Protein complex prediction based on simultaneous protein interaction network

Suk Hoon Jung1, Bora Hyun1, Woo-Hyuk Jang1, Hee-Young Hur1 and Dong-Soo Han2*

1Department of Information & Communications Engineering, Korea Advanced Institute of Science and Technology, 119 Munjiro, Yuseong-gu, Daejeon, 305–714 and 2Department of Computer Science, Korea Advanced Institute of Science and Technology, 335 Gwahangno, Yuseong-gu, Daejeon, 305–710, Korea

Received on February 24, 2009; revised on November 22, 2009; accepted on November 28, 2009
Advance Access publication December 4, 2009

ABSTRACT

Motivation: The increase in the amount of available protein–protein interaction (PPI) data enables us to develop computational methods for protein complex predictions. A protein complex is a group of proteins that interact with each other at the same time and place. The protein complex generally corresponds to a cluster in PPI network (PPIN). However, clusters correspond not only to protein complexes but also to sets of proteins that interact dynamically with each other. As a result, conventional graph-theoretic clustering methods that disregard interaction dynamics show high false positive rates in protein complex predictions.

Results: In this article, a method of refining PPIN is proposed that uses the structural interface data of protein pairs for protein complex predictions. A simultaneous protein interaction network (SPIN) is introduced to specify mutually exclusive interactions (MEIs) as indicated from the overlapping interfaces and to exclude competition from MEIs that arise during the detection of protein complexes. After constructing SPINs, naive clustering algorithms are applied to the SPINs for protein complex predictions. The evaluation results show that the proposed method outperforms the simple PPIN-based method in terms of removing false positive proteins in the formation of complexes. This shows that excluding competition between MEIs can be effective for improving prediction accuracy in general computational approaches involving protein interactions.

Availability: http://code.google.com/p/simultaneous-pin/

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Recent developments in biotechnology have resulted in an increase in the amount of protein–protein interaction (PPI) data. Modeling a PPI network (PPIN) with simple graphs enables many computational methods for the study of protein functions (Brohé and Helden, 2003), LCMA (Local Clique Merging Algorithm; Li et al., 2005), SPC (Super Para-magnetic Clustering; Blatt et al., 1997), RNESC (Restricted Neighborhood Search Clustering; King et al., 2004) and DPClus (Altufiev; Amin et al., 2006; Li et al., 2008), typically focus on the extraction of clusters based on the graph theory.

One specific problem pertaining to conventional methods originates from the fact that with these methods, a PPIN is regarded as a static entity. In reality, a PPIN is not a static but a dynamic entity; the functional state of the network depends on the expression of protein nodes, which is intrinsically controlled by different regulatory mechanisms through time and space (Han et al., 2004; Liang and Li, 2007). In a dynamic network, a protein complex is a group of proteins in which individual proteins interact with each other at the same time and place (Spirin and Mirny, 2008). However, a cluster in a PPIN may include proteins that interact dynamically with each other as well. Conventional approaches based on a simple PPIN cannot properly distinguish protein complexes from interactions that may be activated at a different time and place because they disregard interaction dynamics. This leads to false positive results in protein complex detections (Spirin and Mirny, 2008).

A means of tackling this problem is to use the features of proteins additionally as indirect evidence. Some methods use machine learning methods, and some others enrich the protein interaction network by assigning weights based on functional annotations; gene expression data; or biological, chemical and physical properties (Pei and Zhang, 2006; Qi et al., 2008; Zhang et al., 2006). Given that these features are known to be relevant to protein mechanisms in general, considering them in clustering algorithms may improve the prediction results. However, using indirect evidence is not adequate in itself to pinpoint complexes in PPIN because indirect evidence is not deterministic in complex formations.

Unlike previous approaches, this article focuses on competitive interactions in a PPIN, which are considered to provide more direct and determinative evidence in the identification of proteins in the formation of a complex. Interactions must occur simultaneously in a PPIN, which are considered to provide more direct and determinative evidence in the identification of proteins in the formation of a complex. Thus, competitive interactions in a PPIN should be prudently selected before they are included in a predicted complex.

In addition, many proteins are known to usually have a number of interacting partners, some of which may cooperate or even
A close look into the physical interfaces between interacting proteins provides information on mutual exclusiveness among the interacting partners of a protein, and mutual exclusiveness results in interaction competition. If two or more interaction partners can bind to a common or an overlapping interfacial surface of a protein, the surface is considered to be physically available only for one partner at a given moment. Such interactions are mutually exclusive, as the occurrence of one of these interactions automatically excludes the occurrence of the remaining interactions. A target protein whose partners compete for the interaction is termed the host protein in this article. In addition, the term MEI is used to denote a pair of interactions that is mutually exclusive for a host protein. A case in which more than two interfacial surfaces are overlapped or partially cascaded is represented by a set of MEIs.

