

# Statistical and Integrative Approach for Constructing Biological Network Maps

**Hiroyuki Kurata**

kurata@bio.kyutech.ac.jp

**Natsumi Shimizu**

c673045n@bio.kyutech.ac.jp

**Kanako Misumi**

c673045n@bio.kyutech.ac.jp

Department of Bioscience and Bioinformatics, Kyushu Institute of Technology, 680-4  
Kawazu, Iizuka, Fukuoka 820-8502, Japan

## Abstract

A goal of systems biology is to build a concrete biochemical network map, which provides an important instruction to trace the pathways of interest or to understand the mechanism of a biological system. In the postgenomic era, not only the concrete biochemical maps, but also postgenomic maps (mRNA coexpression and protein-protein interaction networks) have been extensively produced. In the biochemical map, the individual reactions are reliable, but the number of the reactions is limited, because molecular biology requires extensive experiments to verify them. By contrast, postgenomic data provide much information regarding interactions, but are coarse-grained. To expand the biochemical network, an intuitional approach, which superposes postgenomic data on the map one by one, has been carried out, but it is not effective when a large amount of the coarse-grained data is handled. In order to effectively integrate such postgenomic interactions into a biochemical map, a statistical approach would be suitable rather than intuition. In this article, we proposed a novel statistical approach that integrates postgenomic interaction networks into the biochemical network, predicting novel pathways. A statistical correlation for such different types of networks identifies functional modules; subsequently the superposition of the different networks on the functional modules predicts inter-modular relations, which are the key pathways to construct a large-scale biochemical network.

**Keywords:** CADLIVE, biochemical networks, topological overlap, postgenome, cell cycle

## 1 Introduction

Advances in molecular biology with postgenomic technology clarify individual reactions at the molecular interaction level, assembling them as a network. The notation of concrete biochemical networks, which include not only genes and proteins but also complexes and modified molecules, has been proposed [5, 6], which enables constructing a large-scale biochemical network in common consensus. A concrete biochemical network map is an important instruction to understand the exact image of molecular networks. Molecular biologists have extensively focused on relatively local signal transduction pathways, while global regulations remain to be elucidated due to its experimental complexity. Current knowledge has not been guaranteed to be sufficient for *de novo* reconstruction of a large-scale biochemical map, especially global regulations are missed on the map. One still cannot expect enough data to support current maps.

To complement such insufficient maps from a global standpoint, postgenomic networks, such as transcriptome, proteome, and interactome networks, are available, but they provide different levels of networks, and are coarse-grained. We need to make a strategy of how different types of network knowledge are integrated to construct biochemical maps. Identification of functional modules is a key to solve this problem.

Large-scale networks are currently investigated in terms of topology, motifs, correlation structure, and modular properties that are related to function [1, 2, 3]. A functional module is defined as a

group of genes or their products that are related by one or more genetic or molecular interactions. An important property of the module is that its function is separable from other modules and that its members have more relations among themselves than with members of other modules, which is reflected in the network topology. In other words, modules are separable substructures of a network, e.g., the complex of fatty acid synthetase subunits, ribosomes, and APC complexes.

Clustering is a useful way to identify common data patterns out of different types of networks and is the first step of identifying functional modules. The different types of networks are independent but biologically related data sets, such as concrete biochemical networks, mRNA expression networks, and protein-protein interaction maps. To obtain more reliable functional modules than from the individual data sets alone, a statistical integration of the mRNA expression data into the protein-protein interaction network has been performed with correlation mapping and mean Pearson correlations [4, 7]. Integrating the information from the different postgenomic networks may lead to the notion of distinct functional modules.

However, a problem is that postgenomic data often have false positive and false negative interactions and less reliable than the molecular biology-verified reactions, which is an inherent property of current postgenomic technology. Our solution for this problem is the use of the reliable biochemical networks to verify the clustering results. A statistical correlation of the postgenomic data on the biochemical map enhances the correctness for prediction of functional modules.

To build a large-scale biochemical network, we propose a novel integrative strategy for prediction of novel pathways. This strategy aims at expanding the core network by using postgenomic data, and consists of two processes: statistical identification of functional modules by integrating different types of networks, and prediction of inter-modular pathways by superposing postgenomic data into the core biochemical map that represents prior knowledge on a particular biological sub-system. A statistical comparison of different types of networks raises the fidelity of the functional modules, and the superposition of postgenomic data on the core network helps one predict new pathways among the modules. To demonstrate the feasibility of this strategy, we employed the budding yeast cell cycle map.

