A Protein-Protein Interaction Prediction Method Embracing Intra-Protein Domain Cohesion Information

Woo-Hyuk Jang Suk Hoon Jung Bo-ra Hyun Dong-Soo Han*
Department of Information and Communications Engineering, ICC, Korea Advanced Institute of Science and Technology, 119, Munjiro, Yuseong-gu, Daejeon, 305-714, Korea.
torajim@kaist.ac.kr sh.jung@kaist.ac.kr hyunbr@kaist.ac.kr dshan@kaist.ac.kr

Abstract

Recently, many computational methods for predicting protein-protein interaction (PPI) have been developed by utilizing domain-domain interaction or associated information. However, most of the methods lack of reflecting the collaboration effect of multiple domains to the prediction of PPI. In this paper, we develop a computational model that considers not only inter relationship between protein pair but also the intra-domain functional cohesion effect in PPI. In the computational model, a value assigning method to reflect the intra and inter collaboration devised and the computed values are stored in Interaction Significance (IS) matrix. Then an equation for PPI prediction is devised on IS matrix. For S. cerevisiae PPI data from DIP, MINT and IntAct, domain data from Pfam-A, the prediction method achieved 50.19% and 84.57% sensitivity and specificity respectively. Here we also observed that embracing intra-domain effect gives positive result on PPI prediction by inspection of interaction probability change in protein pairs.

1. Introduction

Since the domain based computational model for protein-protein interaction (PPI) prediction method was proposed by Sprinzac[1], many domain or domain combination based PPI prediction methods have been proposed[2-5]. These methods are known to be effective in predicting PPIs by manipulating the PPI and domain data available on the Internet. Especially the domain combination based PPI prediction method proposed by Han et al. [6] was one of well-studied computational approaches for PPI and they achieved high prediction accuracy in Yeast proteins.

*To whom correspondence should be addressed. Tel: +82 42 350 6130; FAX: +82 42 350 6394

The conventional methods usually analyze the domain or Domain Combination (DC) appearance patterns among already-known interacting proteins. These approaches were revealed to be effective to some extents if PPI and domain data would be sufficiently provided. However, it does not mean that the conventional computational models for PPI are complete enough to explain the complex phenomena of PPI in nature.

Most of the conventional PPI prediction methods usually adopt simple computational models which cannot support sufficient understanding of PPI phenomena. Consequently, they often have limitations in providing useful additional supplementary information like which domain or DC pair will act as a main body in a predicted interaction.

In this paper, we develop a computational model that can predict PPIs in a more sophisticated manner than the models used in conventional domain-based PPI prediction methods. Some preliminary knowledge on domains and their appearance patterns in PPIs is essential for the development of such models. Actually, we can classify DC appearance patterns into two groups according to the appearance sites of the DCs in a protein interaction. One group is the appearance patterns of domains confined to the boundary of respective protein, and another group is the appearance patterns of DCs found in inferable DC pairs from interacting protein pairs. We need to consider both groups of the appearance patterns for the prediction of PPI to make the computational models more complete. In order to articulate the differences of the two groups of domain appearance patterns, we introduce notions of intra-protein domain cohesion and inter-protein DC coupling in this paper. Intra-protein domain cohesion is the co-occurrence tendency of domains in respective protein of the whole set of proteins, and inter-protein DC coupling is the co-occurrence tendency of DCs in the inferable DC pairs from the whole set of PPI.

While conventional domain or domain combination based methods consider only inter-protein DC coupling,

our model considers intra-protein domain cohesion as well as inter-protein DC coupling. In that sense, the model is a more sophisticated and advanced computational model than the models used in conventional domain or DC based prediction methods.

In this paper, we also define a DC pair coupling power as the possibility or strength for a DC pair to act as a main body for a protein interaction, and then devise an expression for computing the DC pair coupling power with intra-protein domain cohesion and inter-protein DC coupling information. That is, intra-protein domain cohesion and inter-protein DC coupling are two primary factors for computing the DC pair coupling power. Once the coupling powers for all the DC pairs are computed, the results are stored in a matrix called Interaction Significance (IS) matrix. Then we predict PPIs and identify primary interacting DC pair using the values of IS matrix.

