

タンパク質ドッキングと配列情報特徴解析の融合による タンパク質間相互作用予測の高精度化

小幡 康文^{1,2,a)} 石田 貴士^{1,2,b)} 秋山 泰^{1,2,c)}

概要: タンパク質間相互作用は生命活動を理解する上で重要な現象であり、現在では計算機を用いて相互作用を予測する手法の研究が行われている。これまでに開発されてきた相互作用予測手法は、タンパク質ドッキングによる予測と既知相互作用情報を用いる予測に大きく分類される。タンパク質ドッキングによる予測手法は、未発見の様式を持つ相互作用を発見できる可能性を持つが、偽陽性が多いという欠点があり、既知の相互作用情報を用いた予測手法は、既知情報に類似するタンパク質ペアについては精度良く予測可能であるが、既知情報に類似のものが存在しないペアについては予測することが難しいという欠点がある。そこで、本研究では、タンパク質ドッキングを用いた予測と、タンパク質配列情報の特徴解析による予測を融合し新たな予測モデルを3種類提案した。評価実験の結果、個別の手法と比較して予測精度の改善に成功し、最大で、F 値 0.05 (0.21 → 0.26) の改善を達成した。

キーワード: タンパク質間相互作用予測, タンパク質ドッキング, タンパク質配列相同性, MEGADOCK

Highly precise protein-protein interaction prediction using hybrid prediction model based on *de novo* docking and sequence-based methods.

YASUFUMI OBATA^{1,2,a)} TAKASHI ISHIDA^{1,2,b)} YUTAKA AKIYAMA^{1,2,c)}

Abstract: Protein-protein interaction (PPI) is important for understanding biological functions. Currently, various *in silico* PPI prediction methods have been developed and they are generally classified into two typical approaches: a method based on protein docking and a method based on known PPI information. The result of *de novo* docking based method contains many false-positive predictions, and the known PPI information based method is difficult to predict a novel PPI. In this research, we have proposed new hybrid PPI prediction methods by combining *de novo* docking based prediction and known PPI information based prediction. For evaluating proposed method, we applied the methods to human apoptosis signaling pathway. As results, proposed method achieved an improvement in F-measure (0.21 → 0.26).

Keywords: protein-protein interaction prediction, protein-protein docking, sequence homology search, MEGADOCK

1. 導入

タンパク質は、生体内において他のタンパク質と相互作用を起こし、機能を発揮することが知られている。この相互作用のことをタンパク質間相互作用 (protein-protein interaction, PPI) と呼ぶ。この PPI は、生体機能において中心的な役割を担っており、生命維持において重要な現象である。また、複数のタンパク質間での PPI の関係は PPI

¹ 東京工業大学 大学院情報理工学研究科 計算工学専攻,
Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

² 東京工業大学 情報生命博士教育院,
Education Academy of Computational Life Science, Tokyo Institute of Technology

a) obata@bi.cs.titech.ac.jp

b) ishida@cs.titech.ac.jp

c) akiyama@cs.titech.ac.jp

ネットワークと呼ばれ、生命現象の関係性を表している。PPI ネットワークの解明は、生命現象のより深い理解につながると思われる、生物学の分野においても最も重要な研究課題とされている [1]。PPI の同定には Yeast 2 Hybrid 法 [2] などの実験的手法が用いられ、これまでに同定された PPI は、DIP (the Database of Interacting Proteins) [3] や STRING [4] といったデータベースに集積されている。

しかしながら、タンパク質ペアについて網羅的な実験を行うことは、時間的・金銭的制約から困難である。このため、計算機を用いたハイスループットな PPI 予測手法が研究されている [5]。現在の PPI 予測手法は大きく 2 種類に分類される。1 つ目の手法は、タンパク質ドッキングを用いた予測である。この手法は、既知の PPI ペアとの配列相同性を基にデータベースマッチングを用いた手法 [6], [7], [8] や機械学習を用いた手法 [9] などが提案されている。2 つ目の手法は、タンパク質立体構造情報を用いた、タンパク質ドッキングによる予測 [10], [11], [12], [13] である。既知 PPI 情報を用いた手法は、既知情報との類似度が高いものについては高い精度で予測を行うことが可能である一方、既知情報と大きく離れたものについては予測することが難しい。その一方で、タンパク質ドッキングによる予測は、物理科学的特性を鑑みた予測が行えることから、未発見様式の PPI を発見できる可能性を持つ。しかしながら、偽陽性の検出数が多いという欠点がある。予測後に実験を行う場合、偽陽性の数を減少させる必要が有ると考えられる。