## 2 Methods

### 2.1 CADLIVE

CADLIVE (Computer-Aided Design of Living systems) is a software suite for constructing large-scale models of complicated biochemical reactions such as gene regulations, complicated signal transduction pathways, and metabolic circuit, and for storing their regulator-reaction model in a database [6]. Its graphical user interface enables one to draw gene regulatory/metabolic maps in a simple manner, which eliminates the need for laborious, time-consuming, and annoying activities typically involved in this process. CADLIVE has improved on Kohn's notation for network simulation, which describes complicated signal transduction pathways and metabolic circuits in a form that can be readily processed by both computers and humans. The regulator-reaction equations and their graphical notation are able to represent not only known interactions, but also ambiguous reactions in the form of text and diagram. Using sophisticated notation draws complicated reactions, such as multicomplex formation, protein modification, regulation of transcription, and transportation, in the order of events.

### 2.2 Biological Networks

The use of the CADLIVE editor drew a concrete biochemical map of the budding yeast cell cycle that consists of 184 molecules and 152 reactions [6]. This map was constructed based on the experimental data in molecular biology, and named the CADLIVE map. The protein-protein interaction map was made based on the binding relationships between proteins in the GRID database (

mshri.on.ca/yeast\_grid/servlet/SearchPage), and named the GRID map. The gene regulation map was constructed by applying the Bayesian network inference method to the time course data for mRNA expression (VoyaGene, MKI, Ltd.), and named the BN map.

### 2.3 Graph Representation

In order to enable statistical comparison of different types of networks: the CADLIVE, GRID, and BN maps, we employ a graph representation whose nodes and edges are molecules and their relationships, respectively. The CADLIVE map includes various kinds of molecules, e.g., complexes and modified molecules, while the postgenomic networks (GRID and BN) are described with gene products, mRNAs or proteins. In this analysis, the gene products are regarded as their corresponding proteins, i.e., all the nodes are proteins. The interactions in the postgenomic networks are readily converted into a graph representation by assigning nodes and edges to the proteins and their relations, respectively. By contrast, some devices are required to convert the CADLIVE map in a form suitable for comparison with the postgenomic networks, because it includes complexes, modified molecules, and regulators that act on reactions. We extract the proteins from the CADLIVE map, relating the neighboring proteins by searching all the possible pathways among them. The generated nodes and edges show the proteins and their interactions, respectively.

### 2.4 Clustering

Time course data of mRNAs are clustered by the k-means method to obtain coexpression groups. The resultant coexpression clusters are named as the COE clusters. Since many mRNAs have similar dynamics, it is hard to cluster them by using hierarchical methods. Out of various clustering methods, the k-means method provided consistent results with the intuitive predictions. Thus, we selected the k-means method to cluster the mRNA expression profile. In the cases of the GRID and CADLIVE maps, we reduced the network map to a graph representation, and clustered in a hierarchical structure by calculating their topological overlap matrix  $O_T(i, j)$  [8]. A topological overlap of 1 between proteins  $i$  and  $j$  implies that they are connected to the same proteins, whereas a 0 value indicates that  $i$  and  $j$  do not share links to common proteins among the proteins that they react with. The proteins of highly integrated modules have a high topological overlap with their neighbors. We define the topological overlap:

$$O_T(i, j) = \frac{J_n(i, j) + \theta(l_{i,j})}{\min(k_i, k_j) + 1 - \theta(l_{i,j})}.$$

The function of  $J_n(i, j)$  denotes the number of nodes to which both  $i$  and  $j$  are linked and is provided by:

$$J_n(i, j) = \sum_{l=1}^N (l_{i,l} \cdot l_{l,j}),$$

where  $N$  is the number of nodes.  $l_{i,j} = 1$  when there is an edge between molecules  $i$  and  $j$ ,  $l_{i,j} = 0$  when there is no edge between them. The function of  $\min(k_i, k_j)$  is the smaller of the  $k_i$  and  $k_j$  degrees, where  $k_i$  is the number of the nearest neighbors of a node  $i$ .  $\theta(l_{i,j})$  is the step function. The resultant clusters in the GRID and CADLIVE maps are named as the GRID cluster and CADLIVE cluster, respectively.

