Assessment of high-confidence protein–protein interactome in yeast

Kubra Karagoz, Kazim Yalcin Arga*

Department of Bioengineering, Marmara University, Goztepe 34722, Istanbul, Turkey

A R T I C L E   I N F O

Article history:
Received 9 February 2013
Received in revised form 15 March 2013
Accepted 15 March 2013

Keywords:
Protein interaction network
Protein–protein interactions
High-confidence
Saccharomyces cerevisiae
Yeast

A B S T R A C T

The identification of protein–protein interactions (PPIs) and their networks is vitally important to systemically define and understand the roles of proteins in biological systems. In spite of development of numerous experimental systems to detect PPIs and diverse research on assessment of the quality of the obtained data, a consensus – highly reliable, almost complete – interactome of Saccharomyces cerevisiae is not presented yet. In this work, we proposed an unsupervised statistical approach to create a high-confidence yeast PPI network. For this, we assembled databases of interacting protein pairs for yeast and obtained an extremely large PPI dataset which comprises of 135 154 non-redundant interactions between 6191 yeast proteins. A scoring scheme considering eight heterogeneous biological features resulted with a broad score distribution and a highly reliable network consisting of 29 046 physical interactions with scores higher than the threshold value of 0.85, for which sensitivity, specificity and coverage were 86%, 68%, and 72%, respectively. We evaluated our method by comparing it with other scoring schemes and showed that reducing the noise inherent in experimental PPIs via our scoring scheme further increased the accuracy. Current study is expected to increase the efficiency of the methodologies in biological research which make use of protein interaction networks.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Proteins interact with each other in a highly specific manner that these interactions play key roles in biological systems such as signal transduction pathways and regulation of biological processes. Therefore, comprehensive identification and characterization of protein–protein interactions (PPIs) and their networks is essential in order to systematically define and understand the information transfer within and between biological processes. As a result of the accumulation of genome sequence information and development of different genomic and proteomic experimental technologies for identification of PPIs (Zhu and Snyder, 2003), enormous amounts of data have been generated and databases storing the experimentally achieved protein–protein interactions (Aranda et al., 2010; Ceol et al., 2010; Günden et al., 2006; Salwinski et al., 2004; Stark et al., 2011 and Szklarczyk et al., 2011) have expanded significantly during the last decade. Unfortunately, the accuracy of these data is often criticized, since PPI data obtained from high-throughput experiments is thought to have a large number of false positives, i.e. the interactions that are spurious and do not occur in the cell (Mrowka et al., 2001 and von Mering et al., 2002). Especially the large amount of data produced by the high-throughput experimental methods need to be analyzed and verified before employment in biological research, otherwise these false positives will lead to spurious discoveries that can be potentially costly, e.g. wrong signaling routes, or wrong drug targets for diseases (Arga et al., 2007 and Tekir et al., 2009). Validation of PPI data is a difficult task.

A number of approaches have been proposed to assess the probability that an interaction is a biologically relevant true positive (von Mering et al., 2002; Iossifov et al., 2004; Chen et al., 2002; Deane et al., 2002; Deng et al., 2003; Edwards et al., 2002; Ewing et al., 2007; Giot et al., 2003; Jansen et al., 2002; Krogan et al., 2006; Kuchaiev et al., 2009; Perrish et al., 2007; Patil and Nakamura, 2005; Qi et al., 2005; Saito et al., 2003; Sambourg and Mieg, 2010; Scott and Barton, 2007; Sharaf et al., 2005; Sprinzak et al., 2003; Titz et al., 2008; von Mering et al., 2005; Yamanishi et al., 2004 and Yu and Finley, 2009). First attempts were on designing methods taking into account an individual biological feature. The expression profile reliability method was based on the observation that interacting proteins are co-expressed (Deane et al., 2002). Two distributions of expression distances were predefined for non-interacting and reliably interacting protein sets. The distribution of expression distances for a protein dataset of interest was assumed to be a linear combination of two predefined distributions and the linearity coefficient was used in characterization of the accuracy of a given dataset. However, this method was useful for only datasets of proteins and is not applicable for individual interactions. A paralogues verification method by use of sequence homology information was also proposed. An interaction was judged to be true if the concerned...
proteins had paralogues that interact as well. But their results were limited by the number of proteins that have known paralogues.

