World Academy of Science, Engineering and Technology
International Journal of Bioengineering and Life Sciences
Vol:1, No:1, 2007

Correspondence between Function and Interaction in Protein Interaction Network of *Saccaromyces cerevisiae*

Nurcan Tuncbag, Turkan Haliloglu, Ozlem Keskin

***Abstract*—**Understanding the cell's large-scale organization is an interesting task in computational biology. Thus, protein-protein interactions can reveal important organization and function of the cell. Here, we investigated the correspondence between protein interactions and function for the yeast. We obtained the correlations among the set of proteins. Then these correlations are clustered using both the hierarchical and biclustering methods. The detailed analyses of proteins in each cluster were carried out by making use of their functional annotations. As a result, we found that some functional classes appear together in almost all biclusters. On the other hand, in hierarchical clustering, the dominancy of one functional class is observed. In the light of the clustering data, we have verified some interactions which were not identified as core interactions in DIP and also, we have characterized some functionally unknown proteins according to the interaction data and functional correlation. In brief, from interaction data to function, some correlated results are noticed about the relationship between interaction and function which might give clues about the organization of the proteins, also to predict new interactions and to characterize functions of unknown proteins.

***Keywords*—**Pair-wise protein interactions, DIP database, functional correlations, biclustering.

I. Introduction

PROTEINS are large molecules responsible for executing and regulating various biological functions. Although some protein structures can be functional alone, most of them have to associate with other proteins to act in the biological processes. In other words, they perform their functions by interacting with other proteins. The combination of these different interactions in the organisms results in biological processes. Antigen-antibody recognition, enzyme substrate binding, DNA replication and transcription, RNA splicing and metabolic cycles are some examples of biological processes containing protein-protein interactions. The complex functions in the biological systems are a result of the large network of the proteins formed by pair wise protein – protein interactions. Thus, the network of interactions between the proteins increases the understanding of protein functions and this network controls the lives of cells [1]. Moreover, protein interaction networks provide functional information about the uncharacterized proteins and remote similarities between proteins [2].

Interactions of the proteins are likely to correlate with the functional properties of the proteins. Protein interaction maps are generally used to uncover functionally unknown proteins [1]. Generally, if two proteins interact directly they are likely to be involved in the same biological process or pathway [3].

Protein-protein interactions are obtained from experimental results [4,5] and also from databases (MIPS, DIP, BIND, GRID and Yeast Protein Database) [6,7,8,9,10]. The databases, which catalog the interactions between the proteins, provide quick access to the experimental interaction data and also to the large scale properties of these biological networks. One of the interaction databases is the Database of Interacting Proteins (DIP) [7,11]. In addition to documenting experimentally determined pair wise physical interactions, DIP combines information from a variety of sources to create a single, consistent set of protein-protein interactions. The data within the DIP can be extracted both manually and computationally [7]. DIP interaction participating proteins have a special identity number as "DIP:nnnN". Besides this accession number, DIP provides cross-references to the three major sequence databases, SWISSPROT, PIR and GenBank. By the help of the cross references from DIP to other sequence databases, it is possible to obtain information about general aspects of the proteins. In this way, from the pair wise interaction data in DIP, the functional correlation of two interacting proteins can be found by the cross-references [7,11]. Because experimental interaction data comes with false negatives, we used core interactions in our study [protein function connectivity]. CORE interaction means that the interaction data verified by one or more computational and experimental verification methods [7,11].

Clustering of the proteins in an interaction network can depend on one or more than one properties like pair wise interaction data, evolutionary data or functional description of the proteins. The clustering algorithms deal with the similarities between genes or proteins across all conditions.

Nurcan Tuncbag is M. S. student in Computational Science and Engineering, Koc University 34450, Sariyer, Istanbul, Turkey (email: ntuncbag@ku.edu.tr).

Prof. Turkan Haliloglu is with Chemical Engineering Department, Bogazici University, Bebek, Istanbul, Turkey (e-mail: halilogt@boun.edu.tr).

Assist. Prof. Ozlem Keskin is with Chemical and Biological Engineering Department, Koc University, 34450, Sariyer, Istanbul, Turkey (e-mail: okeskin@ku.edu.tr).

Biclustering takes also the same conditions matrix as an input and tries to find statistically significant sub matrices called biclusters. Generally, a gene or a protein has many different functions. Its similarity to other genes or other proteins can be limited to limited set of conditions. The advantage of the biclustering approach is the fact that genes or proteins which behave similarly under a subset of the conditions but does not share common behavior under other conditions [12].

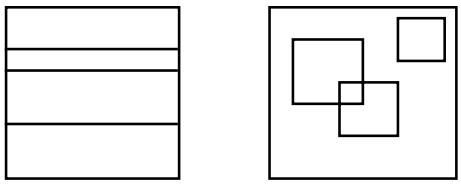

Fig. 1 general representation of the difference between clustering and biclustering.

