Genome Informatics 15(2): 171-180 (2004) 171

PreSPI: Design and Implementation of Protein-Protein Interaction Prediction Service System

Dong-Soo Han'
dshan@icu.ac.kr

Hong-Soog Kiml
kimkk@icu.ac.kr

Woo-Hyuk Jangl
torajim@icu.ac.kr

Sung-Doke Leel
sdlee@icu.ac.kr

Jung Keun Suh2
suhjung@lgls.co.kr

1 School of Engineering, Information and Communications University, 119, Munjiro,
Yuseong-gu, Daejeon 305-714, Korea

2 Proteomics and Bioinformatics Program, LG Life Sciences R&D, 104-1, Munji-dong,
Yuseong-gu, Daejeon 305-380, Korea

Abstract

With the recognition of the importance of computational approach for protein-protein interac-tion prediction, many techniques have been developed to computationally predict protein-protein interactions. However, few techniques are actually implemented and announced in service form for general users to readily access and use the techniques. In this paper, we design and implement a pro-tein interaction prediction service system based on the domain combination based protein-protein interaction prediction technique, which is known to show superior accuracy to other conventional computational protein-protein interaction prediction methods. In the prediction accuracy test of the method, high sensitivity (77%) and specificity (95%) are achieved for test protein pairs con-taining common domains with learning sets of proteins in a Yeast. The stability of the method is also manifested through the testing over DIP CORE, HMS-PCI, and TAP data. The functions of the system are divided into core, subsidiary, and general service function categories. The core func-tion category includes the functions that can be provided only by using the domain combination based protein-protein interaction prediction method. Interaction prediction for a single protein pair and visualization of interaction probability distributions are the functions in this category. The subsidiary function category includes the functions that can be derived from the core functions. Domain combination pair search with high appearance probability and construction of protein in-teraction network are the functions in this category. Lastly, the general service function category
includes the functions that can be implemented by collecting and organizing the protein and do-
main data in the Internet. Performance, openness and flexibility are the major design goals and
they are achieved by adopting parallel execution techniques, Web Services standards, and layered architecture respectively. In this paper, several representative user interfaces of the system are also introduced with comprehensive usage guides.

Keywords: PreSPI, protein-protein interaction prediction, domain combination, web services, AP matrix, primary interaction probability

1 Introduction

With the recognition of the importance of computational approach for protein-protein interaction

prediction, many techniques have been developed to computationally predict protein-protein interac-tion [5, 6, 12, 14] . Finding and analyzing subsequences affecting the protein-protein interactions from raw protein sequence is one approach [7] . Another is to predict protein interactions by analyzing the

physicochemical properties or tertiary structure of proteins [3] . Domain based protein-protein inter-action prediction is also an actively studied approach by several research groups recently [5, 14, 16] .