As learning sets, we collected PPI and domain data from DIP, IntAct, MINT, and Pfam-A respectively, and UniProt was used for data integration. To evaluate a proposed method, sensitivity and specificity were measured in 0.0 to 1.0 thresholds. Our prediction method achieved 50.19% of moderate sensitivity and 84.57% specificity in threshold 0.1.

Interaction probability of true test set was increased about 33.4% while one of false test set increased only 11.5% after embracing intra-protein domain cohesion. We also present an example case which shows primary interaction DC pair is correctly changed after giving intra-protein domain cohesion force.

The rest of this paper is organized as follows. Section 2 defines intra-protein domain cohesion. DC and DC pair are formally introduced, and weighted DC pair is developed on it. Section 3 illustrates inter-protein DC coupling power, and Section 4 introduces IS matrix construction from intra and inter DC collaboration calculation. Evaluation and analysis are presented in Section 4 and 5, and finally we end up with a summary and a future work in Section 6.

2. Intra-protein domain cohesion and weighted domain combination pair

In DC based approach, a DC-pair is considered as a basic unit of PPI instead of single domain pair. This DC concept originates from the idea that domains do not work independently, and neighbor domains sometimes collaborate with each other in a protein. Polypeptide folding is sensitive to the environmental effect to form a tertiary structure and a function. Thus, a domain as a folded polypeptide structure would be influenced by neighboring domains which is one of environmental effects. As a result, it is plausible conjecture that multiple domains in a protein work together to interact with domains in another protein.

As a preliminary step, we define DC and DC pair using formal mathematical notations. Suppose that a protein p has one or more domains, and then we can generate DC by computing a power set of the set of domains. That is, DC, DC_p is defined by

$$DC_p = \{dc | dc \subset PowerSet(domain(p)), dc \neq \emptyset\}, \quad (1)$$

where $domain(p)$ represents a set of domains in protein p and $PowerSet()$ is a function which generates power set of domain set. If a protein p has non-redundant n domain(s), the number of DCs of p is $2^n - 1$. The following equation is the definition of DC pair of two proteins, p, q.

$$DC\text{-}pair(p,q) = \left\{ < dc_i, dc_j > \middle| \begin{matrix} < dc_i, dc_j > \in DC_p \times DC_q \\ or\ DC_q \times DC_p \end{matrix} \right\}, \quad (2)$$

where $dc_i, dc_j \in DC_p\ or\ DC_q$. If two proteins p, q have n, m non-redundant domains respectively, DC based method inspects $(2^n - 1) \times (2^m - 1)$ pairs while only $n \times m$ pairs were considered in the conventional single domain approach.

Though the DC pair is advanced concept, it is not a certain fact that all generated DC pairs always work together. Among the DC pairs, some of them may have higher effect on PPI than others. To differentiate the interaction force of DC pairs, we adopt intra-domain cohesion concept that computes conserved degree of DCs in a protein. When we consider that a protein has been evolved toward a specific function, a domain as a sub structure of protein also work together with other domains in order to realize the function. Naturally, during an evolution, the conserved part may be extended to accommodate a team of DC pairs. In the sense of that, we consider that DC pairs coupled by highly conserved DC may have higher effect on PPI.
In this paper, we compute the conserved degree of DC using an association rule, all-confidence value, and regard the all-confidence values as representing mutual dependencies of domains in a DC. The all-confidence means co-occurrence probability of DC in the entire proteins. We found that there are higher correlations between all-confidence and the similarity of molecular functions of domain than conventional support or confidence based approach [7].