Clustering of the protein data set is significant to characterize the functionally unknown proteins or to discover new interactions and to verify putative interactions. In this way, protein clusters give information about the organization of the network and about the most important node in the network both from functional side and interaction side [1].

In this study, in order to analyze the functional correlations and organization of the proteins, we started with the interaction data in DIP database. We used the contact matrix to represent all pair wise interactions in the yeast interaction network. After obtaining the pair wise cross correlations between the proteins and clustering of them, we continued with functional description of the proteins in each cluster. For this purpose, Gene Ontology project [13] is used to annotate the functional correspondence. In our work, we analyzed experimental interaction data in yeast obtained from DIP in a computational way. As a result, we verified some putative interactions and also we predict the functions of some functionally unknown proteins.

Our work can give some clues about the relationship between interaction and function, and also it predicts unknown protein-protein interactions in the yeast protein network.

II. Methods and Results

The main focus of this paper is to find the set of correlated proteins in the protein-protein network of *S. cerevisiae*; in this way, to observe the correlation between the pair wise physical interactions between *S. cerevisiae* proteins and functional organization between them. For this purpose, a novel method is used in order to find the cross correlations between the *S. cerevisiae* proteins. The flowchart of the observation process is shown in Figure 2.

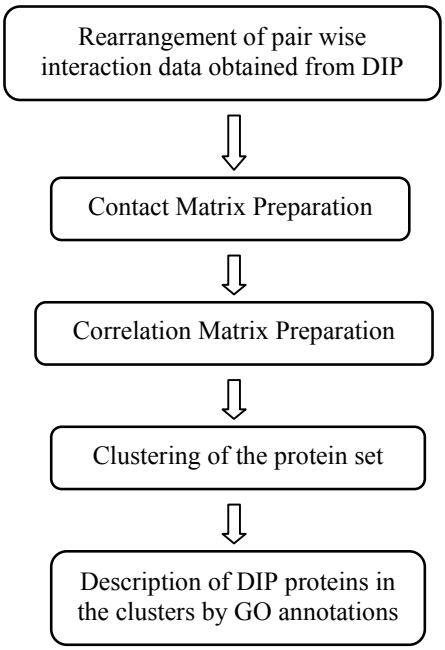

Fig. 2 the flowchart of the project to obtain the correspondence between function and interaction.

The used pair wise interaction data set was obtained from the DIP in December 11, 2005. The data set contains the pair wise CORE interaction (verified one or more than one computational or experimental method) data of the yeast proteins. In the data set, there are totally 2635 proteins and 6342 core interactions observed among these proteins.

The network of protein interactions are represented as an undirected graph with proteins as nodes and interactions as undirected edges [2]. In this paper, we utilized from the graph method used by [14,15,16] to analyze the protein structure and dynamics. A modified version of the same method has been used also for backbone clustering and structure similarity [17]. In the algorithm of the project, firstly, the pair wise protein-protein interaction data available in DIP for yeast is rearranged to prepare a contact matrix to represent the interaction data among the proteins. For this purpose, each protein is referenced to an integer identifier. The protein identifiers are also used as indexes to form the contact matrix (A). By using the model below, the contact matrix is prepared. According to this model, it is controlled whether the protein *(i)* interacts with another protein *(j)* or not. If *i* interacts with *j*, 1 is inserted at the *ij*th element of the matrix, if not, 0 is inserted at this element. And the i^{th} element of the diagonal of the matrix A is the negative of the i^{th} row sum [16].

$$A = \begin{Bmatrix} 1 & if \;\; i \Leftrightarrow j \\ 0 & if \;\; i \Downarrow j \\ -\sum_{j=1}^{n} A_{i,j} & if \;\; i = j \end{Bmatrix} \quad (1)$$

where $\Leftrightarrow$ represents the interaction between two proteins.

There are other methods to define the contact matrix or to cluster the protein data set. There are similar studies about the spectral analysis of the protein interaction network of yeast [18,19,20] However, these studies are significantly different from our work. In our study, the diagonal of the contact matrix is defined as the negative of the summation of that row. Also, the other difference is the clustering method, the biclustering (hierarchical clustering is also used as a trial). In other words, a protein in our data set can be a member of more than one cluster. Our contact matrix is singular and all non zero eigenvalues of this matrix are negative. 92 of all eigenvalues are zero. If the number of zero eigenvalues is high, the connectivity of the nodes in the network is less. It has been observed in Gaussian Network Model and Anisotropic Network Model to study protein structure [14,15,16].

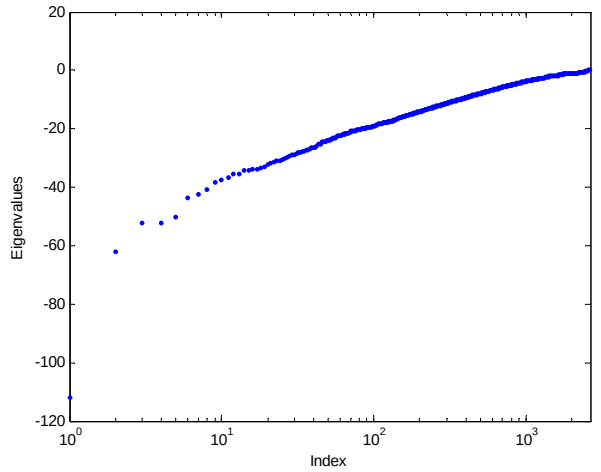

Fig. 3 eigenvalue distribution of the contact matrix obtained from protein interaction data of yeast.

