Vol.2010-BIO-20 No.1
2010/3/4

立体構造情報と機能情報による タンパク質間相互作用予測法の改良

吉川達也[†1] 瀬尾茂人[†1]
竹中要一[†1] 松田秀雄[†1]

相互作用するタンパク質ペアを高精度に識別するために，タンパク質間ドッキングの評価結果に基づいた相互作用予測法を提案する．従来研究では，相互作用に必要な親和性の強さはタンパク質の機能と無関係であると仮定していた．ところが，この親和性の強さは，剛体ドッキングやフレキシブルドッキングのようなドッキング様式によって異なり，この様式はタンパク質の機能に関係していることが分かってきた．本研究では，タンパク質の機能情報に基づいてサンプリングした親和性スコア分布を統計的に解析する新たなスコアリング手法を提案する．すべてのタンパク質ペアを用いて評価した結果，提案手法は予測精度を改善し，提案手法を組み合わせて用いた場合に，さらに高い精度を達成することが検証された．これにより，タンパク質の機能情報を考慮することで偽陽性の発生を抑制して相互作用を予測できることが示された．

Tatsuya Yoshikawa,[†1] Shigeto Seno,[†1]
Yoichi Takenaka[†1] and Hideo Matsuda[†1]

To identify protein–protein interaction pairs with high accuracy, we propose a method for predicting these interactions based on characteristics obtained from protein–protein docking evaluations. Previous studies assumed that the required protein affinity strength for an interaction was not dependent on protein functions. However, the protein affinity strength appears to differ with different docking schemes, such as rigid-body or flexible docking, and these schemes may be related to protein functions. Thus, we propose a new scoring system that is based on statistical analysis of affinity score distributions sampled by their protein functions. As a result, of all possible protein pair combinations, a newly developed method improved prediction accuracy of F-measures. By combining two proposed scoring systems, Receptor-Focused Z-scoring and Ligand-Focused Z-scoring, further improvement was achieved. This result suggested that the proposed prediction method improved the prediction accuracy, with few false positives, by taking biological functions of protein pairs into consideration.

1. Introduction

Most biological functions involve interactions between several proteins in a cell. Therefore, it is important to elucidate biological phenomena, including cell signaling, enzyme reactions, and gene expression regulation, by analysis of protein–protein interactions (PPIs). In the past, several reviews[1)–3)] have examined the interaction characteristics of known protein complexes and PPI maps, which can play a role in discovering protein partners. Several methods have been developed in computational studies for predicting PPIs using genomic information[4),5)]. Recently, research on PPIs has focused on not only ascertaining their roles in living organisms but also applying this knowledge to medicinal fields such as drug design. In essence, the interactive properties associated with protein conformation need to be thoroughly investigated if we are to analyze the relationships between protein structure and function. Ideally, PPI studies for drug design should be based on protein structural information. Structure-based PPI study includes protein–protein docking. Most previous studies on docking have attempted to solve protein-docking problems; the ultimate objective of these studies was to accurately predict the structures of protein complexes from three-dimensional (3D) structures of individual proteins.

In this study, given that protein docking has the potential to decide whether or not a complex actually occurs in nature and for measuring its affinity, we focused on the 3D structures of proteins in order to calculate PPIs via protein–protein docking. This kind of PPI study has been previously discussed by Smith et al.[6)], although they did not clearly demonstrate the functioning of an actual system or analysis of results. Recent studies[7)–10)] involved in predicting PPIs on the basis of shape complementarity docking succeeded in up to 23 of 84 predictions. Sacquin-Mora et al.[11)] successfully predicted 7 out of 10 interaction partners using weighted interaction energies. However, methods for determining protein affinity of only one complex pair have been confined to

†1 大阪大学大学院 情報科学研究科
Graduate School of Information Science and Technology, Osaka University

Vol.2010-BIO-20 No.1
2010/3/4

forecasting a local binding likelihood for each protein–protein pair, therefore have not been considered that the required strength of binding is probably different by biological functions.

In order to solve this problem, we focused on assessing protein interactions by globally considering affinities of one protein with other proteins. In this study, we propose a docking-based prediction method for PPIs by using all possible protein pair combinations. The aim of this study is to improve the prediction accuracy of PPIs.