Figure 1 depicts an example of the modeling of an MEI. The first step in the detection of MEIs is to identify the interface of each protein interaction, which is represented by a set of interfacial residue pairs. In this research, an interface between a protein pair is examined at the level of the protein domain. The protein domain is an evolutionary conserved unit of the structure and function of the protein, and hence, it is regarded as a subunit that mediates PPIs (Boxem et al., 2008).

Figure 2 illustrates the process of MEI extraction using PSIMAP (Gong et al., 2005). PSIMAP provides information pertaining to interfacial residue pairs in physical domain-domain interactions (DDI) based on an analysis of the crystal structures of proteins, the protein interacting pairs and the complex recorded in the PDB (Berman et al., 2000). Similar to PSIMAP, this study adopts the SCOP domain definition. For each domain, we compute overlapping interfacial residues for all possible pairs of partners with which the domain interacts (Fig. 2b and c). In this process, self-pairing of each partner domain should be considered as well because a protein may have several interacting partner proteins mediated by an identical DDI. Another consideration is that a pair of domains can interact through several different interfaces (Aragues et al., 2007; Winter et al., 2006). Hence, although two partner domains seem to have an overlapping binding site on a host domain, they could still bind simultaneously by using disjoint alternative binding sites on the host. Therefore, two partners are recognized to be mutually exclusive if and only if they have no other option but to compete for an overlapping interfacial surface on the host domain.

The next step is protein domain assignment by referring Interpro (Hunter et al., 2009) that offers integrative protein signature data. A DDI interface is used in identifying the interface of a PPI mediated by the corresponding DDI, and MEIs are inferred by referring to the mutually exclusive DDI data that is obtained (Fig. 2d). In this process, the DDI within a protein is ignored because its interface is considered to be already occupied by an alternative binding site on the host. Therefore, two partners are recognized to be mutually exclusive if and only if they have no other option but to compete for an overlapping interfacial surface on the host domain. It is possible to represent an MEI relationship using a Boolean expression. In conventional network model, an interaction is represented with a static edge regardless of the time and/or conditions. With this conjecture, the interaction list of a protein can be represented as a conjunction of all interactions where an interaction has a value of true when it occurs. However, two interactions of a MEI should be connected by XOR (⊕) as both cannot occur simultaneously. Figure 3 illustrates an example of representing MEI information in a simple network. The notation $\text{Xor}_{p}$ is used to represent interactions with
The SPIN is a subnetwork of a PPIN. A SPIN is comprised of a set of non-competitive interactions and all of the proteins inherited from the original network. A non-competitive interaction set selectively includes one of the mutually exclusive pairs of each protein in order to achieve mutual exclusion among the interactions. Therefore, its interactions may be activated simultaneously without competition in nature. SPINs from a PPIN can be viewed as snapshots, each of which represents a possible coactive state that the dynamic network may attain.

Based on the xSPIN, \( P \setminus MEI \), computed from the MEIs, SPINS are extracted from the PPIN based on each conjunctive clause in xSPIN, \( P \setminus MEI \). In Figure 4, the PPIN has two MEIs \( i_5 \), \( i_6 \) and a subnetwork preparation adopts a naive clustering algorithm which is used in PSIMAP, implying that a subnetwork for SPIN construction should cover all of the dense regions. The proposed framework focuses on the extraction of non-competitive sets of proteins in a PPIN; hence, protein complex detection from extracted sets exploits conventional clustering algorithms. In this research, MCODE and LCMA are adopted from among various conventional graph-theoretic clustering algorithms for the evaluation of the framework.
Sub-network Sub-network SPIN Construction Post - clustering

Fig. 5. Outline of the SPIN-based framework.

3 RESULTS

In an effort to evaluate the proposed framework for detecting protein complexes, it was compared with plain PPIN-based clustering methods. Two experiments were performed on a SPIN using the two clustering algorithms MCODE and LCMA. They are considered SPIN-based methods and were termed SPIN_MCODE and SPIN_LCMA, respectively. The same algorithms were also applied to a plain PPIN for comparison, and they are termed PPIN_MCODE and PPIN_LCMA for convenience sake. For fair comparisons, identical parameters for the PPIN- and SPIN-based methods were used in a clustering algorithm.

