Web-Based Computational System for Protein-Protein Interaction Inference

Ki-Bong Kim*

Abstract—Recently, high-throughput technologies such as the two-hybrid system, protein chip, Mass Spectrometry, and the phage display have furnished a lot of data on protein-protein interactions (PPIs), but the data has not been accurate so far and the quantity has also been limited. In this respect, computational techniques for the prediction and validation of PPIs have been developed. However, existing computational methods do not take into account the fact that a PPI is actually originated from the interactions of domains that each protein contains. So, in this work, the information on domain modules of individual proteins has been employed in order to find out the protein interaction relationship. The system developed here, WASPI (Web-based Assistant System for Protein-protein interaction Inference), has been implemented to provide many functional insights into the protein interactions and their domains. To achieve those objectives, several preprocessing steps have been taken. First, the domain module information of interacting proteins was extracted by taking advantage of the InterPro database, which includes protein families, domains, and functional sites. The InterProScan program was used in this preprocess. Second, the homology comparison with the GO (Gene Ontology) and COG (Clusters of Orthologous Groups) with an E-value of 10^{-5}, 10^{-3} respectively, was employed to obtain the information on the function and annotation of each interacting protein of a secondary PPI database in the WASPI. The BLAST program was utilized for the homology comparison

Keywords—Protein-Protein Interactions (PPIs), WASPI (Web-based Assistant System for Protein-protein interaction Inference), InterPro, InterProScan, BLAST

1. INTRODUCTION

As various genome projects have produced enormous amounts of genome sequences, numerous amounts of research have been actively going on to find out the proteins that known genomes contain and along with analyzing their functions. In the analysis of protein function, conventional studies regarded proteins as not being components of a whole interaction network but as isolated entities. However, current studies have focused on examining functions and roles of each individual gene and protein in the context of a protein interaction network. Also, from the aspect of potential practical applications, the understanding of the proteins as components of a whole system helps new drugs that can specifically interrupt or activate protein interactions to

※ This work was supported by Sangmyung University
Manuscript received December 6, 2011; first revision May 15, 2012; accepted June 29, 2012.
Corresponding Author: Ki-Bong Kim
* Dept. of Biomedical Technology, Sangmyung University, Cheonan, Korea (kbkim@smu.ac.kr)

be developed.

To detect the PPI (Protein-Protein Interaction) relationship, biologists have been taking advantage of various experimental methods like two-hybrid systems, Mass Spectrometry, and the protein chip. However, the protein interaction data obtained through experimental methods have not been accurate so far and the quantity of the data has been limited. Therefore, computational methods for predicting and validating the PPIs have been developed. They provide a complementary approach to detecting protein-protein interactions. In general, all computational approaches attempt to leverage the knowledge about experimentally determined known interactions, in order to predict new PPIs. These methods enable one to discover novel putative interactions and often provide information for designing new experiments for specific protein sets. The four main computational methods are as follows: first, the phylogenetic profiles method is based on the pattern of the presence or absence of a given gene in a set of genomes. The similarity of phylogenetic profiles might be interpreted as a symptom of the functional need for corresponding proteins to be simultaneously present in order to perform some function together. Second, the gene neighborhood conservation method depends on the neighborhood relationship and becomes even more useful when it is conserved in multiple species. The adjacency of genes in various bacterial genomes has been used to predict functional relationships between the corresponding proteins. One of the main limitations of this method is that it is only directly applicable to bacteria, in which the genome order is a significant property. Third, the gene fusion method is based on interactions between proteins that can be deduced from the presence in different genomes of the same protein domains, which either form part of a single polypeptide chain (multidomain protein) or act as independent proteins (single domain). Methods based on recursive sequence searches and multiple sequence alignments have been combined in order to detect such domain fusion events [1]. In addition, Support Vector Machine (SVM) was trained to recognize and predict interactions based solely on primary structure and associated physicochemical properties such as charge, hydrophobicity, and surface tension [2]. In order to eliminate a significant amount of false-positive PPIs that are obtained only through full-length protein sequence similarity, Jerome *et al.* combined sequence similarity searches with clustering based on interaction patterns and interaction domain information [3]. Moreover, the association rule, which is one of the methods for data mining, was adopted to discover the general rules of PPI and the method to efficiently visualize PPI data was introduced [4, 5].