If there are N proteins in the interaction network, then the contact matrix is NxN in size and has N eigenvalues and N corresponding eigenvectors. Also, the inverse of the contact matrix gives the cross correlation between proteins. However, it should be noted that the determinant of the matrix A is zero. Because the contact matrix is singular and do not have inverse, the pseudo inverse of the matrix is taken. To find the pseudo inverse, contact matrix (A) is decomposed by singular value decomposition. For this purpose, contact matrix can be written as in Equation(2);

$$A = U \cdot \Sigma \cdot V^T \tag{2}$$

where Σ is a diagonal matrix and contains absolute values of eigenvalues.

The pseudo inverse (A^+) of a matrix is a generalization of the inverse matrix and includes all modes. The computationally simplest way to calculate the pseudo inverse of a matrix is using singular value decomposition (SVD). If $A = U \cdot \Sigma \cdot V^T$ is the singular value decomposition of *A*, then the the psuedoinverse of A is $A^+ = V \cdot \Sigma^+ \cdot U^T$. For a diagonal matrix such as Σ, which consists of the singular values of matrix A, the pseudoinverse of this diagonal matrix is the reciprocal of each non-zero element on the diagonal [7].

The contact matrix can also be represented by spectral theorem as in Equation(3);

$$A = U^T \cdot \Lambda \cdot U \tag{3}$$

where U is a square matrix composed of the eigenvectors, and Λ is a diagonal matrix composed of corresponding eigenvalues on the diagonal.

The inverse of the contact matrix can be represented as in Equation(4) .

$$A^{-1} = U \cdot \Lambda^{-1} \cdot U^{-1} \tag{4}$$

Alternatively, it can be written as the sum of N – 1 matrices like in Equation(5) of size N x N, each representing the contribution of a single internal mode [14,15,16].

$$A^{-1} = \sum_m \left[\lambda_m^{-1} \cdot u_m \cdot u_m^T \right] \tag{5}$$

The size of the contact matrix in the project is 2635 x 2635 – because there are 2635 proteins in the dataset – and the sparsity of it is 0.9978, which means that 99.8 % of the matrix elements are zero. The sparsity pattern of the contact matrix for yeast core interaction data in DIP is shown in Fig. 4.

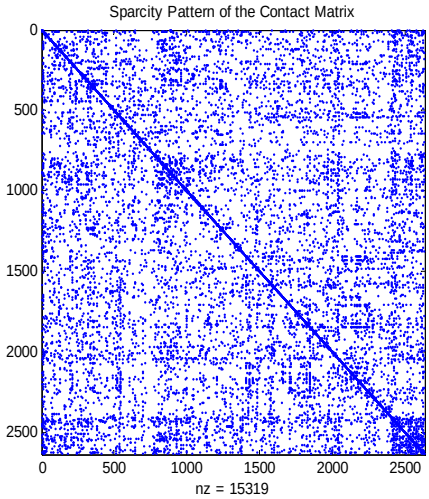

Fig. 4 sparsity pattern of the contact matrix for the pair wise interaction data where nz is the number of non zero elements.

The normalized cross correlations between protein interactions are found by normalization of the pseudo inverse matrix. After taking the pseudo inverse of the contact matrix,

the next step is to prepare the normalized cross-correlation matrix between the proteins. By normalization of pseudo inverse matrix, the cross-correlation matrix is found. Cross-correlations between the DIP proteins are calculated as in Eq. (6). The matrix C gives the normalized cross correlations between the proteins [14,15,16].

$$C(i,j) = -\frac{\left[A^{-1}\right]_{ij}}{\left(\left[A^{-1}\right]_{ii} \cdot \left[A^{-1}\right]_{jj}\right)^{0.5}} \qquad (6)$$

As a result, all the diagonal elements of the cross-correlation matrix (C) are equal to 1 which means that the protein *(i)* is 100 % correlated with itself. In the cross-correlation matrix, the correlation value of each element changes between – 1 and 1. (-) correlation values represent anti-correlation between two proteins, 0 represents no correlation between them and (+) correlation values represents how much correlated are two proteins. To perform all these steps, Python 2.4 Programming Codes has been used.

A. *Hierarchical Clustering*

After calculation of the cross correlations between the proteins, the cross-correlation matrix has been clustered according to the correlation values in it. For first trial in the project, hierarchical clustering was used. By "clusterdata" function in MATLAB, the proteins were clustered hierarchically. This function first computes the Euclidean distance between pairs of objects in the correlation matrix. Then, it creates a hierarchical cluster tree, using the Single Linkage algorithm, and finally, constructs clusters from this hierarchical cluster tree.

