

# MEGADOCKを用いたタンパク質間相互作用予測の ヒトアポトーシスパスウェイ解析への応用

大上 雅史<sup>1,2</sup> 松崎 由理<sup>1</sup> 石田 貴士<sup>1</sup> 秋山 泰<sup>1</sup>

**概要:** 我々が開発している MEGADOCK は、立体構造情報からタンパク質間相互作用 (Protein-Protein Interaction, PPI) を予測するソフトウェアであり、これまで細菌走化性シグナル伝達パスウェイでその性能評価が行われてきたが、その他のパスウェイでの評価はまだ行われていなかった。別パスウェイでの性能評価のために、本研究では MEGADOCK をヒトアポトーシスパスウェイ中のタンパク質群に適用し、PPI 予測を行った。その結果、構造情報の得られた 57 タンパク質から 452 の PPI が予測され、そのうち 88 の PPI が既知 PPI データベース情報として存在することを確認した。既知複合体構造をテンプレート情報として用いる PRISM の予測性能と比較すると、MEGADOCK は PRISM に比べてわずかに低い結果となった。予測結果の内容については、MEGADOCK は PRISM と大きく異なっていることが分かったので、双方のコンセンサスを取り、より信頼性の高い予測を行うことを試みた。その結果、予測の信頼度である Precision 値を改善させることに成功した。コンセンサスによる結果から未知 PPI として予測されたものの中には、実際に相互作用する可能性が高いと考えられるタンパク質ペアがいくつか含まれていることが分かった。

**キーワード:** タンパク質相互作用ネットワーク, アポトーシス, タンパク質立体構造, MEGADOCK .

## An Application of Protein-Protein Interaction Prediction to Human Apoptosis Signal Transduction Pathway Analysis by Using MEGADOCK

MASAHITO OHUE<sup>1,2</sup> YURI MATSUZAKI<sup>1</sup> TAKASHI ISHIDA<sup>1</sup> YUTAKA AKIYAMA<sup>1</sup>

**Abstract:** MEGADOCK is protein-protein interaction (PPI) prediction software developed by our group, and has been only applied to the bacterial chemotaxis signaling pathway and has not been applied to other pathways yet. In this study, MEGADOCK was applied to the human apoptosis pathway for validating its performance on another pathway. As results, 452 PPIs were predicted from 57 proteins taken from the Protein Data Bank. 88 PPIs of them existed in PPI database. As compared with the predicted performance of PRISM which is another PPI prediction system based on known complex structure as template information, MEGADOCK brought a slightly low performance. Moreover, we found that the prediction results of MEGADOCK were widely different from PRISM. So we tried to predict PPI with high reliability using consensus of MEGADOCK and PRISM. As results, we succeeded to improve Precision value corresponding to prediction reliability. Also, the consensus prediction results included several new PPI candidates.

**Keywords:** Protein-Protein Interaction Network, Apoptosis, Protein Tertiary Structure, MEGADOCK.

<sup>1</sup> 東京工業大学 大学院情報理工学専攻 計算工学専攻  
Graduate School of Information Science and Engineering,  
Tokyo Institute of Technology

<sup>2</sup> 日本学術振興会 特別研究員  
Research Fellow of the Japan Society for the Promotion of  
Science

## 1. はじめに

### 1.1 タンパク質間相互作用予測

タンパク質間相互作用は生命現象において中心的な役割を果たしており、相互作用関係の網羅的な理解は、疾病の発症の原因解明、医薬品の開発ターゲットの同定に重要である [1]。現在では実験的手法と並んで計算機による予測手法が盛んに研究されており [2], [3], [4], [5], 中でもタンパク質の構造情報に着目した相互作用予測手法が注目されている [6], [7], [8]。

Matsuzaki らは、タンパク質ドッキング予測ソフトウェアである ZDOCK[9] が出力する複合体構造予測結果に対してクラスター解析を行って相互作用の有無を予測する手法を提案し [6], 細菌走化性のシグナル伝達パスウェイ解析に応用した。また、Ozbabacan らは表面構造の類似性を活用するタンパク質相互作用予測ソフトウェアである PRISM[7] を用い、ヒトのアポトーシスパスウェイへの適用を行っている [10]。このように、タンパク質の構造情報を利用した相互作用予測は近年様々に応用されてきており、予測結果がもたらす構造上の新たな知見は今後ますます蓄積されていくものと考えられる。

我々が開発を続けているタンパク質間相互作用予測システム MEGADOCK[11] も構造情報を利用した予測を用いている。詳細は後の章に記述するが、Matsuzaki らと同様のデータセットを用いて細菌走化性シグナル伝達パスウェイに適用している [11] 他、タンパク質ドッキング計算の枠組みが類似している ZDOCK を用いた結果との比較を行ってきた [12]。ZDOCK との比較においては、同様の構造探索を行っていても構造の評価関数が異なると相互作用の予測結果が異なることが示され、複数のツールによる相互作用予測を行うことが新規相互作用の発見には有用であることが分かってきた [12]。また、PRISM はテンプレートとして利用される複合体の相互作用面の構造情報の質が相互作用予測の結果に強く影響するため、テンプレート構造を必要としない MEGADOCK などの手法による予測結果と組み合わせることは、新規相互作用の発見において重要であると考えられる。