To easy understand, suppose that given proteins *P1, P2, P3* and *P4* have domains like Table 1. In this case all-confidence of DC, *X={d1,d2}* is computed like follows.

$$all\text{-}conf(dc) = \frac{|\{p|p \in P \wedge dc \in PowerSet(domain(p))\}|}{MAX(\{i|\forall l(l \in PowerSet(dc) \wedge i = |\{q|q \in P \wedge l \in PowerSet(domain(q))\}|)\})}. \quad (3)$$

$$WDCP < p,q >_{i,j} = \frac{all\text{-}conf(dc_i)}{\sum_{\forall dc_u \in DC_p} all\text{-}conf(dc_u)} \times \frac{all\text{-}conf(dc_j)}{\sum_{\forall dc_v \in DC_q} all\text{-}conf(dc_v)} . \quad (4)$$

$$DCPPW < p,q >_{i,j} = \frac{Power\ of\ < dc_i, dc_j >}{Power\ of < dc_i, dc_j > + Power\ of\ dc\ pair(p,q) \neq < dc_i, dc_j >}. \quad (5)$$

$$DCPPW < p,q >_{i,j} = \frac{WDCP < p,q >_{i,j} \times |I\text{-}pair(dc_i, dc_j)|}{\sum_{<dc_u, dc_v> \in dc\ pair(p,q)} WDCP < p,q >_{u,v} \times |I\text{-}pair(dc_u, dc_v)|}, \quad (6)$$

Where $I\text{-}pair(dc_i, dc_j) = \{< p_u, q_v > | < dc_i, dc_j > \in dc\text{-}pair(p_u, q_v)\}$

Table 1. Sample proteins and domains information

Protein	Domains
P1	d1, d2
P2	d1, d3, d4
P3	d1, d2, d3, d5
P4	d4, d5

$$all\text{-}conf(X) = \frac{\#\ of\ protein(s) whose\ DCs\ contains\ X}{Max\ \#\ of\ protein(s) whose\ DCs\ contains\ subset\ of\ X}$$

Both $P1$ and $P3$ have $DC, \{d1, d2\}$, so the numerator of $all\text{-}conf(X)$ should be two. Otherwise, among the subset of $X, < \{d1\}, \{d2\}, \{d1, d2\} >, \{d1\}$ is more frequently appeared in all proteins than any other subset of X, and the denominator is three. Thus, all-confidence of DC X is 2/3. This all-confidence means that the probability of collaboration d1 and d2 is at least 2/3. Likewise, we pre-computed all-confidence of each DC extracted from entire protein set. Fig 2 illustrates correlations between all-confidence and the similarity of molecular functions of a DC.

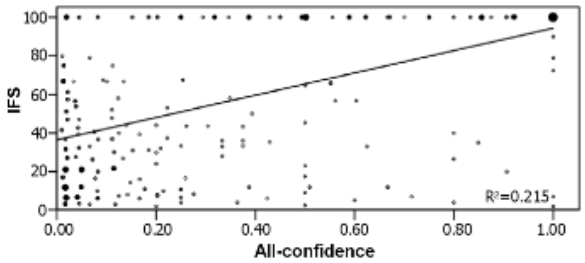

Fig. 2. Correlation distribution between all-confidence of domain combination and Inner Functional Similarity (IFS).

The Inner Functional Similarity, IFS is calculated by Gene Ontology similarity of each domain in a DC. The size of each dot stands for the frequency of a DCs detected in the entire protein. From the result of Pearson correlation coefficient measurement, both X and Y axises are related to each other by 0.01 to 0.464 Pearson value. The formal definition of all-confidence is in Eq. (3) where *P* is a set of all proteins and *MAX()* is a function selecting the maximum number from the set of positive numbers.

3. Inter-protein domain combination coupling power

With all-confidence value, we found functional collaboration degree of DC in a protein. Now, the all-confidence is extended to calculate a potential probability of collaboration from the DC in the single protein to the DC pair in the protein pair. That is, original DC pairs are differentiated by their effect on PPI, and we call them Weighted Domain Combination Pair (WDCP). WDCP of dc_{ij} for a single PPI $< p,q >$ is Eq. (4). Eq. (4) indicates that the weight value of a DC in a single protein is calculated by dividing all-confidence of the DC by the sum of all-confidence of others, and the collaboration probability of DC pair is multiplication of the weight value of each DC. The equation is devised in the intention that the probability 1.0 is translated to one interaction event, and each DC-pair fairly takes the interaction chances.