There are a number of protein interaction databases available, each with different advantages and also limitations. The Database of Interacting Proteins (DIP, http://dip.doe-mbi.ucla.edu/) contains information only on protein-protein interactions but employs rigorous criteria for evaluating the reliability of each interaction [6]. The Molecular INTeraction database (MINT, http://mint.bio.uniroma2.it/) contains additional information on proteins, nucleic acids, and lipid interactions [7]. The Biomolecular Interaction Network Database (BIND, http://bond.unleashedinformatics.com/) has gathered three kinds of data: biomolecular interactions, biomolecular complexes, and pathways [8]. The Biological General Repository for Interaction Datasets (BioGRID, http://thebiogrid.org/) has been developed in order to integrate the databases of PPI and has recently been classified into the three categories of Yeast GRID, Worm GRID, and Fly GRID according to the taxonomy [9]. The STRING database (http://string.embl.de/) is a collection of known and predicted protein interactions [10]. The interactions include direct (physical) and indirect (functional) associations. They are derived from four sources such as the genomic context, high-throughput experiments, co-expression, and previous knowl-

edge.

This work has focused on the fact that PPI is actually derived from the interactions of domains that interacting protein pairs contain. Protein domains are structural or functional units within the proteins. They are usually evolutionarily-conserved modules of amino acid sequences. The existence of certain domains in proteins can suggest the propensity for the proteins to interact or form a stable complex to bring about certain biological functions. Unfortunately, unlike protein-protein interaction detection, high-throughput experimental results for domain-domain interactions are currently unavailable. For this reason, in this work, the information on domain modules of each protein in PPI databases, which was analyzed and collected by the InterProScan program [11], has been employed in order to find out the protein interaction relationship. The system developed here, WASPI (Web-based Assistant System for Protein-protein interaction Inference), has been implemented to provide many functional insights into protein interactions and its domains. To achieve those objectives, several preprocessing steps have been taken. First, the domain module information of interacting proteins was extracted by taking advantage of the InterPro database [12], which includes protein families, domains and functional sites. As mentioned above, the InterProScan program was used in this preprocess. Second, the homology comparison with the GO (Gene Ontology) database [13] and COG (Clusters of Orthologous Groups) database [14] with an E-value of 10^{-5}, 10^{-3} respectively, was carried out to obtain the information on function and annotation for protein sequences in the X-Large, which is a secondary database of the WASPI. The output of the previous preprocess was used to make another secondary database, which is the Function/Annotation DB in the WASPI. The BLAST (Basic Local Alignment Search Tool) program [15] was utilized for the homology comparison. The X-Large database was derived from the three main PPI databases, including DIP, BIND, and BioGRID. The data redundancy of X-Large was eliminated through preprocessing. Finally, we analyzed the relationship of potential interacting domain pairs to make the Domain-Domain Interaction (DDI) database, which is also a secondary database of the WASPI. Therefore, users can search interacting protein pairs from multiple databases at once using the WASPI. In addition, the WASPI facilitates users to predict the interacting relationship of an input sequence based on homology comparison and domain module information.

2. SYSTEM AND METHODS

The main purpose of the WASPI is to integrate and summarize the three main PPI databases into a built-in secondary PPI database (called X-Large) and to make the best use of information on functional annotation and domain modules that have been derived from interacting proteins to infer a PPI relationship. To achieve those objectives, we have taken several preprocessing steps and constructed four kinds of built-in secondary databases that are X-Large, Function/Annotation, Domain, and DDI. Fig. 1 shows the overall workflow and framework of the WASPI system. First, the X-Large database is an integrated and summarized database of DIP, BIND, and BioGRID.

Second, the Function/Annotation DB contains functional information and annotation of protein sequences in X-Large DB. In order to obtain the information on functional annotation, we performed a homology comparison with GO (Gene Ontology) and COG (Clusters of Orthologous Groups) with an E-value of 10^{-5} and 10^{-3} respectively. The BLAST program was used for