172 Han et al.

However, most of the techniques are still in immature state and only a few of them are provided as concrete services for general users. This is due to the fact that protein-protein interaction prediction research is in early stage and thus their prediction accuracy is not good enough to be serviced for general users. Interpom [13] provides useful information on proteins, domains, and protein-protein interactions in integrated manner and it tries to predict protein-protein interactions based on domain information. But the prediction method of Interpom is rather simple and the prediction accuracy of Interpom is not apparent.
Recently, a domain combination based protein-protein interaction prediction method is well studied by Han's group and the prediction accuracy of this method is revealed to be superior to that of other conventional domain based protein interaction prediction methods [9, 10] . The method is more sophisticated and more reliable than conventional domain based protein-protein interaction prediction methods. In this paper, we have measured again the prediction accuracy of the method with different conditions and data from those of the previous measurements to confirm the superiority of the domain combination based prediction method. High sensitivity (77%) and specificity (95%) are achieved for test protein pairs containing common domains with learning sets of protein pairs in a Yeast. The stability of the method is also manifested through testing over DIP CORE, HMS-PCI, and TAP data. Although the test result is limited only to proteins in a Yeast organism, the prediction system with such an accuracy and stability can provide valuable information to biologists.
Then we design and implement a protein-protein interaction prediction system, named PreSPI (Prediction System for Protein Interaction), using the technique. For the implementation of PreSPI, we first list up the service functions and extract design goals of PreSPI. Then the software architecture of PreSPI is devised to effectively implement the service functions and to achieve the design goals. The major function of PreSPI is to predict the interaction possibility of proteins. But it is also equipped with various auxiliary functions for researchers on proteins or protein interactions. PreSPI's functions are largely divided into core functions, subsidiary functions, and general service functions. In this paper, the details of definitions and usage guides of the functions in each category are described with the introduction of representative user interfaces of PreSPI.
In general, computational approach for protein-protein interaction requires huge amount of com-putations on volatile distributed data. Moreover, the service result of PreSPI usually provides only clues for biologists to conduct various attempts rather than the final expected answer. Thus the ser-vice system should be equipped with various means for external applications or systems to effectively access its services. This becomes clear when we consider the situation that one wants to prepare an application that asks a service to PreSPI, wait the result, receive the result, and send the result as an input for other service requests.
Performance, openness, and flexibility are extracted as the three major design goals that PreSPI has to achieve to meet above requirements. The performance goal of PreSPI is achieved by executing most time consuming services in parallel. Although many services of PreSPI requires large amount of computation, we found that they can be executed in parallel without much of additional programming effort. The openness goal of the system is supported by adopting Web Services standards [17, 18, 19] when converting PreSPI service functions into public service. External applications or systems can readily call and use the APIs of PreSPI on Web services infrastructure. For the support of extensibility or flexibility of PreSPI, layered architecture is adopted. Data module and service module are explicitly divided. This layered architecture is appropriate specially for PreSPI. When raw data is newly updated, it can be reflected to the data module of the system without influencing the service module of the system.
This paper is organized as follows. In Section 2, we briefly explain the domain combination based protein-protein interaction prediction method. In Section 3, we explain the major functions and structures of PreSPI. In Section 4, we describe the usage guide for several representative user interfaces of PreSPI. Finally, we draw conclusion in Section 5.

Design and Implementation of Protein-Protein Interaction Prediction Service System 173

2 Domain Combination Based Protein-Protein Interaction Predic-tion Method

In this section, we briefly introduce the background, framework, and validation result of the domain combination based protein-protein interaction prediction method to be implemented in this work. More details of the method can be found in [9] and [10] .

2.1 Motivation

The domain combination based protein-protein interaction prediction method originated from the domain based protein-protein interaction method [5, 14, 16] and some of the drawbacks of domain based approach are eliminated in the domain combination based approach. Most domain based protein-protein interaction prediction methods share the conjecture that protein-protein interaction is the result of domain-domain interaction. Those methods infer domain-domain interacting information from protein-protein interaction and then try to predict protein interactions based on the inferred domain-domain interacting information. But previous domain based researches usually considers only the interactions of single domain pairs. They even assume that the interactions of single domain pairs are independent of one another for computational convenience. As a result, the prediction accuracy of conventional domain based approaches was not so good for the methods to be used in research or industrial fields.

There could be many other reasons for the limitations of conventional domain based approaches but the assumption that single domain pair is the basic unit of protein interaction seems to be the major reason of the limitations. To overcome these limitations, domain combination based approach introduces the notion of domain combination and domain combinations pair (dc-pair). The term domain combination is used to represent the set of domains. Domain combination based approach interprets protein-protein interaction as the result of the interactions of multi-domain pairs or the interaction of groups of domains, i.e., the domain combination based protein-protein interaction pre-diction model considers dc-pair as a basic unit of protein interactions. The clear contrast of domain combination based approach and conventional domain based approach can be found in [9] .