The structural information of proteins was also employed to assess the quality of interactions (Edwards et al., 2002). Although this method has a high reliability, it is strictly limited by the number of 3-D protein structures available in databanks. von Mering and co-workers (von Mering et al., 2002) identified the interactions that are observed in more than one high-throughput experiment and used them to estimate the fraction of true positives in the datasets. They hypothesized that if an interaction is supported by several methods, it should be more likely to be true. However, this idea results in ignoring true interactions that are only supported by one technique, e.g. transient and rare interactions. In addition, the overlaps between data-sets were surprisingly small to come to a conclusion. Protein localization method was also proposed (Sprinzak et al., 2003), which defined true positives as interacting proteins that are localized in the same cellular compartment and/or interacting proteins that are annotated to have a common cellular role. Their results showed that the accuracy of the data strongly depends on the experimental technique. In the following years, network topology was also used in identification of true interactions (Jossifov et al., 2004; Chen et al., 2002 and Saito et al., 2003). Although these methods have low false positive rates, they also have low true positive rate, due to the scale-free topology of protein interaction networks, i.e. the number of proteins with more than one interaction partners is relatively few. Since the performance of studies integrating individual genomic features with protein interaction data were not satisfactory, it has been shown that the use of a combination of genomic features, such as sequence, structure and functional annotation information, within a Bayesian network (Jansen et al., 2002 and Patil and Nakamura, 2005) or logistic regression modeling framework (Yu and Finley, 2009) results in a more accurate prediction of the true interactions in high throughput datasets. Moreover, MINT (Ceol et al., 2010), STRING (Szklarczyk et al., 2011), and HitPredict (Patil et al., 2011) databases provide scores that were calculated by using several biological features.

A disadvantage of most of the reported scoring schemes is that they work within a dataset obtained from a single type of experimentation or database. The scores designed for different datasets are mostly not comparable to each other, and for the interactions provided at other databases, a score assignment is not possible. Another disadvantage of these scoring approaches is that they use training data that consists of interactions that are only assumed to be true. Since these datasets are constructed based on simple assumptions, the approaches may produce training sets that are biased. A major disadvantage of all reported scoring approaches is that the high confidence data-sets resulting from these scoring schemes are poor in terms of coverage and completeness, i.e. these data-sets do not represent an interacting network, but represent numerous clusters of interacting proteins.

In spite of various studies on assessing the quality of protein–protein interaction data, a consensus – highly reliable, almost complete – PPI network of yeast is not presented yet. A dataset of high accuracy will not be very useful in biological research if its coverage is low. The main objective of the present study was to fill up this gap by integrating multiple heterogeneous biological features and linking them onto a huge set of PPIs obtained from assembling of different publicly available databases.

2. Materials and methods

2.1. Collection of PPI data

Five public databases, IntAct (Aranda et al., 2010), MINT (Ceol et al., 2010), MIPS (Güldener et al., 2006), DIP (Salwinski et al., 2004), and BioGRID (Stark et al., 2011), use the results of high-throughput experiments, and collect experimentally or computationally determined PPIs from the literature as well. We have assembled databases of physically interacting protein pairs for Saccharomyces cerevisiae from these public databases as available on April 5, 2011. To eliminate the uncertainties originated from the use of different identifiers in databases, yeast systematic names were used as reference. Totally 135 154 pairwise interactions between 6191 proteins were acquired after elimination of the overlaps and self-interactions.

2.2. Biological features

A total of eight heterogeneous genomic and proteomic features were used to quantify the probability that an interaction is a biologically relevant true positive: (1) The number of databases reporting the interaction, (2) The frequency of detection by different experimental systems, (3) The number of published papers reporting the interaction, (4) Functional coherence, (5) Biological process coherence, (6) Co-localization, (7) Structural conformity, and (8) Gene expression correlation between interacting proteins.