2. Protein–Protein Interaction Problem

2.1 Protein–Protein Interaction Prediction Method

Most of the previous studies in computational approach developed the genome-scale techniques, whereas recently the structure-based studies are considered as an effective mean of applying to drug design. This approach is based on predicting PPIs by protein–protein docking. This method consists of three procedures that are outlined in **Fig. 1**. 3D structures of a receptor and ligand are the input data flow, and the predicted result of the interaction is the output. The squares with "Protein Docking," "Affinity Evaluation," and "Interaction Prediction" are the key procedures, and each input/output port is shown in parentheses.

2.1.1 Protein–Protein Docking

Protein–protein docking is performed for calculating the 3D structure of a protein complex, starting from individual structures of constituent proteins. That is, the general aim of this study is to predict the near-native complex structure of two proteins, which is different from that of a PPI prediction problem that determines whether or not the proteins interact. The reason for considering protein docking as a procedure in a PPI prediction problem is based on the following assumption. It is assumed that protein affinity plays a role in deciding whether or not proteins interact only when their binding affinity can be calculated accurately. That is, if the affinity of proteins is equal to or more than a certain threshold, then they can be computationally regarded as a protein pair that interacts. Based on this assumption, PPI prediction can be realized by statistically analyzing docking scores from protein docking programs.

A rigid-body model for expressing molecules is exhaustively screened in a six-

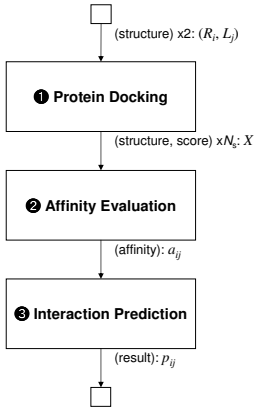

Fig. 1 Flowchart for PPI prediction by protein–protein docking.

dimensional rotation and translation space. The procedure starts by rotating a ligand as a probe protein. When the rotational angular step Δ is equal to the widely used 15 degrees, λ (=3600) poses are sampled because of rotational symmetry. A target protein (receptor) and 3600 probe proteins (ligands) are discretized into 3D grids n^3 by a certain grid pitch γ (e.g., 1.2 Å), where n is the number of grid points in each coordinate. Each grid point is assigned to a structural and chemical property value based on geometric characteristics (i.e., core, surface and cavity areas) and free energies, respectively. When the rotated ligand is translated with respect to the receptor, the docking algorithm calculates the product sum of assigned property values, which is referred to as the docking score (s). Given the grid size of a receptor n^3 and the number of sampled ligand poses λ, λn^3 docking scores can be obtained by an exhaustive search for only one protein pair. The top-ranked N_s ($N_\mathrm{s} \geq 1$) scores are generally sampled as candidates because of a large number of docking results. The results $X = \{x_k \mid 1 \leq k \leq N_\mathrm{s}\}$ include not only the docking scores s_k but also the structural information of complex candidates:

$$x_k = \left(s_k, t_k\left(t_x, t_y, t_z\right), r_k\left(r_\theta, r_\phi, r_\psi\right)\right), \tag{1}$$

where $t_k(t_x, t_y, t_z)$ and $r_k(r_\theta, r_\phi, r_\psi)$ are the translational distance and the rotational

angle of the ligand, respectively.

2.1.2 Affinity Evaluation

Based on the docking results, affinity evaluation plays a role in calculating the binding likelihood. The aim is to assess how strong the interactions are. The simplest way is to utilize the maximum value among N_s docking scores as protein affinity such that

$$a = max\big(s_k\big)_{k=1}^{N_s}. \tag{2}$$

An alternative is to use the statistical characteristics provided by clustering the docking results according to the score or structural similarities between candidates.

2.1.3 PPI Prediction

Interaction prediction makes the final decision as to whether or not proteins interact. The primitive threshold-based approach is used to determine the interactions as follows:

$$p = \begin{cases} 1 & \text{if } a \geq \tau \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where p is the prediction result that includes the Boolean values (i.e., 1 and 0 indicating positive and negative PPI, respectively), and τ is the affinity threshold for deciding whether or not proteins interact. Here, the threshold τ has to be decided using biological knowledge or statistical characteristics of affinity values, etc.