Additionally, we performed experiments based on random SPIN in a comparison to determine whether or not our improvement stems from using structural MEI information. In these experiments, the same procedure used with the SPIN framework was utilized; however, after a subnetwork preparation step, each prepared subnetwork was assigned with randomly generated MEIs with the same number as its structural MEIs.

The following subsections present the results of the experiments, explain the relationship between the clusters in the SPINs and a plain PPIN, compare the prediction results with known complexes, and discuss the effect of the SPIN-based framework.

3.1 Reference sets

Experiments were performed on the *S. cerevisiae* (yeast) interactome downloaded from the MIPS MPact database (Guldener et al., 2006). After removing all the self-interactions, the final network contained 15,524 interactions among 4579 yeast proteins. Two clustering algorithms on two base networks, PPIN and SPIN, generated four predicted cluster sets, and these prediction results were compared with known protein complexes recorded in the MIPS yeast complex database (Guldener et al., 2006). There were 267 manually annotated complexes that were considered as gold standard data.

3.2 MEI extraction

From 14,594 multi-domain PDB entries (release date December 5, 2008), PSIMAP extracted 4948 DDIs with 64,985 examples of interface evidence among 2527 domains. Among them, 1842 domains were revealed to have at least one pair of partner domains that were mutually exclusive, and the number of mutually exclusive pairs was 6174 in total. Supplementary Table 1 lists the mutually exclusive DDIs, overlapping residue indexes and PDB evidence.

In PPIN network, it was found that 100 proteins had at least one MEI competing for the interaction with them, and there were 458 MEIs in the network. Supplementary Table 2 lists the MEIs on each host protein along with the mutually exclusive DDI pair to which the MEIs refer.

As discussed an actual complex should not have MEIs within it, we investigated the occurrence of MEIs in MIPS complexes. There were 14 MEIs in six out of 267 MIPS complex data. We hypothesized that there might be incomplete interface data. Specifically, host proteins of the 14 MEIs might have an unknown alternate binding site which allow for the MEIs to occur simultaneously.

3.3 The relationship between the clusters in PPIN and SPIN

A SPIN is constructed by refining a PPIN, and proteins in the refined network cannot be more interactive compared with those in the PPIN. Additionally, SPIN-based methods use the same clustering algorithm as comparison methods with the same parameters. Therefore, a SPIN cluster must be a subgraph of the corresponding PPIN cluster.

Table 1 shows a summary of the prediction results of the four methods. LCMA extracts a much larger number of clusters compared with MCODEs because, unlike MCODE, LCMA finds loosely connected clusters that may be overlapped.

Applying the SPIN concept increases the number of predicted clusters in both MCODE and LCMA. However, the number of distinct proteins in SPIN clusters is fewer than that in PPIN clusters. This indicates that the clustering on the SPINs results in the removal of proteins in the original clusters. As many interactions and all of the proteins appear in common in the SPIN and the PPIN, many SPIN clusters are identical to PPIN clusters. On the other hand, the occurrence of MEIs creates a difference between the results of the SPIN- and PPIN-based methods. PPIN_MCODE generates nine unique clusters that have MEIs. When the SPIN is constructed,
enforcing mutual exclusion eliminates some of the interactions in the network regions where nine PPIN clusters are located. As a result, SPIN_MCODE generates 40 unique clusters that were also subgraphs of the corresponding PPIN clusters. Likewise, 142 clusters of PPIN_LCMA were redefined into 519 clusters by the SPIN-based method.

### 3.4 Comparison with known complexes

The results were assessed using an evaluation metric used in earlier studies (Altaf-Ul-Amin et al., 2006; Bader and Hogue, 2003; Li et al., 2005, 2008) to determine how effectively a predicted cluster matches a known complex, and vice versa. Equation 1 calculates the overlapping score $OS(p, m)$ between a predicted cluster $p \in P$ and a known complex $m \in M$, where $P$ is the set of predicted clusters and $M$ is the set of known complexes as recorded in MIPS.

$$OS(p, m) = \frac{|V_p \cap V_m|^2}{|V_p| \times |V_m|} \quad (1)$$

In Equation (1), $|V_p \cap V_m|$ is the size of the intersection protein set of the predicted cluster and the known complex, $|V_p|$ is the number of proteins in the predicted cluster and $|V_m|$ is the number of proteins in the known complex. A known complex and a predicted cluster are considered as a match if their overlapping score is equal to or larger than a specific threshold. Conventionally, a predicted cluster and a known complex are considered to a match if $OS(p, m) \geq 0.2$ (Altaf-Ul-Amin et al., 2006; Bader and Hogue, 2003; Li et al., 2005, 2008).