2.2 Prediction Method

In domain combination based protein interaction prediction method, the appearances of domain com-bination pairs of interacting and non-interacting set of protein pairs are registered in matrices. The matrix is called AP (Appearance Probability) matrix. Then a probability equation that maps a pro-tein pair to a real number in the range of 0 to 1 is devised based on the information stored in the matrices. The real number is called PIP (Primary Interaction Probability) value in this paper. When the equation is applied to every protein pair in interacting and non-interacting set of protein pairs, two distributions of PIP values are obtained. Using the two PIP distributions, for an unknown pro-tein pair, its PIP value is computed and the interaction possibility of the protein pair is predicted by deciding to which distribution the PIP value belongs. The schematic view of this process and the details of the method are described in [9] .

2.3 Validation Result

According to [9, 10], the domain combination based protein-protein interaction prediction method shows remarkably better prediction accuracy than conventional domain based prediction methods. But the previous validations of the method missed the point that when there is no common domain between a testing protein pair and the constructed AP matrix, the application of the method is meaningless. That is, when there is no overlapping domain between a protein pair and AP matrix, the domain or domain combination based protein-protein interaction method should not be applied. So in the validation of the method, we eliminated all the protein pairs which contain no overlapping domains with AP matrix in the test set of protein pairs.

174 Han et al.

We used exactly the same way as the previous validations in preparing learning sets of interaction and non-interacting protein pairs. That is, two sets of protein pairs were used. One is the interacting set of protein pairs acquired from DIP database (http://dip.doe-mbi.ucla.edu/) [15], where 15,174 interacting protein pairs in a Yeast organism were prepared. Since not all the proteins in the protein pairs have domain information, only 7,500 interacting protein pairs could be used in the validation. The domain information for the proteins is extracted from PDB (http://www.ebi.ac.uk/proteome/) [1, 2].
On the other hand, the non-interacting set of protein pairs is artificially generated by randomly pairing the reported proteins with domain information in a Yeast organism. Note that there is no publicly announced information on the non-interacting set of protein pairs. Approximately 6,000 proteins are known from Yeast. Among them, 2,700 proteins have domain information and they can be used in the creation of non-interacting sets of protein pairs. 127,700 protein pairs were generated by randomly paring the 2,700 proteins. Then the negative sets of protein pairs were created by randomly selecting required amount of protein pairs from the prepared set when necessary. Since interacting protein pairs could be included as well in the prepared set of protein pairs, we eliminated interacting protein pairs when selecting protein pairs for the preparation of non-interacting sets of protein pairs. For test of prediction accuracy of the method, we divided the interacting and non-interacting sets of protein pairs into learning and testing sets of protein pairs, respectively. Among the data, 80% is used for learning sets and 20% is reserved for test. For the precise evaluation of our protein-protein interaction prediction method, we increased the size of the non-interacting set of protein pairs because it is more natural to assume that there are more non-interacting protein pairs than interacting protein pairs. Note that the protein pairs without overlapping domains in AP matrix are not included in the test data in the measurement.

Table 1 shows the sensitivities and specificities of each test group depending on the ratios of interacting and non-interacting set of protein pairs. The data in each test group is divided further into two subgroups; one group is the test set of protein pairs which has a matching PIP value in PIP distributions and the other group is the test set of protein pairs without matching PIP value in PIP distribution. As shown in Table 1, very high sensitivities and specificities were achieved for the test groups with matching PIP values, whereas moderate sensitivities and specificities were achieved for the test groups without matching PIP values. In the test, it was revealed that protein pairs with common domains in AP matrix are amenable to have matching PIP values in the PIP distributions. Only less than 5% of the protein pairs with common domains in AP matrix had no matching PIP value in the PIP distributions.

As well, the overall prediction accuracy was improved as the relative size of non-interacting set of protein pairs in the training sets was increased. When the size of the non-interacting set of protein pairs was 10 times bigger than that of the interacting set of protein pairs, 77% sensitivity and 95% specificity were achieved for the test protein pairs with common domains in AP matrix.

Table 1: The change of sensitivities and specificities by the ratios of interacting to non-interacting sets of protein pairs in training sets.