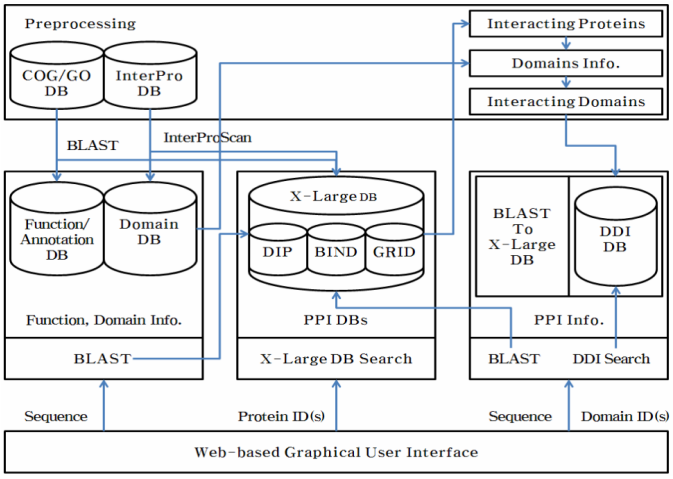
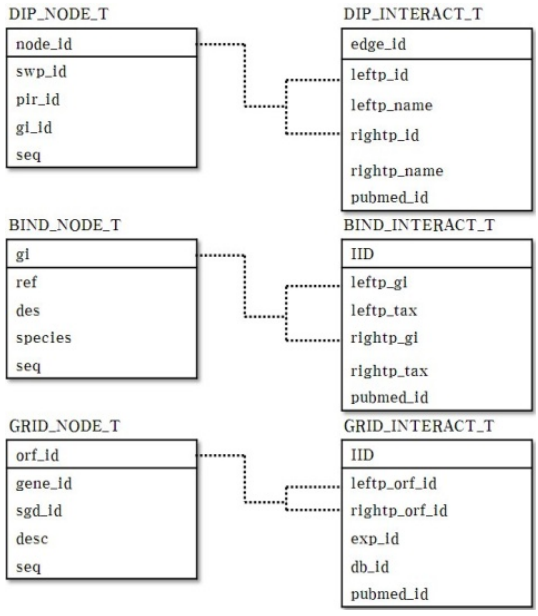


Fig. 1. Overall Workflow and Framework of the WASPI System

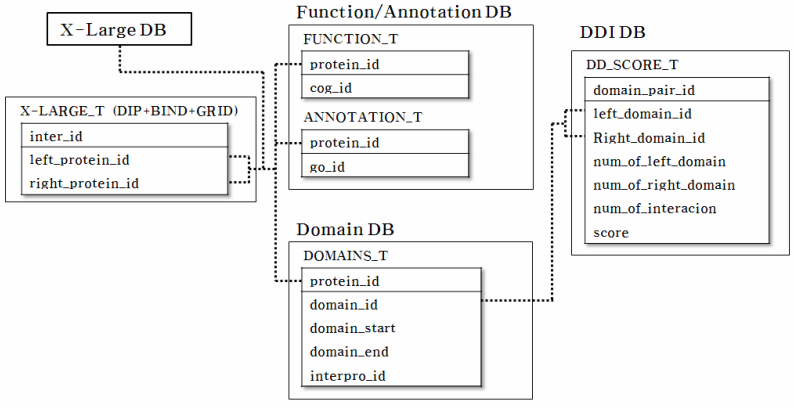
homology comparison in this work. BLAST is one of the most widely used database search program in the bioinformatics field. It relies on finding the core similarity, which is defined by a window size that has a preset size (called a "word") with an amino acid similarity score that is above the given threshold for proteins. The BLAST calculates an E-value, which is the number of alignments with a score of at least S that would be expected to occur by chance alone in searching a complete database of n sequences. These E-values can vary from 0 to n. Third, Domain DB has been created to offer information on domain modules that interacting proteins include. Because domains are structural or functional units within the proteins, which are usually evolutionarily-conserved modules of amino acid sequences, they are very important factors in predicting a PPI relationship. We have collected the domain module information of each interacting protein of X-Large by taking advantage of InterProScan, which is an application for retrieving the domain module of input proteins from InterPro.