The cluster numbers are found according to the cutoff value. The cutoff value is a threshold for cutting the hierarchical tree generated by linkage into clusters [21]. The optimum cluster number is found by the graph "cutoff value vs. cluster number", shown in Fig. 3. In the graph, while the cutoff value increases, the number of clusters decreases and goes to 1 cluster. According to this graph, the optimum cutoff value is chosen as 1.154. The number of clusters for this optimum cutoff value is 507.

In the hierarchical clustering results, the cluster sizes are small; also some of them have only one member. In the analysis of the clusters in the functional perspective, the clusters which have less than and equal to 5 members are eliminated. As a result, cluster numbers decrease from 507 clusters to 93 clusters. The cluster size distribution after elimination of the redundant clusters is shown in Fig. 6.

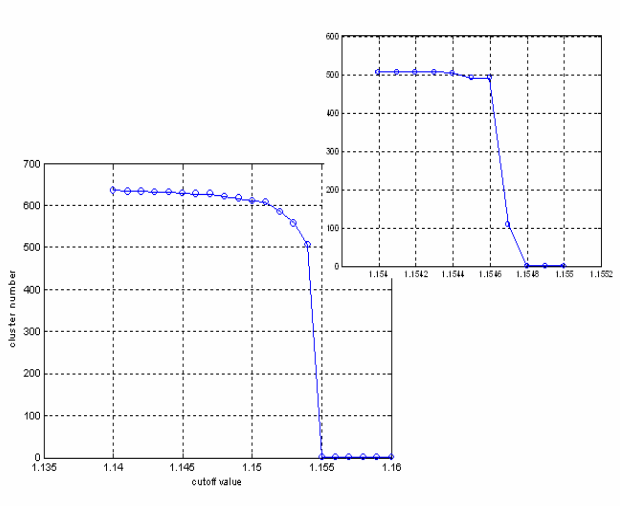

Fig. 5 the cutoff value vs cluster number graph of hierarchical clustering by MATLAB.

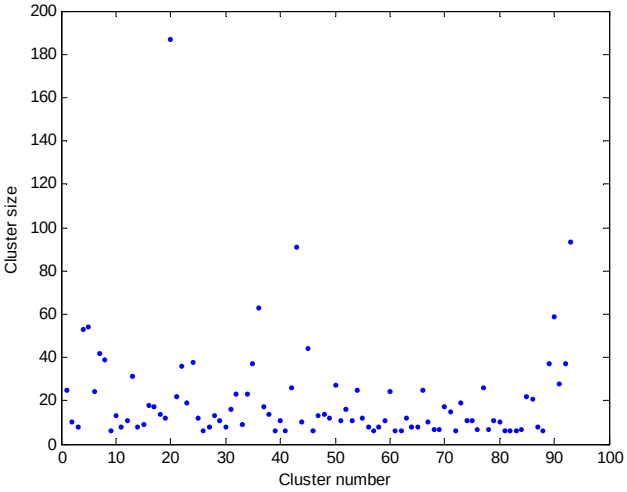

Fig. 6 cluster size distribution for hierarchical clustering. This distribution obtained after elimination of the clusters which have less than or equal to 5 members. As a result, there is 93 clusters available coming from hierachical clustering.

B. *Biclustering*

Hierarchical clustering is one of the mostly used methods for clustering a data set in biological systems. However, in biological systems, a protein can function in more than one process. In other words, one protein can be put more than one cluster. Because of this situation, biclustering method seems more appropriate for biological systems and for clustering these proteins.

Biclustering algorithm does not force the proteins to belong to one cluster. To bicluster the proteins in the dataset effectively, the EXPANDER software [22] was used. EXPANDER is a package for the analysis of gene expression data, contains various data analysis algorithm implementations. One of them is biclustering analysis. The biclustering tool of the EXPANDER uses SAMBA algorithm

to bicluster the data set [22]. The detailed information about biclustering and EXPANDER software can be found in http://www.cs.tau.ac.il/~rshamir/expander/expander.html. In this work, the normalized cross-correlation matrix (C) is biclustered. Firstly, the matrix is loaded in the EXPANDER software, and then the SAMBA [Tanay 2004] algorithm is runned. As a result, there is 344 biclusters found. However, lots of clusters have same members at high ratios. For example, some biclusters have 70 % or more similarity, means that 70% of members of a biclusters are same with another bicluster. Because of this situation, the biclusters are associated according to their similarity ratios. As a result, for 70 % similarity, the bicluster number decreases from 344 to 222. The cluster size distribution after association of the clusters is shown in Fig. 7.