### 2.5 Correlation Mapping of a Pair-wise Relationship on 2-D Cluster Matrix

To investigate a potential correlation between different level clusters, we employed the correlation mapping of a binary relationship on a two-dimensional (2-D) cluster matrix [4], as shown in Fig. 1 We generate a 2-D matrix by organizing the clusters derived from a set of related networks into two identical axes. We then arrange in the matrix pairs of proteins that can interact, according

to the cluster to which each protein belongs. For  $n$  clusters, the matrix arrangement results in  $n^2$  square, with each square representing all pair-wise combinations of proteins either in a single cluster or between two different clusters. Thus, for a matrix of  $n$  clusters, pairs of interactions can be assigned to the corresponding intra-cluster or inter-cluster squares. For each square, we calculated an index of interaction density as the ratio for the number of interaction pairs to the total number of possible pair-wise combinations of protein pairs. Significantly higher protein interaction density ( $PID$ ) for intra- versus inter-cluster squares reveals a correlation between the two clusters at a different level. The 2-D matrix shows all pair-wise combinations between the clusters. The numbers assigned to each cluster are indicated on the corresponding rows and columns of the matrix along with the number of proteins each cluster contains. The matrix shows interaction pairs with the cluster to which the corresponding proteins belong. Each square of a correlation matrix is defined by a  $(k_1, k_2)$  pair, where  $k_1$  and  $k_2$  refer to the numbers assigned to each of the clusters organized in the matrix. For each square, we calculated  $PID$  as  $IP/PP$ , where  $IP$  is the observed number of protein interaction pairs and  $PP$  is the number of all pair-wise combinations of proteins.  $PP$  is provided by:

$$PP(k_1, k_2) = n_{k_1}n_{k_2} \text{ (for intercluster squares, } k_1 < k_2)$$

$$PP(k_1, k_2) = \frac{n_{k_1}(n_{k_1} + 1)}{2} \text{ (for intercluster squares, } k_1 = k_2),$$

where  $n_{k_1}$  and  $n_{k_2}$  are the number of genes that belong to  $k_1$  and  $k_2$  clusters, respectively. A control can be generated by the same approach from randomized protein pairs. The average  $PIDs$  from all intra-cluster and inter-cluster squares can be calculated from the correlation maps.

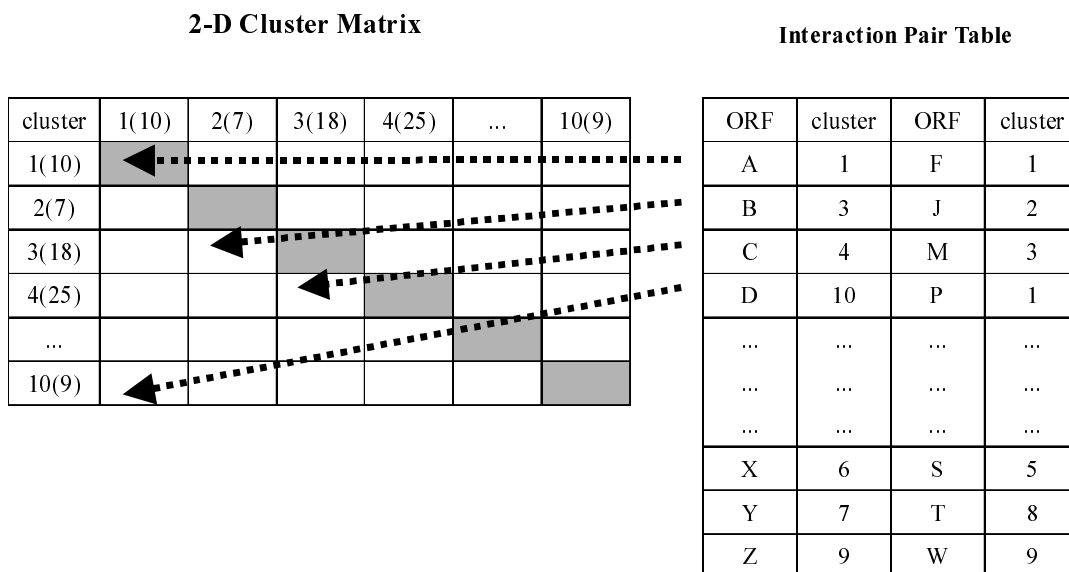


Figure 1: Mapping of the pair-wise relations of a network on the 2-D cluster matrix of another network. We generated the 2-D matrix by organizing the clusters derived from a network map into two identical axes. We then arranged in the matrix pairs of proteins that interacted according to the cluster to which each protein belonged.