### 1.2 アポトーシスパスウェイの概略

アポトーシスは主に動物で起こるプログラム細胞死の一種であり、個体の状態を良好に保つための積極的に引き起こされる制御された細胞自殺機構である。パスウェイ図を図 1 に示す。アポトーシスの減少に起因する疾病として癌や自己免疫疾患があり、アポトーシスの増加に起因するものとしては AIDS や神経変性疾患が挙げられ、これらの疾病に関わる創薬ターゲットとしてアポトーシスパスウェイ中のタンパク質が注目されている [13]。

アポトーシスパスウェイは、細胞外のシグナルタンパク

質が細胞表面の細胞死受容体に結合して起動する外部経路と、DNA 損傷や種々の細胞ストレスを受けて細胞内部から活性化される内部経路が存在する。外部経路では Fas 細胞死受容体に Fas リガンドが結合すると、FADD タンパク質が結合し、それに開始カスパーゼタンパク質 (Casp8) が結合し活性化する。この開始カスパーゼタンパク質が活性化すると、下流のアポトーシス実行カスパーゼタンパク質 (Casp3, Casp6, Casp7) が連鎖的に活性化され、これらの実行カスパーゼタンパク質によって細胞内の主要なタンパク質が切断され、細胞を死に至らす。

一方内部経路では、ミトコンドリアから放出されるシトクロム *c* などのプロアポトーシス分子によって起動され、Apaf1 タンパク質に結合し多量体を形成する (アポトソーム)。Apaf1 は開始カスパーゼ (Casp9) と結合し、これが実行カスパーゼに結合してアポトーシスが実行される。

このように、外部経路と内部経路は関わるタンパク質が異なるが、経路内で活性化の促進・抑制を行うタンパク質が様々な存在し、一部は両方の経路に関わっているものもあるため、外部経路が内部経路を誘発してカスパーゼ連鎖を増幅させるものもある。その一例が BID タンパク質である。BID は外部経路において Casp8 に切断されるが、切断された BID がミトコンドリアに移行し、アポトーシスを抑制する働きを持つ BCL-2 タンパク質や BCL-XL タンパク質を阻害したり、シトクロム *c* などのタンパク質を放出させたりすることで内部経路を誘発する。

### 1.3 研究目的

本研究は、これまで細菌走化性シグナル伝達系でのみ性能評価が行われていた MEGADOCK を、新たにヒトアポトーシスパスウェイに適用し、評価を行うことを目的とする。また、対象のヒトアポトーシスパスウェイは既に PRISM によって予測されていることを受け、双方の予測結果を比較し、また予測結果を組み合わせることで、より信頼性の高い PPI 予測を行うことを試みる。

## 2. 関連研究

Ozbabacan らの行った研究 [10] では、PRISM[7] を用いてヒトアポトーシスパスウェイへの適用を行っている。PRISM は、(1) ターゲットタンパク質とテンプレートセットの相互作用面構造を剛体比較し、(2)FiberDock[15] によるエネルギースコアを用いた構造リファインメントを行い、(3) 結合エネルギーの計算と予測された複合体構造を出力する、というものである。

テンプレートセットには Protein Data Bank (PDB)[16] から構築された相互作用面のデータベース [17] から、ホモロジーなどの冗長性を省いて 1,146 個の相互作用面構造を取得し用いている。ターゲットタンパク質は KEGG[18] の human apoptosis pathway (KEGG pathway ID: *hsa04210*)

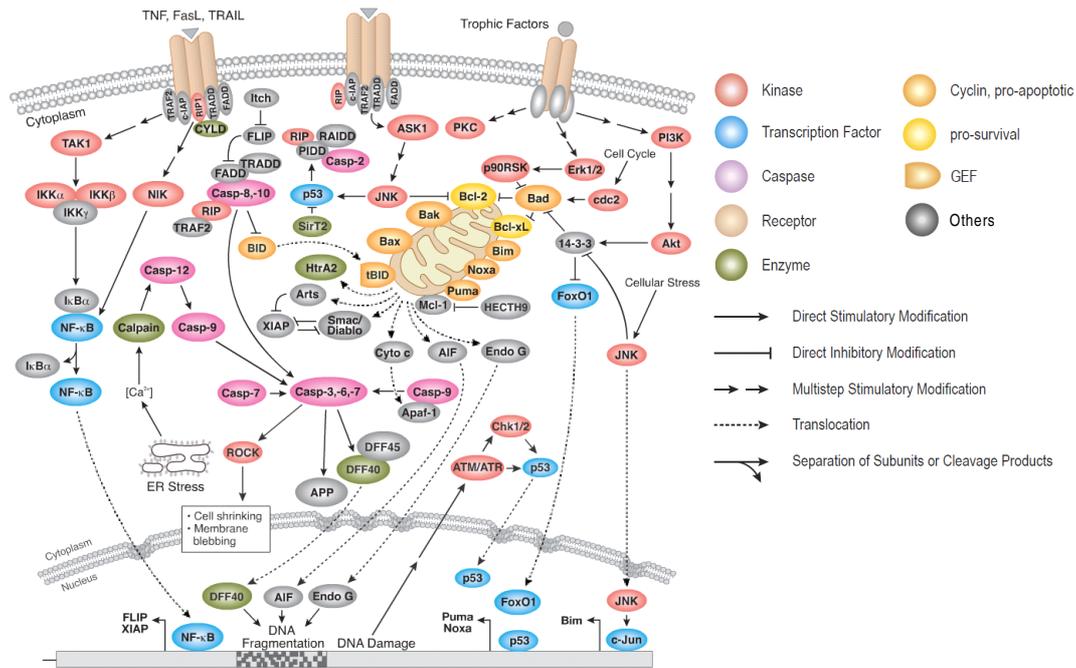


図 1 アポトーシスパスウェイの概要 ([14] より引用).  
Fig. 1 The overview of apoptosis pathway[14].