これらの欠点を補い、予測精度を向上させるためにこれら 2 種類の予測手法を融合する研究が近年試みられている。先行研究として、Ohue らによって提案された手法 [14] が挙げられる。Ohue らはタンパク質ドッキングを用いた予測と、既知相互作用面の立体構造情報に基づく予測によって得られた予測結果に対して、論理積を用いることで新たな PPI 予測を行っている。この手法では、各手法単独で予測を行った場合と比較して、適合率が向上し、より正確な予測を実現している。しかし、Ohue らの手法では、タンパク質配列は考慮されず、加えて相互作用面として利用可能な情報は 2 万件ほどである。既知の PPI 情報として蓄積されているタンパク質配列の情報量は 7 万件以上の情報が蓄積されており、Ohue らの手法ではこの多数蓄積されている情報を利用することは出来ない。また、Ohue らの手法では、融合する前の個別の予測における予測精度 (F 値) が、既知相互作用面を用いた予測において 0.296 であるにもかかわらず、融合手法 (論理積) においては、0.285 と低下しており、予測精度については改善を行っていない。そのため融合手法について、Ohue らの提案した論理積や論理積以外の融合手法が適用可能であり、個別の予測における予測精度を上回る融合予測モデルを構築可能であると考えられる。

この問題に対して、本研究では、タンパク質ドッキング

を用いた予測と、タンパク質配列情報の特徴解析による予測との融合手法を提案する。既知 PPI 情報の中でも蓄積量の多いタンパク質配列情報を基にした予測とタンパク質ドッキングを用いた予測とを融合することで、より既知情報を有効に利用した予測手法を提案する。また、これらの手法を融合した予測モデルを 3 種類提案し、PPI 予測の精度改善を目指した。

2. 提案手法

本研究では、2 つの PPI 予測手法の予測結果を融合し、新たな予測モデルを作成することで PPI 予測の予測精度の向上を図る。本研究では、配列情報特徴解析による予測の重要な手法の一つである配列相同性を基にした予測と、タンパク質ドッキングを用いた予測との融合予測モデルを提案する。

2.1 ベースとなる予測手法

2.1.1 タンパク質ドッキングを用いた予測

本研究では、タンパク質ドッキングを用いた予測として、Ohue らによって提案された予測手法 [10] を用いる。この手法では、タンパク質ペアの個々の立体構造情報を入力とし、タンパク質ドッキングによって予測されたタンパク質複合体構造を用いて、PPI 評価値を算出する。PPI 評価値は、予測された複合体構造のエネルギーに基づいて算出される単位を持たない値である。この値を基準として、PPI 予測を行う。

この手法では、如何なるタンパク質ドッキングソフトウェアでも用いることが可能である。本研究では、タンパク質ドッキングソフトウェアとして、MEGADOCK [15] を用いる。MEGADOCK は、大量のドッキング演算を行うことが可能な高速性を有するドッキングソフトウェアである。現在発見されているタンパク質数は 50 万種類以上になっており、生体機能の解明を考えた場合、MEGADOCK の高速性は非常に重要であると考えられることから、MEGADOCK を用いる。

2.1.2 配列相同性を基にした PPI 予測

配列相同性を基にした PPI 予測として、本研究では Folador らが研究で用いた手法 [8] を用いる。本手法では、タンパク質ペアの個々のタンパク質配列を入力とし、既知 PPI のデータベース中のタンパク質配列との配列相同性を評価して得られたスコアを基に PPI 予測を行う。本研究では、Folador らの手法とは異なり、HSSP-value [16] を配列相同性評価のスコアとして用いる。これは、HSSP-value が、タンパク質配列の一致率 (identity) とアライメントされた配列の長さを考慮することが出来るスコアであり、有用であると考えられるためである。