The information on experimental system used in the detection of PPIs and the publication data (i.e. the PubMed identifiers) were taken directly from its source, i.e. the databases (Aranda et al., 2010; Ceol et al., 2010; Güldener et al., 2006; Salwinski et al., 2004 and Stark et al., 2011). Totally 6230 PubMed identifiers and 75 different experimental systems were employed. In case of uncertainty, the experimental technique used in detection of the PPI was manually checked from their publications by use of PubMed identifiers.

Functional coherence, biological process coherence and co-localization analyses were determined based on Gene Ontology (GO) terminology. The GO terms of yeast proteins were obtained from the consortium website as available on April 5, 2011.

To analyze structural conformity of interacting pairs, the protein–domain relationship for yeast proteins were obtained from PFAM (Bateman et al., 2000) and the domain–domain interaction data was directly taken from a previous study (Deng et al., 2002).

In determination of the correlation between the gene expression levels of interacting proteins, a large gene expression dataset was obtained from the SPELL tool (Hibbs et al., 2007) of Saccharomyces Genome Database (Cherry et al., 2012). The Pearson correlation coefficient for each protein pair, which ranges from –1.0 to 1.0, was used to enumerate the correlation.

2.3. Computation of individual association scores and confidence scores

Considering the assembled dataset of PPIs, the association scores, \( S_i \), of four features, i.e. the number of databases reporting the interaction, the frequency of detection by different experimental systems, the number of published papers reporting the interaction, and structural conformity, were calculated independently using a 4-step methodology. In Fig. 1, this methodology is illustrated by calculation of the association scores for one of the biological features (i.e. the number of databases reporting the interaction) as an example. Initially, the frequency distribution that represents the number of feature value associated with the full set of interaction pairs was obtained (Fig. 1A). For example, 95 261 PPIs (70.48%) were reported in a single database, whereas 544 PPIs (0.40%) were reported in all of the databases considered in the study. In the next step, via nonlinear regression, an exponential model was fitted, which represent the probability that a randomly chosen interaction pair has the corresponding feature value, i.e. the probability that a randomly chosen protein–protein interaction is reported in a certain number of databases (Fig. 1B). Since the reliability increases with increasing feature value, the preliminary score distribution was considered
as the complement of the probability distribution (Fig. 1C), and the score distribution was normalized to ensure that the minimal value of the feature gives a zero score, i.e. no contribution to confidence (Fig. 1D).

The association scores for three biological features, i.e. functional coherence, biological process coherence and co-localization, were determined based on GO semantic similarity analysis (Yu et al., 2010) which computes the similarity of the GO terms annotated to interacting genes/proteins. This method takes into account the overall structure of the GO graph to assess the specificity of shared annotations and computes scores between 0 and 1. In addition, the value of Pearson correlation coefficient was directly used as the score of gene expression correlation.

Naïve Bayesian approach is employed to calculate confidence score, CS, for any interacting pair of proteins so it is a function of the individual association scores:

\[ CS = 1 - \prod (1 - S_i) \quad (1) \]

Note that, the confidence scores obtained by Eq. (1) are always greater than the corresponding association scores, which signify increased confidence when an interaction is supported by various types of evidence.

2.4. Assessment of confidence scores

We evaluated our scoring results under sensitivity–specificity–coverage trade-off. Sensitivity and specificity is the ability to identify a true positive and true negative in a data set, respectively. Coverage indicates the completeness of the network in terms of the fraction of yeast proteins covered by the network. For this purpose, previously reconstructed gold-standard datasets of PPIs (Arga et al., 2007) were updated and employed (Supplementary Table S1): The gold standard positive dataset of 2530 interactions was composed of (i) physical interactions with very high confidence score (>0.999) in STRING (Szklarczyk et al., 2011), (ii) MIPS dataset consisting of interactions obtained by small-scale experimentation (Güldener et al., 2006), and (iii) core network of DIP (Salwinski et al., 2004). The gold standard negative dataset consisted of 2994 interactions obtained from protein co-localization data in yeast (Jansen and Gerstein, 2004) under the assumption that if the proteins do not exist in the same sub-cellular compartment, they cannot be interacting (Arga et al., 2007 and Jansen and Gerstein, 2004).

