

タンパク質間相互作用予測における 可視化手法の開発および予測精度の改善

山本航平^{†1} 大上雅史^{†1,†2} 内古閑伸之^{†1}
松崎由理^{†1} 石田貴士^{†1} 秋山泰^{†1}

タンパク質の構造情報を利用したタンパク質間相互作用 (Protein-Protein Interaction, PPI) 予測を行うために、我々はタンパク質ドッキング計算に基づく PPI 予測手法の開発を行ってきた。本研究ではタンパク質ドッキング計算によって得られた複合体候補構造群について、それらがどのように 3 次元空間上で分布しているかを可視化する手法と、タンパク質の各残基の相互作用している傾向を可視化する手法の 2 種類を開発した。また、可視化結果から得た知見をもとに、相互作用予測結果から、本来は相互作用しないのに“相互作用する”と判定された結果を識別する評価値を導入し、それを利用した PPI 予測精度の改善手法を開発した。

A visualization method and accuracy improvement for protein-protein interaction predictions

KOHEI YAMAMOTO,^{†1} MASAHIITO OHUE,^{†1,†2}
NOBUYUKI UCHIKOGA,^{†1} YURI MATSUZAKI,^{†1}
TAKASHI ISHIDA^{†1} and YUTAKA AKIYAMA^{†1}

To predict all-to-all the protein-protein interactions (PPIs) from protein monomeric structures, we have been developing a PPI prediction method based on a protein docking calculation. In this study, we developed two methods which visualize complex candidates generated by protein docking calculations. The first method visualizes 3D distribution of the complex candidates in space. And the second one visualizes PPI tendency of each residue by coloring them according to the interaction frequency. By using these visualization methods, we also developed a new score to identify false positives and improved the accuracy of a PPI prediction.

1. はじめに

タンパク質間相互作用 (Protein-Protein Interaction, PPI) は生命現象において、中心的な役割を果たしていることが知られている。計算機による相互作用予測手法には、タンパク質のアミノ酸配列モチーフを用いた手法¹⁾⁻³⁾ や、既知のドメイン間相互作用情報などのデータベースに基づく手法^{4),5)} などが用いられてきた。我々はタンパク質の構造情報が相互作用予測に重要であると考え、タンパク質同士の形状相補性に基づくタンパク質ドッキング計算を利用した PPI 予測手法の研究を行ってきた⁶⁾⁷⁾。タンパク質ドッキングを利用した手法では、ドッキング計算によって得られた大量の複合体候補構造を用いて相互作用するかどうかの予測を行う。予測精度の改善のためにこれまで様々な後処理の手法が検討されてきたが、さらなる予測精度の向上が望まれている。

本稿では大きく分けて 2 つのを行う。第 1 に予測精度改善手法開発の補助ためにタンパク質ドッキング結果の可視化手法の提案を行う。予測精度改善のためにはドッキング計算で出力される複合体候補構造から知見を得ることが重要であると考えているが、1 回のドッキング計算で出力される複合体候補構造の数は最大で 10,800 個と膨大であり、ここから直視的に知見を得ることは難しい。ドッキング計算結果を可視化することは、大量のデータから新しい知見を得る補助となることが期待される。第 2 に可視化から実際に得た知見をもとにして新たな評価値を導入し、それを利用した予測精度の改善手法を示す。

2. タンパク質ドッキングを用いた PPI 予測手法

本稿では我々が開発したタンパク質ドッキング計算を用いた相互作用予測手法について説明する。本稿でいうタンパク質ドッキングとは、タンパク質を剛体とみなし、複合体を形成する際に形を変えないと仮定した上で、主にタンパク質表面の形状相補性を利用して計算する複合体構造予測手法である。タンパク質ドッキングを行う既存ソフトウェアとしては MolFit⁸⁾ や FTDock⁹⁾、ZDOCK¹⁰⁾⁻¹²⁾ などが知られている。しかしこれらのソフトウェアは 2 つのタンパク質間での複合体予測を目的に開発されたものである。一方 PPI 予測では、予測対象が相互作用ネットワークなど大規模な系となり、多対多のタンパク質について