In the single domain based PPI model illustrated by Fig. 3 (a), two DC pairs get one-half possibility to work together. Fig. 3 (b), DC methods have three DC pairs, so the possibility is equally divided by three. Otherwise in our approach, it is possible to assign different possibility to each DC pair like Fig. 3 (c) by using all-confidence. Since we assign 1 to the weight of single domain, the total weights of DCs made by two or more domains in a protein should not exceed the value 1. It is rational because the number of single domain interaction is still larger than that of multi domain associated interaction.

$WDCP < p,q >_{i,j}$ explains how strongly domains in the DC pair work together to realize a specific function. However, the equation cannot be directly applicable for computing actual interacting DC pair because

the equation calculates collaboration degree only in one protein.

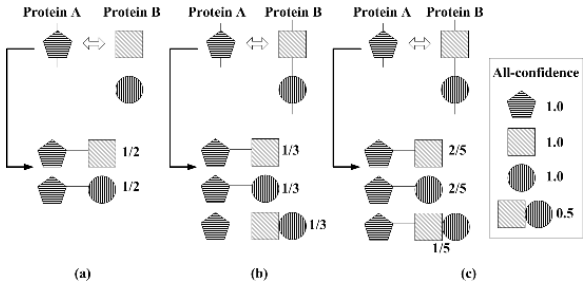

Fig. 3. The inter-protein DC coupling power calculation in a single PPI. Each figure shows the probability based on single domain (a), domain combination (b) and weighted domain combination (c) respectively.

We need an equation to compute the actual interacting DC pair in a PPI. Suppose that there is a set of PPIs, $\{< A, D >, < B, D >, < C, D >\}$ and each protein has domains like *domain(A)*={a, b}, *domain(B)*={a, c}, *domain(C)*={a, d}, *domain(D)*={e}. If we assign equivalent weights to DCs for convenience, the collaboration degrees of DC pair, <{a}, {e}> and <{b}, {e}> become 1/3. However, intuitively none of collaboration degrees of <{b}, {e}>, <{c}, {e}>, <{d}, {e}> should not exceed that of <{a}, {e}> because <{a}, {e}> appears more frequently than others. Thus, we must consider the entire PPI data to figure out which DC pair plays a key role in an actual interaction. When a DC pair $< dc_i, dc_j >$ appears more frequently in PPI, it has more power in an interaction, so the power of $< dc_i, dc_j >$ is computed by multiplication of $WDCP < p, q >_{i,j}$ and the appearance frequency. The interaction power of inter-protein DC pairs is defined by domain combination pair coupling power (DCPPW) in Eq. (5). Eq. (5) means that actual interaction possibility of inter-protein DC pair is calculated by the ratio of the DC pair's coupling power against the negative effect of other pairs' coupling power. We can redefine DCPPW with WDCP like Eq. (6) finally.

4. Interaction significance matrix construction and PPI prediction

Same DC pairs may have different DCPPW value in different PPI. Interaction Significance (IS) matrix accumulates average DCPPWs of all possible DC pairs from entire PPI in the learning set. Both columns and rows of IS matrix correspond to the union of DC of the entire proteins, so it is a symmetric matrix. When a DC pair is detected from interaction data, the contribution of that pair is computed based on intra-protein domain cohesion and inter-protein DC coupling, and the result is stored corresponding element of the matrix. The contribution value of $< dc_i, dc_j >$ is computed by Eq. (7)

$$IS(dc_i, dc_j) = \frac{\sum_{\forall (p_u, q_v) \in I-pair(dc_i, dc_j)} DCPPW < p_u, q_v >_{i,j}}{|I\text{-}pair(dc_i, dc_j)|}. \quad (7)$$