## 2.6 Statistical Analysis

To determine whether the enrichment of protein interaction pairs in the intra-cluster squares is statistically significant, we used the cumulative binominal distribution given by the formula:

$$P(i > i_0) = \sum_{i=i_0}^I p^i(1-p)^{I-i} \left[ \frac{I!}{i!(I-i)!} \right].$$

The  $P$  value is the probability that the observed strength for the intra-cluster or inter-cluster correlation was obtained by random coincidence.  $I$  is the total number of protein interaction pairs sampled,  $i_o$  is the number of protein interaction pairs falling in the intra-cluster regions and  $p$  is the probability of a protein interaction falling in the intra-cluster regions, assuming that protein interaction pairs are randomly distributed. Then, the probability  $p$  is provided by:

$$p = \frac{\sum_{k=1}^K n_k(n_k + 1)/2}{T(T + 1)/2},$$

where  $K$  is the total number of clusters,  $n_k$  is the number of genes in cluster  $k$ , and  $T$  is the total number of gene in all clusters.

## 3 Results and Discussion

### 3.1 2-D Mapping of Different Types of Networks

To investigate a potential correlation between different types of networks, we employed the correlation mapping of binary relationships on a two-dimensional (2-D) cluster matrix. The binary relationships between proteins are readily identified from the CADLIVE, GRID and BN maps. The use of a hierarchical clustering method classified the CADLIVE and GRID maps into 12 and 9 groups, respectively, as shown in Fig. 2. We will explain how to determine the number of the clusters later.

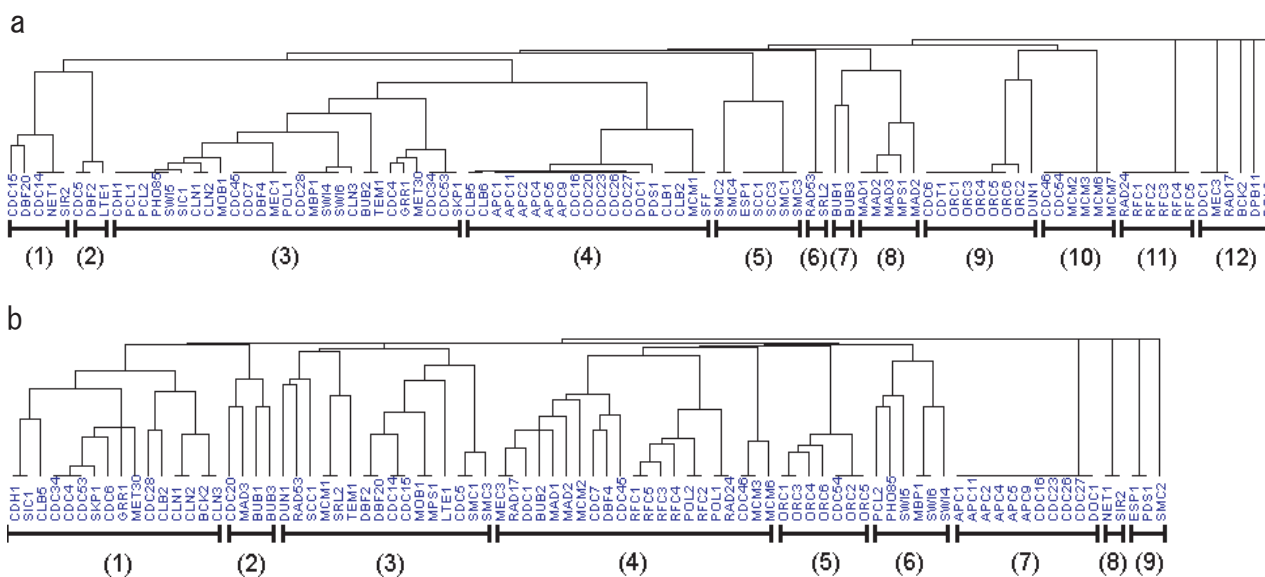


Figure 2: Hierarchical clustering for the budding yeast cell cycle map based on the topological overlap. **a.** Concrete biochemical map (CADLIVE), **b.** Protein-protein interaction map (GRID).