本研究では、タンパク質相同性を探索するために BLAST [17] シリーズの内、タンパク質アミノ酸配列同

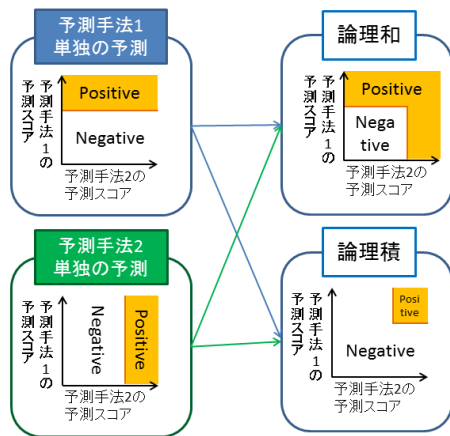


図 1 論理和・論理積を用いた融合の概念図

士の比較を行う BLASTP を使用する。配列相同性評価スコアの HSP-value は BLASTP によって算出される値を用いて計算が可能である。

2.2 融合予測モデル

「論理和・論理積を用いた予測」と「多段階予測」, 「スコア線形和を用いた予測」の3種類のモデルを提案する。本研究で三手法を提案するのは、各手法それぞれに長所・短所があることから、それを鑑みて、利用する場面で適切な手法を選ぶことが必要であると考えられるためである。

2.2.1 論理和・論理積を用いた融合

タンパク質ドッキングを用いた PPI 予測手法と配列情報特徴解析を用いた PPI 予測手法は、それぞれ個別に予測結果を出力する。本モデルでは、この予測結果の論理和・論理積を用いて予測を行う。このモデルは、モデルの構築が非常に単純な為、適用可能な範囲が広く、容易に用いることができる点が大きな利点である。

本モデルの概念図を図 1 に示す。この図において、縦軸がある予測手法のスコアの値、横軸も一方の予測手法のスコアの値を示している。論理和を用いた場合、PPI 有り (positive) と判定される区域が広がることから再現率の向上が期待される。一方で、論理積を用いた場合は positive と判定される区域が狭まり、加えて各手法で高スコアを獲得しているタンパク質ペアのみが判定されることより、適合率の向上が期待される。

2.2.2 多段階予測

論理和・論理積を用いた融合は、単純なモデルである点が大きな利点であるが、その単純さゆえに、新しいモデルについて調整を行うことが出来ない点が欠点である。これに対して多段階予測では、ある予測において判定境界に近い部分に存在するタンパク質ペアについて、別の手法によって再判定を行うことで、タンパク質ペアの選別を行うことで予測精度の向上を目指す。本モデルの概念図を図 2 に示す。

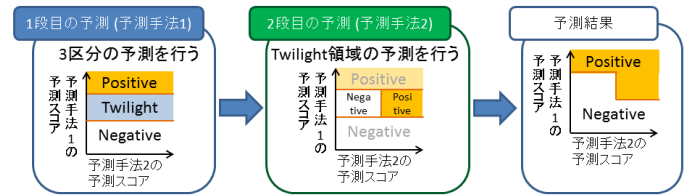


図 2 多段階予測の概念図

多段階予測では、予測を多段階に適用する順番によって、期待される効果が異なる。まず、タンパク質ドッキングを用いた予測を第 1 段階とした場合について述べる。タンパク質ドッキングを用いた予測の欠点として、偽陽性の多さが挙げられる。この偽陽性となるタンパク質ペアの多くは、PPI 予測の閾値周辺に存在するペアであると考えられる。そこで、タンパク質ドッキングによる予測を第 1 段階の予測として用い、スコアが閾値周辺に存在するペアについて、配列相同性による予測で再予測することでタンパク質ペアの選別を行う。これによって、偽陽性を減少させ適合率が向上することが期待される。次に、配列相同性を用いた予測を第 1 段階とした場合について述べる。配列相同性を用いた予測では、配列類似性の僅かな差で偽陰性となっているペアが存在する。このようなペアについて、ドッキングによる予測を用いて物理化学的特徴から PPI の可能性を再検討することで、偽陰性ペアが減少し、再現率が向上することが期待される。

本モデルでは、第 1 段階における予測を 3 区分の予測 (positive・twilight・negative) へと拡張する。twilight 領域は、PPI 有り (positive) と PPI 無し (negative) の間に存在する領域である。twilight 領域に含まれるタンパク質ペアは、2 区分で PPI 予測を行った場合、positive と negative の閾値近くに存在するタンパク質ペアであり、第 1 段階の手法において予測の確度が高くないペアであると考えられる。この twilight 領域に含まれるペアについてのみ、第 2 段階での PPI 予測を行いタンパク質ペアの選別を行う。そして、これらの予測結果の統合を行う。本予測モデルは第 1 段階で 3 分類の予測を行う事が出来れば、如何なる予測手法であっても適用可能である。