^{†1} 東京工業大学 大学院情報理工学研究科 計算工学専攻

Graduate School of Information Science and Engineering, Tokyo Institute of Technology

^{†2} 日本学術振興会特別研究員

Research Fellow of the Japan Society for the Promotion of Science

の網羅的な予測が必要となる。そこで我々は大規模で網羅的な計算に適したドッキングエンジンを含む PPI 予測手法の開発を行っており、このシステム全体を MEGADOCK⁽⁶⁾⁷⁾と呼んでいる。

MEGADOCK による予測システムは、ドッキング計算を行う部分と相互作用予測を行う部分に分かれている。ドッキング計算部では、入力されたタンパク質ペアの一方を回転させながらもう一方のタンパク質の周囲を移動させ、形状相補性と静電相互作用を考慮した独自のドッキングスコア⁽⁶⁾⁷⁾を計算し複合体候補構造を生成する^{*1}。この計算をタンパク質の角度を少しずつ変えながら行い、大量の複合体候補構造群を出力する。

相互作用予測部では、各入力タンパク質ペアについて、それが相互作用するかどうかを、ドッキング計算によって得られた複合体候補構造のドッキングスコアに基づいて判定をする。ドッキングスコアをそのまま使った相互作用予測の精度は充分なものではなく、構造クラスタリング^{(13),(14)}や、ZRANK スコアに基づくドッキング結果の再評価（リランキング）などの手法が、予測精度改善のために試されてきた⁽⁶⁾⁷⁾。現在の MEGADOCK では、擬似結合エネルギーを高速に計算するソフトウェアである ZRANK⁽¹⁵⁾を用いて、ドッキング計算で得られる複合体候補構造のリランキングを行い、MEGADOCK ドッキング計算部で用いられるドッキングスコアを利用したときよりも精度の高い順位付けを実現することで、PPI 予測精度の改善に成功している。その一方、MEGADOCK の予測システムを実問題に適用するにはより高い感度が要求されており、さらに良い精度が得られることが望ましい。予測精度を向上させるアプローチはいくつか考えられるが、その中で我々は、ドッキング計算によって出力される複合体候補構造が膨大であり、複合体候補構造群に含まれる有用な情報を十分に活用しきれていないことが問題であると考えている。そこで、複合体候補構造群から効率的に情報を取り出すための補助の目的で、次章でドッキング計算結果の可視化手法を提案する。

3. タンパク質ドッキング結果の可視化

3.1 可視化手法

MEGADOCK による PPI 予測精度のさらなる向上のために、タンパク質ドッキング計算によって生成される大量の複合体候補構造情報を利用することを考える。MEGADOCK では 1 組のタンパク質ペアのドッキング計算において、複合体候補構造は最大で 10,800 個

出力される。我々はこの膨大な数の複合体候補構造から相互作用予測精度の改善に有効な情報を得ることができると考えているが、複合体候補構造の数が非常に多く、直視的に有効な情報を得ることは困難である。そこでドッキング計算結果に含まれる情報を効率よく得るためにドッキング計算結果の可視化を行う。可視化を行うことで大量の情報を簡便かつ有効に得られることが期待される。本稿では次の 2 つの可視化手法を提案する。

- ベクトルプロット法
- コンタクトカラーリング法

3.1.1 ベクトルプロット法

ベクトルプロット法はドッキング計算によって出力される複合体候補構造中のリガンドがレセプターの周囲にどのように分布しているかを知ることを目的とし、リガンド重心を固定したレセプターの周囲にプロットする手法である。各リガンドがどのような向きでレセプターの周囲に分布しているかを知るため、**図 1**に示すように、リガンドをその重心を始点としたベクトル（棒）としてプロットする。