$IS(dc_i, dc_j)$ is an average of $DCPPW < p_u, q_v >_{i,j}$ in all protein pairs, $< p_u, q_v >$ with the DC pair, $< dc_i, dc_j >$. The interaction probability of an unknown protein pair is highly dependent on the interaction of DoDC of each protein. In this prediction method, we assume that PPI occurs only when there is at least one interaction among the DC pairs. Based on IS matrix, the interaction probability of an unknown protein pair $< p, q >$ is computed by Eq. (8).

$$IP(p, q) = 1 - \prod_{<dc_i, dc_j> \in dc\text{-}pair(p,q)} (1 - IS(dc_i, dc_j)). \quad (8)$$

$IP(p, q)$ is a probability that protein *p, q* have an interaction, and this is equivalent to a probability that there is at least one interaction among the DC pairs of protein *p, q*. Each element in IS matrix stands for an interaction probability of DC pair, so the probability that there is at least one interaction among the DC pairs is equal to subtracting a probability that all DC pairs have no interaction from 1.0. For example, if protein pair <*p, q*> has three DC pairs and DCPPWs of each pair are X_1, X_2, X_3, $(1 - X_1)(1 - X_2)(1 - X_3)$ is the probability that there is no interaction. Thus, the probability $IP(p, q)$ which means there is at least one interaction is $1 - (1 - X_1)(1 - X_2)(1 - X_3)$.

5. Validation and result

For training, three PPI databases (DIP: http://dip.doe-mbi.ucla.edu[8], IntAct: http://www.ebi.ac.uk/intact[9], MINT: http://mint.bio.uniroma2.it/mint[10]) were integrated. In order to use up-to-dated data, 57,036 PPIs of 20,442 proteins were extracted from DIP on Oct. 14th, 2008. IntAct released on Nov. 11th, 2008 had 115,331 PPIs with 55,036 proteins, and 95,256 PPIs from 29,148 proteins exists in MINT released on Apr. 14th , 2009. In this validation, we selected 65,902 *S. cerevisiae* PPIs out of 197,165 integrated PPIs in which both protein have domain information. The domain information for the proteins was extracted from Pfam (http://pfam.sanger. ac.uk)[11] released on July, 2008. We used Pfam-A families as domain database because Pfam families contain only non-overlapped

protein sequences. Unlike hierarchical database like InterPro[12] we didn't need to eliminate redundancies in Pfam families. We manually removed PPIs missing Pfam-A domain information and finally prepared 45,892 PPIs (69.63% coverage) for the validation. SwissProt and TrEMBL IDs extracted from UniProt[13] release 14.4(Nov. 4th, 2008) were used for the integration of all database. We measured prediction accuracy with five fold cross-validation method, and analyzed it through sensitivity and specificity. Note that the non-interacting test set was artificially generated by randomly pairing from the *S. cerevisiae* proteins which have domain information, and removed manually if random pair exists in the interaction test set. We measured prediction accuracies increasing thresholds of IP value by 0.1 in the interval from 0 to 1. We assumed that an interaction occurs if a pair has higher IP value than the threshold, otherwise it does not occur. Fig. 4 shows prediction accuracy curve.

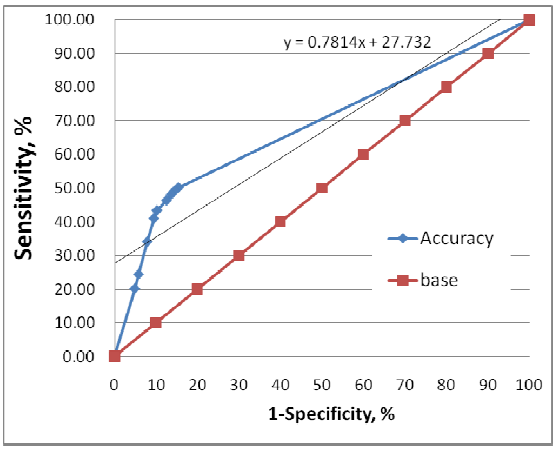

Fig. 4. ROC curve of prediction accuracy.