2.2.3 スコア線形和による融合

多段階予測では、パラメータが 3 つ存在することから、その探索に時間がかかってしまうという欠点がある。また、論理和・論理積による融合や多段階予測では、各予測結果を個別に用いていた。これに対して、本予測モデルではドッキングによる予測と配列相同性を基にした予測それぞれで得られる予測スコアの線形和を用いて融合し予測を行う。本モデルの概念図を図 3 に示す。

論理和・論理積を用いた融合と多段階予測においては、各予測手法における個別の予測結果を用いて融合を行っていたが、スコア線形和を用いた融合では、融合によって各

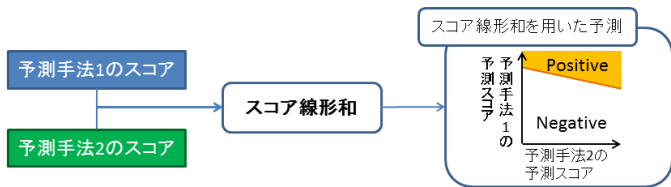


図 3 スコア線形和を用いた融合の概念図

予測の信頼度に応じたスコアが新たに計算されることから、信頼度の高い予測を中心としながら、信頼度が劣る予測手法を考慮した予測が可能となる。本手法では、予測においてスコアを出力することが可能な予測手法のみに適用することができる予測手法であり、本研究で提案した3種類の予測手法の中では最も適応範囲の狭い予測手法である。

3. 評価実験

本研究では、提案手法の予測精度を評価するために、ヒトのアポトーシスパスウェイに関連するPPIを扱ったデータセットを用いて、予測精度の評価実験を行った。

3.1 テストセット

本実験では、Ohue らが用いたヒトアポトーシスパスウェイに関連するタンパク質のデータセット [14] を用い、予測精度の検証を行う。

本データセットには、計 57 種類のタンパク質が含まれている。相互作用予測を行うタンパク質ペアは、1596 ペアである。この内、相互作用することがわかっているペアは、137 ペアとなっている。よってこのデータセットに含まれている positive ペアの割合は、約 8.6% である。テストセットに含まれるタンパク質の立体構造情報と配列情報は、タンパク質立体構造データベースである PDB[18] より取得する。本データセットには、1つのタンパク質に複数の PDB ファイルが対応している場合があり、このような場合は、全 PDB ファイルを用いて予測を行い、その予測結果について論理和を取ることで、対応するタンパク質の予測結果とする。

3.2 配列相同性を基にした PPI 予測で使用するデータベース

配列特徴解析を用いた PPI 予測において用いる既知の PPI データベースとして、本研究では DIP[3] を用いる。DIP は 720 種の生物における PPI データを集積したデータベースである。本研究では、2014 年 10 月 1 日更新分のデータベースを用いる。このデータベースには、計 74,078 件の PPI 情報が集積されている。このデータベースから、3.1 節で述べたテストセットに含まれるタンパク質と類似しているタンパク質の除去を行う。今回は、タンパク質配列間の identity (アミノ酸配列の文字一致率) を基準とし、

表 1 データベースに含まれる情報

	オリジナル	id 40	id 90
# PPI	74,078	68,378	70,629
# タンパク質	26,790	24,192	24,770

表 2 予測結果の分類

		正解	
		Positive	Negative
予測結果	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

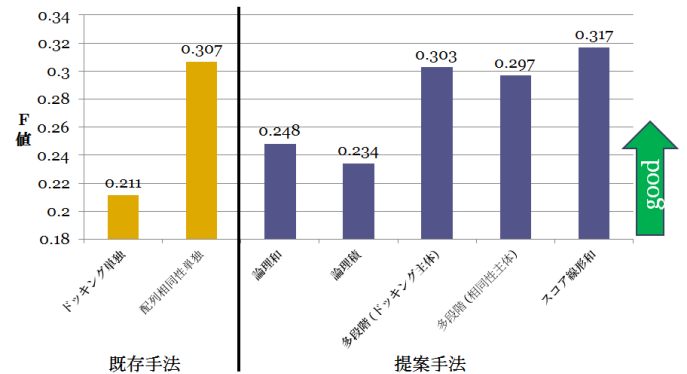


図 4 id 90 を用いた場合の予測結果。

その基準を超えるものについては、データベースから削除した。本研究では、identity 40%を基準とした現実的な状況のデータベース (id 40) と、基準を identity 90%としたやや極端で、配列相同性を基にした予測において理想に近い状況のデータベース (id 90) を用いる。データベースに含まれる情報量を表 1 にまとめて示す。