We generated the 2-D matrix by organizing the clusters derived from a network map into two identical axes [4]. We then arranged in the matrix pairs of proteins that interacted according to the cluster to which each protein belonged (Fig. 1). The intra-cluster or inter-cluster value of  $PID$  characterizes the correlation of the clusters derived from the two networks; the high value of the intra-cluster  $PID$  indicates that the clusters from the different networks overlap very much. Clusters of proteins showing the intra-cluster relations have a high probability that they perform the same function, which helps one finding functional modules. When proteins are randomly distributed into a cluster matrix, the  $PIDs$  of both inter-cluster and intra-cluster relations are the same. As a control, we estimate the probability ( $P$ ) that this correlation strength is found by chance. The lower the  $P$  value decreases, the more the correlation increases.

Table 1: Correlation of different types of networks by the 2-D matrix mapping. All pair-wise interactions are mapped on the 2-D cluster matrixes. Concrete biochemical network cluster (CADLIVE), mRNA coexpression cluster (COE), and protein-protein interaction cluster (GRID).

Pair	2-D cluster	<i>PID</i>		<i>P</i>	
		Intra	Inter	Intra	Random
CADLIVE	CADLIVE	18.10	0.30	3.75E-94	0.59
	COE	5.02	1.41	1.82E-11	0.58
	GRID	18.05	0.48	3.67E-72	0.60
GRID	CADLIVE	26.27	2.15	3.73E-92	0.55
	COE	8.43	4.11	4.41E-09	0.54
	GRID	40.82	0.71	2.73E-172	0.56

Table 2: 2-D cluster matrix of all pair-wise protein-protein interactions (GRID) between the CADLIVE clusters. The numbers assigned to each cluster are indicated corresponding rows and columns of the matrix along with the number of proteins each cluster contains. The number in the parentheses indicates the number of proteins in the CADLIVE cluster.

Cluster	1 (5)	2 (3)	3 (27)	4 (19)	5 (7)	6 (2)	7 (2)	8 (5)	9 (9)	10 (6)	11 (6)	12 (6)
1 (5)	4											
2 (3)	3	0										
3 (27)	7	11	73									
4 (19)	0	0	8	69								
5 (7)	0	2	2	2	3							
6 (2)	0	0	3	1	1	0						
7 (2)	0	0	1	0	0	0	3					
8 (5)	0	0	7	2	0	0	1	0				
9 (9)	0	1	7	0	1	1	0	0	10			
10 (6)	0	0	7	0	0	0	0	0	2	3		
11 (6)	0	0	4	0	0	0	0	0	1	0	23	
12 (6)	0	0	9	0	0	0	0	0	0	0	2	8

Table 1 shows the result of the correlation mapping, where we calculate the correlation strength of the protein groups that are identified as the CADLIVE, COE, and GRID clusters. The intra-cluster *PIDs* between the COE cluster and the CADLIVE or GRID cluster were higher than the inter-cluster *PIDs* under the low *P* value. Thus, the COE clusters correlated the clusters of the CADLIVE or GRID map, which is consistent with the general observation that mRNA coexpression correlates protein-protein interactions [4, 7]. The correlation of coexpressed proteins significantly overlapping with the GRID cluster appears to be a logical consequence that coexpression is required to tightly interact functionally dependent proteins.

The clusters of the CADLIVE map strongly correlated those of the GRID map. The *PID* of the intra-cluster regions was >10 times as high as that of the inter-cluster regions under the very low value of *P*. A statistically significant correlation existed between clusters across the CADLIVE map and the GRID map. This strong correlation between them helps one identifying reliable functional modules. As shown in Table 2, we mapped the binding relations for the GRID map on the 2-D clusters of the CADLIVE map. Overall, the resultant correlation map shows high-density squares along the diagonal, indicating that the combination of proteins from the same clusters results in a higher intracluster *PID*. Out of the 12 diagonal squares, nine clusters have the overlapped proteins from the two networks, although three clusters had no proteins. The reason why the non-overlapped squares appear is as follows. The CADLIVE map has not only protein-protein interactions, but also enzyme reactions and protein synthesis, whereas the GRID map has only protein binding reactions. In addition, the GRID map may have some binding relations that have not been verified in the current concrete biochemical

map.

