Plant Biotechnology 26, 495–501 (2009)

Original Paper

Special Issue

Predicting conformation of protein complexes by determining statistically significant domain–domain interactions

Kensaku Nishikata[1,2], Masayoshi Wada[1], Hiroki Takahashi[1], Kensuke Nakamura[1], Shigehiko Kanaya[1,3], Md Altaf-Ul-Amin[1,*]

[1] Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma, Nara 630-0101, Japan; [2] VALWAY Technology Center, NEC Soft Ltd., Tokyo 136-8627, Japan; [3] JST, CREST, Kawaguchi, Saitama 332-0012, Japan

* Email: amin-m@is.naist.jp Tel: +81-743-72-5257 Fax: +81-743-72-5258

Received September 18, 2009; accepted October 19, 2009 (Edited by D. Ohta)

Abstract The present study proposes a method to predict the conformation of protein complexes by using statistically significant domain–domain interactions (DDIs). High-throughput methods for detecting protein interactions generate a significant number of false-positives, and especially the combinatorial method of protein-complex purification and mass spectrometry detect both direct and non-direct interactions i.e. "bait–prey" and "prey–prey" interactions making it difficult to predict the conformation of complexes. Therefore in this work we utilized the DDIs as a means to support the interactions and subsequently to predict the conformation of complexes. As the first step, we extracted 312 statistically significant DDIs out of 1,162 DDIs underlying 3, 118 protein–protein interactions (PPIs) of *Arabidopsis thaliana* by using Fisher's exact test. Next, 67 protein complexes were obtained by applying a graph clustering algorithm to the PPI network. Finally, we discussed the conformation of protein complexes based on DDI information extracted in the first step. Information on significant DDIs can also be utilized to annotate unknown function proteins and to predict localization of proteins with confidence.

Key words: Conformation of protein complexes, domain–domain interaction (DDI), protein–protein interaction (PPI).

More or less over the past 10 years, vast amount of protein–protein interaction data have been generated by high-throughput methods for detecting protein interactions, such as the combinatorial method of protein-complex purification with subsequent analysis by mass spectrometry (MS) and the yeast two-hybrid (Y2H) system (Uetz et al. 2000; Ito et al. 2001; Ho et al. 2002; Gavin et al. 2002). However, there being no complete and accurate detection method, each experimental strategy generates a significant number of not only false-negatives but also false-positives (Titz et al. 2004). False-positives are usually a more serious problem because they cause erroneous results and misleading conclusions, making PPI analysis complicated and difficult.

The methods such as protein-complex purification and MS, and Y2H analysis have propensity to detect different kinds of protein interactions. For example, the combinatorial method of pull-down assay and MS identifies stable interactions such as those in protein complexes, whereas Y2H more often find transient interactions. In case of the PPI detection method using pull-down assay and MS, protein complexes are often isolated in an affinity purification experiment in which a single protein (the "bait") is provided with a molecular tag such as FLAG (Ho et al. 2002) , TAP (Gavin et al. 2002) or His-tag (Arifuzzaman et al. 2006), then proteins ("preys") in isolated complexes are identified using subsequent analysis by MS. Isolation of protein complexes in this procedure allows the purification of the "bait" together with all of the "prey" proteins that belong to the same multi-protein complex. The problem seems to lie in the fact that every "prey" protein doesn't directly interacts with the "bait" protein; rather, the topology of the complex will include both "bait–prey" and "prey–prey" type interactions (Hakes et al. 2007). It implies that the conformation, i.e., the true topology of the protein complexes cannot be determined from the individual experiments only. In case that proteins *a*, *b* and *c* are identified in isolation as a protein complex using protein 'a' with a tag as a "bait", three PPIs can be obtained (*a*-*b*, *a*-*c*, and *b*-*c*), but it is difficult to conclude what is the true conformation of the complex out of the four possible cases as shown in i)~iv) in Figure 1. Though understanding the conformation of protein complexes is necessary in order to obtain useful information about them, little attention has been given by the bioinformatics researchers to decipher the true topology of the protein complexes. Also the

This article can be found at http://www.jspcmb.jp/

experimental technologies used to detect PPI do not focus on this matter.