I: Protein pairs with matching PIP values.
II: Protein pairs without matching PIP values.

Design and Implementation of Protein-Protein Interaction Prediction Service System 175

Table 2: The sensitivities and specificities from the experiments using DIP, DIP CORE, HMS-PCI and TAP data.

In order to ascertain that the method provides stable prediction accuracies for other data sets,

its prediction accuracy was also measured using DIP CORE [4], HMS-PCI [11] and TAP [8] data. Table 2 shows the sensitivities and specificities of each test group. Only the case when there was matching PIP values in the PIP distributions was considered. As shown in Table 2, quite stable and high prediction accuracies were obtained irrespective of data sets. When the ratio is 10, the accuracy of using DIP data is under those of other cases. This indirectly indicates DIP data contains more erroneous data than the other data sources. On the other hand the prediction accuracy of using TAP data was almost perfect. From this result, we can conclude that the domain combination based protein-protein interaction prediction service system can provide quite reliable computationally predicted protein-protein interaction information for the protein pairs with overlapping domains in AP matrix.

3 PreSPI

In order to make the domain combination based protein-protein interaction method be used by general users, we need to develop a service system that can be easily accessible from the users. In this section, we introduce a protein-protein interaction prediction service system that implements the method in Section 2 and provides some other useful facilities for protein and protein interactions.

3.1 Functions of PreSPI

As explained in introduction part, the functions of PreSPI are largely divided into core functions, subsidiary functions, and general service functions. The core functions are the functions that can be provided only by PreSPI. They include functions such as the interaction prediction for single protein pair, visualization of PIP distribution, category determination of multiple protein pairs, and ranking the interaction possibilities of protein pairs. The subsidiary functions include functions such as search of dc-pairs with high appearance probability on AP matrix, construction of protein interaction network, and report of the prediction system's accuracy. The general service functions include functions such as retrieval of domain information for a given protein and inter-translation of accession Ids (DIPID, SWISSPROTID, PIRJD). While the core functions and subsidiary functions are implemented using the techniques developed by Han et al. [10], the general service functions are the useful functions that can be implemented by collecting and organizing the protein and domain data on the Internet. We leave the details of the functions to PreSPI web site (http : //silver . icu . ac . kr : 8080/toraj im/ index. html).

3.2 Architecture of PreSPI

In general, computational approaches of protein-protein interaction prediction system uses continually updated distributed data and requires huge amount of computations. In PreSPI, several AP matrices with more than one billion entries should be prepared, and dozens of PIP distributions should be obtained based on the AP matrices. Once AP matrices and PIP distributions are prepared, the service functions introduced in the previous subsection can be implemented. But the service time is

176 Han et al.

different from depending on the type of service. Some services take couple of seconds, whereas some services take several hours.

Meanwhile, the service result of PreSPI provides clues for biologists to conduct various attempts rather than the final expected answer. Thus the service system should be equipped with various facilities to effectively access the services from external applications or systems.
In order to meet above requirements, performance, openness, and flexibility are adopted as design goals of PreSPI. Performance design goal is achieved by processing several time taking services in parallel. Interacting protein search is one of the services that can be processed in parallel by dividing a set of target proteins into multiple groups and processing each group independently. We found that many services in PreSPI can be processed in parallel through multi-job parallel processing tech-niques. Note that multi-job parallel processing is usually can be embodied without so much additional programming efforts.
For the openness design goal, PreSPI uses Web Services standards when opening its services
to public. External applications and systems can easily access and use the services as long as it follows web services standards. For the support of flexibility and extensibility, PreSPI adopts layered architecture. Data and service modules are explicitly divided and again service module is divided into three sub-layers. Figure 1 shows the architecture of PreSPI. Data module is responsible for preparing

Figure 1: The software architecture of PreSPI.