2.2 Previous Studies

2.2.1 Outline of ZDOCK

A previous method for PPI prediction, which we have designated as ZDOCK hereafter, uses docking scores from the protein–protein docking program ZDOCK 3.0.1[12]. This tentative method was used for comparing the prediction method in Section 2.2.2. The ZDOCK 3.0.1 program can assess structural and chemical complementarity between proteins. It enables us to find binding sites and complex structures using a FFT-based search algorithm with a scoring function that is based on pairwise shape complementarity, electrostatics, and explicit interface atomic contact energies. In affinity evaluation, the maximum score from only one docking simulation among 2000 docking scores was used to simply assess the affinity of a protein pair. The setting values defined in Section 2.1.1 are as follows: $\Delta = 15$, $\gamma = 1.2$, and $N_s = 2000$, where Δ is the rotational angular step, γ is the grid pitch, and N_s is the number of samplings for candidates. Here, when applying the ZDOCK 3.0.1 program execution option, all default values (i.e., –N (=2000), –S (=no at randomization), and –D (=none)) were used. The threshold τ was determined to maximize the F-measure by ROC analysis described in Section 3.2.3.

2.2.2 Outline of Affinity Evaluation and Prediction (AEP)

A previous study[10] had predicted the interactions by assessing the statistical significance of binding likelihood based on shape complementarity characteristics between protein pairs: *affinity evaluation and prediction* (AEP). In their protein–protein docking procedure, the original docking program with a scoring function of pair-wise shape complementarity was developed.

Here, setting values were $\Delta = 15$, $\gamma = 1.2$, and $N_s = 512$. In the affinity evaluation procedure, the statistical characteristics of N_s candidates were used to assess protein affinity. The docking scores were classified into several clusters according to the structural similarities between candidates. Protein affinity in previous studies was calculated by the distribution of docking scores of a representative of each cluster and that of the cluster density. Key parameter settings were also important for determining structural similarities and extracting cluster characteristics in this procedure. In the PPI prediction procedure, the affinity threshold τ was obtained by the receiver operating characteristics (ROC) procedure. The values of key parameters in affinity evaluation and τ are optimized so that the F-measure is maximized.

Using the above procedures, the previous study successfully predicted 23 interaction pairs out of 84. The study assumed that the required protein affinity strength for the interaction was not dependent on the function of proteins. However, the protein affinity strength appears to differ by different docking schemes (i.e., rigid-body or flexible docking), and the scheme may be related to protein functions.

2.2.3 Protein-Pair Data Set

In order to evaluate the performances of ZDOCK and AEP, 168 bound proteins derived from 84 co-crystallized complex structures by Protein–Protein Docking Benchmark 2.0[13),14)] were used. All proteins were classified as either receptors (R) or ligands (L) according to Weng's dfinition, which resulted in 84 proteins of each type. When the receptor and ligand were derived from the same complex, the receptor molecule was

Table 1 Protein-pair data set consisting of 84 receptors and 84 ligands.

PDB ID (Index Number)
Antibody–Antigen (A)
1AHW(1), 1BGX(2), 1BVK(3), 1DQJ(4), 1E6J(5), 1JPS(6), 1MLC(7), 1VFB(8) 1WEJ(9), 2VIS(10)
Bound Antibody–Antigen (AB)
1BJ1(11), 1FSK(12), 1I9R(13), 1IQD(14), 1K4C(15), 1KXQ(16), 1NCA(17), 1NSN(18) 1QFW(19), 2HMI(20), 2JEL(21), 2QFW(22)
Enzyme–Inhibitor or Substrate (E)
1ACB(23), 1AVX(24), 1AY7(25), 1BVN(26), 1CGI(27), 1D6R(28), 1DFJ(29), 1E6E(30) 1EAW(31), 1EWY(32), 1EZU(33), 1F34(34), 1HIA(35), 1KKL(36), 1MAH(37), 1PPE(38) 1TMQ(39), 1UDI(40), 2MTA(41), 2PCC(42), 2SIC(43), 2SNI(44), 7CEI(45)
Others (O)
1A2K(46), 1AK4(47), 1AKJ(48), 1ATN(49), 1B6C(50), 1BUH(51), 1DE4(52), 1E96(53) 1EER(54), 1F51(55), 1FAK(56), 1FC2(57), 1FQ1(58), 1FQJ(59), 1GCQ(60), 1GHQ(61) 1GP2(62), 1GRN(63), 1H1V(64), 1HE1(65), 1HE8(66), 1I2M(67), 1I4D(68), 1IB1(69) 1IBR(70), 1IJK(71), 1K5D(72), 1KAC(73), 1KLU(74), 1KTZ(75), 1KXP(76), 1M10(77) 1ML0(78), 1N2C(79), 1QA9(80), 1RLB(81), 1SBB(82), 1WQ1(83), 2BTF(84)

always larger than the ligand. Here, these 84 complexes were classified as follows: 10 pairs of antibody–antigen (functional category A), 12 pairs of bound antibody–antigen (AB), 23 pairs of enzyme–inhibitor or substrate (E), and others (O). Each complex was assigned with both a protein data bank (PDB) ID and an index from 1 to 84.