The bioinformatics analysis of PPI has mainly followed two different approaches after the high throughput experiments started to produce huge amount of data. One of the approaches is the analysis of protein interaction networks based on graph theory, aiming to detection of protein complexes from PPIs networks (Bader and Hogue 2003; Altaf-Ul-Amin et al. 2006). These studies have reported that the densely connected regions in a network correspond to known protein complexes or protein functional units. The other approach is the analysis of DDIs coming out from PPI data by statistics or machine learning intended to predict unknown PPIs (Sprinzak and Margalit 2001; Riley et al. 2005; Singhal and Resat 2007; Liu et al. 2009). These studies have shown that the concept of DDIs statistically extracted from large-scale PPI data can explain the makeup of PPIs to some extent.

Taking these two recent approaches into account, the present study focuses on statistically significant DDIs i.e. the direct interactions to predict the conformation of protein complexes by avoiding the effects of false-positives or non-direct interactions. Corresponding to Figure 1, if DDI analysis supports that protein *b* and *c* directly interacts with different domains of protein *a* then we would predict the conformation of ii) and on the other hand if protein *b* and *c* directly interacts with an identical domain of protein *a* competitively then we would predict the conformation of iv). Further detail prediction of interactions in protein complexes were performed based on DDIs.

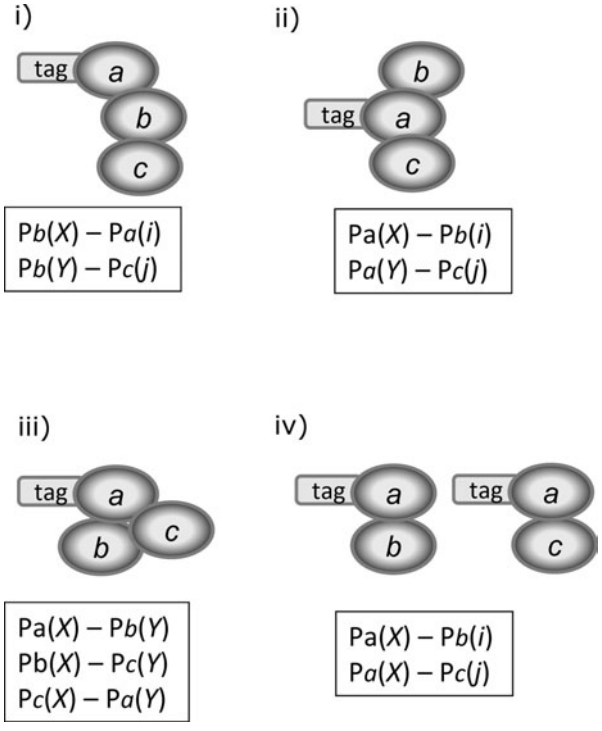

Figure 1. Conformation diversity of protein complexes in case that proteins *a*, *b* and *c* are identified as a complex by MS.

Methods

In the present study, we propose a procedure for inferring conformation of protein complexes with reliable PPIs by using information of DDIs (Figure 2), which comprises three steps, (Step 1) Detection of domains in proteins, (Step 2) Extraction of statistically significant DDIs, and (Step 3) validation of protein complexes based on extracted DDIs.

Step 1: Detection of domains in proteins

The InterProScan, which is a tool that combines different protein signature recognition methods into one resource, can detect protein families, domains, repeats and functional sites containing post-translational