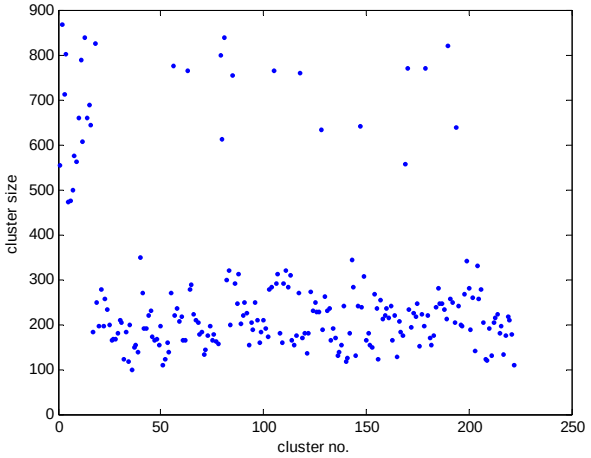

Fig. 7 cluster size distribution of the biclusters. This distribution obtained after merging of 70% identical biclusters.

C. Functional Annotation of the Proteins

After the correlation matrix is clustered by hierarchical clustering and by biclustering, the proteins in the clusters and biclusters are interpreted according to the molecular functions, whether there are functional correlations between the proteins in the clusters. For this purpose, each protein in the clusters are described with the Gene Ontology (GO) Annotations and analyzed whether there is functional relationship between them.

GO gives consistent descriptions of gene products in different databases. The GO annotations describe the gene products from three ways: (i) *cellular component* is the component of the cell; for instance, ribosome, nucleus etc.; (ii) *biological process* is series of events accomplished by one or more ordered assemblies of molecular functions; for example cellular physiological process, RNA metabolism or signal transduction; (iii) *molecular function* describes activities, such as catalytic or binding activities at the molecular level. The GO terms goes from broad terms to more specific terms. For example, "binding" is a broad GO term, at the second level the GO term below binding, "protein binding" give more specific information. Each term in GO have a unique numerical identifier like (GO:nnnnnnn) [13].

The GO ID's of each protein is found by cross references from DIP to GO. There is no direct way to get GO annotations from DIP database as shown in Fig. 8. Firstly, the cross-reference between the DIP database and SWISSPROT is used. Each DIP name is changed by SWISSPROT ID's. However, the SWISSPROT ID's of 306 of the 2635 proteins in the dataset are not found in the cross reference between DIP and SWISSPROT. Then, the cross-references between the SWISSPROT and GO annotations are found. By this way the function, in which the protein participate is found for the proteins in each clusters.

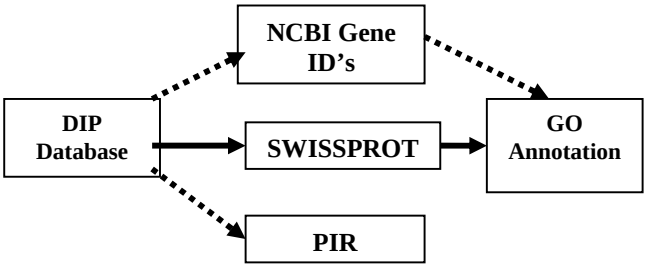

Fig. 8 cross – reference from DIP database to GO annotations. The dashed lines represents the possible ways to go from one database to another. The straight line represents the way used in this study to go from DIP to GO annotations.

D. The distribution of the functional classes in the dataset

The first level functions are used from GO annotations. At the first level GO annotations, there are 19 different function classes which are shown in Table 2. 12 of these functional categories are occupied by the proteins in the core interaction data set. Each protein is assigned to one or several of the 12 functional classes.

TABLE I
THE NUMBER OF FUNCTIONS IN THE DATASET

GO Function (first level)	ID	# of proteins participate at that function
antioxidant activity	1	6
binding	2	1512
catalytic activity	3	1092
chaperone regulator activity	4	2
chemoattractant activity	5	0
chemorepellant activity	6	0
energy transducer activity	7	0
enzyme regulator activity	8	152
molecular function unknown	9	0
motor activity	10	12

nutrient reservoir activity	11	0
obsolete molecular function	12	0
protein tag	13	3
signal transducer activity	14	41
structural molecule activity	15	139
transcription regulator activity	16	136
translation regulator activity	17	37
transporter activity	18	155
triplet codon-amino acid ad. act.	19	0

The large amounts of the proteins in the interaction dataset have binding and catalytic activity as seen in Table 1. They are excluded from the functional categories, since they would over amplify the results. Also, the proteins in the data set are not functioning in the functional categories chemoattractant activity (#5), chemorepellant activity (#6), energy transducer activity (#7), nutrient reservoir activity (#11), obsolete molecular function (#12) and triplet codon-amino acid adaptor activity (#19). When we focus on the functions, the proteins in the data set are occupied in the functional classes of antioxidant activity (#1), chaperone regulator activity (#4), enzyme regulator activity (#8), motor activity (#10), protein tag (#13), signal transducer activity (#14), structural molecule activity (#15), transcription regulator activity (#16), translation regulator activity (#17) and transporter activity (#18).