から 41 個の KEGG Protein ID を取得し、対応する 57 個の Protein Name を用いて PDB 構造を抽出、90%以上の配列類似度を持つものを除外し、124 個の PDB 構造をターゲットタンパク質とした。

これらのデータセットを用いて PRISM による予測を行った結果は、KEGG パスウェイ情報と、PPI データベースである STRING データベース [19] (以下 STRING DB と呼ぶ) の情報\*1 を利用して検証されており、KEGG および STRING に含まれる相互作用を正例とし、それ以外のタンパク質の組み合わせは負例として、True Positive(TP), False Positive(FP), False Negative(FN), True Negative(TN) が求められている。これらの値の定義は表 1 の通りである。結果として、TP = 56, FP = 186, FN = 81, TN = 1,273, F-measure = 0.296 という結果を得た。ただし、KEGG pathway や STRING には自分自身との PPI の情報は含まれていないため、予測対象となるタンパク質ペアは  ${}_{57}C_2 = 1,596$  ペアである。

### 3. 方法

#### 3.1 MEGADOCK によるタンパク質間相互作用予測

MEGADOCK は我々が開発している大規模 PPI ネットワークの予測のためのソフトウェアである。予測方法は、(1)3-D グリッドモデルを利用したタンパク質剛体ドッキング計算 [20] を行い、得られた複数の複合体候補構造群から、(2) 候補構造群のドッキングスコア分布に基づく相互

表 1 混合行列と Precision, Recall, F-measure の定義。

Table 1 The definition of confusion matrix, Precision, Recall and F-measure.

	DB に存在する	DB に存在しない
PPI 有と予測	TP	FP
PPI 無と予測	FN	TN

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F-measure} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

作用判定を行う、というものである。

今回使用した MEGADOCK は、MEGADOCK Ver. 2.6.2[21] である。MEGADOCK 2.6.2 は、剛体ドッキング計算での複合体候補構造の評価関数に、形状相補性 (rPSC スコア [11]), 静電相互作用, 疎水性相互作用 (Atomic Contact Energy[22] による) を用いており、複素数を用いてこれら 3 つの評価関数を同時に高速に計算する工夫を行っている。

本研究では、計算する複合体候補構造を各タンパク質ペアに対して 10,800 個とした。候補構造それぞれにドッキングスコア  $S_i^{(p_1, p_2)}$  ( $i = 1, \dots, 10,800$ ) と ZRANK エネルギースコア [23] が付与されている。タンパク質ペア ( $p_1, p_2$ ) に対し、ZRANK エネルギースコアの最も良い候補構造のドッキングスコア  $S_{\text{top}}^{(p_1, p_2)}$  を用いて、

\*1 STRING の PPI は、パラメータである confidence score を 0.5 としたものを利用している。

$$E^{(p_1, p_2)} = \frac{S_{\text{top}}^{(p_1, p_2)} - \mu^{(p_1, p_2)}}{\sigma^{(p_1, p_2)}}$$

$$\text{PPI}^{(p_1, p_2)} = \begin{cases} \text{True} & E^{(p_1, p_2)} > E^* \\ \text{False} & \text{otherwise} \end{cases}$$

として PPI の有無を予測する．ここで  $\mu^{(p_1, p_2)}$ ,  $\sigma^{(p_1, p_2)}$  は  $S_i^{(p_1, p_2)}$  ( $i = 1, \dots, 10,800$ ) の平均, 標準偏差である． $E^*$  の値はタンパク質ペア  $(p_1, p_2)$  によらず一定の値を用いる．本研究では 0.1 刻みで  $E^*$  を振り, 既知の PPI に対する F-measure の値が最も良くなる値を用いることとした．

### 3.2 データセット

データセットには表 2 に示す Ozbabacan らが用いた 124 個の PDB 構造を用いた．ただし, 複数の chain を持つ PDB 構造については, 同一 chain を省き, Ozbabacan らの用いた Protein Name と対応しない chain は省いた．結果として得られた 158 個の構造の全組み合わせ ( $158 \times 158 = 24,964$  通り) を MEGADOCK の入力として用いた．得られた PDB chain と Protein Name の対応を表 3 に示す．

### 3.3 評価方法

個別のタンパク質ペアごとの MEGADOCK の PPI 予測結果に対し, STRING DB を用いて予測結果の評価を行った．STRING と KEGG の相互作用情報からは, 全  ${}_{57}C_2 = 1,596$  ペアに対し, 正例として 137 ペアが得られている．また, 自分自身との相互作用や, STRING に含まれていない相互作用の可能性を加味し, シンガポール科学技術研究庁 (A\*STAR) の Dr. Vachirane Limviphuvadh より提供頂いた PPI データベース (以下 LIM DB と記述) も併用した．LIM DB は, 既存文献を元にした PPI データベースである BIND[24], BioGRID[25], DIP[26], HPRD[27], IntAct[28], MINT[29], MPact[30], MPPI[31] の結果を統合し, 検出された PPI の中からハイスループットな PPI 検出実験の結果を取り除いたものである．