We show how we determine the number of clusters in the different networks. We investigated how the number of clusters affected their correlation and the biological meanings of the resultant clusters. In terms of biological significance, approximately ten modules were reasonable [6], where each module can be annotated distinctly. Thus, we optimized the statistical correlations between the networks around 10 clusters. Nine clusters of the GRID network and 12 clusters of the CADLIVE map showed very high correlation (the lowest value of  $P$ ), suggesting that the CADLIVE map is well described in terms of protein interactions.

### 3.2 Mapping on Core Pathways

It is a challenging task to interpret the modules (GRID clusters) in the context of the known core pathways (the CADLIVE map). In this analysis, mapping the functional modules on the core pathways assigns biological significances to the modules, and determines the relationship among them. As shown in Fig. 3, we mapped the GRID clusters on the CADLIVE map, and named the clusters in terms of the molecular biology-annotated cell cycle events, e.g., starting process, DNA replication, chromatin formation, spindle formation, and checkpoints. According to such conventional annotations, the GRID clusters on the CADLIVE map were named as follows: (1) global regulation that includes *cdc28*, (2) spindle formation and chromatin checkpoint, (3) chromatin condensation and spindle checkpoint, (4) DNA replication and DNA checkpoint, (5) complex for DNA replication, (6) starting process, (7) APC complex, (8) RENT complex, (9) chromatin separation.

In the GRID clusters, some events that have been conventionally annotated based on molecular biology are combined through protein binding. For example, the central regulator of *cdc28* put together several cyclins that govern various cell cycle modules, such as DNA replication and spindle formation, in the GRID module (1), spindle formation and chromatin checkpoint are coupled in the GRID module (2), DNA replication and DNA checkpoint are also coupled in the GRID module (4). These GRID-derived functional modules are not perfectly consistent with the conventional annotations, although they can be explained in terms of biology. The inconsistency will provide a new insight for reconstructing the functional modules of the cell cycle map.

### 3.3 Prediction of Novel Pathways of Inter-Cluster Modules

In order to explore the inter-cluster relations among the GRID-derived functional modules, we mapped the binary relations of the BN map on the 2-D GRID clusters (Table 3). The modules of the GRID network strongly correlates the CADLIVE map and are biologically annotated, but intrinsically excludes information of gene regulations. Thus, we employed the BN network to predict the inter-cluster relationship among the functional modules. As shown in Table 3, out of 12 intra-cluster regulations, 10 regulations are traced in the CADLIVE map. By contrast, out of 35 inter-cluster regulations, 27 regulations are not found in the CADLIVE map, suggesting that the inter-modular relations remain to be elucidated. The regulations that have not been found in the CADLIVE map are the key to relate the functional modules. For example, 17 inter-modular regulations between the chromatin condensation and spindle checkpoint module (3) and the modules (1, 4-7) including global regulations, DNA replication, DNA checkpoint, complex for DNA replication, starting process, and APC complex are predicted to control cell the cycle events. The predicted pathways are the relations for DUN1-CLN1, DUN1-CLN2, TEM1-CLN1, TEM1-CLN2, TEM1-CLB5, MCM6-DBF2, RFC2-TEM1, ORC1-MOB1, ORC3-TEM1, ORC1-CDC5, MPS1-SWI4, SMC1-PCL2, SWI4-CDC5, PHO85-MOB1, APC1-DBF2, APC1-RAD53, and APC5-DBF20. The superposition of the BN map on the functional modules suggests that many inter-modular regulations have not been discovered in the current budding yeast cell cycle map (the CADLIVE map). It may show that molecular biology has focused on the local networks within the functional modules, but has not elucidated the inter-modular pathways extensively. These inter-modular regulations are the key to construct the exact process flow of the cell cycle.