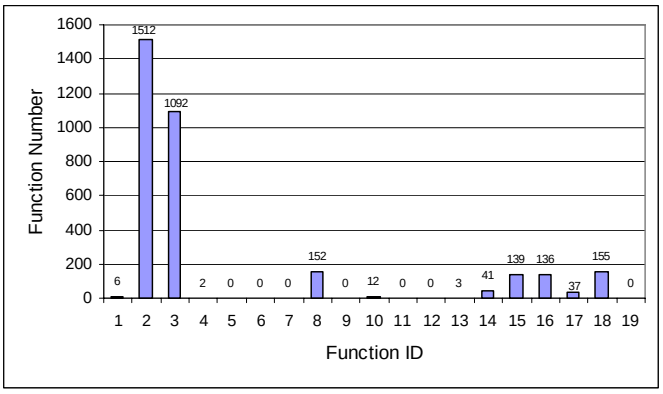

Fig. 9 functional distribution of the all yeast proteins in the data set

After each protein is described by its functional annotations, the hierarchical clusters and biclusters are checked separately, whether they are correlated or they are not. When the functional annotations are analyzed (excluding the protein binding (#2) and catalytic activity (#3), since because they are in all clusters at high ratios normally) generally almost in every bicluster the functional classes, enzyme regulator activity (#8), signal transducer activity (#14), structural molecule activity (#15), transcription regulator activity (#16), translation regulator activity (#17) and transporter activity (#18), take place as blocks. However, in the clusters of hierarchical clustering, these functional class blocks are not seen. The functional classes of binding (#2) and catalytic activity (#3) exists dominantly in all clusters in hierarchical clustering like biclusters, but we do not see the same functional behavior of the proteins and 6 functional category blocks. Generally, in hierarchical clusters, one functional class is dominated except the classes #2 and #3. Both in small hierarchical clusters and small biclusters, one functional class is dominant and separate from others. To be more specific about it, we choose one of the 344 biclusters.

E. Case Studies for Analysis of the Clusters

1) Bicluster #40

The bicluster #40 is chosen to be observed more detailed. This bicluster has 350 proteins; because one protein can be assigned one or several functional classes, there are formed totally 473 functions and 84 % of these functions are in the class of binding (#2) and catalytic activity (#3).

TABLE II
THE FUNCTIONAL CLASSES OF THE BICLUSTER # 40.

Functional Class	**# of proteins participate at that function**
binding	210
catalytic activity	188
enzyme regulator activity	15
signal transducer activity	4
structural molecule activity	12
transcription regulator activity	15
translation regulator activity	12
transporter activity	17

When binding and catalytic activity are disregarded, again the 6 functional groups are observed together in bicluster #40 as in almost the rest of the data set. For biclustering results, we can conclude that these 6 functional groups are working collectively in the yeast. Because we started from interaction data, it can be suggested that these functional grouping of the proteins shows the correlation between interaction and function. In Figure 8, the partition of the functional categories in bicluster 40 is shown.

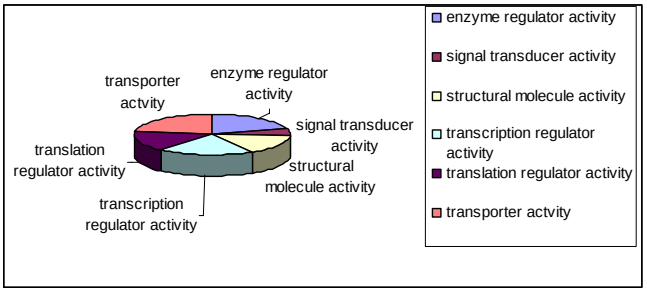

Fig. 10 the graph of the 6 functional classes of bicluster 40.

By the usage of the server in Yeast Genome Database (www.yeastgenome.org), the GO annotations tree for the bicluster #40 is drawn from the process side [10]. For bicluster #40, the biological processes are identified and we observed that only 2 of the 12 processes are occupied by the proteins in the bicluster 40 as seen in Fig. 11 which is physiological process and cellular process. In Figure 11, only the upper levels of the GO annotation tree are shown.

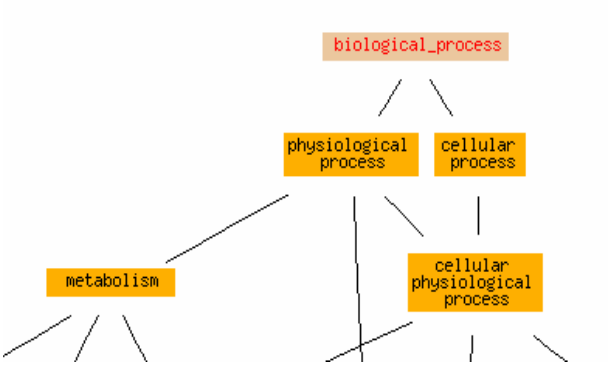

Fig. 11 First two level GO process tree of the bicluster 40. The proteins in this cluster are participating into 2 processes out of 12 first level GO processes.