各タンパク質ペアについて, STRING DB と LIM DB のどちらかに正例として登録されているものは正例, 両方ともに登録されていないものを負例とした (図 2)．これにより, 対象となるタンパク質ペアは  ${}_{57}C_2 + 57 = 1,653$  ペア, 正例は 187 ペアとなった．また, 比較のため STRING DB の登録情報のみを用いた場合の評価も行った．

今回用いたデータセットでは表 3 に示したように 1 つのタンパク質に対して対応する構造が複数ある場合がある．そこで, 複数の構造を持つタンパク質ペアについては, Matsuzaki ら [6] の評価方法と同一方法をとった．これは, 1 つでも PPI 有と予測された構造ペアが存在すればそのタンパク質ペアは PPI 有と予測する, という方法である．

また, MEGADOCK の計算するドッキングスコア  $S$  はタンパク質の入力順に対して非対称の形をしているため,

表 3 使用した KEGG pathway から得られた PDB chain 構造 (158 PDB chain)．‘\_’ 以前の 4 文字コードは PDB ID を, ‘\_’ 以降の 1 文字コードは chain 名を表す．

Table 3 PDB chains of human apoptosis pathway protein from *hsa04210* KEGG pathway (158 chains).

Name	PDB ID_chain
AIF	1M6L_A
AKT1	1UNQ_A, 3CQW_A, 3O96_A
AKT2	1MRV_A, 1O6K_A, 1O6L_A, 1P6S_A
AKT3	2X18_A
APAF1	1CY5_A, 1Z6T_A, 2YGS_A, 3IZA_A, 3YGS_C
Bax	1F16_A, 2G5B_L, 2XA0_C, 3PK1_B
BCL-2	2W3L_A, 2XA0_A
BCL-XL	2B48_A, 3FDL_A
BID	2BID_A, 2KBW_B
Calpain1	1ZCM_A
Calpain2	1KFU_L, 2NQA_A
CASP3	1RHQ_A, 1RHQ_B, 2DKO_A, 2DKO_B, 2J32_A
CASP6	2WDP_A
CASP7	1F1J_A, 1I4O_A, 1I51_A, 1I51_B, 2QL9_A, 2QL9_B
CASP8	1QTN_A, 1QTN_B, 2FUN_B, 3H11_B
CASP9	1JXQ_A, 1NW9_B, 3D9T_C, 3YGS_P
Cn(CHP)	2E30_A
Cn(CHP2)	2BEC_A
Cn(PPP3CA)	1AULA, 1MF8_A, 2R28_C, 3LL8_A
Cn(PPP3R1)	1AULB, 1MF8_B, 3LL8_B
CytC	1J3S_A
DFF40	1IBX_A
DFF45	1IBX_B, 1IYR_A
FADD	1A1W_A, 2GF5_A, 3EZQ_B
Fas	3EWT_E, 3EZQ_A
FLIP	3H11_A
IAP(BIRC2)	3D9T_A, 3M1D_A, 3MUP_A
IAP(BIRC3)	2UVL_A, 3EB5_A, 3EB6_A, 3M0A_D, 3M0D_D
IAP(BIRC4)	1G73_C, 1I4O_C, 1I51_E, 1NW9_A, 2ECG_A, 2KNA_A, 2POLA, 3CM7_C
IκBα	1IKN_D, 1NFL_E
IKK	2JVX_A, 3BRT_B, 3BRV_B, 3CL3_D, 3FX0_A
IL-1(A)	2ILA_A
IL-1(B)	1ITB_A, 2NVH_A, 3O4O_A
IL-1R(1)	1ITB_B
IL-1R(RAP)	3O4O_B
IL-3	1JLI_A
IL-3R	1EGJ_A
IRAK2	3MOP_K
IRAK4	2NRU_A, 3MOP_G
MyD88	2J57_A, 3MOP_A
NF-κB(NFKB1)	1IKN_C, 1NFL_B, 1SVC_P, 2DBF_A
NF-κB(RELA)	1IKN_A, 1NFL_A
NGF	1WWW_V, 2IFG_E
PI3K(PIK3CA)	2ENQ_A, 2V1Y_A, 3HHM_A
PI3K(PIK3CG)	1E8Y_A
PI3K(PIK3R1)	1A0N_A, 1H9O_A, 1PBW_A, 2IUG_A, 2V1Y_B, 3HHM_B, 3I5R_A
PI3K(PIK3R2)	2KT1_A, 2XS6_A, 3MTT_A
PRKACA	3AGM_A
PRKAR2A	2IZX_A
TNFα	1A8M_A, 4TSV_A
TNF-R1	1EXT_A, 1ICH_A
TP53	1AIE_A, 1OLG_A, 1XQH_B, 1YC5_B, 2B3G_B, 2FOO_B, 2GS0_B, 2K8F_B, 2VUK_A, 3D06_A, 3DAB_B, 3LW1_P, 1F3V_A
TRADD	1CZZ_A, 1D00_A, 1F3V_B, 3KNV_A, 3M0A_A, 3M0D_A
TRAF2	1D4V_B, 1DG6_A, 1DU3_D
TRAIL	1D4V_A, 1DU3_A
TRAIL-R	1HE7_A, 1SHC_B, 1WWW_X, 2IFG_A
TrkA	

一般に  $\text{PPI}^{(p_1, p_2)}$  と  $\text{PPI}^{(p_2, p_1)}$  は等しいとは限らない．本研究では, この場合もどちらかで PPI 有と判定されれば, そのタンパク質ペアは PPI 有と予測した．