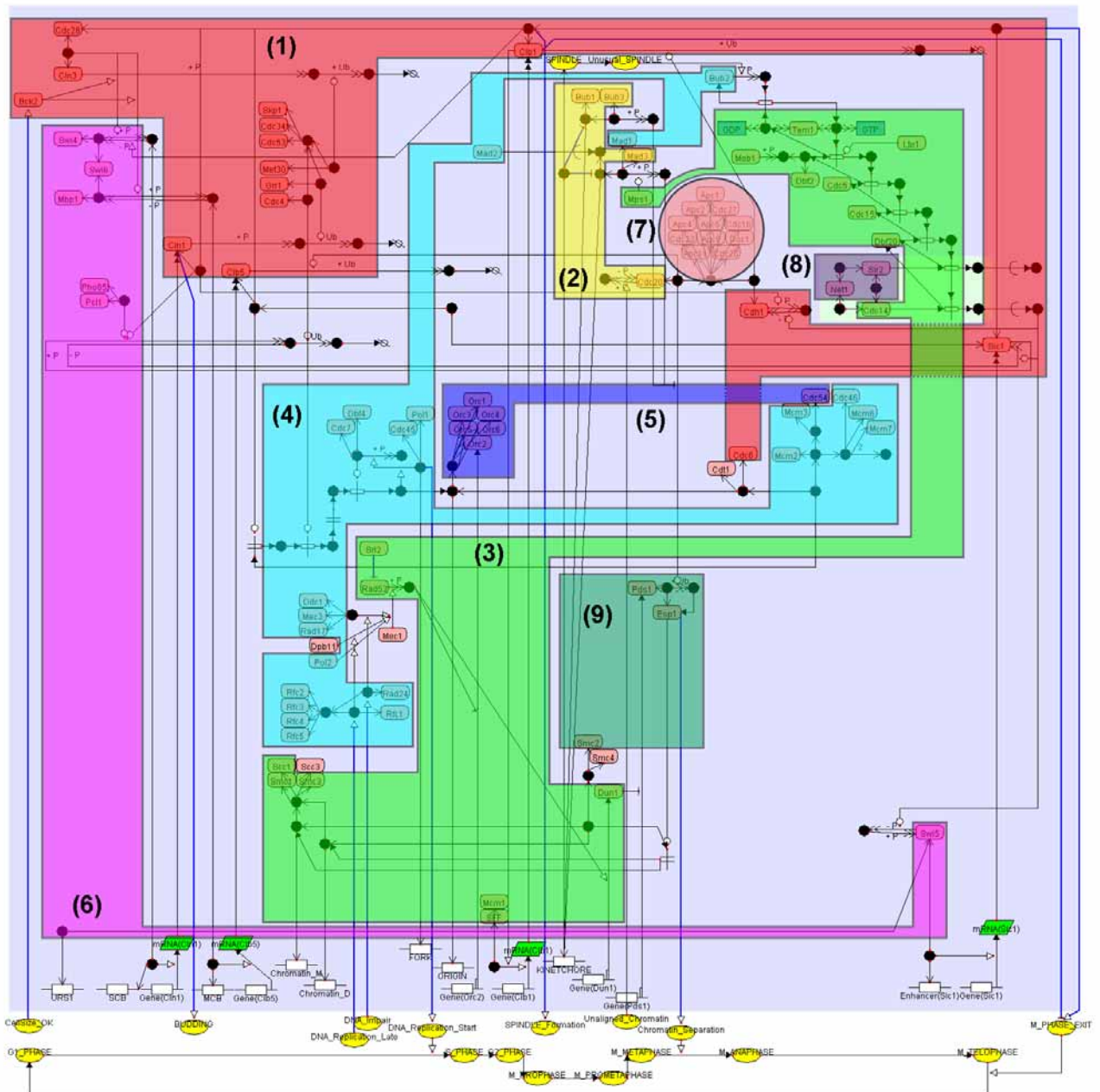


Figure 3: Mapping the functional modules (GRID clusters) on the budding yeast cell cycle map (CADLIVE). The closed squares are the functional modules derived from GRID database. On the map not only proteins but also their related complexes and modified molecules are grouped together as the GRID clusters. The genes and events are excluded. The proteins that do not belong to the cluster are not included in the GRID database. The CADLIVE map is manually constructed by the CADLIVE system based on the literatures, and consists of 184 molecules and 152 reactions in a really compact space. The molecules and reactions are placed from the left to the right according to the event, and located from the bottom to the top according to the gene and protein signal transduction layers. The round rectangles are proteins, the rectangles metabolites, the parallelograms mRNAs, and the ovals events. The solid black circles are complexes or modified molecules. The various kinds of arrows indicate reactions (binding, modification, conversion, transcription, translation, and transportation) or regulations (by an enzyme, activator, or inhibitor).