3.3 評価基準

本実験では、F 値 (F-measure) を用いて、手法の評価を行う。F 値は、適合率と再現率の調和平均である。式 (1) に F 値の定義を示す。

$$F\text{-measure} = \frac{TP}{\frac{1}{2} \times \{(TP + FP) + (TP + FN)\}} \quad (1)$$

式中の TP, FP, FN, TN はそれぞれ、予測結果と正解の関係から定義されるもので、式においては分類された予測結果の個数を示す。その予測結果の分類を表 2 に示す。

3.4 実験結果

3.4.1 実験結果 (id 90 を用いた場合の結果)

配列相同性を基にした PPI 予測において理想に近いデータベース (id 90) を用いた際の予測精度を図 4 に示す。図中のラベルのカッコ内は多段階予測において第 1 段階で使用される予測手法を示す。この結果より、ドッキングによる予測では予測精度が 0.211 である一方、配列相同性を基にした予測において 0.307 と高い予測精度を達成している。提案手法の内、論理和・論理積による融合では、配列相同

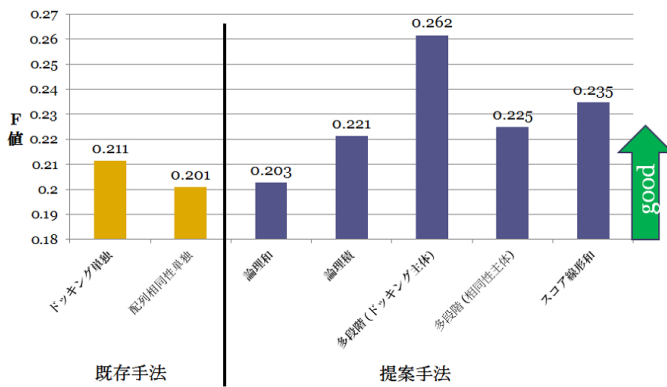


図 5 id 40 を用いた場合の予測結果.

性を基にした予測と比較して予測精度は低下した。一方で、多段階予測の内配列相同性を主体とした多段階予測の予測精度は、配列相同性による精度より 0.01 低い 0.297 である。また、ドッキング主体の多段階予測では配列相同性による予測と同等の予測精度 (0.303) であった。更に、スコア線形和による融合による予測では配列相同性による予測の精度を上回る 0.317 を達成した。これらのことから、データベースが理想的な状況に近い場合、スコア線形和を用いた融合によって、精度の高い予測を行うことができるといえる。

3.4.2 実験結果 (id 40 を用いた場合の結果)

配列相同性を基にした PPI 予測において現実に近いデータベース (id 40) を用いた際の予測精度を図 5 に示す。データベースに類似度の高いタンパク質が存在しないため、配列相同性による予測精度は理想的な状況に近いデータベース (id 90) を用いた場合と比較して低下し、ドッキングを用いた予測の予測精度 (0.211) と同程度の 0.201 となっている。提案手法の内、論理和を用いた融合では予測精度は向上しなかったが、論理積を用いた融合では個別の予測と比較して 0.01 高い予測精度 (0.221) となった。また、配列相同性を基にした予測を主体とした多段階予測でも論理積を用いた融合と同程度の予測精度 (0.225) であり、スコア線形和を用いた融合では予測精度は 0.235 となった。提案手法の中でも、ドッキングを主体とした多段階予測では、予測精度は各手法の中でも最も高い 0.262 を達成した。このドッキングを主体とした多段階予測では、個別の予測の予測精度と比較して、0.05 の予測精度向上を達成した。これらのことから、データベースに類似したタンパク質が存在しない現実的な状況において、ドッキングを主体とした多段階予測によって、精度の高い予測を行うことが可能であるといえる。