To evaluate the prediction method more exhaustively, N_{p} ($N_{\mathrm{R}} \times N_{\mathrm{L}} = 7056$) possible pairs were constructed by combining N_{R} (=84) receptors and N_{L} (=84) ligands, where N_{p} is the total number of receptor–ligand pairs, and N_{R} and N_{L} are the numbers of receptors and ligands, respectively. Therefore, the data set included 84 pairs that have been previously identified experimentally as forming complexes and 6972 others. Because receptor–receptor or ligand–ligand protein interactions are believed not to occur under normal biological conditions, we only employed receptor–ligand pairs. Here, in order to directly compare our method with previous results, the bound structures of proteins were used.

2.3 Performance Measures

All protein pairs in the data set were classified into either 84 pairs whose interactions have been previously experimentally detected or 6972 others whose interactions have not been detected. Using the binary values of 1 or 0, the prediction results suggested whether or not proteins interacted, indicating positive or negative for PPIs. **Table 2** shows the logical combinations of experimental interactions and prediction results as a

Table 2 Logical combinations of experimental interactions and prediction results.

		Experimental interaction	
		Detected	Not detected
Prediction result	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

2×2 contingency table. The four logical combinations (i.e., TP, FN, FP, and TN) are defined in Table 2, and these numbers are represented as tp, fn, fp, and tn, respectively.

In general, performance measures for information retrieval are used for assessing the prediction accuracy of a binary classification problem. Many of the different measures, such as sensitivity ($sens$), recall (rec), precision ($prec$), and F-measure (F), are given by

$$rec(= sens) = \frac{tp}{tp + fn}, \quad prec = \frac{tp}{tp + fp}, \quad F = \frac{2 \cdot prec \cdot rec}{prec + rec}. \tag{4}$$

F-measure is defined as the harmonic mean of recall and precision. Its value increases significantly as the values of both these factors increase. Considering the attributes of the used criteria, we employed the F-measure to assess prediction results. Because the F-measure can quantitatively gauge the prediction accuracy relative to the prevalence of a problem, we can evaluate recall and precision as trade-off in the form of a combined value.

3. Method

3.1 Overview

In order to solve the problem of the previous study, we focused on the docking scores of functionally classified pairs. This is because when distributions of scores of pairs with the same functions were checked, the following findings can be revealed: (1) potential protein affinities exist that represent protein bindings, and (2) affinity thresholds were different between particular functionally classified pairs, which determine whether or not proteins interact. **Figure 2** shows a flowchart of the proposed prediction method. The method consists of four procedures, including three key procedures, as shown in Fig. 1, and an additional procedure by ROC analysis. In the ROC procedure, the optimal affinity threshold τ is determined by ROC analysis. Although the flow outline is similar to that of a previous prediction method (i.e., AEP), repeated procedures are

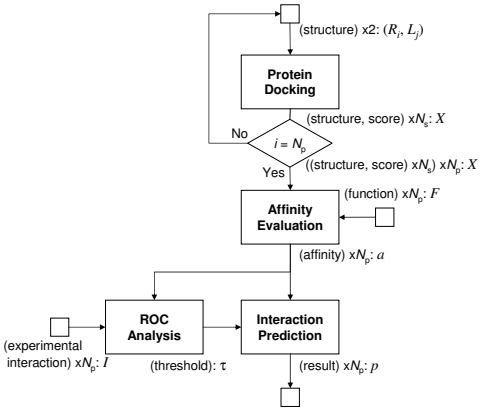

Fig. 2 Flowchart of the proposed prediction method.

quite different. Compared with "Protein Docking" and "Affinity Evaluation" procedures in the previous method, only the "Protein Docking" procedure is repeated N_{p} (i.e., the number of all protein pairs) times. Since the proposed protein affinity is defined not only by docking scores of one pair but also by that of others, the iteration in the "Protein Docking" procedure needs to be completed before the "Affinity Evaluation" procedure. In addition, calculation methods in each procedure and score definitions are different.