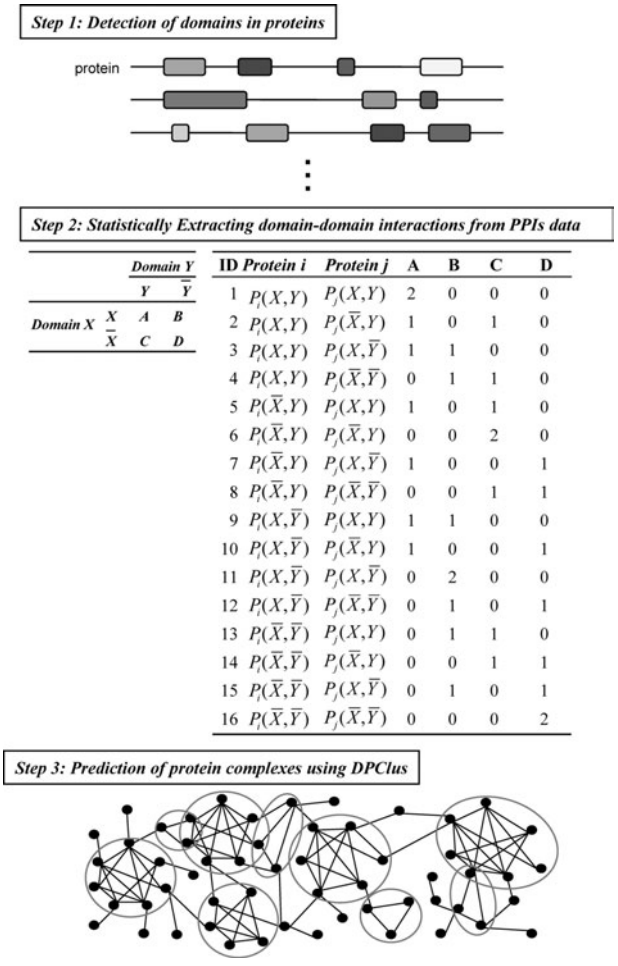

Figure 2. Procedure to predict conformation of protein complexes based on DDI, Firstly, domains were extracted by InterProScan (Step 1), then statistically important co-occurred domain-domain pairs were extracted by Fisher's exact test using 2×2 contingency table (Step 2), and, in Step 3, validation of protein complexes were carried out using the domain-domain pairs recognized in Step 2.

modification sites (Zdobnov and Apweiler 2001). In the present study the domains present in a protein were detected by InterProScan. The protein signature by InterProScan has the hierarchical structure, that is, a parent/child relationship between two signatures is defined in the output of InterProScan, and the parent is the entry containing a more general signature, while the children are more specific to certain members of the signature. In the context of protein interaction, domains or smaller peptide motifs act as recognition elements, therefore domains, repeats and functional sites of second depth in hierarchy but not families by InterProScan were simply used as domains in this DDI analysis.

Step 2: Statistically extracting domain–domain interactions from PPI data

Statistical analysis based on 2×2 contingency table was applied to detect significant relation between a domain pair by judging their presence or absence in a set of interacting protein pairs (Figure 2). Concerning the presence or absence of two domains say, *X* and *Y* in two interacting proteins say, *i* and *j* there could be 16 combinations and Figure 2 shows how we counted *A*, *B*, *C* and *D* of the contingency table corresponding to each combination. The null hypothesis is that the occurrence of domain *X* in a protein and the occurrence of domain *Y* in the other protein in a PPI are independent of each other. So, the test of independence between domain *X* and domain *Y* was performed using Fisher's exact test with significance level $\alpha = 0.01$, taking multiple hypotheses into consideration, that is, Bonferroni's correction was adopted in order to avoid statistical significance that might occur by chance. To determine significant DDIs, we statistically tested potential DDIs, containing self-DDIs, for which a protein has a domain and the other protein has another domain in at least one PPI, i.e., the count *A* on the contingency table in Figure 2 is not lower than one.

Step 3: Validation of protein complexes generated by DPClus

Firstly, protein complexes were predicted by applying DPClus (Altaf-Ul-Amin et al. 2006) to the whole PPI network. DPClus detects densely connected regions of a graph comprising nodes and edges as clusters which correspond to protein complexes in case of a PPI network. The PPIs in the complexes were validated by significant domain–domain pairs obtained in Step 2 which in turn helps to cast insight into probable conformation of the complexes. In summary we propose that to predict the conformation of a complex the significant DDIs should be given priority over the PPIs.

Results and discussion

Collection and integration of PPIs data

The PPIs of the *Arabidopsis* interactome were collected by following two procedures: collecting from public PPI databases and manual collecting from research papers. Any computationally predicted PPI was excluded from this study. In the first procedure, the PPI data was assembled from BIND (Bader et al. 2000; 2003), DIP (Xenarios et al. 2002), MINT (Zanzoni et al. 2002; Chatr-aryamontri et al. 2007), HPRD (Peri et al. 2003) and IntAct (Hermjakob et al. 2004) which are major PPI data resources accepting experimentally determined PPIs from research papers. In the second procedure, 946 PPIs were manually gathered by reading experimental research papers. PPI data redundancies were removed by mapping PPI information onto the Arabidopsis gene codes (AGI codes), and, as a result an integrated PPI data, 3,118 PPIs composed of 1,302 Arabidopsis proteins was obtained.