In comparison of the proposed scoring scheme with the other PPI scoring methods to illustrate the predictive power of the scoring scheme, the Receiver Operating Characteristic (ROC) curve was used. ROC curve is a graphical representation used to assess the discriminatory ability of a dichotomous classifier by showing the tradeoffs between sensitivity and specificity (Lin et al., 2004). The predictive power of the classifiers, i.e. the scoring methods, was quantified by calculation of the area under curve (AUC) metric.

2.5. Statistical analyses

To statistically describe and analysis of confidence score distributions within biological features, box-plot representation was employed. Each box represents the 25%–75% range of the confidence score distribution. The median are demonstrated by horizontal lines in the boxes. The lowest and highest data without outliers are shown by “Whiskers” and the stars express mild and extreme outliers. Since the score distributions do not follow the normality assumption, statistical analyses for the differences between the medians were performed by the Wilcoxon–Mann–Whitney test for
pairwise comparisons and the Kruskal–Wallis ANOVA test for multiple comparisons.

3. Results and discussion

3.1. Collection of PPI data and evaluation on biological features

The budding yeast, *S. cerevisiae*, is probably the most intensely studied model organism for functional interactions in eukaryotes, and therefore a huge number of experimental and computational studies have been conducted to detect PPIs in yeast. Five public databases, IntAct ([Aranda et al., 2010], MINT ([Coel et al., 2010], MIPS ([Güldener et al., 2006], DIP ([Salwinski et al., 2004]), and BioGRID ([Stark et al., 2011]), use the results of high-throughput experiments and collect experimentally/computationally determined PPIs from the literature as well. Since the entire PPI data reported in literature is not covered by a single database, in order to achieve the largest possible coverage we have assembled databases of interacting protein pairs for *S. cerevisiae* from these public databases as available on April 5, 2011 and obtained an extremely large PPI dataset which comprises of 135 154 non-redundant physical interactions between 6191 yeast proteins.

The analysis of the sources of the PPIs in the dataset showed that less than 0.40% of the all interactions (544 interactions) were reported in all of the databases considered. On the other hand, the majority of PPIs (70.5% of all) were stored by only one of these databases. These findings indicated a significant difference between the PPI databases, which arise mostly from the use of different resources, confidence sets and threshold values ([Reguly et al., 2006]), and confirmed the need for integration of different repositories to increase the coverage.

Generally, it is assumed that the interactions that have been validated by multiple experimental techniques and/or more than one research groups have a reasonable confidence ([Reguly et al., 2006]). However, majority of the interactions (77%) in the dataset were explored by a single experimental system. In addition, 37 interactions (0.03% of all) were reported in 10 or more articles in literature, whereas the majority of the interactions (86.5%) were reported in a single publication. Analysis of the similarity between interacting protein pairs in terms of GO terms indicated significantly low semantic scores; for example, zero-scores were obtained for almost 66% of PPIs when GO molecular function terms were considered. Similarly, the semantic scores of 18.5% of all interactions were calculated as zero in terms of GO biological process and 12.3% of all interactions displayed a zero semantic score when GO cellular component terms were considered. Since proteins interact with each other through their specific domains, domain–domain interaction information also gives clues on the reliability of the PPIs ([Chen et al., 2002; Edwards et al., 2002; Deng et al., 2002 and Rain et al., 2001]). Considering the 516 domains and 891 domain–domain interactions identified for *S. cerevisiae*, only 2.3% of all interactions were assigned by at least one domain–domain interaction. In terms of the correlation between gene expression profiles of interacting proteins, 4446 interacting pairs (3.3% of all) indicated a correlation coefficient higher than 0.60. All these preliminary analysis results on heterogeneous biological features indicated that each biological feature has its own pros and cons, but none of them is sufficient to define the accuracy of the PPI alone. Thus, the predictive power of the scoring system came from the integration of the multiple features.