the databases that are needed in service module. There are three modules in the data module. One is protein-domain dictionary building sub-module, another is AP matrix construction sub-module and the last is PIP value distribution generation sub-module. Protein-domain dictionary building sub-module gathers information of proteins and domains from external databases on the Internet and correlates them in order to link domain information with protein information. The service module responds to the service requests from users. AP matrix construction sub-module generates two AP matrices corresponding to interacting and non-interacting set of protein pairs respectively. PIP value distribution generation sub-module applies the PIP function to all elements of AP matrices and stores the resulting PIP values into the database. Once the three databases are prepared by data module, PreSPI is ready to service through service module.
Based on the information in databases prepared by data module, the service module provides the

Design and Implementation of Protein-Protein Interaction Prediction Service System 177

functions listed in Table 3. The main function of the service module is to receive and respond to service requests from users. The service module has three layers: UI (user interface) layer, connection layer, and service library routine layer. The UI layer accepts user's service requests, hands them over to the lower function connection layer, and then visualizes the results from the function connection layer. The function connection layer provides channels between UI layer and service library routines. It locates the appropriate library routines for the service request from UI layer. The service library routine layer is the set of various service routines. The separation of service module and data module enhances flexibility and reduces maintenance overhead of PreSPI. New types of service requests can be easily and transparently handled by adding corresponding routines in the service library routine layer, and new function connection routines and UI routines in corresponding layers. When data or information is newly added or updated, it can be easily accommodated in the data module without modifying the service module.

4 Implementation
PreSPI was implemented using Python, Java, and Web services technologies. Data module was imple-mented in Python 2.2.2 and MySQL is used for the database construction. Web interface was used for the access of UI layer and the service module was implemented mainly using Java, Java applet, JSP, and Jython. Jython is used for calling the functions implemented in Java from the code in Python. JDBC driver (mysql-connector-java-3.0.8-stable) was used for accessing MySQL database from Java code. Jakarta-tomcat was used in Apache web server and Axis-1_1 was used as container for the support of web Services. All the functions of PreSPI listed in the previous section were converted into the form of Web Services for public access. Since we adopted layered architecture, this converting was relatively easy and flexible.
In this section, we introduce several representative user interfaces and usage guides of PreSPI for the thorough understanding of PreSPI. Each service web page uniformly contains service description part and usage guide part. Service description part explains the services provided in a page and usage guide part explains how users can receive the service.
4.1 Visualization of PIP Distributions

This page shows the PIP value distributions of interacting and non-interacting set of Yeast protein pairs. Blue lines represent the distribution of non-interacting set of protein pairs and red lines are used to represent the distribution of interacting set of protein pairs. Figure 2 shows a snapshot of regular-interval PIP distributions. Regular-Interval PIP distribution lines up the PIP values of interacting and non-interacting set of protein pairs and shows the frequencies of interacting and non-interacting protein pairs for each PIP value. The intervals between the adjacent PIP values are equivalent irrespective of the adjacent PIP values. Consequently, when there are 10,000 different PIP values in the distributions, the distance between the adjacent PIP values uniformly becomes 1/10,000. PreSPI provides this function because the classification power of using PIP values in distinguishing interacting and non-interacting protein pairs is illustrated better by regular-interval PIP distribution. Although there exists some overlapping between the two distributions, most of the PIP values from non-interacting set of protein pairs are detected near 0, and most of the PIP values from interacting set of protein pairs are detected near 1. From this, we can conclude that PIP value can be used as a classifier for interacting and non-interacting sets of protein pairs. In order to look at absolute value PIP distributions instead of regular-interval PIP distributions, users can select 80% PIP distribution' or 100% PIP distribution' button. 80% PIP distribution and 100% PIP distribution illustrate the

PIP distributions when 80% and 100% of the interacting set of protein pairs in training sets are used respectively for the creation of the distributions.