Significant domain–domain interactions extracted from PPI data

Proteins must physically bind to other proteins, either stably or transiently, to perform their functions. Interaction specificity results from the binding of a modular domain to another domain or smaller peptide motif in the target protein (Pawson and Nash 2003). For example, some cytoskeletal proteins bind to actin through their modular gelsolin repeat domains (McGough et al. 2003), and Src-homology 3 domains (SH3) bind to proline rich peptides that have a PxxP consensus sequence (Lim et al. 1994). In the context of protein interaction, such domains and peptides act as recognition elements; we refer to these binding domains or recognized peptides simply as 'domains' in this study. Over the past few years with developments of high-throughput PPI detection technologies, many researchers have shown an interest in extracting domain–domain interactions (DDIs) from large-scale PPI data by statistical methods, demonstrating that the idea of DDIs explain the cause of PPIs in some measure (Sprinzak and Margalit 2001; Riley et al. 2005; Singhal and Resat 2007; Liu et al. 2009). Here, we statistically extracted DDIs from integrated PPI data of Arabidopsis by following a procedure described in the 'Method' section. Total 312 significant DDIs were obtained (Fisher's exact test, $\alpha = 0.01$ with Bonferroni correction) out of 1,162 potential DDIs for which a protein has a domain and the other protein has another domain in at least one PPI.

PPI detection technologies experimentally provide the information about existence of interaction, but usually no direct information about the domains and peptides which act as recognition elements or binding sites, and determining binding domains and peptides in proteins

requires further analysis. Therefore, as a benchmark for true-positive DDIs, we used pairs of domains reported to interact in determined structures of protein complexes in iPfam (Finn et al. 2005). In iPfam, two domains are defined as interacting if they are close enough to form at least one interaction based on available PDB structures. It should be noted that known set of interacting domain pairs determined from structures are only a small fraction of all DDIs that may exist, i.e. though these are the gold standard DDIs, it is certainly possible that predicted DDIs are also true and structures which contains predicted DDIs have not been determined yet. According to a recent study (Itzhaki et al. 2006), DDIs in iPfam and 3DID (Stein et al. 2005) databases could explain no more than 20% of the PPIs for any of the *E. coli*, *S. cerevisiae*, *C. elegans*, *D. melanogaster*, and *H. sapien*, suggesting that the number of known DDIs is rather small. So we used iPfam to assess true-positive rate with respect to P-value after Bonferroni's correction for DDIs and 70% of the known DDIs have been recognized within the threshold of 0.01 showing the effectiveness of our method (Figure 3A). Using this threshold, 312 statistically significant pairs of interacting domains were obtained (Figure 3B) and if we add 20 other gold DDIs not detected by our method then the number of total significant DDIs is 332. Now 20% of the PPIs can be explained by 66 DDIs and if we consider a linear relation then 330 DDIs are required to explain 100% PPIs which almost matches with the number 332 implying that the proposed DDI prediction method is a good one.