After all known complexes and predicted clusters have their best match calculated according to their OS scores, three evaluation criteria are applied to quantify the quality of the protein complex detection methods:

- **Precision ($p$):** measures the fraction of the predicted clusters that match the positive complexes among all predicted clusters.
- **Recall ($r$):** measures the fraction of known complexes matched by predicted clusters, divided by the total number of known complexes.
- **F1:** the F1 score combines the precision and recall scores. It is defined as $2pr/(p+r)$.

**Recall** quantifies the extent to which a prediction set captures the known complexes. **Precision** measures the exactness or fidelity of the prediction set. The F1 measure provides a reasonable combination of both precision and recall. All three values range from 0 to 1, with 1 being the best score. These three criteria are frequently used in many computational areas including protein complex detection (Qi et al., 2008). Here, because our reference set MIPS is incomplete, some predicted clusters which are more likely true complexes will be regarded as false positives if they do not match the current MIPS complexes well. As such, the F-measure of the algorithms should not be taken at their absolute values but only as comparative measures.

The performance comparison is presented in Table 2. For each method, we report the **precision**, **recall** and **F1**, with the threshold OS $\geq 0.2$. As can be seen, our methods based on SPIN dominate PPIN-based methods in all measures. In terms of the F1 measures, SPIN_MCODE achieved a 23% higher value compared with the PPIN_MCODE value. When using the LCMA algorithm, SPIN_LCMA achieved a 31% higher F1-value compared with the PPIN_LCMA result.

In contrast with SPIN-based methods, the experiments based on random SPIN showed minor changes compared with the PPIN-based results in all three measures. This indicates that the improvements of the SPIN-based methods stem from the use of structural MEI information.

As the proposed framework aims to exclude superfuous proteins in the formation of complexes, the overlapping score of a known complex with a SPIN cluster is expected to be equal to or greater than the score of the corresponding PPIN cluster. Table 3 shows the number of known complexes matched by the clusters extracted by MCODE and LCMA from the PPIN, the SPIN and the random SPIN with respect to different thresholds. The word **loss** in parentheses refers the number of complexes that are matched by PPIN clusters without distinction, all unique PPIN clusters have MEIs.

### Table 1. Summary of the prediction results from the four methods

<table>
<thead>
<tr>
<th></th>
<th>MCODE</th>
<th>SPIN_MCODE</th>
<th>LCMA</th>
<th>SPIN_LCMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Predicted clusters</td>
<td>140</td>
<td>171</td>
<td>1696</td>
<td>2073</td>
</tr>
<tr>
<td>(b) Proteins in clusters</td>
<td>620</td>
<td>543</td>
<td>3731</td>
<td>2967</td>
</tr>
<tr>
<td>(c) Identical clusters</td>
<td>131</td>
<td>1554</td>
<td>142</td>
<td>519</td>
</tr>
<tr>
<td>(d) Unique clusters</td>
<td>9</td>
<td>0</td>
<td>147</td>
<td>1274</td>
</tr>
<tr>
<td>(e) MEIs included</td>
<td>147</td>
<td>0</td>
<td>1274</td>
<td>0</td>
</tr>
</tbody>
</table>

(a) the number of predicted clusters. (b) The number of distinct proteins of the predicted clusters. (c) The number of clusters that are predicted by the naive- and SPIN-based method in common. (d) The number of unique clusters (a–c). (e) The number of MEIs included in clusters without distinction, all unique PPIN clusters have MEIs.

### Table 2. Performance comparison between the methods based on PPIN, SPIN and random SPIN (Ran_SPIN)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Network</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PPIN</td>
<td>0.213</td>
<td>0.314</td>
<td>0.254</td>
</tr>
<tr>
<td></td>
<td>SPIN</td>
<td>0.243</td>
<td>0.441</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td>Ran_SPIN</td>
<td>0.199</td>
<td>0.358</td>
<td>0.255</td>
</tr>
<tr>
<td>MCODE</td>
<td>PPIN</td>
<td>0.401</td>
<td>0.098</td>
<td>0.158</td>
</tr>
<tr>
<td></td>
<td>SPIN</td>
<td>0.528</td>
<td>0.128</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>Ran_SPIN</td>
<td>0.438</td>
<td>0.094</td>
<td>0.155</td>
</tr>
<tr>
<td>LCMA</td>
<td>PPIN</td>
<td>0.528</td>
<td>0.128</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>SPIN</td>
<td>0.528</td>
<td>0.128</td>
<td>0.207</td>
</tr>
<tr>
<td></td>
<td>Ran_SPIN</td>
<td>0.438</td>
<td>0.094</td>
<td>0.155</td>
</tr>
</tbody>
</table>
The number of known complexes matched by predicted clusters from PPIN, SPIN and random SPIN with respect to different thresholds