実際にベクトルプロット法で可視化を行う際は、各複合体候補構造のリガンドの重心を計算し、次に重心からリガンド中の特定の原子を終点とするベクトルを計算を行い、それを棒として描画する。通常、棒の終点となる原子は、タンパク質の構造情報を表すデータ中で 1 番最初に現れる原子を用いている。可視化を行った例は**図 2**の通りである。

3.1.2 コンタクトカラーリング法

我々は、複合体中でレセプターとリガンドの相互作用に関わっている残基は特に重要視すべきものであると考える。そのため特に相互作用する可能性の高い残基を可視化することは、ドッキング計算結果から相互作用に関わる情報を得る上で有用であると考えている。そこで次で定義する**相互作用頻度**をもとにタンパク質表面に色付けを行い、相互作用している残基を可視化するコンタクトカラーリング法を提案する（**図 3**）。

相互作用頻度

まず、“相互作用している残基”を、“相手タンパク質との残基間距離が閾値以下となっているレセプターまたはリガンドの残基”と定義する。残基間距離は、残基を構成する全ての原子同士の間で計算するものとし、それらの原子間距離の値で最小のものを残基間距離とする。

レセプター（リガンド）の各残基について、何個の複合体候補構造がその残基を相互作用残基としているかを求め、それを複合体候補構造の数で割った値、すなわち複合体候補構造全体に対する割合がその残基についての相互作用頻度となる。

*1 このとき、移動するタンパク質を**リガンド**、もう一方の固定したタンパク質を**レセプター**と呼ぶ。

色 付 け

どの残基が強い相互作用の可能性を示すかを明確に理解するために、タンパク質に色付けを行い可視化を行う。色付けは残基単位で行い、相互作用頻度に応じて濃淡をつける。相互作用頻度が大きくなるほど色が濃く表示されるものとする。相互作用頻度は複合体候補構造数に対する割合で表されるため、複合体候補構造の数によらず可視化結果を比較できる利点がある。コンタクトカラーリング法で可視化をした例を図4に示す。

3.2 可視化結果の考察

ベクトルプロット法は、複合体候補構造中リガンドの空間的分布を直感的に理解できる点が優れている。またリガンドの向きも比較することが可能であり、同じような位置に分布しつつも全く相互作用している部位が異なるリガンドを検出することが可能である。しかし空間的な分布は把握できるものの、実際にレセプタータンパク質表面のどの位置にどの程度の頻度で相互作用を行っているかという情報は乏しい。

一方コンタクトカラーリング法はベクトルプロット法で表現できないレセプタータンパク質表面の相互作用している残基の情報を可視化している。また各残基が相互作用を介する可能性を色の濃度によって表現しており、生物学的に重要な相互作用残基の情報をよく表していると言える。ただし、タンパク質のサイズが小さいほど同じ領域が相互作用に関わる可能性が高く、色の濃淡が複合体候補構造の分布と必ずしも一致するわけではない。

ベクトルプロット法とコンタクトカラーリング法は可視化する情報が異なっているため、双方を目的に応じて使い分けることが望ましい。

4. 相互作用頻度による相互作用予測結果の解析

4.1 ドッキング計算結果の可視化による傾向

コンタクトカラーリング法を用いて、MEGADOCKによる44×44の網羅的なタンパク質ドッキング計算結果⁶⁾⁷⁾の可視化を行うことで、MEGADOCKによる相互作用予測で相互作用すると判定されたタンパク質ペアのうち、真に相互作用するペア (True Positive, TP) と、実際は相互作用しないにも関わらず相互作用すると判定されたペア (False Positive, FP)の間ではタンパク質表面の色の濃淡が明らかに異なっている傾向を発見した (図5)。これは、予測結果がTPのタンパク質ペアとFPのタンパク質ペアの間で相互作用に使われる残基の分布に違いがあることを意味する。このことから、相互作用予測を行ったペアの間には、相互作用頻度について何らかの差があると考えられる。そこで相互作用頻度を利用した評価値を導入し、相互作用予測されたタンパク質ペアについて解析を行う。