3.2 Algorithm

Input and output of the proposed method are as follows:

Input:

3D structures and functions of all receptor–ligand pairs,

Output:

prediction results of computational protein–protein interactions.

The method consists of four steps, as outlined in **Fig. 2**.

Step 1 (Protein Docking):

Search the structures of complex candidates with docking scores using the 3D structure of one receptor–ligand pair. Step 1 is repeated for all protein pairs.

Step 2 (Affinity Evaluation):

Evaluate protein affinities by statistically inspecting docking scores of pairs using functional information.

Step 3 (ROC Analysis):

By ROC analysis, detect the optimal threshold τ for determining whether or not protein pairs interact.

Step 4 (Interaction Prediction):

Predict computational interactions from protein affinities and functions of pairs.

In order to describe the method, a protein is modeled according to the computational information that represents the 3D structure and biological function, where the protein structure includes the 3D coordinates of its constituent atoms, and biological functions are as follows:

$$F = \{\text{“}A\text{”}, \text{“}AB\text{”}, \text{“}E\text{”}, \text{“}O\text{”}\}. \tag{5}$$

In addition, the experimental interaction of each receptor–ligand pair for determining the threshold τ and measuring the performance of the prediction method was previously detected by

$$I = \begin{cases} 1 & \text{if } \textit{experimental interaction is detected} \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

3.2.1 Protein–Protein Docking (Protein Docking, Step 1)

In Step 1, we use the protein–protein docking program ZDOCK 3.0.1 to assess structural and chemical complementarity between proteins. As described in Section 2.2.1, ZDOCK 3.0.1 enables us to find binding sites and complex structures using a FFT-based search algorithm with a scoring function based on pair-wise shape complementarity, electrostatics, and explicit interface atomic contact energies. The setting values defined in Section 2.1.1 are as follows: $\Delta = 15$, $\gamma = 1.2$, and $N_{\mathrm{s}} = 2000$. Then, all default values (i.e., –N (=2000), –S (=no at randomization), and –D (=none)) of execution options are used. This procedure outputs the top N_{s} scores for each docking score s_{ij} of a receptor–ligand pair (R_i, L_j). By repeating Step 1, docking scores of all N_{p} (= $N_{\mathrm{R}} \times N_{\mathrm{L}}$) pairs are calculated.

3.2.2 Affinity Evaluation (Step 2)

In Step 2, protein affinities are assessed statistically based on functionally classified

docking scores. In order to estimate the affinity of a target pair, docking scores of all pairs having the same function are considered. We propose the following Z-scoring systems: *Receptor-Focused Z-scoring* (RFZ) and *Ligand-Focused Z-scoring* (LFZ). The aim of this method was to statistically evaluate each receptor- or ligand-focused group of docking scores and then convert the docking score into an affinity based on the Z-score. That is, RFZ collects those ligands that have scores similar to a specific receptor, while LFZ collects those receptors that have scores similar to a specific ligand. Therefore, either receptor or ligand functions are used for affinity evaluation. When RFZ evaluates protein affinity based on docking score distribution of receptors having the same functions, functional information of ligands is not used. Since the key steps for RFZ and LFZ are the same, the details for only RFZ are shown below.

Input and output of the procedure are as follows:

Input:

docking scores,

Output:

protein affinities.

The method consists of three steps.

Step 2.1 (Preparation of docking scores):

Calculate the maximum score among $N_{\rm s}$ scores for each receptor and ligand pair and obtain a set of docking scores $\{s_{ij} \mid R_i \in R,\ L_j \in L\}$ for assessing protein affinity.

Step 2.2 (Grouping of docking scores):

In this step, the available information for grouping the docking scores depends on receptor–ligand pairs. In this study, the following cases can be considered: (1) both structural and functional information of the confirmed pairs having same functions, and (2) only the structural information of all pairs consisted of proteins with varying functions.

All $N_{\rm p}$ docking scores are classified according to pair functions. For example, a set of receptors in functional category A (i.e., antibody–antigen) is given by:

$$R^{\rm A} = \{R_i \mid R_i \in R,\ F(R_i) = \text{“}A\text{”}\}, \tag{7}$$

where $F(R_i)$ denotes the biological function of receptor R_i.