## 4. 結果と考察

### 4.1 PPI データベースによる予測結果の検証

MEGADOCK によるアポトーシスパスウェイの予測結果を表 4 に示す．MEGADOCK の予測精度は, STRING DB で評価すると  $F\text{-measure} = 0.220$  となり, PRISM の  $F\text{-measure} = 0.296$  に少し劣る結果となった．しかし, STRING DB と LIM DB 両方に基づく評価では,  $F\text{-measure} = 0.277$  となり, 予測対象の数は異なるものの, PRISM に近い予測精度が得られた．

なお, STRING DB での評価において, TP の数を PRISM の 56 例に揃えた場合の結果を表 4 の “MEGADOCK (TP = 56)” の行に示した．このとき, MEGADOCK による予測結果は PRISM に比べて FP の数が約 1.8 倍多く

表 2 KEGG pathway (KEGG pathway ID: *hsa04210*) から得られた PDB 構造 (124 個) .

Table 2 PDB IDs of human apoptosis pathway protein from *hsa04210* KEGG pathway (124).

1A0N, 1A1W, 1A8M, 1AIE, 1AUI, 1CY5, 1CZZ, 1D00, 1D4V, 1DG6, 1DU3, 1E8Y, 1EGJ, 1EXT, 1F16, 1F1J, 1F3V, 1G73, 1H9O, 1HE7, 1I4O, 1I51, 1IBX, 1ICH, 1IKN, 1ITB, 1IYR, 1J3S, 1JLI, 1JXQ, 1KFU, 1M6I, 1MF8, 1MRV, 1NFI, 1NW9, 1O6K, 1O6L, 1OLG, 1P6S, 1PBW, 1QTN, 1RHQ, 1SHC, 1SVC, 1UNQ, 1WWW, 1XQH, 1YC5, 1Z6T, 1ZCM, 2B3G, 2B48, 2BEC, 2BID, 2DBF, 2DKO, 2E30, 2ECG, 2ENQ, 2FOO, 2FUN, 2G5B, 2GF5, 2GS0, 2IFG, 2ILA, 2IUG, 2IZX, 2J32, 2JS7, 2JVX, 2K8F, 2KBW, 2KNA, 2KT1, 2NQA, 2NRU, 2NVH, 2POI, 2QL9, 2R28, 2UVL, 2V1Y, 2VUK, 2W3L, 2WDP, 2X18, 2XA0, 2XS6, 2YGS, 3AGM, 3BRT, 3BRV, 3CL3, 3CM7, 3CQW, 3D06, 3D9T, 3DAB, 3EB5, 3EB6, 3EWT, 3EZQ, 3FDL, 3FX0, 3H11, 3HHM, 3I5R, 3IZA, 3KNV, 3LL8, 3LW1, 3M0A, 3M0D, 3M1D, 3MOP, 3MTT, 3MUP, 3O4O, 3O96, 3PK1, 3YGS, 4TSV
--

表 4 アポトーシスパスウェイの予測結果 . “PRISM” の行は文献 [10] の結果である .

Table 4 The prediction results of human apoptosis pathway. The row of “PRISM” shows results of [10].

Method	Database	TP	FP	FN	TN	Precision	Recall	F-measure
MEGADOCK	STRING	62	365	75	1,094	0.145	0.453	0.220
MEGADOCK (TP = 56)	STRING	56	338	81	1,121	0.142	0.409	0.211
MEGADOCK	STRING+LIM	88	364	96	1,105	0.195	0.478	0.277
PRISM	STRING	56	186	81	1,273	0.231	0.409	0.296
コンセンサス (AND) 予測	STRING	30	63	107	1,396	0.322	0.219	0.261
OR 予測	STRING	88	488	49	971	0.153	0.642	0.247

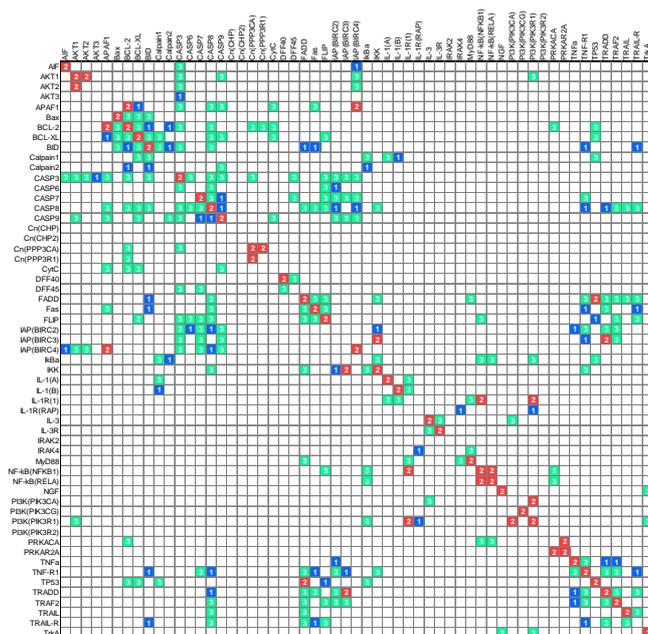


図 2 STRING DB と LIM DB の PPI . 色のついているセルが相互作用するタンパク質ペアを表す . ‘1’ のセル (青) は STRING DB, ‘2’ のセル (赤) は LIM DB, ‘3’ のセル (緑) は両方のデータベースに共通して含まれる PPI である .