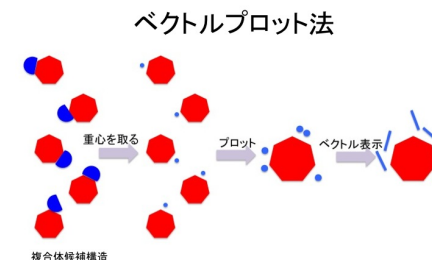


図1 ベクトルプロット法の概略図

Fig.1 A schematic view of vector plot method

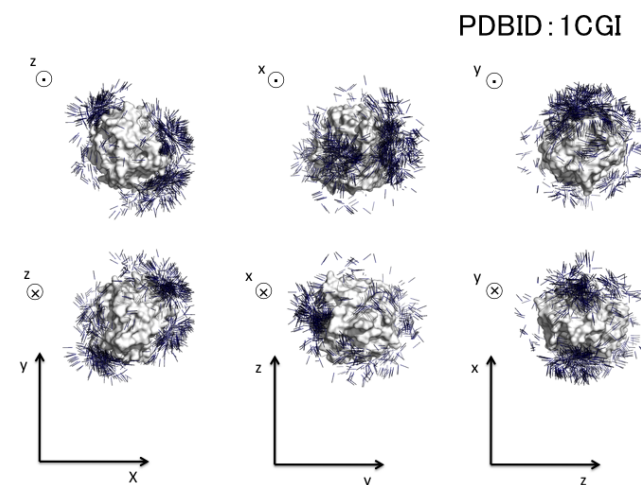


図2 ベクトルプロット法の実例

Fig.2 An example of vector plot method

4.2 相互作用頻度を利用した評価値

ドッキング計算によって出力される複合体候補構造の各残基についての相互作用頻度が、レセプター (またはリガンド) が同じ様々なタンパク質ペアの間でどの程度の差があるかを表現する評価値を定義する。このとき各残基の相互作用頻度の傾向が他のタンパク質ペアと

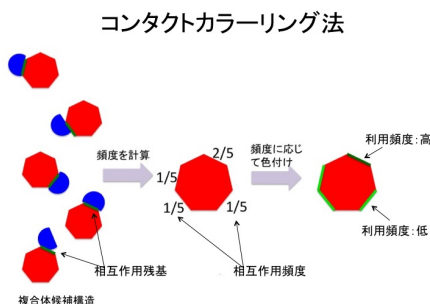


図3 コンタクトカラーリングの概略図

Fig.3 A schematic view of contact coloring method

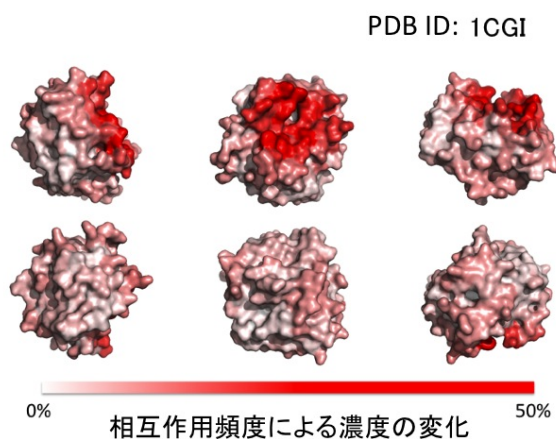


図4 コンタクトカラーリング法の実例

Fig.4 An example of contact coloring method

異なるものを特異的なタンパク質ペアと呼ぶ。

評価値を求める際は、1つのタンパク質と異なる複数のタンパク質の間でそれぞれドッキング計算 (1 vs. N のドッキング計算) を行う。評価値は、1 vs. N のドッキング計算で生成されるそれぞれの複合体候補構造群の間での、残基の相互作用に使われる傾向の差を表現す