Next, $N_{\rm RA} \times N_{\rm L}$ docking scores are split into $N_{\rm RA}$ subset of scores of pairs that have the same receptor such that:

$$s_i^{\rm A} = \{s_{ij} \mid L_j \in L\},\ R_i \in R^{\rm A}, \tag{8}$$

where $N_{\rm RA}$ denotes the number of receptors with the function A among R. Similarly, for other functions AB, E, and O, other subsets $s_i^{\rm AB}$, $s_i^{\rm E}$, and $s_i^{\rm O}$ are obtained from s_{ij}.

Step 2.3 (Z-score calculation):

In this step, the Z-score of each score subset is calculated as follows:

$$a_{ij}^{\rm A} = \frac{s_{ij} - E(s_i^{\rm A})}{\sigma(s_i^{\rm A})}, \tag{9}$$

where $E(s_i^{\rm A})$ and $\sigma(s_i^{\rm A})$ denotes the mean and standard deviation of scores $s_i^{\rm A}$, respectively. Thus, the affinities of all pairs, a_{ij}, are determined.

3.2.3 Receiver operating characteristic analysis (ROC Analysis, Step 3)

In Step 3, the affinity threshold τ is decided to maximize the F-measure by ROC analysis. When the optimal τ value is used for determinating PPIs, many prediction results of the 84 protein pairs (R_i, L_j) with experimentally validated interactions (i.e., $I_{ij} = 1$) are correctly evaluated as positives for PPIs (i.e., $p_{ij} = 1$), and the other 6972 pairs whose interactions have not been experimentally detected (i.e., $I_{ij} = 0$) are correctly predicted as negatives (i.e., $p_{ij} = 0$). In ROC analysis, a cut-off value is generally used as the criterion for separating the two classes (i.e., positives and negatives for PPIs). When the cut-off value is changed by a certain incremental amount from minimum to maximum of all affinities (a), recall and precision values for each cut-off value are obtained. Therefore, the F-measure is obtained from recall and precision values. Although there are various ways of obtaining the optimal threshold, we employed a method based on the balance of recall and precision, as the objective of this study is to maximize the F-measure. The threshold τ was determined so that the value of $\{\text{recall}^2 + (1 - \text{precision})^2\}$ becomes maximum.

The values of τ vary with prediction methods and functional categories. That is, $\tau_{\rm RFZ}^{\rm A}$ for predicting the interactions of pairs in functional category A by RFZ is different from $\tau_{\rm RFZ}^{\rm AB}$ for predicting AB pairs by the same method. In addition, $\tau_{\rm RFZ}^{\rm A}$ is not equal to $\tau_{\rm LFZ}^{\rm A}$ in spite of same functional categories. When the prediction method uses a combination of RFZ and LFZ, called RFZ×LFZ and RFZ+LFZ (see Section 3.2.4 for

the definitions of RFZ×LFZ and RFZ+LFZ), a threshold $\tau^{\mathrm{A}}_{\mathrm{RFZ\times LFZ}}$ for protein pairs in A from RFZ×LFZ shows a pair of thresholds for both RFZ and LFZ such as ($\tau^{\mathrm{A}}_{\mathrm{RFZ}}$, $\tau^{\mathrm{A}}_{\mathrm{LFZ}}$).

In the previous method, the entire protein-pair data set, including 84 target protein pairs, was used for determining the threshold τ by ROC analysis. In order to directly compare the previous results, the entire data set was also used in this study.

3.2.4 Interaction Prediction (Step 4)

In Step 4, the interactions of all receptor–ligand pairs are finally predicted. Protein affinities are assessed by comparisons with the optimal threshold τ. If the affinity value is more than or equal to τ, the prediction result p_{ij} is set to 1, indicating "positive;" Otherwise, the value is set to 0, indicating "negative," such as

$$p_{ij} = \begin{cases} 1 & \text{if } a_{ij} \geq \tau \\ 0 & \text{otherwise} \end{cases} . \tag{10}$$

Here, we propose additional scoring systems, RFZ×LFZ and RFZ+LFZ, which predict the interactions by combining the results of RFZ and LFZ. That is, the prediction result of each protein pair, $p^{\mathrm{RFZ\times LFZ}}_{ij}$ is defined by the logical AND operation of p^{RFZ}_{ij} and p^{LFZ}_{ij}, while $p^{\mathrm{RFZ+LFZ}}_{ij}$ is defined by the logical OR operation.