When the tree is examined at the low levels, we have seen that all of the proteins in this cluster is participating into the RNA metabolism. The low levels of the tree are represented with a simplified scheme in Figure 12. The entire of the GO annotation tree of this bicluster is available in http://home.ku.edu.tr/~ntuncbag/yeastclusters with the gene names at each subprocesses. This situation shed light on the hypothesis that the interacting proteins in the same bicluster are involved in the same biological process.

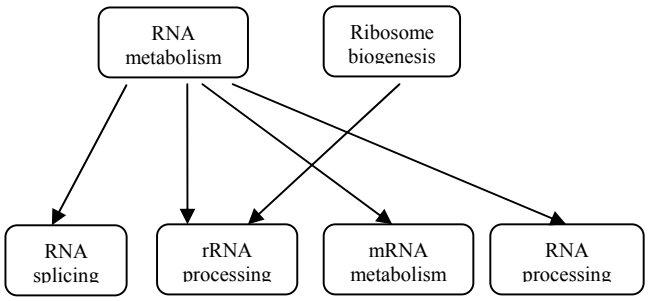

Fig. 12 General representation of the organization of the bicluster40 at low levels. Almost all genes in this bicluster are participating in the RNA metabolism. The detailed and colored version of the tree with the gene names on it is available in http://home.ku.edu.tr/~ntuncbag/yeastclusters.

2) *Hierarchical Cluster #20*

The same procedure used for biclusters is followed also for the hierarchical clusters. The cluster #20 is chosen for this procedure. This cluster has 187 members. As seen in Table 4, there was only one functional class except binding and catalytic activity. The functional class transcription regulator activity (#16) is the single category, when category #2 and #3 are disregarded.

TABLE III

THE FUNCTIONAL CLASSES OF THE CLUSTER # 20.

Functional Class	**# of proteins participate at that function**
binding	95
catalytic activity	105
transcription regulator activity	19

The complete list of the hierarchical clusters and biclusters with its functional annotations is available at home.ku.edu.tr/~ntuncbag/yeastclusters.

3) *Bicluster #334*

In another case for analysis of the clusters in a detailed way, we chose a cluster to find new interactions and to characterize unknown proteins. For this purpose, we selected the bicluster #334. In the interaction network of the bicluster #334, there is one large and two small fully connected networks and some single proteins. In Figure 6, representation of the large network in the bicluster 334 is available.

We verified the putative interactions according to the hypothesis that if the proteins are in the same bicluster, they possibly interact and they would function in the same process. In Figure 13, the blue straight edges represent the core interactions. The red dashed edge represents the verified interaction in this study between DIP: 3842 and DIP: 701N. In the core interaction data set, there was no interaction between protein DIP: 701N and DIP: 3842N, but it is given as a possible interaction in the DIP database. In brief, here we verified the interaction between the proteins 701N and 3842N computationally.

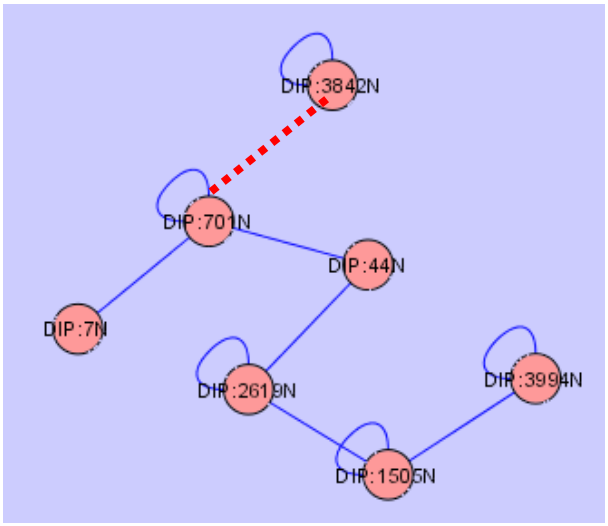

Fig. 13 one of the small networks in the bicluster 334. In this study, the interaction between the protein 3842N - single - and the protein 701N - in the network – is verified.

When we analyze the GO annotations of these single proteins, we see a correspondence between them. Most of them are processing in the transcription process. From this correspondence we can conclude that there are possible undiscovered interactions between these single proteins.

III. Conclusion

The starting point of this paper was the hypothesis that interacting proteins have a high probability to belong to same functional class. For this purpose, after obtaining cross correlations between yeast proteins from interaction data, two clustering methods were used and at the end, two different results were obtained. As a result of biclustering, we observed the collective existence of same functional classes. Moreover, after observation of one bicluster in the view of processes, dominancy of one process was observed in the entire of the bicluster. On the other hand, in hierarchical clustering the dominancy of one functional class is noticed, especially in the small sized clusters. Also, some unverified interactions in DIP are verified according to being in the same bicluster.

In the future, after detailed analysis of the clusters and biclusters, more verifications and new interactions would be found. Moreover, the functionally unknown proteins could be characterized according to being with high possibility in the same biological process with other interacting pairs of them.

Appendix

Part A

The Websites Used In The Project.