Finally, the DDI (Domain-Domain Interaction) DB was constructed by making full use of both the X-Large DB and Domain DB. If two proteins have an interactive relationship, we can assume that domains, which those interacting proteins contain, also have an interactive relationship. In this work, the following computational method was considered to obtain domain-domain interaction information. If two proteins are known to have an interaction relationship, we can infer that the domains from the two proteins could potentially be interacting. In other words, if the two proteins P_1 and P_2 are known to interact with each other, we infer that domain $d_{1,i}$ potentially interacts with domain $d_{2,j}$ with a minimal probability of $1/mn$, where m and n are the number of domains in proteins P_1 and P_2 respectively, and $d_{1,i}$ and $d_{2,j}$ are the i-th and j-th domains of proteins P_1 and P_2 respectively. Actually, Wan and Park [16] have proposed a method to extract the domain module information of interacting proteins and for classifying TID (Truly Interacting Domain) pairs from PID (Potentially Interacting Domain) pairs. From this point of view, this paper tried to make the best use of domain module information, which interacting protein pairs contain, in order to find out the protein interaction relationship. Fig. 2 shows the schemata of four secondary databases of the WASPI.

By retrieving the information from those four secondary databases, users can gain the syn-



(a)



(b)

Fig. 2. Database Schema of the WASPI System. (a) Database Schema of X-Large DB. (b) Database Schema of Other Databases (DDI, Function/Annotation, and Domain) Except for X-Large DB

thetic information of an input sequence, such as the function and annotation, the domain modules and the interactive relationship of not only proteins but also domains. WASPI has 3-tier architecture which is composed of a client, an application server, and back-end databases (Fig.1). The application server consists of the BLAST module and the InterProScan module. As mentioned before, they are adopted to measure the similarity of the input sequence and to obtain the information on domain modules of the input sequence respectively. As stated above, the back-end databases are composed of four new secondary databases, that is, the X-Large DB, Function/Annotation DB, Domain DB, and DDI DB.

3. MAIN FEATURES OF THE WASPI

3.1 Searching the X-Large DB

WASPI has a useful and easily accessible X-Large DB that has been created by integrating and concisely summarizing diverse PPI databases such as DIP, BIND, and BioGRID. Such databases are widely used by many researchers for benchmarking the prediction of protein-protein interactions. Therefore, the WASPI provides a unitary channel or an interface for three heterogenous PPI databases.

3.1.1 Searching the DIP Database

The DIP database is composed of two main identifiers such as the node ID and edge ID. First, the node ID is a unique identifier for each protein described within the DIP database and has the format <DIP:nnnN>. It also has cross-references to, at least, one of the major protein databases of PIR, SWISSPROT, and GenBank. Second, the edge ID is a unique identifier for each interaction described within the DIP database and has the format <DIP:nnnE>. DIP can be searched in various ways. Users can search for the function, the annotation, and the interactions of a specific

1. Node

Node ID	1000N	SWP ID	P12689
PIR ID	S67255	GI ID	2133088

2. COG

COG ID	KOG2093
COG Desp	Translesion DNA polymerase - REV1 deoxycytidyl transferase

3. GO

Num.	Name	Type	Acc.
1	nucleotidyltransferase	molecular_function	GO:0016779
2	DNA repair protein	molecular_function	GO:0003685
3	response to DNA damage	biological_process	GO:0006974
4	DNA repair	biological_process	GO:0006281
5	extrachromosomal circular DNA	cellular_component	GO:0005727

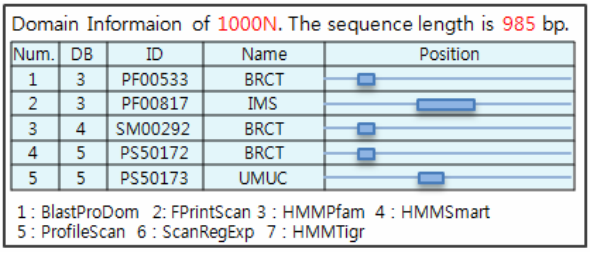
4. Domains

Num.	DB	ID	Name	Start	End	InterPro ID
1	3	PF00533	BRCT	163	249	IPR001357
2	3	PF00817	IMS	361	737	IPR001126
3	4	SM00292	BRCT	163	239	IPR001357
4	5	PS50172	BRCT	161	249	IPR001357
5	5	PS50173	UMUC	358	554	IPR001126

5. Interacting Proteins

Edge	Pro ID	Pro Name	Pro Name	Function	GO
857E	663N	SNP1	YIL061C	KOG0113	GO:0016779 GO:0003685 GO:0006974 GO:0006281 GO:0005727

Fig. 3. Result of an X-Large DB Search


Domain Informaion of 1000N. The sequence length is 985 bp.