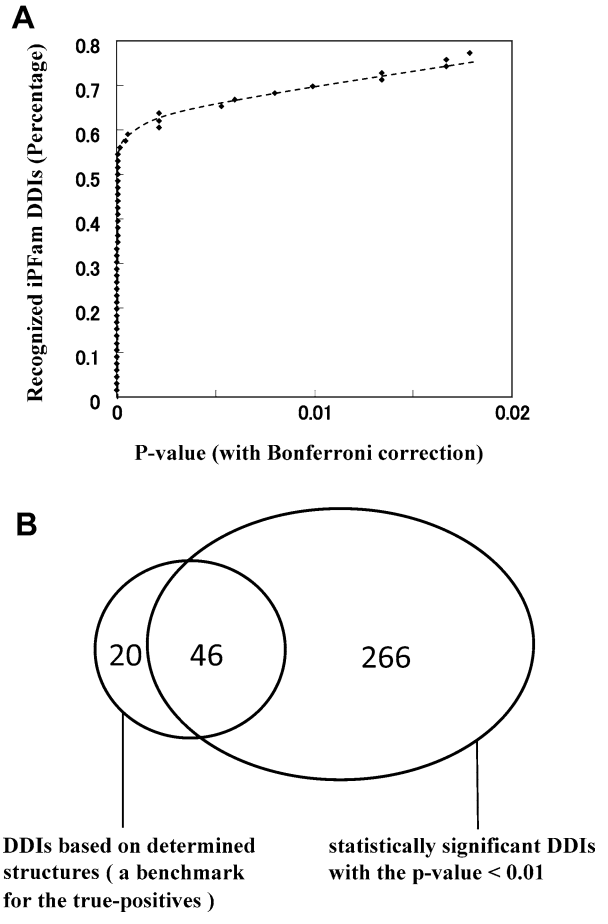

Figure 3. Validation of statistically extracted DDIs using iPfam. (A) True-positive DDIs with the p-value; (B) True-positive DDIs among the statistically significant DDIs. By iPfam, 66 reported DDIs were obtained, of them, 46 DDIs were included in statistically significant domain–domain pairs determined by the proposed method (Step 2 in Figure 2). So true-positive rate is estimated as 0.70.

Prediction of protein complexes

Proteins interact with other proteins in complexes to perform cellular functions. In the past, we developed an algorithm called "DPClus", which extracts the densely connected regions in a network and demonstrated that many of these densely connected regions correspond to known protein complexes or protein functional units (Altaf-Ul-Amin et al. 2006). DPClus is a robust algorithm not affected by high rate of false positives in data from high-throughput interaction-detection techniques. While predicting the protein complexes by DPClus, we adopted the "overlapping clustering mode," which allows identical proteins to be classified into different clusters, because it is biologically well established that proteins can be present in multiple complexes at different times and locations. By setting three parameters as 0.7 for network density, 0.5 for cluster property and 3 for least number of members in a cluster, 67 protein complexes were obtained from 3,118 PPIs (Figure 4). In the present study, 1,629 out of 3,118 PPIs were supported by statistically significant domain-domain pairs (Appendix Table 1). Additionally, using significant DDI information, we predicted the conformation of all protein complexes detected by DPClus. Figure 5 shows some examples, in which each node and each edge represent a protein and an interaction respectively: blue and red edges represent interactions determined by PPI experiments and interactions supported by statistically significant domain–domain pairs (suggesting direct interaction) respectively. The complex represented by cluster 41 in Figure 5 is composed of At1g16970, At1g48050 and At4g13870, and according to the present DDI analysis, domain IPR005160 of At1g16970 and domain IPR006164 of At1g48050 contact with identical domain IPR002562 of At4g13870. Thus this complex is explainable by the competitive interactions of two proteins, At1g16970 and At1g48050, with domain IPR002562 of At4g13870. The cluster 23 consists of 5 proteins and 3 of its 7 PPIs are supported by significant DDIs. At1g16240 has two domains (IPR000727 and IPR010989) to interact with At1g28490 and At5g46860, respectively, and At1g28490 also has two domains (IPR000727 and IPR010989) to