4.2 Interaction Prediction for Single Protein Pair
In this page, users can get the results of interaction possibility for an input protein pairs. Protein ID should be used for the input of a protein pair. The PIP value is computed and the result PIP

178 Han et al.

Figure 2: Regular-interval PIP distributions.

value is illustrated in PIP distributions. Some additional annotations - whether or not the pair was already confirmed as an interaction pair through experiments and if the computed PIP value has a matching value in PIP distributions - is also provided. The matching value information is important because when there is a matching PIP value in the distributions, the prediction result is more reliable. Accession ID, SWISSPROTID and PIRID are also allowed to be used for the specification of a protein.
Figure 3 shows the prediction result when protein pair <6500N (GID Number), 5307N> is sub-mitted for prediction. From the result, we can figure out that protein 6500N has a domain IPR001126 and protein 5307N is reported to contain 5 domains: IPR002314, IPR002320, IPR004095, IPR004154 and IPR006195. The PIP value of <6500N, 5307N> is computed to 1.0 and we can figure out that this value has a matching PIP value in PIP distributions because the value of In-pip-distribution field is true. This page also shows that the protein pair is not yet confirmed to have interaction through experiment and the computationally predicted interaction probability by PreSPI is 90.34%. Even though those values are not 100% reliable, users can get useful preliminary information on proteins 6500N, 5307N and their interactions in summarized form.

Figure 3: Interaction prediction for single protein pair.

Design and Implementation of Protein-Protein Interaction Prediction Service System 179

4.3 Interaction Prediction for Plural Number of Protein Pairs

When plural number of protein pairs have to be tested for the investigation of their interaction possibilities, the function provided in this page is useful. This page not only predicts the interaction probabilities of input protein pairs but also provides various means to compare the protein pairs. Besides, by allowing plural number of protein pairs to be tested as a bundle, the efforts of users to test protein pairs one by one can be relieved drastically. Users may directly input protein pairs into the table in the page or upload protein pairs from an input file in predefined format.
The protein pairs in the input file are tested either one by one or in parallel, and the results are summarized and listed in the table of the page. In PreSPI, users can change the order of the list by designating a field for sorting in the table.

4.4 Interacting Protein Search
Typically, biologists concentrate on specific proteins in their researches. In that case, the information of proteins that have high probabilities to interact with a specific protein is quite useful. In this page, users submit a specific protein and get a protein list that contains proteins that are computationally expected to interact with the input protein. The interaction probabilities of protein pairs are computed and the protein pairs with high interaction probabilities are listed on top of the list with the interaction probabilities. Currently PreSPI tests approximately 2,700 protein pairs in Yeast when a protein is submitted. Since all the protein pairs should be tested and 5-8 seconds are taken for testing each protein pair, it takes more than 6 hours to completely test the total protein pairs. Nevertheless, biologists can get useful information from this service.
Some other services of PreSPI are not introduced in this paper because of space limitations. For example, domain and domain combination information of a protein also can be retrieved in PreSPI, but we leave the details of the services to PreSPI web site (http : //silver . icu . ac . kr : 8080/toraj im/ index . html) .

5 Conclusion

In this paper, we have designed and implemented a domain combination based protein interaction prediction service system with performance, openness, and flexibility design goals. As expected, several services of the system often require huge amount of computation time and we have confirmed that the performance of some services can be easily enhanced by applying typical parallel processing techniques.
We have found that the Web Services standard is quite useful in achieving openness goal of the system. By adapting the system to the Web Services standard, it can be easily integrated with other systems or applications.
Meanwhile, PreSPI has to reflect the continually updated remote data, which is usually distributed on the Internet, to its database. In order to support this, the system adopted layered architecture and we have found that layered architecture is well suited to this situation. Even though PreSPI is designed to implement the domain combination based protein interaction method, its architecture can be used as a reference model for other systems in similar situation.
Currently only Yeast proteins can be handled in PreSPI and it still lacks many useful functions for biologists. In future, we are planning to extend PreSPI so that it can be able to handle proteins of other organisms such as C. elegance, Drosophila, E. coli., Mouse, and Human. Receiving feedbacks from the actual users of PreSPI and reflecting the requests to the system is essential for the success of PreSPI.