Num.	DB	ID	Name	Position
1	3	PF00533	BRCT	
2	3	PF00817	IMS	
3	4	SM00292	BRCT	
4	5	PS50172	BRCT	
5	5	PS50173	UMUC	

1 : BlastProDom 2: FPrintScan 3 : HMMPfam 4 : HMMSmart
5 : ProfileScan 6 : ScanRegExp 7 : HMMTigr

Fig. 4. Specific Domain Module Information

protein by its gene name or its accession code in the GenBank, PIR, or SWISSPROT. The search result displays all of the interactions of DIP, which the protein is involved in. In addition, it provides all of the information on the protein in terms of the functional context of COG, the annotation of GO, domain modules and sequence data. The search result is shown in Fig. 3. Fig. 3 describes that protein 1000N has an interactive relationship with protein 663N and the fact that 1000N and 663N have the same GO identifier can increase the reliability of interaction. In order to confirm the locations and detailed information of domain modules, the user can click the icon next to "Domains". The domain module information is shown like it is in Fig. 4.

3.1.2 Searching the BIND Database

BIND is a collection of records describing molecular interactions. BIND has 3 classifications for molecular associations: molecules that associate with each other to form an interaction, molecular complexes that are formed from one or more interactions, and pathways that are defined by a specific sequence of two or more interactions. BIND has a GI as a unique identifier and also has references to the PubMed's REF identifier. GI or PubMed's REF ID can search BIND. Like the result from a DIP search, the result of a BIND search returns all of the interactions recorded in BIND, in which that protein participates. In addition, it provides the information on the functional COG context, the GO annotation, the domain modules, and the sequence data of a specified protein.

3.1.3 Searching the BioGRID Database

BioGRID has been developed to combine and archive existing physical, genetic and functional interaction data. Any valid gene name or ORF (Open Reading Frame) name or SGD (Saccharomyces Genome Database) ID can be searched for to yield a comprehensive list of known interactions and associated annotations. The search result displays all of the interactions recorded in BioGRID in which that protein participates. Moreover, it provides the information on the functional COG context, the GO annotation, the domain modules, and the sequence data of a specified protein.

3.2 Prediction of Protein-Protein Interaction Relationship

In this work, two kinds of methods are used to predict the interactive relationship between proteins. The first method is associated with the BLAST search, which is based on a homology comparison. And the second method is an application of domain module information for predicting PPI.

3.2.1 Homology Comparison Based PPI Prediction

In order to infer the PPI relationship of an input protein sequence, the BLAST program is run to do the homology comparison of the protein sequence with all of the protein sequences of X-Large DB. If an input protein has a homology with some of the proteins in X-Large DB, we can assume that the input protein has the same interactive relationship with the homologous proteins. This method has a weak point because it supposes that an interactive relationship is completely dependent on all of the sequence data. However, as a homology comparison analysis is an essential process of all the kinds of the sequence analysis, it is also applicable to the prediction of the PPI.

The WASPI provides two types of PPI prediction based on a homology comparison. First, by entering one sequence as an input data, users can examine all of the interacting proteins of the input sequence. Second, by entering two sequences, users can ascertain whether the two proteins have an interactive relationship or not. The basic assumption of this approach is that the reliability of a given protein interaction can be evaluated by the presence of paralogous interactions. The basis for this is that if two proteins are paralogs then the proteins that they are observed to interact with are often also paralogs. This observation is related to the notion of interologs that Vidal and co-workers proposed [17].

The web interface of PPI prediction based on homology comparison is shown in Fig. 5. Fig. 6 explains that two proteins, for when users want to know about interactive relationships, have paralogous interactions. In addition, the fact that proteins with paralogous interactions are detected in the same organism can increase the reliability of their interactive relationship.