Table 3: Superposition of the gene regulation pairs (BN network) on the protein-protein interaction matrix (GRID cluster). The Bayesian network inferred the gene regulation network that consisted of the pair-wise relationships between mRNAs. The number in the parentheses indicates the number of proteins in the GRID clusters. In each square, the number of unknown pairs / the number of total inferred pairs in each cluster is indicated. The unknown pair indicates that it is not found in the CADLIVE map.

GRID Cluster	1 (16)	2 (4)	3 (16)	4 (21)	5 (7)	6 (6)	7 (11)	8 (2)	9 (3)
1 (16)	0 / 1								
2 (4)	0 / 1	0 / 1							
3 (16)	5 / 7	2 / 2	2 / 6						
4 (21)	2 / 2	0	2 / 3	0 / 3					
5 (7)	0	0	3 / 3	1 / 1	0 / 1				
6 (6)	0 / 1	1 / 1	4 / 5	0	0	0			
7 (11)	0 / 1	0	3 / 3	3 / 3	0	0	0		
8 (2)	1 / 1	0	0	0	0	0 / 1	0	0	
9 (3)	0	0	0	0	0	0	0	0	0

### 3.4 Core Expansion

In order to construct a large-scale biochemical network map, we propose a novel statistical/integrative strategy that expands a known core map for concrete biochemical reactions by integrating postgenomic networks. Identification of functional modules plays an important role for predicting inter-modular pathways. This approach consists of two processes: (i) statistical identification of reliable functional modules by integrating different types of networks, and (ii) prediction of novel pathways by superposing inferred gene regulations on functional modules. The statistical integration of different networks helps one finding reliable functional modules. In this analysis, 12 clusters of the CADLIVE map strongly correlated 9 clusters of the GRID map, showing that the core pathway map is well described in terms of protein interactions. The superposition of different types of networks suggests that most inter-modular pathways are not found in the current budding yeast cell cycle map, predicting novel pathways among the components that are located in different modules. To construct a biochemical network of the cell cycle, biological predictions would require exploring inter-modular relations rather than focusing on local signal transduction pathways within a module intensively.

## Acknowledgments

This study was carried out as a part of The Project for Development of a Technological Infrastructure for Industrial Bioprocesses on R&D of New Industrial Science and Technology Frontiers by Ministry of Economy, Trade & Industry (METI), and entrusted by New Energy and Industrial Technology Development Organization (NEDO). This work was also partially supported by a Grant-in-Aid for Scientific Research on Priority Area by of the Ministry of Education, Culture, Sports, Science and Technology, Japan

## References

- [1] Asthana, S., King, O.D., Gibbons, F.D., and Roth, F.P., Predicting protein complex membership using probabilistic network reliability, *Genome Res.*, 14:1170–1175, 2004.

- [2] Bar-Joseph, Z., Gerber, G.K., Lee, T.I., Rinaldi, N.J., Yoo, J.Y., Robert, F., Gordon, D.B., Fraenkel, E., Jaakkola, T.S., Young, R.A., and Gifford, D.K., Computational discovery of gene modules and regulatory networks, *Nat. Biotechnol.*, 21:1337–1342, 2003.
- [3] Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Zhang, J., Sun, S., Ling, L., Zhang, N., Li, G., and Chen, R., Topological structure analysis of the protein-protein interaction network in budding yeast, *Nucleic Acids Res.*, 31:2443–2450, 2003.
- [4] Ge, H., Liu, Z., Church, G.M., and Vidal, M., Correlation between transcriptome and interactome mapping data from *Saccharomyces cerevisiae*, *Nat. Genet.*, 29:482–486, 2001.
- [5] Kohn, K.W., Molecular interaction map of the mammalian cell cycle control and DNA repair systems, *Mol. Biol. Cell*, 10:2703–2734, 1999.
- [6] Kurata, H., Matoba, N., and Shimizu, N., CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle, *Nucleic Acids Res.*, 31:4071–4084, 2003.
- [7] Pereira-Leal, J.B., Enright, A.J., and Ouzounis, C.A., Detection of functional modules from protein interaction networks, *Proteins*, 54:49–57, 2004.
- [8] Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L., Hierarchical organization of modularity in metabolic networks, *Science*, 297:1551–1555, 2002.