Fig. 2 The PPIs from STRING DB and LIM DB. Colored cells show interacted protein pairs. ‘1’ (blue) cells are from STRING DB, ‘2’ (red) cells are from LIM DB and ‘3’ (green) cells are both from STRING DB and LIM DB.

なった . PRISM が用いている複合体結晶構造によるテンプレート構造情報の活用が , FP の数の減少に寄与しているものと考えられる .

#### 4.2 2手法のコンセンサスによる結果

MEGADOCK と PRISM の結果を比較すると , TP や FP の内容が異なっていることが分かった (図 3) . そこで , MEGADOCK と PRISM の予測結果に対し , 両方の手法で PPI 有と予測されたペアのみ相互作用するという予測をする手法 (コンセンサス (AND) 予測) と , どちらか一方の手法で PPI 有と予測されれば相互作用するという予測をする手法 (OR 予測) を適用し , 結果を調べた . 結果は表 4 の下部 2 行に示した .

コンセンサス予測では FP の数が PRISM に比べて 26 ペア減少したが , FP が 123 ペア減少し , 結果として Precision が 0.231 から 0.322 に増加した . 一方 OR 予測では Recall が PRISM や MEGADOCK に比べて大きく増加した . 対象とする問題にもよるが , 今回のような規模の大きなパスウェイからこれまでに知られていない興味深い相互作用の発見を試みるという目的には , 予測された相互作用がどれだけ信頼できるかが重要となる . Precision 値はこの予測された相互作用の信頼性を表すものであり , 本研究におけるコンセンサス予測による相互作用予測の結果は , MEGADOCK や PRISM を単体で用いたときの予測結果よりも信頼性が高いと言えます , 現在のデータベースでは False Positive と判定されたものの中にも , 実際には相互作用するものが存在する可能性がより高いと考えられる .

#### 4.3 コンセンサス予測結果のうち FP とされた相互作用について

先述のとおり , コンセンサス予測は高い Precision 値を示しており , FP と判定されたものの中に新たな相互作用

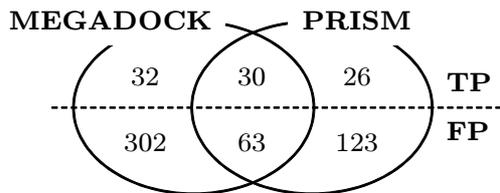


図 3 MEGADOCK と PRISM の予測結果のベン図

Fig. 3 Venn diagram of MEGADOCK and PRISM predictions.

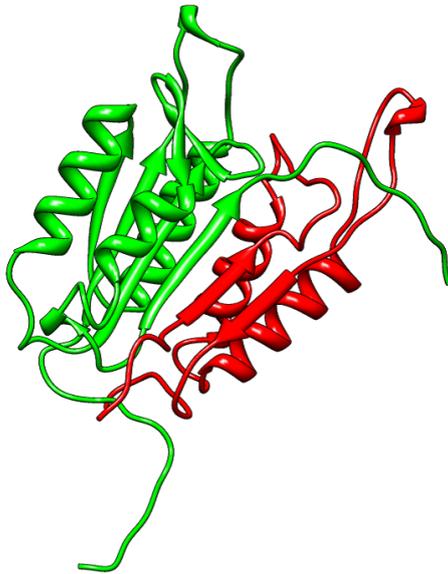


図 4 MEGADOCK による CASP3 と CASP7 の複合体予測結果 . 緑のタンパク質は CASP3 (PDB: 2DKO\_A), 赤のタンパク質は CASP7 (P10 subunit, PDB: 2QL9\_B) である .

Fig. 4 The predicted complex structure of CASP3 and CASP7 by MEGADOCK. Green colored protein is CASP3 (PDB: 2DKO\_A), red colored protein is CASP7 (P10 subunit, PDB: 2QL9\_B).

が予測できていることが考えられる . そこでコンセンサスによって予測された結果のうち , STRING DB に登録されていない ( False Positive とされた ) タンパク質ペアのいくつかについて考察を行う .

#### (a) CASP3, CASP7

CASP3 と CASP7 はどちらもアポトーシスを起こす実行カスペーゼタンパク質である . 開始カスペーゼによって活性化される実行カスペーゼではあるが , 自身も他の実行カスペーゼの活性化を行うと考えられており , 事実 , CASP3 と CASP6 はどちらも実行カスペーゼであるが相互作用をすることが知られている [32] . CASP3, CASP6, CASP7 は機能は異なるものの [33], [34], 相互に活性化を行う可能性は十分考えられる . 図 4 に MEGADOCK によって予測された CASP3-CASP7 複合体の予測構造を示す .