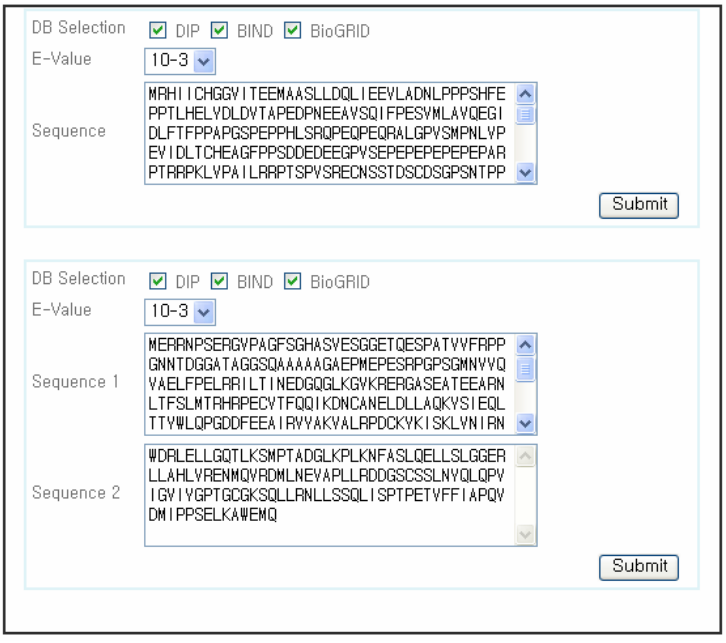
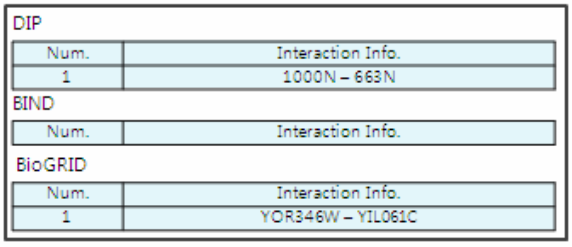


Fig. 5. Interface for PPI Prediction Based on a Homology Comparison



DIP

Num.	Interaction Info.
1	1000N – 663N

BIND

Num.	Interaction Info.

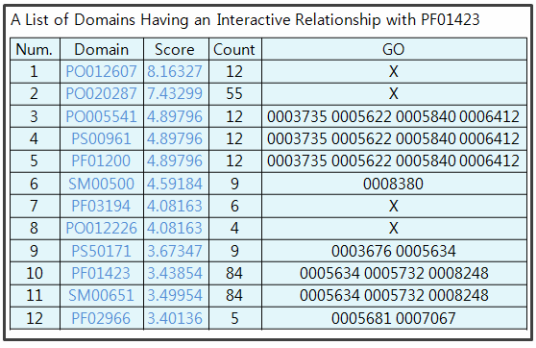
BioGRID

Num.	Interaction Info.
1	YOR346W – YIL061C

Fig. 6. Result of the Homology Comparison

3.2.2 Domain Modules Based PPI Prediction

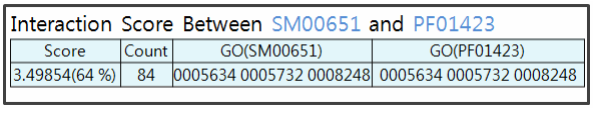
WASPI's main objective is to make the best use of domain module information to predict a PPI relationship. WASPI offers two types of methods that are based on domain modules for the PPI prediction, like the homology comparison method. First, with a domain identifier (Pfam, Smart, ProSite, etc.), users can obtain all of the domains having an interactive relationship with input domain. Fig. 7 shows a list of domains having an interactive relationship with PF01423, and PF01423 has the highest interaction reliability with PD012607. However, 5 domains such as PD005541, PS00961, PF01200, PF01423, and SM00651 that share GO identifier with PF01423 can be considered to have a reliable probability of interaction. Fig. 8 displays the score of the interaction between SM00651 and PF01423.



A List of Domains Having an Interactive Relationship with PF01423

Num.	Domain	Score	Count	GO
1	PO012607	8.16327	12	X
2	PO020287	7.43299	55	X
3	PO005541	4.89796	12	0003735 0005622 0005840 0006412
4	PS00961	4.89796	12	0003735 0005622 0005840 0006412
5	PF01200	4.89796	12	0003735 0005622 0005840 0006412
6	SM00500	4.59184	9	0008380
7	PF03194	4.08163	6	X
8	PO012226	4.08163	4	X
9	PS50171	3.67347	9	0003676 0005634
10	PF01423	3.43854	84	0005634 0005732 0008248
11	SM00651	3.49954	84	0005634 0005732 0008248
12	PF02966	3.40136	5	0005681 0007067

Fig. 7. Search Results of the Domain-Domain Interaction when a Search is done with a Domain ID