TP(1UDI-1UDI)	FP(1UDI-1EAW)	平均(1UDI-全リガンド)

図5 予測結果が True Positive (TP) , False Positive (FP) の間でのレセプター表面の色の濃度の違い
Fig.5 Difference of the surface color of a receptor between two prediction results, True Positive and False Positive

るものとする。

具体的には、一方のタンパク質 (レセプター) を固定し、他方 (リガンド) を替えながらドッキングさせたときの、リガンド全体に対する平均的なレセプター残基の使われ方を表す $\overline{U_R}$ と、あるタンパク質ペアのレセプター残基の使われ方を表す U_R を求め、その差を評価値とする。

あるタンパク質ペアのレセプター R に対するリガンド L のスコア S_R を、相互作用頻度とその平均とのずれの二乗和として次のように定義する。

$$S_R = \sum_m (U_R(m, L) - \overline{U_R(m, L)})^2$$

ただし、 $U_R(m, L)$ はレセプター R とリガンド L のドッキング計算をしたときの、レセプター R の残基 m についての相互作用頻度である。また、 $\overline{U_R(m, L)}$ はレセプター R の残基 m についての“平均の”相互作用頻度である。“平均の”相互作用頻度とは、1つのレセプターに対して複数種類のリガンドをドッキングさせることによって作成される全ての複合体候補構造を利用して求めた相互作用頻度のことである。

このスコア S_R が大きくなるほど、リガンド L は他のリガンドと比べて、レセプターに対して特異的にドッキングしていると言える。そこで、スコア S_R を以後**相互作用特異度スコア**と呼ぶ。

スコアの正規化

相互作用特異度スコアはこのままではレセプタータンパク質のサイズが小さいほど1つの残基を使用する複合体候補構造の数が多くなりやすいことが考えられるため、大きさの異なるタンパク質間での比較に用いることができない。そこで Z -score を利用して相互

作用特異度スコアの正規化を行っている。なお、Z-score は次の式で表される。

$$Z\text{-score} = \frac{S_R - \mu}{\sigma}$$

ただし、 μ , σ はそれぞれ、1つのレセプター R の様々なリガンド L に対する相互作用特異度スコア S_R の平均と標準偏差であり、1つのレセプターの各リガンドに対する S_R の値を正規化している。以後、正規化した S_R の値を S_R^* と表すとする。

4.3 相互作用特異度スコアを用いた解析

相互作用予測結果によって、このタンパク質ペアの相互作用特異度スコアの値がどのように変化するかを調べる。

データセット

我々が以前に行った、Protein-protein docking benchmark 2.0¹⁶⁾ の44個のレセプターと44個のリガンドのペアに対する、 $44 \times 44 = 1,936$ 通りの MEGADOCK による網羅的な相互作用予測結果⁶⁾⁷⁾ に対して相互作用特異度スコアについての解析を行った。このときの相互作用予測結果は表1の通りである。なお、Precision, Recall, F-measure については以下の式で表されるものである。また、TP, FP, TN, FN はそれぞれ、True Positive, False Positive, True Negative, False Negative を意味する。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

表1 MEGADOCK による 44×44 の網羅的相互作用予測結果
Table 1 Results of PPI predictions 44×44 protein-protein pairs using MEGADOCK

TP (個数)	FP (個数)	TN (個数)	FN (個数)	Precision	Recall	F-measure
17	21	1871	27	0.45	0.39	0.41

解析手順

解析は以下の手順で行う。

- (1) 全てのレセプターとリガンドのペアについて、予測結果を元に TP, FP, TN, FN のカテゴリ分けを行う。
- (2) 各レセプターとリガンドのペアについて、レセプターの相互作用特異度スコアを求めて、各ラベルごとスコアの平均を求める。