#### (b) Akt, Bax

Bax はミトコンドリアからのシトクロム *c* 放出を促進する刺激因子である . Akt は Bax を抑制すると考えられており [35], 細胞死への連鎖を抑制するため , 生存シグナ

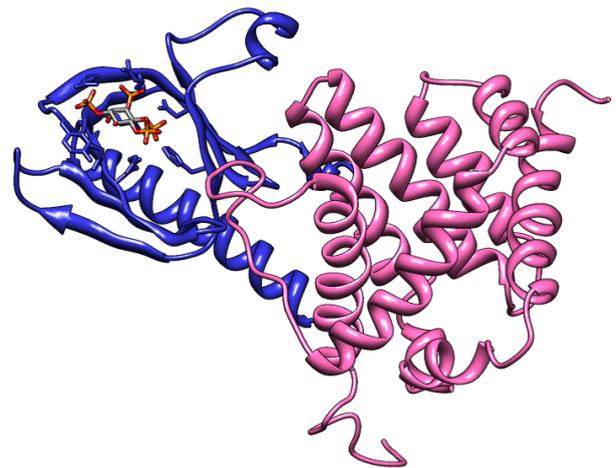


図 5 MEGADOCK による Akt1 と Bax の複合体予測結果 . 青色のタンパク質は Akt1 (PDB: 1UNQ\_A), ピンク色のタンパク質は Bax (PDB: 1F16\_A) である .

Fig. 5 The predicted complex structure of Akt1 and Bax by MEGADOCK. Blue colored protein is Akt1 (PDB: 1UNQ\_A), pink colored protein is Bax (PDB: 1F16\_A).

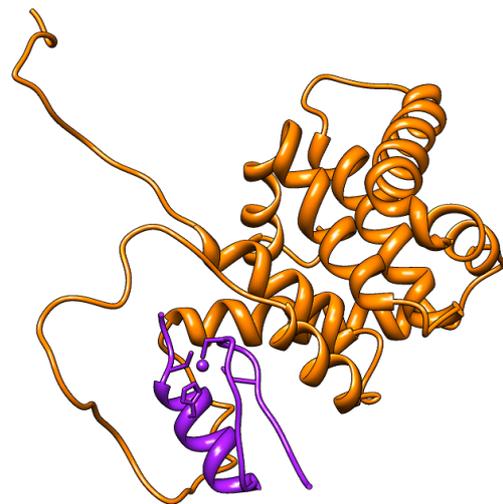


図 6 MEGADOCK による BID と IKK の複合体予測結果 . 橙色のタンパク質は BID (PDB: 2BID\_A), 紫色のタンパク質は IKK (PDB: 2JVX\_A) である .

Fig. 6 The predicted complex structure of BID and IKK by MEGADOCK. Orange colored protein is BID (PDB: 2BID\_A), purple colored protein is IKK (PDB: 2JVX\_A).

ル伝達因子と呼ばれている . 図 5 に MEGADOCK による Akt1-Bax 複合体の予測結果を示す . なお Akt1 は Akt のアイソフォームの 1 つである .

#### (c) BID, IKK

BID はアポトーシスの内部経路と外部経路のクロストークを担うタンパク質であるが , IKK (I $\kappa$ K キナーゼ) を含むタンパク質複合体と相互作用することが知られている [36] . 図 6 に MEGADOCK による BID-IKK 複合体の予測結果を示す .

## 5. おわりに

タンパク質間相互作用予測ソフトウェア MEGADOCK 2.6.2 によって、新たにヒトアポトーシスパスウェイの予測を行い、予測性能の評価を行った。MEGADOCK による予測結果は F-measure = 0.277 という結果となり、既存の PRISM による予測結果 (F-measure = 0.296) と比べるとわずかに劣る結果となった。MEGADOCK は PRISM のように既知のテンプレート構造を用いていないことから、この予測精度の差異はテンプレート構造の利用の有無によるものと考えられる。また、コンセンサスによる 2 手法の統合結果が高い Precision 値 (0.322) を示すことを確認した。これにより、PPI 有と予測された結果の信頼性を高めることができ、新規相互作用の発見にコンセンサスが有用であることが分かった。さらに、コンセンサスによる予測結果のうち既存の PPI データベースに含まれていなかった予測結果については、CASP3-CASP7, Akt-Bax など、実際に相互作用可能性のあるタンパク質ペアが存在することを確認した。

## 謝辞

PPI データベースの情報を提供頂いたシンガポール科学技術研究庁 (A\*STAR) の Dr. Vachiranee Limviphuvadh に感謝の意を表する。

本研究は、科研費 (特別研究員奨励費 23-8750)、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」、文部科学省 博士課程教育リーディングプログラム「情報生命博士教育院」の支援を受けて行われたものである。