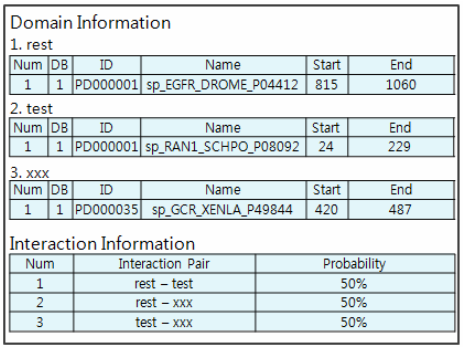
Interaction Score Between SM00651 and PF01423

Score	Count	GO(SM00651)	GO(PF01423)
3.49854(64 %)	84	0005634 0005732 0008248	0005634 0005732 0008248

Fig. 8. Search Result of the Domain-Domain Interaction when a Search is done with two Domain IDs

3.3 Prediction of the Best Pair from Multiple Sequences

Recently, although high throughput experimental methods have produced a great amount of PPI data, biologists cannot be confident of the reliability of experiment data. When a protein has many protein candidates having an interaction relationship, WASPI returns the most probable



Domain Information

1. rest

Num	DB	ID	Name	Start	End
1	1	PD000001	sp_EGFR_DROME_P04412	815	1060

2. test

Num	DB	ID	Name	Start	End
1	1	PD000001	sp_RAN1_SCHPO_P08092	24	229

3. xxx

Num	DB	ID	Name	Start	End
1	1	PD000035	sp_GCR_XENLA_P49844	420	487

Interaction Information

Num	Interaction Pair	Probability
1	rest – test	50%
2	rest – xxx	50%
3	test – xxx	50%

Fig. 9. Results of the Best Pair Prediction

interacting protein candidate. WASPI extracts the domain module information from multiple proteins and calculates the interaction score between the domains of the corresponding proteins. Therefore, users can evaluate the most potential protein pair (Fig. 9).

4. RESULTS AND DISCUSSION

PPIs play a critical role in many cellular functions. A number of experimental techniques have been applied to discover PPIs. However these techniques are expensive in terms of time, money, and expertise. As mentioned before, there are also large discrepancies between the PPI data that is collected by the same or different types of techniques for the same organism. In this respect, this paper turns to a computational technique for the prediction of PPIs. The collection, indexing, validation, analysis, and extrapolation of PPI and its related data were carried out for making the WASPI system which provides the access to easy-to-use PPI-related secondary databases and facilitates the prediction of the PPIs. The WASPI has three main features: efficient integration of PPI databases, prediction of the PPI relationship based on the homology comparison and domain module information, and the inference of the best pair from among multiple protein sequences. In addition, it provides synthetic information such as function/annotation (GO, COG description) and the domain modules of structural or functional units. As stated above, we have taken advantage of the InterPro database, in order to extract domain module information about proteins in the X-Large DB. As a result, 83% (5,534/6,652), 76% (3,414/4,480), and 75% (4,417/5,888) proteins contain domain modules in DIP, BIND, and BioGRID respectively. We derived putative domain-domain interactions using PPI information, and calculated a score to measure the reliability of domain-domain interaction. The score is calculated according to Equation 1:

$$P(d_m, d_n) = \left(\frac{k_{mn}}{k_m \times k_n}\right) \times 100 \tag{1}$$

where k_{mn} is the number of interactions of d_m and d_n, k_m is the number of proteins that contain at least one domain d_m, and k_n is the number of proteins that contain at least one domain d_n. The maximum number of k_{mn} is 98 for the interaction between PS00017 and PS50099. And the

Table 1. Example Scores of Domain-Domain Interaction Pairs

No.	Domain (d_m)	Domain (d_n)	k_m	k_n	k_{mn}
1	PF02984	PD005152	26	4	10
2	PF02984	PF01111	26	4	10
3	PF02984	PS00944	26	4	10
4	PF02984	PS00945	26	4	10

maximum score of all of the domain pairs is 9.61538. Table 1 shows the domain-domain interaction pairs with the maximum score of 9.61538. The minimum score is 0.00122 for a domain-domain pair between PS00017 and TIGR00879. However, the remarkable fact is that PF02984 is included in all of the interactions with the highest score. PF02984 is described as the Cyclin, C-terminal domain. Cyclins regulate cyclin dependent kinases (CDKs). UNG2 HUMAN is a Uracil-DNA glycosylase that is related to other cyclins. Cyclins contain two domains of a similar all-alpha fold, of which this family corresponds with the C-terminal domain. This result provides the evidence that the output of the WASPI system is very reliable.