4. Results and Discussion

For comparison with previous methods, we employed the protein-pair data set in Section 2.2.3. At the first evaluation for improving prediction accuracy using the proposed method, the entire data set ($84 \times 84 = 7056$ protein pairs) was used. **Table 3** shows comparison of prediction accuracy for the proposed scoring systems (RFZ and LFZ) and previous methods (AEP and ZDOCK). AEP indicates the previous method detailed in Section 2.2.2, and the best performance of AEP is given by optimizing key parameters, as shown in Table 3. ZDOCK is a simple PPI prediction method that directly uses ZDOCK 3.0.1 for protein-protein affinity calculations, as described in Section 2.2.1. The proposed RFZ and LFZ greatly improved prediction accuracy, with the obtained F-measures higher than those obtained by previous methods. This improvement arose as the proposed method could reduce many false positives; the number of false positives obtained by RFZ was only 19 among all 6972 negative examples, compared with 629 by AEP. As a result, using only structural information of proteins, RFZ and LFZ correctly predicted the interactions of 17 and 14 pairs among 84, respectively.

Table 3 Comparison of proposed methods, RFZ and LFZ, and previous methods.

Method	Performance measures						
	F-measure	Recall	Precision	*tp*	*fp*	*fn*	*tn*
RFZ (τ = 2.984)	28.3	20.2	47.2	17	19	67	6953
LFZ (τ = 3.103)	21.9	16.7	31.8	14	30	70	6942
AEP (τ = 4.520)	6.3	27.4	3.5	23	629	61	6343
ZDOCK (τ = 1364.018)	4.0	73.8	2.1	62	2941	22	4031

Table 4 Comparison of proposed methods, RFZ and LFZ, and previous methods in respective functional categories.

Method	Performance measures						
	F-measure	Recall	Precision	*tp*	*fp*	*fn*	*tn*
(a) Antibody–Antigen (10 × 10)							
RFZ (τ = 2.001)	33.3	20.0	100.0	2	0	8	90
LFZ (τ = 0.445)	24.2	40.0	17.4	4	19	6	71
AEP (τ = 4.630)	33.3	30.0	37.5	3	5	7	85
ZDOCK (τ = 1489.529)	23.7	90.0	13.6	9	57	1	33
(b) Bound Antibody–Antigen (12 × 12)							
RFZ (τ = 1.751)	47.1	33.3	80.0	4	1	8	131
LFZ (τ = 0.949)	31.6	50.0	23.1	6	20	6	112
AEP (τ = 4.590)	22.2	25.0	20.0	3	12	9	120
ZDOCK (τ = 1831.057)	29.8	58.3	20.0	7	28	5	104
(c) Enzyme–Inhibitor or Substrate (23 × 23)							
RFZ (τ = 1.536)	36.7	47.8	29.7	11	26	12	480
LFZ (τ = 1.751)	37.2	34.8	40.0	8	12	15	494
AEP (τ = 4.530)	20.9	30.4	15.9	7	37	16	469
ZDOCK (τ = 1320.669)	16.2	73.9	9.1	17	170	6	336
(d) Others (39 × 39)							
RFZ (τ = 3.131)	40.0	25.6	90.9	10	1	29	1481
LFZ (τ = 3.483)	39.2	25.6	83.3	10	2	29	1480
AEP (τ = 3.970)	7.1	66.7	3.8	26	665	13	817
ZDOCK (τ = 1359.312)	10.9	74.4	5.9	29	463	10	1019

For predictions using both structural and functional information of proteins, **Table 4** summarizes the comparison results of prediction methods in each functional category. The previous methods, ZDOCK and AEP, indicate prediction results only using protein pairs having same functions. Each threshold τ was determined so as to maximize the F-measure by ROC analysis. As shown in Table 4, both proposed scoring systems improved the performance in F-measures in almost all cases compared to that by previous methods. Only in the case of LFZ, the 10 × 10 subset of functional category A

was less than that of AEP (Table 4(a)). As shown in Table 4, LFZ gave a prediction accuracy of 24.2% with an F-measure <33.3% of AEP because of the increase from 5 false positives by AEP to 19 by LFZ, whereas the number of true positives was slightly increased by one. In contrast, RFZ provided maximum performance in each category. The considerable improvement by RFZ was owing to the decrease of many false positives. This is supported by the fact that precision was greater than recall in three of four categories. In addition, although there was naturally a trade-off between recall and precision, LFZ in the AB category achieved the best recall (=sensitivity) that was 50.0% more than AEP (Table 4(b)). These results indicate that the proposed scoring system could provide a significant improvement in prediction accuracy using both structural and functional information of proteins.