Acknowledgments

This work was funded in part by the Korea Institute of Science and Technology Information (Grant No. K-04-BI-16-01S-4) and LG Life Science. Dr. J. K. Suh was supported partly by the 21C Fron-

180 Han et al.

tier Microbial Genomics and Application Center Program, Ministry of Science & Technology (Grant No. MG02-0302-003-1-0-0), Republic of Korea. We are also grateful to Ms. Smita Kurkuri for her help in English correction. Comments from reviewers were also great helpful to us in improving the presentation of this paper.

References

[1] Apweiler, R., et al., The InterPro database, an integrated documentation resource for protein families, domains and functional sites, Nucleic Acids Res., 29:37-40, 2001.

[2] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E., The protein data bank, Nucleic Acids Res., 28:235-242, 2000.

[3] Bock, J.R. and Gough, D.A., Prediction of protein-protein interaction from primary structure, Bioinformatics, 17:455-460, 2001.

[4] Deane, C.M., Salwinski, L., Xenarios, I., and Eisenberg, D., Protein Interactions: Two methods for assessment of the reliability of high throughput observations, Mol. Cell. Proteomics, 1:349-356, 2002.

[5] Deng, M., Mehta, S., Sun, F., and Chen, T., Inferring domain-domain interactions from protein-protein interactions, Genome Research, 12:1540-1548, 2002.

[6] Enright, A.J., Iliopoulos, I., Kyrpides, N.C., and Ouzounis, C.A., Protein interaction maps for complete genomes based on gene fusion events, Nature, 402:86-90, 1999.

[7] Enright, A.J. and Ouzounis, C.A., Molecular cloning manual, Protein-Protein Interactions, Cold Spring Harbor Laboratory Press, 2002.

[8] Gavin, A., et al., Functional organization of the yeast proteom by systematic analysis of protein complexes, Nature., 415:141-147, 2002.

[9] Han, D., Kim, H., Seo, J., and Jang, W., A domain combination based probabilistic framework for protein-protein interaction prediction, Genome Informatics, 14:250-259, 2003.

[10] Han, D., Kim, H., Jang, W., and Lee, S. Domain combination based protein-protein interaction possibility ranking method, Proc. 4th IEEE Sym. Bioinfo. Bioeng., 434-441, May, 2003.

[11] Ho, Y., et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature, 415:180-183, 2002

[12] Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O., and Eisenberg, D., Detecting protein function and protein-protein interactions from genome sequences, Science, 285:751-753, 1999.

[13] Ng, S., Zhang, Z., and Tan, S., Interpom: A databse of putative interacting protein domains for validating predicted protein interactions and complexes. Nucleic Acids Res., 31:251-254, 2003.

[14] Ng, S., Zhang, Z., and Tan, S., Integrative approach for computationally inferring protein domain interactions. Bioinformatics, 19:923-929, 2003.

[15] Xenarios, I., Fernandez, E., Salwinski, L., Duan, X.J., Thompson, M.J., Marcotte, E.M., and Eisenberg, D., DIP: The database of interacting proteins: 2001 update, Nucleic Acids Res., 29:239-241, 2001.

[16] Wojcik, J. and Schachter, V., Protein-Protein interaction map inference using interacting domain profile pairs, Bioinformatics, 17 Suppl:S296-S305, 2001.

[17] W3C Web Services Architecture Working Group, Web Services Architecture, World Wide Web Consortium, http://www. w3.org/TR/ws-arch/, Feb., 2004.

[18] W3C Web Services Architecture Working Group, Web Services Architecture Requirements, World Wide Web Consortium, http://www.w3.org/TR/wsa-reqs/, Feb., 2004.

[19] W3C Web Services Architecture Working Group, Web Services Architecture Usage Scenarios, World Wide Web Consortium, http://www.w3.org/TR/ws-arch-scenarios/, Feb., 2004.