## 参考文献

- [1] Ravasi T, Suzuki H, Cannistraci CV, *et al.* An Atlas of Combinatorial Transcriptional Regulation in Mouse and Man, *Cell*, 140, 744–752, 2010.
- [2] Tuncbag N, Kar G, Keskin O, Gursoy A, Nussinov R. A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Briefings in Bioinformatics*, 10(3), 217–232, 2009.
- [3] Valencia A, Pazos F. Prediction of protein-protein interactions from evolutionary information. *Structural Bioinformatics, Second Edition*, Wiley and sons: New York, 617–634, 2009.
- [4] Shen J, Zhang J, Luo X, *et al.* Predicting protein-protein interactions based only on sequences information. *Proc Natl Acad Sci USA*, 104, 4337–4341, 2007.
- [5] Wass MN, David A, Sternberg MJE. Challenges for the prediction of macromolecular interactions. *Curr Opin Struct Biol*, 21(3), 382–390, 2011.
- [6] Matsuzaki Y, Matsuzaki Y, Sato T, Akiyama Y. *In silico* screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis. *J Bioinform Comput Biol*, 7(6), 991–1012, 2009.
- [7] Tuncbag N, Gursoy A, Nussinov R, Keskin O. Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM. *Nature Protocols*, 6, 1341–1354, 2011.
- [8] Wass MN, Fuentes G, Pons C, Pazos F, Valencia A. Towards the prediction of protein interaction partners using physical docking. *Mol Syst Biol*, 7, 1–8, 2011.
- [9] Mintseris J, Pierce B, Wiehe K, *et al.* Integrating statistical pair potentials into protein complex prediction. *Proteins*, 69(3), 511–520, 2007.
- [10] Ozbabacan SEA, Keskin O, Nussinov R, Gursoy A. Enriching the human apoptosis pathway by predicting the structures of protein-protein complexes. *J Struct Biol*, 179(3), 338–346, 2012.
- [11] 大上雅史, 松崎由理, 松崎裕介, 佐藤智之, 秋山泰. MEGADOCK: 立体構造情報からの網羅的タンパク質間相互作用予測とそのシステム生物学への応用. 情報処理学会論文誌 数理モデル化と応用, 3(3), 91–106, 2010.
- [12] Matsuzaki Y, Ohue M, Uchikoga N, Akiyama Y. Protein-protein interaction network prediction by using rigid-body docking tools: application to bacterial chemotaxis. *Protein & Peptide Letters*. (submitted)
- [13] Oltersdorf T, Elmore SW, Shoemaker AR, *et al.* An inhibitor of Bcl-2 family proteins induces regression of solid tumours. *Nature*, 435(7042), 677–681, 2005.
- [14] Cell Signaling Technology, Inc. Overview: Regulation of Apoptosis. <http://www.cellsignal.com/pathways/apoptosis-signaling.jsp>.
- [15] Mashiaeh E, Nussinov R, Wolfson HJ. FiberDock: Flexible induced-fit backbone refinement in molecular docking. *Proteins*, 78(6), 1503–1519, 2010.
- [16] Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. *Nucleic Acids Res*, 28(1), 235–242, 2000. <http://www.rcsb.org/pdb/>.
- [17] Tuncbag N, Gursoy A, Guney E, Nussinov R, Keskin O. Architectures and functional coverage of protein-protein interfaces. *J Mol Biol*, 381, 785–802, 2008.
- [18] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28, 27–30, 2000. <http://www.genome.jp/kegg/>.
- [19] Szklarczyk D, Franceschini A, Kuhn M, *et al.* The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*, 39, D561–568, 2011. <http://www.string-db.org>.
- [20] Katchalski-Katzir E, Shariv I, Eisenstein M, *et al.* Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci USA*, 89, 2195–2199, 1992.
- [21] Ohue M, Matsuzaki Y, Ishida T, Akiyama Y. Improvement of the Protein-Protein Docking Prediction by Introducing a Simple Hydrophobic Interaction Model: an Application to Interaction Pathway Analysis. *The 7th IAPR International Conference on Pattern Recognition in Bioinformatics (PRIB2012), Lecture Note in Bioinformatics*, 7632, 178–187, Springer Heidelberg, 2012.
- [22] Zhang C, Vasmatzis G, Cornette JL, DeLisi C. Determination of atomic desolvation energies from the structures of crystallized proteins. *J Mol Biol*, 267(3), 707–726, 1997.
- [23] Pierce B, Weng Z. ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins*,

- 67(4), 1078–1086, 2007.
- [24] Bader GD, Betel D, Hogue CW. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res*, 31(1), 248–250, 2003.
  - [25] Stark C, Breitkreutz BJ, Chatr-Aryamontri A, *et al.* The BioGRID Interaction Database: 2011 update. *Nucleic Acids Res*, 39, D698–704, 2011.
  - [26] Salwinski L, Miller CS, Smith AJ, *et al.* The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res*, 32, D449–451, 2004.
  - [27] Keshava Prasad TS, Goel R, Kandasamy K, *et al.* Human Protein Reference Database–2009 update. *Nucleic Acids Res*, 37, D767–772, 2009.
  - [28] Kerrien S, Aranda B, Breuza L, *et al.* The IntAct molecular interaction database in 2012. *Nucleic Acids Res*, 40, D841–846, 2012.
  - [29] Ceol A, Chatr-Aryamontri A, Licata L, *et al.* MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res*, 38, D532–539, 2010.
  - [30] Guldener U, Munsterkotter M, Oesterheld M, *et al.* MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Res*, 34, D436–441, 2006.
  - [31] Pagel P, Kovac S, Oesterheld M, *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics*, 21(6), 832–834, 2005.
  - [32] Allsopp TE, McLuckie J, Kerr LE, *et al.* Caspase 6 activity initiates caspase 3 activation in cerebellar granule cell apoptosis. *Cell Death Differ*, 7, 984–993, 2000.
  - [33] Slee EA, Adrain C, Martin SJ. Executioner caspase-3, -6, and -7 perform distinct, non-redundant roles during the demolition phase of apoptosis. *J Biol Chem*, 276(10), 7320–7326, 2001.
  - [34] Walsh JG, Cullen SP, Sheridan C, *et al.* Executioner caspase-3 and caspase-7 are functionally distinct proteases. *Proc Natl Acad Sci USA*, 105(35), 12815–12819, 2008.
  - [35] Sadidi M, Lentz SI, Feldman EL. Hydrogen peroxide-induced Akt phosphorylation regulates Bax activation. *Biochimie*, 91(5), 577–585, 2009.
  - [36] Yeretssian G, Correa RG, Doiron K, *et al.* Non-apoptotic role of BID in inflammation and innate immunity. *Nature*, 474(7349), 96–99, 2011.