Sensitivity rapidly descends from the best 50.19% with the increasing threshold values. In contrast, specificity showed marginal changes between 84% and 95% irrespective of the changes of threshold values.

We also inspected intra-protein domain cohesion effect on interacting and non interacting test set. Firstly, we divided each test set into three different categories. If all DC pairs of a protein pair have DCPPW greater than zero, we call it "fully overlapped to interaction DC pair pattern". If some of DC pairs have zero DCPPW, it can be partial overlapped protein pair. Likewise we picked out no overlapped protein pairs whose whole DC pairs have zero DCPPW. As shown in Table 2 and 3, all-confidence gives more effect to interaction test set. Especially, if a protein pair has fully overlapped DC pairs, its IP value usually increased. Since we assumed IP value is the interaction possibility, the increase of IP in the interaction test set means all-confidence gives good effect to prediction method. Otherwise, we found that the increase is relatively small in non interaction test set.

Table 2. IP change by adding all-confidence in interaction test set.

	$IP_{new} > IP_{old}$	$IP_{new} = IP_{old}$	$IP_{new} < IP_{old}$
Fully	2182.8	1238.4	84.2
Partially	729.4	168.2	14.6
No	0	4305.4	0
Total	2912.2	5712	98.8
Portion %	33.39	65.48	1.13

Table 3. IP change by adding all-confidence in non interaction test set.

	$IP_{new} > IP_{old}$	$IP_{new} = IP_{old}$	$IP_{new} < IP_{old}$
Fully	602.4	264.6	10.2
Partially	397.4	100.4	1.4
No	0	7346.6	0
total	999.8	7711.6	11.6
Portion %	11.46	88.41	0.13

6. Discussion

In the validation process, we found that DC pairs detected in experimental PPI pairs had conservational fixed patterns. Non-interacting protein pairs prepared by randomly pairing of proteins seldom overlapped with DC pair patterns of the learning sets, and thus they usually showed low IP value. However, DC pair patterns detected in the interacting test set were more frequently found in the learning sets than the case of the non-interacting test set. However, sensitivity stopped around 50%. We can explain it in the following ways. Firstly, the error data included in the learning sets is one of hindrances of increasing the sensitivity. PPI data used in this study is obtained through high-throughput experiments, and it is known that many errors are usually included in the data from high-throughput experiments. Secondly, the learning sets have yet to be sufficient enough. Domains detected in current PPIs only cover a part of the domains in the nature. Thus if more large number of domains and PPIs are included, prediction accuracies will be improved further. Lastly, despite diverse features of protein are usually engaged in the protein interactions, only domain interaction is considered in our PPI prediction method. This is because we assumed that domain interaction plays decisive role in protein interactions, and thus the prediction method of this paper may not be sufficient to explain the entire phenomena of protein

interaction. However, we observed that DCPPW gives strong hints to protein interaction from the prediction result of proposed method.

Table 4. DCPPW change between domain pairs due to the intra-protein domain cohesion computation in protein, P07258↔P07259

Domain Pair	DCPPW (all-conf)	DCPPW (No weight)	DOMINE Conf.
PF00117_PF00117	0.557	0.282	GS
PF00117_PF00185	0.081	0.007	LC
PF00117_PF00289	0.559	0.220	HC
PF00117_PF00988	0.108	0.009	GS
PF00117_PF02142	0.497	**0.295**	HC
PF00117_PF02729	0.081	0.007	LC
PF00117_PF02786	**0.559**	0.220	GS
PF00117_PF02787	0.389	0.162	GS

Database of Protein Domain Interaction (DOMINE)[14] provides domain-domain interaction information annotated different confidence level such as Gold Standard(GS), high(HC), medium(MC) and low(LC). As PPI pair in Table 4, the primary DCPPW of DC pair, PF00117_PF02142 was changed to the DC pair, PF00117_PF02786 by embracing intra-protein domain cohesion effect. We expect that this phenomena means intra DC cohesion force positively contributes to our PPI prediction. To convince this, we will report detail analysis in near future work.