- (3) 全てのタンパク質ペアの間でカテゴリごとに相互作用特異度スコア S_R^* の平均をとり、値を調べる。

以上の手順を、MEGADOCK ドッキングスコア、または ZRANK スコアの上位何%までの複合体候補構造を相互作用特異度スコアの計算に使うかを変化させて行う。なお、複合体候補構造の数は100%の時で6,000個である。

予測結果毎の値の比較

予測結果毎の値について、主に2つの傾向がみられた。第1に予測結果がTPである場合の値は、他の予測結果(FP, FN)の値に比べ大きい値を示している。第2に、相互作用特異度スコアの計算に使用する複合体候補構造の数を減らすほど、TPの場合の値が大きくなる。この2つの傾向は、複合体候補構造の数の削減の基準を MEGADOCK ドッキングスコア、ZRANK スコアのどちらにした場合でも確認することができる。しかしTPの値そのものは、ZRANK スコアを複合体候補構造の数の削減に用いたときの方が大きい値を示しており、また ZRANK スコアを用いるとTPほどではないものの、FNの値も他と比べて大きくなっている。すなわち ZRANK スコアを数の削減に用いた場合、本来相互作用するペアに関する相互作用特異度スコアの値が、相互作用しないペアのそれよりも大きい値を示すことが確認された。以上のことから、ZRANK スコアを数の削減の基準とした相互作用特異度スコアを用いた方が、より予測精度の改善に有効であると考え、次章の予測精度改善手法にはこちらの相互作用特異度スコアを利用する。

なお、解析を行った ZRANK スコアを用いたときの具体的な値は表2の通りである。またTNの値が常にほぼ0であるが、これは全てのタンパク質の組み合わせのうち、ほとんどがTNのラベルをもっており、その平均の値はZ-scoreの平均の値に近づくためと思われる。

表2 ZRANK スコアで複合体候補構造の数を削減したときのカテゴリごとのスコア S_R^* の平均
Table 2 Score average of each category with reduction by ZRANK pseudo energy

候補を使用した割合	TP	FP	TN	FN
100%	0.17	-0.02	0.00	-0.18
50%	0.11	-0.03	-0.00	-0.01
30%	0.17	-0.03	-0.00	0.09
10%	0.37	-0.04	-0.00	0.17
5%	0.52	-0.02	-0.01	0.11

5. 相互作用特異度スコアを用いた相互作用予測結果の改善

前節の解析結果から予測ラベルが TP のタンパク質ペアと FP のタンパク質ペアの間で相互作用特異度スコアに差があることが示された。そこで、このことを利用した MEGADOCK の PPI 予測結果を改善する手法を提案する。

提案する手法では、MEGADOCK によって相互作用予測を行ったタンパク質ペアに対して相互作用特異度スコアの値を求め、その値が閾値以下でかつ、“相互作用する”と予測されているタンパク質ペアは、“相互作用しない”と予測しなおすことで F-measure の値の改善を試みる。理想的には FN のペアを相互作用すると予測しなおすことも必要だが、表 2 から分かるように、FN の場合の S_R^* の値は TP に比べ差が大きくあるわけではないため、FP のペアを予測しなおすことのみを目的とした手法を提案する。

5.1 評価実験

実験を行うデータセットには引き続き Protein-protein docking benchmark 2.0 について MEGADOCK によって予測を行った結果を用いる。前節にて、ZRANK スコアに従って上位を選択し複合体候補構造の数を 10%以下に削減したときの相互作用特異度スコアは TP と FP の間で特に大きな差が表れている。使用する複合体構造の数を上位 10% まで減らしたときに、TP の場合の値が大きく上昇していることから、提案手法には ZRANK スコアで候補構造の数を上位 10% 以下に制限したときの相互作用特異度スコア S_R^* を用いるものとする。また S_R^* に対する閾値の値は、F-measure を最大にする最適な値を網羅的に探索する。