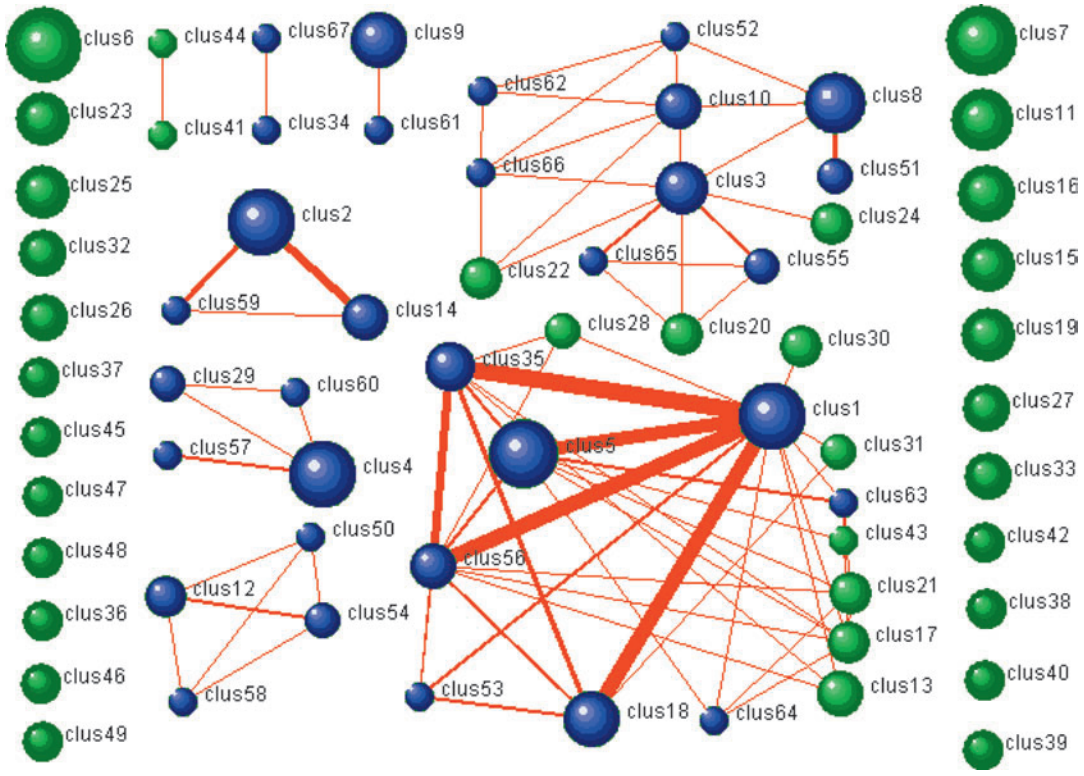

Figure 4. Complexes determined by DPClus in Arabidopsis PPI network. The more the number of proteins in a complex the bigger is its size and the more the number of interactions between two complexes the thicker is the corresponding edge. Complexes of blue color contain one or more proteins that are shared by more than one complex while green complexes do not contain such proteins.

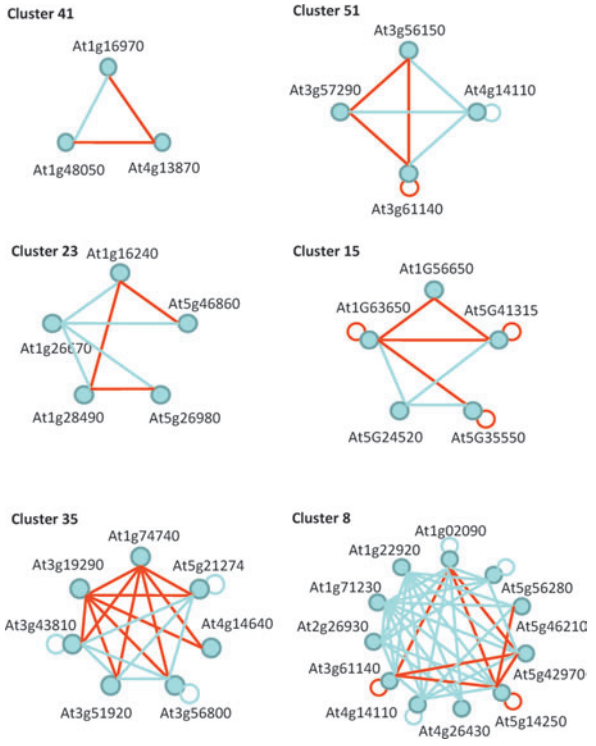

Figure 5. Prediction of the conformation in protein complexes using significant DDI information. The nodes and edges represent proteins and interactions respectively. Blue edges show original interactions, and red edges show interactions supported by significant DDIs.

interact with At1g16240 and At5g26980. This can be explained by the consecutively ordered interaction of four proteins At5g46860, At1g16240, At1g28490 and At5g26980. In the complex 35, it is presumed from the graph that At1g74740 and At3g19290 play a vital role. In fact, in At1g74740, three domains related to interaction with each of the other proteins in the complex were detected whereas, At3g19290 has only one domain to interact with other proteins in the complex, suggesting that other proteins in the complex competitively interact with At3g19290 case by case in different conditions. Complexes 51, 15 and 8 also contain reasonable number of significant domain–domain interactions and such information might be used to predict their 3D conformations. Thus, further detail prediction of interactions in protein complexes were performed based on DDIs.