Previous experiences on testing the quality of protein–protein interaction data indicated that as a result of the hidden physiochemical constraints inherent to each experimental method, conceptually different methods identify different type of interactions ([Yu et al., 2008 and Harrington et al., 2008]). For example, interaction assays such as yeast two-hybrid system are better at identifying binary interactions; however, purification assays such as tandem affinity purification are better at identifying co-complex interactions, i.e. all proteins that are part of the same complex. Therefore, these experimental methods should not be compared, but should be considered as complementary ([Harrington et al., 2008]). In addition, PPI data acquired from high-throughput approaches are reported to include large number of false positives ([von Mering et al., 2002 and Li et al., 2008]). Thus, the interactions that were validated by multiple experimental methods are believed to have a reasonable confidence. Moreover, it is assumed that the interactions published in the literature are reliable since many of them have been validated by multiple experimental techniques and/or more than one research groups ([Reguly et al., 2006]). However, Cusick and co-workers ([Cusick et al., 2009]) suggested that the quality of literature-curated datasets may not be as high as commonly noticed. The extraction of correct information out of long publications is extremely hard and structured vocabularies have not been applied in publications.

The domain–domain interaction data provide significant information to predict unknown PPIs ([Deng et al., 2002 and Rain et al., 2001]). Nevertheless, two proteins may not interact each other although they have potentially interacting domains because their localization in the cell may be different or the expression of them during cell cycle may be different. Moreover, the incompleteness of current domain–domain interaction data also makes it difficult to predict PPI ([Edwards et al., 2002 and Deng et al., 2002]).

3.2. Assessment of confidence scores

Considering the assembled dataset of PPIs, the association scores, $S_i$ of eight biological features (Fig. 2) were calculated independently, and integrated in a Bayesian approach to calculate confidence scores, $CS$, of all PPIs (see Section 2.3). The confidence scores indicated a broad distribution (Fig. 3). Due to the nature of the scoring function, the range of the scores was varying between zero (non-significant interaction indicator) and 1 (interaction with high-confidence), and the confidence scores were always greater than the corresponding association scores, which indicates increased confidence when an interaction is supported by various types of evidence (Supplementary Table S2). The confidence scores of 631 interactions (0.47% of all) were almost 1. These include binary interactions which were supported by almost all of the biological features. 29 046 (21.50% of all) interactions were greater than 0.85, whereas scores of 86 118 (63.72% of all) interactions were greater than 0.50.

To assess the efficiency of our methodology in score calculation and to identify a high-confidence interaction dataset through the large PPI dataset, the resultant confidence scores were assessed in terms of a trade-off between the sensitivity (i.e. the ability to identify true-positives), the specificity (i.e. the ability to identify true-negatives) and the coverage (i.e. the fraction of all proteins covered by the dataset). The traditional method for determination of sensitivity and specificity is the employment of training data, i.e. the gold standard datasets, which consists of interactions that are only assumed to be true or false.

On the other hand, construction of these datasets is a difficult task and requires attention. Since these datasets are constructed based on simple assumptions, the approaches may produce training sets that are biased toward particular type of interaction or highly conserved interactions or particularly well-studied pathways. Previously, the gold standard datasets of PPI were constructed to test the efficiency of the signaling pathway reconstruction schemes ([Arga et al., 2007]). To determine sensitivity and specificity, in the previous study, these datasets were updated and employed: The gold standard positive dataset consists of 2530 interactions and the
gold standard negative dataset includes 2994 PPIs ( Supplementary Table S1). The gold standard negative dataset consisted of interactions obtained from protein co-localization data in yeast (Jansen and Gerstein, 2004) under the assumption that if the proteins do not exist in the same sub-cellular compartment, they cannot be interacting ( Arga et al., 2007 and Jansen and Gerstein, 2004). The gold standard positive dataset was composed of physical interactions with very high confidence score (>0.999) in STRING (Szklarczyk et al., 2011), a set of interactions obtained by small-scale experimentation ( Güldener et al., 2006), and interactions from the core network of DIP (Salwinski et al., 2004). Although each of the three subsets incorporated in construction of the positive dataset were considered as highly confident in literature, the intersection of all datasets constitute only 7.5% of all ( Fig. 4).