Actually, in *S. cerevisiae* case, about 2000 domains are detected, and the possible number of domain combinations reaches about ${}_{2000}C_2 \approx 2{,}000{,}000$ combinations. Even if we consider only single domain pair, the number of combinations is quite huge in case the proteins contain 10 domains. The overlap of DC pair patterns in 35,000 PPI strongly implies that there are fixed and duplicated DC pair patterns in protein interaction.

Our prediction method adopts intra-protein domain cohesion to compute PPI interaction possibility. We can expect several benefits from it. First of all, we can improve the completeness of conventional computational models for the prediction of PPIs. This enhances our understanding of proteins and PPIs. Second, we can improve the prediction accuracy of prediction methods by using the intra-protein domain cohesion information additionally. Lastly, the new model is useful in providing supplementary information on PPIs like primary interacting DC pair. This information is sometimes useful in further study of PPI such as protein structure analysis in residue level. Our further process will be analysis of primary interacting DC pair by comparing it to reported crystal structure.

7. Conclusion

In this paper, we developed a computational model that improves the understanding of PPI by considering not only coupling forces between inter-protein domain combination pairs but also the functional cohesion of intra-protein domain combinations. We expect that the prediction accuracy will be improved further as we increase the size of reliable experimental data in the learning sets. Our computational model is certainly advanced because it provides one of feasible ways to reflect the collaboration effects of multiple domains to the prediction. The prediction method considered that all protein pairs with non-zero IP value may interact with each other. We need an additional study on whether the higher IP value is relevant with the higher possibility of actual protein interaction. This study is possible by analyzing actual domain binding information from PDB and comparing it with DCPPW of DC pairs and final IP value. We leave this to our future work as well.

References

[1] Sprinzak E. and Margalit H., "Correlated sequence-signatures as markers of protein-protein interaction", *Mol. Biol*., 311, 2001, pp. 681-692.

[2] Marcotte E. et. al., "Detecting protein function and protein-protein interactions from genome sequences", *Science*, 285, 1999, pp. 751-753.

[3] Deng M. et. al., "Inferring domain-domain interactions from protein-protein interactions", *Genome Res.*, 12, 2002, pp. 1540-1548.

[4] Szilagyi A. et. al., "Prediction of physical protein-protein interactions", *Phys. Biol.*, 2, 2005, S1-S16.

[5] Chen L. et. al., "Inferring protein interactions from experimental data by association probabilistic method", *Proteins*, 62, 2006, pp. 833-837.

[6] Han D. et. al., "PreSPI: a domain combination based prediction system for protein-protein interaction", *Nucl. Acids Res.*, 32, 2004, pp. 6312-6320.

[7] SH Jung et. al., "Identification of conserved domain combinations in S. cerevisiae Proteins", *Proc. Bioinformatics and bioengineering, BIBE*, 2007, pp. 14-20.

[8] Salwinski L. et. al., "The Database of Interacting Proteins", *Nucleic Acids Research*, 32, 2004, Database issue: D449-51.

[9] Kerrien et. al., "IntAct-Open Source Resource for Molecular Interaction Data", *Nucleic Acids Research*, 35, 2007, Database issue: D561-D565.

[10] Andrew C. et. al., "MINT: the Molecular INTeraction database", *Nucleic Acids Research*, 2006, doi: 10.1093/nar/gkl950.

[11] Finn et. al., "The Pfam protein families database", *Nucleic Acids Research*, 36, 2008, Database Issue: D281-D288.

[12] Sarah Hunter et. al., "InterPro: the integrative protein signature database", *Nucleic Acids Research*, 37, 2009, D211-D215; doi:10.1093/nar/gkn785.

[13] The UniProt Consortium., "The Universal Protein Resource (UniProt)", *Nucleic Acids Research*, 36, 2008, Database issue: D190-D195.

[14] Raghavachari B et. al., "DOMINE: A database of protein domain interactions", *Nucleic Acids Research*, 36 (Database Issue), 2008, D656-661.