4. 考察

ドッキング主体の多段階予測は、現実的な状況を想定したデータベース (id 40) を用いた際に、最も向上幅の大きい予測精度改善を達成した。本節では、ドッキング主体の

表 3 多段階予測 (ドッキング主体) の予測結果

手法	適合率	再現率	F 値
多段階 (ドッキング主体)	0.276	0.248	0.262
ドッキング単体	0.139	0.445	0.211
配列相同性単体	0.146	0.321	0.201

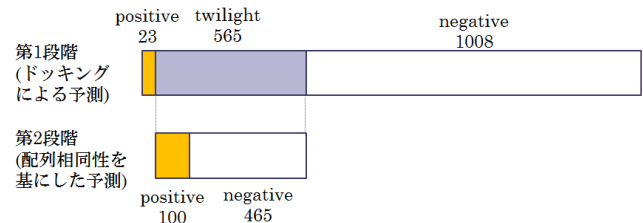


図 6 各段階における予測結果確定数.

多段階予測をより詳細に調査し、予測精度向上の内容を考察する。

まず、表 3 に現実的な状況を想定したデータベース (id 40) を用いた際のドッキング主体の多段階予測と融合した手法個別の予測結果における適合率、再現率、F 値を示す。表 3 より、適合率は単独の場合の値である 0.139 及び 0.146 と比較して向上し、多段階予測では 0.276 を記録している。その一方で、再現率は単独の場合の再現率 0.445 及び 0.321 と比較して低下し、多段階予測では 0.248 となっている。これらの予測傾向の変化の内、適合率の向上の影響により、F 値は 0.201, 0.211 から約 0.05 向上し、0.262 を達成した。

ドッキングを主体とした多段階予測は、ドッキングを用いた予測における偽陽性の減少を目的に提案した予測手法である。ドッキングによる予測の閾値近辺のタンパク質ペアについて、配列相同性を基にした予測を用いてタンパク質ペアの選別を行うことで、偽陽性となるタンパク質ペアの数を減少させることが期待された。そこで、多段階予測の各段階において、予測の確定がどのように行われたのか、実際の予測における予測結果確定の内容を調査した。

各段階における予測結果の確定の様子を図 6 に示す。図 6 より、多段階予測において、第 1 段階で全タンパク質ペアの内、約 65% の予測結果が確定していることが分かる。これらの内、多くが negative ペアである。また、残り 35% が第 2 段階の予測の対象となっており、第 2 段階の予測において全 positive ペアの約 8 割が確定している。図 6 及び表 3 より、第 1 段階のドッキングによる予測で twilight と判定されたペアについて、第 2 段階の配列相同性による予測で選別が行われたと言える。その結果、ドッキングによる予測で問題とされていた偽陽性の数を減少させ、適合率を上昇させ、予測精度である F 値の向上につながったといえる。これは、ドッキングを主体とした多段階予測で期待される効果と同じであり、期待された融合の効果を発揮し、予測精度の向上を達成することができたと考えられる。

5. まとめ

5.1 結論

本研究では、タンパク質ドッキングによる予測と配列相同性を基にした予測を融合した新しい融合予測モデルとして、「論理和・論理積を用いた融合」、「多段階予測」、「スコア線形和を用いた融合」の3種の融合予測モデルを提案した。提案した予測モデルの予測精度を評価するためにヒトアポトーシスパスウェイに関連するタンパク質を対象とした評価実験を行った。その結果、既知PPIデータベースに類似度の高いタンパク質が存在する、理想に近い状況においては、スコア線形和を用いた融合において、既存の予測手法の予測精度(0.211及び0.307)よりも高い予測精度(0.317)を達成した。また、既知PPIデータベースに類似度の高いタンパク質が存在しない現実的な状況においては、ドッキングを主体とした多段階予測において、個別の予測の予測精度と比較して予測精度0.05の向上(0.211→0.262)を達成した。このドッキング主体の多段階予測における予測精度0.05の向上は、本研究において最も大きい精度向上である。また、この精度向上はタンパク質ドッキングによる予測で得られた閾値近辺のペアについて、配列相同性を用いた予測で選別を行うことで、適合率の向上を実現しており、多段階予測において期待された効果を発揮できたとと言える。