As a result of the trade-off analysis for different score thresholds ( Table 1), we determined the confidence score of 0.85 as an optimum threshold for identifying a high confidence dataset (HCS) of PPIs in yeast ( Supplementary Table S3). This dataset, consisting of 29 046 binary PPIs, constitute only 21.50% of all interactions in the large PPI dataset and indicates equilibrium between sensitivity, specificity and coverage (86%, 68% and 72%, respectively). When the association scores of these interactions in the HCS dataset analyzed, it was found that these interactions were supported by multiple evidences, i.e. supported by at least five of the eight heterogeneous features considered in scoring scheme. In addition, a low confidence dataset (LCS) consisting of 49 036 (36.28%) binary interactions with confidence scores lower than 0.50 was also determined ( Supplementary Table S4). These interactions were supported by weakly at most three of the eight biological features considered in the study.

To analyze the contribution of biological features into the final confidence scores, the association score distributions of each biological feature within the whole dataset, within the high-confidence dataset (HCS) and within the low-confidence dataset (LCS) were statistically analyzed (see Section 2.5 and Supplementary File). The contributions of three biological features ( i.e. the number of databases reporting the interaction, the structural conformity, and the frequency of detection by different experimental systems) to the confidence scores of interactions within the HCS were significantly higher than those within the whole dataset ( p-values <10−5). In addition, these biological features had no contribution to the confidence scores of interactions within the LCS. The contributions of three other biological features ( i.e. the number of published papers reporting the interaction, biological process

Fig. 2. The score distributions of association scores for each heterogeneous feature. The association score (Sf) distributions, which were calculated using a 4-step methodology, for: (A) the number of databases reporting the interaction (Sa), (B) the frequency of detection by different experimental systems (Sb), (C) the number of publications papers reporting the interaction (Sc), and (D) structural conformity (Sd). The association score distributions, which were determined based on GO semantic similarity analysis, are represented by cumulative frequencies for: (E) functional coherence (Sh), (F) biological process coherence (Sk), and (G) co-localization (Sl). (H) The association score distribution for gene expression correlation (SPOP) as a linear function of Pearson correlation coefficients.

Fig. 3. Cumulative distribution of confidence scores. Interactions with scores above 0.85 constitute only 21.5% of all and could be considered high-confidence interactions. According to the design of the scoring function, such score implies that the interaction is supported by multiple evidences.

Fig. 4. The distribution scheme of the interactions in the gold standard positive dataset to their sources. Although each of the three datasets incorporated in construction of the positive set were considered as highly confident in literature, the intersection of all datasets constitute only 7.5% of all.
coherence, and co-localization) to the confidence scores of interactions within the HCS were significantly higher than those within the whole dataset and within the LCS (p-values < 10^{-5}). However, there were no significant differences between their contributions within the whole dataset and those within the LCS (p-values > 0.01). Similarly, the contribution of the biological features, the gene expression correlation and functional coherence, to the confidence scores of interactions within the HCS were higher than those within the whole dataset and within the LCS (p-values < 10^{-3}) and there were no significant differences between their contributions within the whole dataset and those within the LCS (p-values > 0.01). The results indicated that all of the biological features considered in the present scoring scheme had significant contributions in the reconstruction process of the high-confidence dataset (HCS).

To illustrate the predictive power of the scoring scheme, the Receiver Operating Characteristic (ROC) curve analysis was employed to give a more complete picture of the performance of the present scoring scheme when compared to some previous schemes (Fig. 5).

The present scoring scheme indicates an excellent discrimination in terms of the AUC value of 0.84 according to general rules used in interpretation of AUC values (Lin et al., 2004). The IntScore scoring scheme also indicates excellent discrimination with an AUC value of 0.84; other scoring schemes represent acceptable discrimination with AUC values between 0.59 (HitPredict) and 0.67 (Mint).

The extreme AUC value of the present scoring scheme is a result of many advantages of the present scoring scheme with respect to the reported scoring systems. First of all, it is more comprehensive than the others since an extremely large dataset of PPIs, which is the combination of widely used five PPI repertoires, was employed in the present scoring methodology. The other scoring schemes were based on a single dataset achieved from a single type of experimentation or database. Thus, the designed score for PPIs are mostly not analogous to each other, and for the interactions provided at other databases, a score assignment is not possible.