Name	URL	Content
Database of Interacting Proteins (DIP)	dip.doe-mbi.ucla.edu	Pair wise protein-protein interactions database.
SWISSPROT	www.expasy.org/sprot	Sequence database
Gene Ontology (GO)	www.geneontology.org	Describes how gene products behave in a cellular context.
SGD	www.yeastgenome.org	Database of the molecular biology and genetics of the yeast *Saccharomyces cerevisiae*
Expander	http://www.cs.tau.ac.il/~rshamir/expander/expander.html	Biclustering tool

Acknowledgment

We thank Attila Gursoy for his useful discussions during this study. Also, we thank The Scientific and Technological Research Council of Turkey (TUBITAK).

References

[1] S.-H. Yook, Z. N. Oltvai, A. L. Barabasi "Functional and topological characterization of protein interaction networks" in *Proteomics*, 2004, pp. 928–942.
[2] A.-L. Barabasi, Z. N. Oltvai "Network Biology: Understanding the Cell's Functional Organization" in *Genetics*, 2004, pp. 101–111.
[3] A. Walhout, R. Sordella, X. Lu, "Protein interaction mapping in C. elegans using proteins involved in vuval development", Science, pp.116-122, 2000.
[4] Uetz,P. et al., "A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae.", in *Nature*, 403, pp.623–627, 2000.
[5] T. Ito, K. Tashiro, S. Muta, R. Ozawa, "A comprehensive system to examine two-hybrid interactions in all possible combinations between yeast proteins", in PNAS, vol.97, pp.1143-1147, 2000.
[6] P. Pagel, S. Kovac, et al., "The MIPS mammalian protein-protein interaction database", in *Bioinformatics*, vol.21, 2005.
[7] L. Salwinski, C.S. Miller, A.J. Smith, F.K. Pettit, J.U. Bowie, D. Eisenberg "The Database of Interacting Proteins: 2004 update" in *Nucleic Acids Research*, 2004, pp.D449-51
[8] G.D. Bader, D. Betel, C.W. Hogue,"BIND:the Biomolecular Interaction Network Database", in *NAR*, vol.31, pp.248-250, 2003.
[9] B.J. Breitkreutz, C. Stark, M. Tyers, "The GRID: The General Repository for Interaction Datasets", in *Genome Biology*., vol.4, 2003.
[10] J.M. Cherry, C. Adler, C. Ball et al., "SGD: Saccharomyces Genome Database", in NAR, vol.26, pp.73-79.
[11] L. Salwinski, C.S. Miller, A.J. Smith, F.K. Pettit, J.U. Bowie, D. Eisenberg "The Database of Interacting Proteins: 2004 update" in *Nucleic Acids Research*, 2004, pp.D449-51
[12] R. Sharan, "Analysis of Biological Networks: Network Modules-Clustering Biclustering", 2005.
[13] Gene Ontology Consortium, "Gene Ontology (GO) project in 2006 ", in *Nucleic Acids Research*, 2006, pp. D322–326.
[14] O. Keskin, "Comparison of full-atomic and coarse grained models to examine the molecular fluctuations of c-AMP dependent protein kinase", in *Journal of Biomolecular Structure&Dynamics*, Vol.20, pp.1-13, 2002.
[15] A.R. Atilgan, S.R. Durell, R. L. Jernigan, M.C. Demirel, O. Keskin, I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model", in *Biophysical Journal*, vol. 80, pp.505-515, January 2001.
[16] P. Doruker, R. L. Jernigan, I. Bahar, "Dynamics of large proteins through hierarchical levels of coarse-grained structures", in *Journal of Comp Chem*, vol. 23, pp.119-127, 2002.
[17] S.M. Patra, S. Vishveshwara, "Backbone cluster identification in proteins by a graph theoretical method", in *Biophysics*, vol. 84, pp 13-25, 2005.
[18] D. Bu, Y. Zhao, L. Cai, H. Xue, et al., "Topological structure analysis of the protein-protein interaction network in budding yeast", in *Nucleic Acids Research*, vol. 31, pp.2443-2450, 2003.
[19] T.Z. Sen , A. Kloczkowski, R. Jernigan, "Functional clustering of yeast proteins from the protein-protein interaction network", BMC Bioinformatics, in 2006.
[20] Tanay,A. et al. "Revealing modularity and organization in the yeast molecularnetwork by integrated analysis of highly heterogeneous genome wide data." ,in *Proc.Natl. Acad. Sci. USA*, 101, pp.2981–2986, 2004.
[21] D. Raicu. (2004, February). Statistics with MATLAB. Available: http://facweb.cs.depaul.edu/Dstan/teaching/tutorials/Statistics%20with%20Matlab.pdf
[22] R. Shamir, A. Maron-Katz, A. Tanay, Chaim Linhart, I. Steinfeld, R. Sharan, Y. Shiloh, R. Elkon " EXPANDER – an integrative program suite for microarray data analysis", in *BMC Bioinformatics*, 2005, 6:232