5.2 結果と考察

複合体構造構造の数と閾値を変化させた結果のうち、特に上位 10%まで複合体候補構造の数を削減したものと 5%まで削減したものをを用いたときの結果を図 6 に示した。上位 10%まで数の削減をした複合体候補構造を使ったとき、最大の F-measure = 0.43 を得た。この時の、TP, FP, TN, FN の数の内訳は表 3 の通りである。表 4 に示すように、F-measure の改善は 0.02 程度とごくわずかだが、TP の数を減らすことなく Precision の値を改善させており、FP のタンパク質ペアのみを相互作用しないと判定しなおすことができている。

F-measure を最大とする相互作用特異度スコア S_R^* の閾値は -0.9 前後であり、その後 F-measure が落ちていく傾向がみられる。これは FP の他に TP のタンパク質ペアも相互作用しないと判定されているためである。このことは、相互作用特異度スコアがある値以下では相互作用すると判定されたタンパク質ペアのうち FP のタンパク質ペアしか存在しない

ことを示唆している。

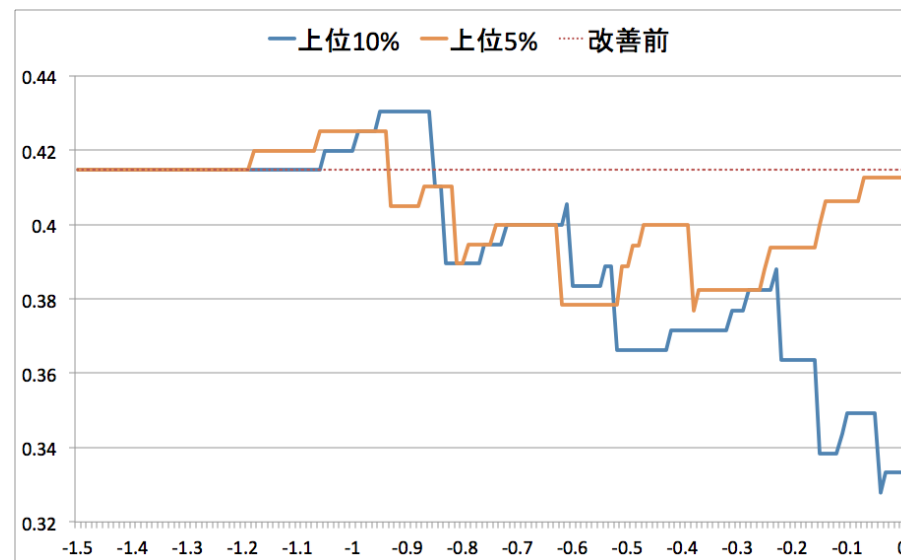


図 6 使用する複合体候補構造の数（上位 10%, 5%）と閾値に対する F-measure の変化
Fig.6 F-measure for each threshold

表 3 F-measure 最大の時の TP, FP, TN, FN の個数
Table 3 Number of TP, FP, TN and FN at the best F-measure

TP	FP	TN	FN
17	18	1874	27

6. おわりに

6.1 可視化手法

本研究では、PPI 予測精度を改善するための補助となるように、タンパク質ドッキング結果を可視化する手法を開発した。可視化手法は、タンパク質周囲にリガンドの分布をプロットする手法と、相互作用している残基に色をつける手法の 2 種類を提案した。タンパク

表 4 予測精度改善手法適用前後の F-measure, precision 及び recall の値
Table 4 A change of F-measure, precision and recall