A major problem with most of the published scoring procedures was the use of training sets, which include only the interactions supposed to be true/false. For instance, Yu and Finley (Yu and Finley, 2009) synthesized four different gold positive training sets based on interactions that (i) are associated with at least 10 PubMed identifiers in the interaction databases; (ii) are putative conserved interactions, which are those found in common between interaction sets for any two species (fly, human, and yeast); (iii) are high-throughput interactions reported to have high confidence by the original publications; and (iv) have expression correlation higher than 0.6. For each training set, a negative set of equal size was synthesized by drawing random samples from the list of all interactions, excluding those in that positive set. On the other hand, these training sets were doubtful since they might be biased toward highly conserved interactions or interactions from particularly well-known pathway or due to random sampling. In this point of view, a major advantage of the present scoring system is that the association scores were calculated considering the full set of PPIs in the large dataset, instead of reconstructing training sets.

Another major problem with other scoring schemes taking into account different types of biological features is the determination of weights assigned to the association scores of each feature in the calculation of the overall confidence score. However, in the present scoring system no artificial weights were assigned, i.e. weights were intrinsic to the scoring function. Application of the methodologies used in the calculation of association score of each feature resulted with different maxima for association scores of each feature (Fig. 2), which can be considered as weights for each feature.

### 3.3. Reliability of experimental systems

The present scoring scheme also allowed us to discuss the reliability of experimental systems used in detection of PPIs. Previously, it was claimed that the high-throughput methods, such as protein arrays and yeast-two-hybrid screens require the assessment of the reliability since 30–50% of the high throughput interactions were known to be biologically relevant (Jossifov et al., 2004). On the other hand, low throughput methods, such as X-ray crystallography and nuclear magnetic resonance, were believed to give the most detailed information about protein–protein interactions (Shoemaker and Panchenko, 2007). Recently, Schaefer and co-workers (Schaefer et al., 2002) assigned scores to various PPI detection methods via manual curation, and in vitro techniques like X-ray crystallography were assigned to the highest score that is 10. Complementation based assays and affinity based technologies were roughly equally scored with an average value of 5. The scores were slightly increased for those methods, such as FRET, that are used generally in homologous, more physiological setups. Methodologies that do not directly provide evidence for interaction, such as co-localization or co-sedimentation, are scored with lower scores. These reliability scores assigned by manual curation necessarily reflect the individual opinion of researchers to these experimental techniques, therefore we evaluated our resultant confidence scores from this perspective and constructed a box-plot representing the
distribution of confidence scores of all interactions characterized by the corresponding experimental system (Fig. 6).

The score distributions of small-scale experimental systems (i.e. nuclear magnetic resonance, co-crystal structure, X-ray crystallography and surface plasmon resonance) possessed narrow distributions with significantly higher median values (p-value < 10^{-4}), when compared to the score distributions of high-throughput experiments (i.e. yeast-two-hybrid screen, tandem affinity purification and protein array), which possessed broad score distributions with significantly lower median scores with respect to the threshold score (p-value < 0.01). These results were consistent with the previous discussions (Lossifov et al., 2004 and Shoemaker and Panchenko, 2007) and with the reliability scores of Schaefer and co-workers (Shoemaker and Panchenko, 2007), where nuclear magnetic resonance, co-crystal structure, X-ray crystallography and surface plasmon resonance methods were assigned to the highest score of 10; on the other hand, complementation-based assays and affinity based technologies, such as yeast-two-hybrid screen, tandem affinity purification and protein array, were roughly equally scored with an average value of 5.

4. Conclusion

Comprehensive identification and characterization of PPIs and their networks is essential in order to systematically define and understand the information transfer within and between biological processes. In the present study, a scoring scheme considering eight heterogeneous biological features was proposed and a high-confidence protein–protein interactome of yeast S. cerevisiae was reconstructed. The proposed methodology is expected to increase the efficiency of the methodologies in biological research which make use of protein interaction networks in several aspects: The proposed scoring scheme will be adapted to the interactome of other species and the reconstructed high-confidence interactome will prevent from spurious discoveries that can be potentially costly, e.g. wrong signaling routes, or wrong drug targets for diseases.

Acknowledgements

Financial support to KYA by Marmara University Research Fund through project FEN-A-130511-0163, and The Scientific and Technological Research Council of Turkey (TUBITAK) through project 110M613 is greatly acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compbiolchem.2013.03.002.

References


