のチェイン構造をとる．
- パーキンソン病の KEGG パスウェイ図において “HTRA2” の上方に “DJ-1” で表されるタンパク質が存在する（図 **11**）．

(2) MEGADOCK-Web-Mito で遺伝子名が “DJ-1” である遺伝子から合成されるタンパク質を検索すると，以下のことが分かる．
- 遺伝子名が “HTRA2” である．
- `PDB ID` が “3BWE”，`chain ID` が “A” のチェイン構造をとる．

(3) `PDB ID` が “5FHT”，`chain ID` が “A” のチェイン構造と，`PDB ID` が “3BWE”，`chain ID` が “A” のチェイン構造とのタンパク質間相互作用予測結果が MEGADOCK-Web-Mito で閲覧できる（図 **12**）．

5.　まとめ

本研究では既存の MEGADOCK-Web 1.0 のフレームワークをベースにミトコンドリアに関連するタンパク質間相互作用予測データベースである，MEGADOCK-Web-Mito を開発した．

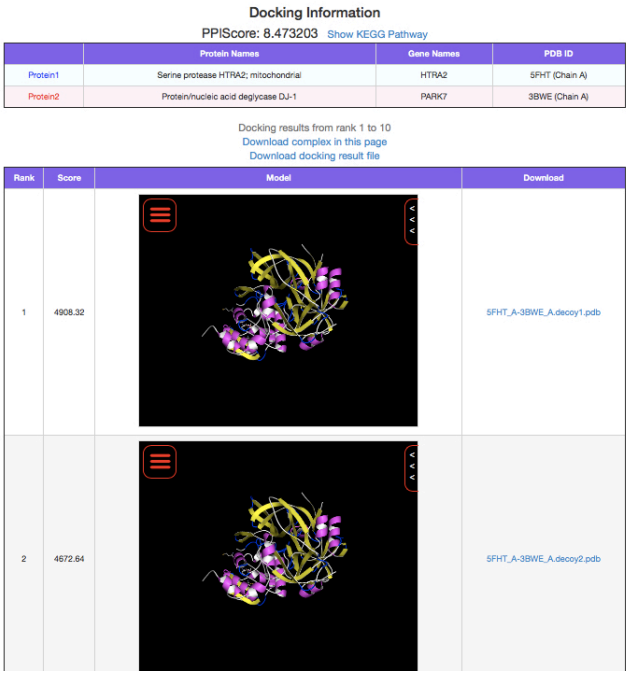


図 12　5FHT のチェイン A と 3BWE のチェイン A の予測複合体一覧

今後の課題として以下の 3 点が挙げられる．

(1) 本研究では，MEGADOCK-Web-Mito に収録するタンパク質の情報を手動の方法により取得したが，今後タンパク質の情報取得手順を自動化する必要がある．

(2) 本研究では，MEGADOCK-Web-Mito をローカル環境に試験的に構築しているが，今後 MEGADOCK-Web-Mito をサーバ環境上に構築し，公開する必要がある．

(3) MEGADOCK-Web-Mito に収録されているタンパク質はいずれもヒトのミトコンドリアに局在するタンパク質であるが，タンパク質間相互作用予測結果のさらなる充実のため，ヒトのタンパク質と立体構造が類似するマウスやラットなどのミトコンドリアに局在するタンパク質もタンパク質間相互作用予測の対象とする必要がある．

謝辞　本研究の一部は，JSPS 科研費（15K16081, 17H01814, 18K18149），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グデータの基盤技術」(JPMJCR1303)，JST リサーチコンプレックス推進プログラム，文部科学省 地域イノベーション・エコシステム形成プログラム，AMED BINDS (JP17am0101112) の支援を受けて行われた．

© 2018 Information Processing Society of Japan

参考文献

[1] A. H. V. Schapira. Mitochondrial diseases. *The Lancet*, Vol. 379, No. 9828, pp. 1825–1834, 2012.

[2] J. de Las Rivas, C. Fontanillo. Protein-protein interactions essentials: Key concepts to building and analyzing interactome networks. *PLoS Computational Biology*, Vol. 6, No. 6, Article No. e1000807, 2010.

[3] T. Oltersdorf *et al.* An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*, Vol. 435, No. 7042, pp. 677–681, 2005.

[4] G. M. Popowicz *et al.* Structures of low molecular weight inhibitors bound to MDMX and MDM2 reveal new approaches for p53-MDMX/MDM2 antagonist drug discovery. *Cell Cycle*, Vol. 9, No. 6, pp. 1104–1111, 2010.

[5] A. Bateman *et al.* UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, Vol. 45, No. D1, pp. D158–D169, 2017.

[6] R. Chen, L. Li, Z. Weng. ZDOCK: An initial-stage protein-docking algorithm. *Proteins*, Vol. 52, No. 1, pp. 80–87, 2003.

[7] Y. Murakami, K. Mizuguchi. Homology-based prediction of interactions between proteins using Averaged One-Dependence Estimators. *BMC Bioinformatics*, Vol. 15, No. 1, Article No. 213, 2014.

[8] H. Zhou, E. Jakobsson. Predicting protein-protein interaction by the mirrortree method: Possibilities and limitations. *PLoS ONE*, Vol. 8, No. 12, Article No. e81100, 2013.

[9] M. Ohue *et al.* MEGADOCK 4.0: An ultra-high-performance protein-protein docking software for heterogeneous supercomputers. *Bioinformatics*, Vol. 30, No. 22, pp. 3281–3283, 2014.

[10] H. M. Berman *et al.* The protein data bank. *Nucleic Acids Research*, Vol. 28, No. 1, pp. 235–242, 2000.

[11] T. Hayashi *et al.* MEGADOCK-Web : an integrated database of high-throughput structure-based protein-protein interaction predictions. Vol. 19, No. Suppl 4, Article No. 62, 2018.

[12] D. Cotter *et al.* MitoProteome: mitochondrial protein sequence database and annotation system. *Nucleic Acids Research*, Vol. 32, pp. D463–D467, 2004.

[13] P. Guda, S. Subramaniam, C. Guda. Mitoproteome: human heart mitochondrial protein sequence database. *Methods of Molecular Biology*, Vol. 357, pp. 375–383, 2007.

[14] M. Kanehisa *et al.* KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, Vol. 45, No. D1, pp. D353–D361, 2017.

[15] Reducing proteome redundancy. `http://www.uniprot.org/help/proteome_redundancy`.

© 2018 Information Processing Society of Japan