	F-measure	Precision	Recall
改善前	0.41	0.45	0.39
改善後	0.43	0.49	0.39

質ドッキング計算結果から重要な情報を得る上で有効に働くことが期待される。

6.2 予測精度改善手法

実際に可視化から得た知見をもとに、タンパク質ペアの相互作用している残基の特異性をあらわす評価値“相互作用特異度スコア”を導入し、相互作用予測結果毎の相互作用特異度スコアの値の違いを確認した。その後、相互作用特異度スコアの値の差を利用した予測精度改善手法を提案し、わずかだが精度の向上を確認した。この手法は1つのデータセットにしか適用していないため、異なるデータセットに本手法を適用することで本稿で示した手法の汎用性を確認することが今後の課題に挙げられる。

謝 辞

本研究は、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」、および科学研究費補助金（基盤研究(B)19300102、特別研究員奨励費 23-8750）の支援を受けて行われたものである。

参 考 文 献

- 1) H.X. Zhou, Y. Shan: “Prediction of protein interaction sites from sequence profile and residue neighbor list”, *Proteins*, 44(3): 336–343, 2001.
- 2) Y. Ofra, B. Rost: “Predicted protein-protein interaction sites from local sequence information”, *Federation of European Biochemical Societies Letters*, 544(1-3): 236–239, 2003.
- 3) A. Koike, T. Takagi: “Prediction of protein-protein interaction sites using support vector machines”, *Protein Engineering, Design & Selection*, 17(2): 165–173, 2004.
- 4) M. Deng, S. Mehta, F. Sun: “Inferring domain-domain interactions from protein-protein interactions”, *Genome Research*, 12: 1540–1548, 2002.
- 5) R. Jansen, H. Yu, D. Greenbaum, Y. Kluger, N.J. Krogan, S. Chung, A. Emili, M. Snyder, J.F. Greenblatt, M. Gerstein, “A bayesian networks approach for predicting protein-protein interactions from genomic data”, *Science*, 302(5644): 449–453, 2004

- 6) Y. Akiyama, T. Sato, Y. Matsuzaki, Y. Matsuzaki: “MEGADOCK - A rapid screening system for all-to-all protein docking analysis with pre-calculated Fourier library of protein structures”, *Proceedings of the 2008 Annual Conference of the Japanese Society for Bioinformatics*: P032, 2008.
- 7) 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.
- 8) E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser: “Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques”, *Proceedings of the National Academy of Sciences of the United States of America*, 89(6): 2195–2199, 1992.
- 9) H.A. Gabb, R.M. Jackson, M.J.E. Sternberg: “Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information”, *J Mol Biol*, 272: 106–120, 1997.
- 10) R. Chen, Z. Weng: “Docking Unbound Proteins Using Shape Complementarity, Desolvation, and Electrostatics”, *Proteins*, 47(3): 281–294, 2002.
- 11) R. Chen, L. Li, Z. Weng: “ZDOCK: An Initial-stage Protein-Docking Algorithm”, *Proteins*, 52(1): 80–87, 2003.
- 12) R. Chen, Z. Weng: “A Novel Shape Complementarity Scoring Function for Protein-Protein Docking”, *Proteins*, 51(3): 397–408, 2003.
- 13) Y. Matsuzaki, Y. Matsuzaki, T. Sato, Y. Akiyama: “Development of post-docking system for protein-protein interaction prediction”, *1st Joint Workshop on Computational Science*, Saitama, Japan, 2008.
- 14) Y. Matsuzaki, Y. Matsuzaki, T. Sato, Y. Akiyama: “In silico screening of protein-protein interactions with all-to-all rigid docking and clustering: an application to pathway analysis”, *Journal of Bioinformatics and Computational Biology*, 7(6): 991–1012, 2009.
- 15) B. Pierce, Z. Weng: “ZRANK: Reranking Protein Docking Predictions with an Optimized Energy Function”, *Proteins*, 67(4): 1078–1086, 2007.
- 16) J. Mintseris, K. Wiehe, B. Pierce, R. Anderson, R. Chen, J. Janin, Z. Weng: “Protein-Protein Docking Benchmark 2.0: an update”, *Proteins*, 60(2): 214–216, 2005.