5.2 今後の課題

今後の課題として、今回提案した予測モデルの更なる改良が挙げられる。例えば、融合を図る予測手法数をさらに増やし、多段階予測を3段階以上にした融合予測モデルや複数のスコア線形和を用いた融合予測モデルが考えられる。しかしながら、3段階以上の多段階予測においては、融合予測モデルの構成が複雑になり、モデルの作成が難しいと考えられる。また、複数のスコア線形和を用いる場合には、パラメータ数が本研究で提案した手法よりも増加することが予想されるため、これらのパラメータの探索に時間が必要となってしまうと考えられる。そのため、何らかの工夫を行うことで、これらの問題に対処し、予測モデルの改良を行う必要があると考えられる。

謝辞

本論文を終えるに当たり、タンパク質間相互作用予測に関して様々な助言を賜りました、東京工業大学秋山研究室の角田 将典博士研究員、大上 雅史博士研究員に深くお礼申し上げます。本研究は、文部科学省 博士課程教育リサーチプログラム「情報生命博士教育院」の支援を受けて行われたものです。

参考文献

- [1] Stelzl, U. et al.: A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome Max-Planck-Institute for Molecular Genetics., *Cell*, Vol. 122, No. 6, pp. 957–968 (2005).
- [2] Fields, S. et al.: A novel genetic system to detect protein-protein interactions., *Nature*, Vol. 340, No. 6230, pp. 245–246 (1989).
- [3] Salwinski, L. et al.: The Database of Interacting Proteins: 2004 update., *Nucleic Acids Res.*, Vol. 32(Database issue), pp. D449–D451 (2004).
- [4] Szklarczyk, D. et al.: STRING v10: protein-protein interaction networks, integrated over the tree of life., *Nucleic Acids Res.*, Vol. 43(Database issue), pp. D447–D452 (2015).
- [5] Liu, Z. P. et al.: Proteome-wide prediction of protein-protein interaction from high-throughput data., *Protein Cell*, Vol. 3, No. 7, pp. 508–520 (2012).
- [6] Chen, C. C. et al.: PPIsearch: a web server for searching homologous protein-protein interactions across multiple species., *Nucleic Acids Res.*, Vol. 37(Web Server issue), pp. W369–W375 (2009).
- [7] Garcia-Garcia, J. et al.: BIPS: BIANA Interolog Prediction Server. A tool for protein-protein interaction inference., *Nucleic Acids Res.*, Vol. 40(Web Server issue), pp. W147–W151 (2012).
- [8] Folador, E. L. et al.: An improved interolog mapping-based computational prediction of protein-protein interactions with increased network coverage., *Integr. Biol. (Camb.)*, Vol. 6, No. 11, pp. 1080–1087 (2014).
- [9] Martin, S. et al.: Predicting protein-protein interactions using signature products., *Bioinformatics*, Vol. 21, No. 2, pp. 218–226 (2005).
- [10] Ohue, M. et al.: MEGADOCK: an all-to-all protein-protein interaction prediction system using tertiary structure data., *Protein Pept. Lett.*, Vol. 21, No. 8, pp. 766–778 (2014).
- [11] Wass, M. N. et al.: Towards the prediction of protein interaction partners using physical docking., *Mol. Syst. Biol.*, Vol. 7, No. 469 (2011).
- [12] Zhang, Q. C. et al.: Structure-based prediction of protein-protein interactions on a genome-wide scale., *Nature*, Vol. 490, No. 7421, pp. 556–560 (2012).
- [13] Zhang, C. et al.: Discovery of binding proteins for a protein target using protein-protein docking-based virtual screening., *Proteins*, Vol. 82, No. 10, pp. 2472–2482 (2014).
- [14] Ohue, M. et al.: Highly precise protein-protein interaction prediction based on consensus between template-based and de novo docking methods., *BMC Proc.*, Vol. 7(Suppl 7), p. S6 (2013).
- [15] Ohue, M. 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).
- [16] Rost, B.: Twilight zone of protein sequence alignments., *Protein Eng.*, Vol. 12, No. 2, pp. 85–94 (1999).
- [17] Altschul, S. F. et al.: Basic local alignment search tool., *J. Mol. Biol.*, Vol. 215, No. 3, pp. 403–410 (1990).
- [18] Berman, H. M. et al.: The Protein Data Bank., *Nucleic Acids Res.*, Vol. 28, No. 1, pp. 235–242 (2000).