Conclusion and remarks

This work has proposed a method to predict the conformation of protein complexes by using domain–domain interactions (DDIs). Significant DDIs were determined by statistical analysis using Fisher's exact test based on 2×2 contingency tables. In the present study, 312 significant DDIs underlying 3, 118 protein–protein interactions (PPIs) were obtained and 1,629 out of 3,118

PPIs were supported by statistically significant domain–domain pairs. Furthermore, we generated 67 protein complexes composed of 1,302 proteins by applying a graph clustering algorithm to the protein interaction network and provided explanation to predict the conformation of the complexes in view of significant DDIs. Mainly, to interpret the conformation of protein complexes we configured the interactions supported by DDIs. Information on significant DDIs can be utilized to annotate unknown function proteins and to predict localization of proteins with confidence. A further study of conformation of protein complexes from PPI data should be conducted, which can also help computer simulation of protein complexes to develop new drugs. The more understanding of conformation of protein complexes would give new clues to development of drugs.

Acknowledgements

This work was supported in part by the Japan Science and Technology Agency, CREST (Project name "Elucidation of Amino Acid Metabolism in Plants based on Integrated Omics Analyses").

References

Altaf-Ul-Amin M, Shinbo Y, Mihara K, Kurokawa K, Kanaya S (2006) Development and implementation of an algorithm for detection of protein complexes in large interaction networks. *BMC Bioinformatics* 7: 207

Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, et al. (2006) Large-scale identification of protein–protein interaction of Escherichia coli K-12. *Genome Res* 16: 686–691

Bader GD, Betel D, Hogue CW (2003) BIND: the Biomolecular Interaction Network Database. *Nucl Acids Res* 31: 248–250

Bader GD, Hogue CW (2000) BIND-a data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics* 16: 465–477

Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2

Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G (2007) MINT: the Molecular INTeraction database. *Nucl Acids Res* 35: D572–574

Chen XW, Liu M (2006) Domain Based Predictive Models for Protein–Protein Interaction Prediction. *EURASIP J on Applied Signal Processing:* 1–8

Deng M, Mehta S, Sun F, Chen T (2002) Inferring domain–domain interactions from protein–protein interactions. *Genome Res* 12: 1540–1548

Finn RD, Marshall M, Bateman A (2005) iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics* 21: 410–412

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, et al. (2008) The Pfam protein families database. *Nucl Acids Res* 36: D281–288

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141–147

Gong S, Park C, Choi H, Ko J, Jang I, Lee J, Bolser DM, Oh D, Kim DS, Bhak J (2005) A protein domain interaction interface database: InterPare. *BMC Bioinformatics* 6: 207

Guimarães KS, Jothi R, Zotenko E, Przytycka TM (2006) Predicting domain–domain interactions using a parsimony approach. *Genome Biol* 7: R104

Hakes L, Robertson DL, Oliver SG, Lovell SC (2007) Protein interactions from complexes: a structural perspective. *Comp Funct Genomics*: 49356

Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, et al. (2004) IntAct: an open source molecular interaction database. *Nucl Acids Res* 32: D452–455

Hishigaki H, Nakai K, Ono T, Tanigami A, Takagi T (2001) Assessment of prediction accuracy of protein function from protein–protein interaction data. *Yeast* 18: 523–531

Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et al. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 415: 180–183

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* 98: 4569–4574

Itzhaki Z, Akiva E, Altuvia Y, Margalit H (2006) Evolutionary conservation of domain–domain interactions. *Genome Biol* 7: R125

Jothi R, Cherukuri PF, Tasneem A, Przytycka TM (2006) Co-evolutionary analysis of domains in interacting proteins reveals insights into domain–domain interactions mediating protein–protein interactions. *J Mol Biol* 362: 861–875

Kann MG, Jothi R, Cherukuri PF, Przytycka TM (2007) Predicting protein domain interactions from coevolution of conserved regions. *Proteins* 67: 811–820