5. Conclusions

We have proposed a method for predicting protein interactions from the docking scores of protein pairs that have same functions. We developed PPI prediction scoring systems, RFZ and LFZ, to statistically evaluate separate receptor- or ligand-focused groups of docking scores and convert the docking score into protein affinity based on the Z-score. The proposed method assessed the improvement in prediction accuracy using a protein-pair data set. By an analysis of biological functions of protein pairs, the prediction accuracy is significantly improved without changing the prediction algorithm itself. Further refinement of prediction accuracy was achieved by combining RFZ and LFZ.

In the near future, we plan to study the identification of protein functions using the existing docking programs with various scoring functions. We also aim at extending the new findings for application to related research such as drug design.

References

1) Jones, S. and Thornton, J. M.: Principles of protein-protein interactions, *Proc. Natl. Acad. Sci. U. S. A.*, Vol.93, No.1, pp.13–20 (1996).
2) Lo Conte, L., Chothia, C. and Janin, J.: The atomic structure of protein-protein recognition sites, *J. Mol. Biol.*, Vol.285, No.5, pp.2177–2198 (1999).
3) Norel, R., Petrey, D., Wolfson, H.J. and Nussinov, R.: Examination of shape complementarity in docking of unbound proteins, *Proteins*, Vol.36, No.3, pp.307–317 (1999).
4) Galperin, M. Y. and Koonin, E. V.: Who's your neighbor? New computational approaches for functional genomics, *Nat. Biotechnol.*, Vol. 18, No. 6, pp. 609–613 (2000).
5) Huynen, M., Snel, B., Lathe, W. 3rd, and Bork, P.: Predicting protein function by genomic context: quantitative evaluation and qualitative inferences, *Genome Res*, Vol.10, No.8, pp.1204–1210 (2000).
6) Smith, G.R. and Sternberg, M. J.: Prediction of protein-protein interactions by docking methods, *Curr. Opin. Struct. Biol.*, Vol.12, No.1, pp.28–35 (2002).
7) Tsukamoto, K., Yoshikawa, T., Hourai, Y., Fukui, K. and Akiyama, Y.: Development of an affinity evaluation and prediction system by using the shape complementarity characteristic between proteins, *J. Bioinf. Comput. Biol.*, Vol.6, No.6, pp.1133–1156 (2008).
8) Yoshikawa, T., Tsukamoto, K., Hourai, Y. and Fukui, K.: Parameter tuning and evaluation of an affinity prediction using protein-protein docking, *MMACTEE '08: Proc. 10th WSEAS International Conference on Mathematical Methods and Computational Techniques in Electrical Engineering*, World Scientific and Engineering Academy and Society (WSEAS), pp.312–317 (2008).
9) Tsukamoto, K., Yoshikawa, T., Yokota, K., Hourai, Y. and Fukui, K.: The development of an affinity evaluation and prediction system by using protein-protein docking simulations and parameter tuning, *Adv. Appl. Bioinf. Chem.*, Vol. 2, pp. 1–15 (2009).
10) Yoshikawa, T., Tsukamoto, K., Hourai, Y. and Fukui, K.: Improving the accuracy of an affinity prediction method by using statistics on shape complementarity between proteins, *J. Chem. Inf. Model.*, Vol.49, No.3, pp.693–703 (2009).
11) Sacquin-Mora, S., Carbone, A. and Lavery, R.: Identification of protein interaction partners and protein-protein interaction sites, *J. Mol. Biol.*, Vol.382, No.5, pp. 1276–1289 (2008).
12) Mintseris, J., Pierce, B., Wiehe, K., Anderson, R., Chen, R. and Weng, Z.: Integrating statistical pair potentials into protein complex prediction, *Proteins*, Vol.69, No.3, pp.511–520 (2007).
13) Chen, R., Mintseris, J., Janin, J. and Weng, Z.: A protein-protein docking benchmark, *Proteins*, Vol.52, No.1, pp.88–91 (2003).
14) Mintseris, J., Wiehe, K., Pierce, B., Anderson, R., Chen, R., Janin, J. and Weng, Z.: Protein-Protein Docking Benchmark 2.0: an update, *Proteins*, Vol.60, No.2, pp. 214–216 (2005).