As stated above, the WASPI is very useful for users who want to obtain synthetic information such as functional information, genomic annotation, the domain modules that a protein contains, and the interactive relationship of the input protein sequence. In other words, it provides a complementary approach to detecting protein-protein interactions. It attempts to leverage the knowledge of experimentally-determined previously known interactions, in order to predict new PPIs. This system enables one to discover novel putative interactions and is able to provide information for designing new experiments for specific protein sets. However, although the WASPI system has a scoring system for assessing the reliability of domain-domain interactions, this scoring system is not enough to estimate the probability of domain interaction more accurately. In addition, as PPI data has been accumulated exponentially by various experiments, diverse PPI databases are updated very often. In order to catch up with frequent updates, we have to find an appropriate way for automatic updates. These kinds of limitations should be overcome when further work is carried out.

REFERENCES

[1] V. Alfonso and P. Florencio, "Computational methods for the prediction of protein interactions", *Current Opinion in Structural Biology*, Vol.12, 2002, pp.368-373.

[2] Joel, *et al.*, "Predicting protein-protein interactions from primary structure", *Bioinformatics*, Vol.17, No.5, 2001, pp.455-460.

[3] W. Jerome and S. Vincent, "Protein-protein interaction map inference using interacting domain profile pairs", *Bioinformatics*, Vol.17, 2001, pp.296-305.

[4] Oyama, *et al.*, "Extraction of knowledge on protein-protein interaction by association rule discovery", *Bioinformatics*, Vol.18, No.5, 2002, pp.705-714.

[5] M. Ralf, "Java applet for visualizing protein-protein interaction", *Bioinformatics*, Vol.17, No.7, 2001, pp.669-670.

[6] Salwinski, et. al., "The Database of Interacting Proteins: 2004 update" *Nucleic Acids Research*, Vol.32 (Database issue), 2004, pp.449-451.

[7] Chatr-aryamontri, *et. Al*., "MINT: the Molecular INTeraction database", *Nucleic Acids Research*, Vol.35 (Database issue), 2007, pp.572-574.

[8] Bader, *et. al*., "BIND : the Biomolecular Interaction Network Database", *Nucleic Acids Research*, Vol.31, No.1, 2003, pp.248-250.
[9] Stark, *et. al*., "The BioGRID Interaction Database: 2011 update", *Nucleic Acids Research*, Vol.39, 2011, D698-704.
[10] Szklarczyk, *et. al*., "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored", *Nucleic Acid Research*, Vol.39 (Database issue), 2011, pp.561-568.
[11] N. J. Mulder and R. Apweiler, "The InterPro Database and Tools for Protein Domain Analysis", *Current Protocols in Bioinformatics*, 2008, Chap. 2, Wiley Online Library.
[12] S. Hunter, *et. al.*, "InterPro: the integtative protein signature database", *Nucleic Acids Research*, Vol.37, No.1, 2009, pp.211-215.
[13] Gene Ontology Consortium, "The Gene Ontology (GO) database and informatics resource", *Nucleic Acid Research*, Vol.32, 2004, pp.D258-D261.
[14] Tatusov, *et al*., "The COG database: new developments in phylogenetic classification of proteins from complete genomes", *Nucleic Acids Research*, Vol.29, 2001, pp. 22-28.
[15] Altschul, *et. al.,* "Basic local alignment search tool", *J. Mol. Biol*., Vol.215, 1990, pp.403-410.
[16] K. Wan and J. Park, "Large Scale statistical prediction of protein-protein interaction by Potentially Interacting Domain (PID) pair", *Genome Informatics*, Vol.13, 2002, pp.42-50.
[17] Walhout, *et. al.*, "Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm", *Yeast*, Vol.17, 2000, pp.88-94.

Ki-Bong Kim

He received a Ph.D. degree in Computer Engineering from Chungnam National Univ. in 2003. He has been a professor at Sangmyung Univ. since 2003. His research interests are in the area of Bioinformatics, topics of which are Biodata Mining, Gene and Promoter Prediction, Protein-Protein Interaction, MicroRNA Gene and Target Recognition, Regulatory Network, and Biologically Significant Pattern Recognition.