Kemmeren P, van Berkum NL, Vilo J, Bijma T, Donders R, Brazma A, Holstege FC (2002) Protein interaction verification and functional annotation by integrated analysis of genome-scale data. *Mol Cell* 9: 1133–1143

Lee H, Deng M, Sun F, Chen T (2006) An integrated approach to the prediction of domain–domain interactions. *BMC Bioinformatics* 7: 269

Lim WA, Richards FM, Fox RO (1994) Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains. *Nature* 372: 375–379

Littler SJ, Hubbard SJ (2005) Conservation of orientation and sequence in protein domain–domain interactions. *J Mol Biol* 345: 1265–1279

Liu M, Chen XW, Jothi R (2009) Knowledge-guided inference of domain–domain interactions from incomplete protein–protein interaction networks. *Bioinformatics* 25: 2492–2499

Liu Y, Liu N, Zhao H (2005) Inferring protein–protein interactions through high-throughput interaction data from diverse organisms. *Bioinformatics* 21: 3279–3285

McGough AM, Staiger CJ, Min JK, Simonetti KD (2003) The gelsolin family of actin regulatory proteins: modular structures, versatile functions. *FEBS Lett* 552: 75–81

Mrowka R, Patzak A, Herzel H (2001) Is there a bias in proteome research? *Genome Res* 11: 1971–1973

Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, Lewis J, Serrano L, Russell RB (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol* 3: e405

Ng SK, Zhang Z, Tan SH (2003) Integrative approach for

computationally inferring protein domain interactions. *Bioinformatics* 19: 923–929

Nye TM, Berzuini C, Gilks WR, Babu MM, Teichmann SA (2005) Statistical analysis of domains in interacting protein pairs. *Bioinformatics* 21: 993–1001

Pagel P, Wong P, Frishman D (2004) A domain interaction map based on phylogenetic profiling. *J Mol Biol* 344: 1331–1346

Pawson T, Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* 300: 445–452

Pazos F, Valencia A (2008) Protein co-evolution, co-adaptation and interactions. *EMBO J* 27: 2648–2655

Pereira-Leal JB, Teichmann SA (2005) Novel specificities emerge by stepwise duplication of functional modules. *Genome Res* 15: 552–559

Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, et al. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* 13: 2363–2371

Raghavachari B, Tasneem A, Przytycka TM, Jothi R (2008) DOMINE: a database of protein domain interactions. *Nucl Acids Res* 36: D656–661

Riley R, Lee C, Sabatti C, Eisenberg D (2005) Inferring protein domain interactions from databases of interacting proteins. *Genome Biol* 6: R89

Schuster-Bockler B, Bateman A (2007) Reuse of structural domain–domain interactions in protein networks. *BMC Bioinformatics* 8: 259

Shoemaker BA, Panchenko AR, Bryant SH (2006) Finding biologically relevant protein domain interactions: conserved binding mode analysis. *Protein Sci* 15: 352–361

Singhal M, Resat H (2007) A domain-based approach to predict protein–protein interactions. *BMC Bioinformatics* 8: 199

Sleno L, Emili A (2008) Proteomic methods for drug target discovery. *Curr Opin Chem Biol* 12: 46–54

Sprinzak E and Margalit H (2001) Correlated sequence-signatures as markers of protein–protein interaction. *J Mol Biol* 311: 681–692

Stein A, Russell RB, Aloy P (2005) 3 did: interacting protein domains of known three-dimensional structure. *Nucl Acids Res* 33: D413–417

Titz B, Schlesner M, Uetz P (2004) What do we learn from high-throughput protein interaction data? *Expert Rev Proteomics* 1: 111–121

Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. (2000) A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. *Nature* 403: 623–627

von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417: 399–403

Wang H, Segal E, Ben-Hur A, Li QR, Vidal M, Koller D (2007) InSite: a computational method for identifying protein–protein interaction binding sites on a proteome-wide scale. *Genome Biol* 8: R192

Werner F (2007) Structure and function of archaeal RNA polymerases. *Mol Microbiol* 65:1395–1404

Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucl Acids Res* 30: 303–305

Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, Cesareni G (2002) MINT: a Molecular INTeraction database. *FEBS Lett* 513: 135–140

Zdobnov EM and Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848