Table 3. The number of known complexes matched by predicted clusters from PPIN, SPIN and random SPIN with respect to different thresholds

<table>
<thead>
<tr>
<th>MCODE</th>
<th>PPIN (gain, loss)</th>
<th>SPIN (gain, loss)</th>
<th>Ran_SPIN (gain, loss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS ≥ 0.1</td>
<td>133 (0, 0)</td>
<td>133 (0, 0)</td>
<td>128 (0, 5)</td>
</tr>
<tr>
<td>OS ≥ 0.2</td>
<td>88 (3, 0)</td>
<td>91 (3, 0)</td>
<td>80 (6, 14)</td>
</tr>
<tr>
<td>OS ≥ 0.3</td>
<td>43 (1, 0)</td>
<td>51 (8, 0)</td>
<td>40 (1, 4)</td>
</tr>
<tr>
<td>OS ≥ 0.4</td>
<td>33 (5, 0)</td>
<td>38 (5, 0)</td>
<td>30 (0, 3)</td>
</tr>
<tr>
<td>OS ≥ 0.5</td>
<td>28 (4, 0)</td>
<td>32 (4, 0)</td>
<td>28 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.6</td>
<td>17 (6, 0)</td>
<td>23 (6, 0)</td>
<td>17 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.7</td>
<td>11 (5, 0)</td>
<td>16 (5, 0)</td>
<td>11 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.8</td>
<td>10 (4, 0)</td>
<td>14 (4, 0)</td>
<td>10 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.9</td>
<td>7 (2, 0)</td>
<td>9 (2, 0)</td>
<td>7 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 1.0</td>
<td>0 (0, 0)</td>
<td>2 (0, 0)</td>
<td>2 (0, 0)</td>
</tr>
</tbody>
</table>

V.S. Jung et al.

The values of gain in parentheses are the number of true positives found in addition to the result of the PPIN-based method.

The values of loss were all zero for the SPIN-based methods.

The table discards the number of known complexes in the case that OS = 0, which are matched by no predicted cluster. OS > 0 indicates that the known complex has a matching predicted cluster in that it shares at least one protein. As a SPIN cluster is a subgraph of a PPIN cluster, the number of complexes matched by SPIN clusters cannot exceed that matched by SPIN clusters at the threshold of OS > 0. However, for the remaining thresholds, SPIN-based methods show better results than PPIN-based approaches.

The values of loss were all zero for the SPIN-based methods. This finding indicates that the results of the SPIN-based methods perfectly covered all the true positive matches from the PPIN-based methods with the thresholds listed in the table while also generating additional true positives. Unlike SPIN constructed using structural MEIs, alternating with random MEIs results in some loss of known complexes as well as additional gains. This result was possible because randomizing the MEI information may remove true and false positive proteins all together as results.

Our model may incorrectly remove true positive protein members, although it generates no loss of matched complexes. In this experiment, SPIN_MCORE removed only false positive proteins, whereas SPIN_LCMA showed two cases of protein loss as it removed two true positive members for matching with known complexes. (See MIPS complexes 410.20 and 160 in Supplementary Table 4.) However, in these cases, the known complexes had a higher OS with the SPIN cluster compared with those that used the PPIN cluster, as the SPIN framework removed many superfluous proteins. This result indicates that the proposed network model using structural MEI information can be successfully applied to graph-theoretic clustering methods for complex predictions with few faults.

3.5 The effect of the SPIN construction

The proposed network model refines a PPIN by excluding interaction competitions and it generates several subnetworks that represent possible coactive states in a process of interaction dynamics.

Table 3. The number of known complexes matched by predicted clusters from PPIN, SPIN and random SPIN with respect to different thresholds

<table>
<thead>
<tr>
<th>MCORE</th>
<th>PPIN (gain, loss)</th>
<th>SPIN (gain, loss)</th>
<th>Ran_SPIN (gain, loss)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OS ≥ 0.0</td>
<td>133 (0, 0)</td>
<td>133 (0, 0)</td>
<td>128 (0, 5)</td>
</tr>
<tr>
<td>OS ≥ 0.1</td>
<td>88 (3, 0)</td>
<td>91 (3, 0)</td>
<td>80 (6, 14)</td>
</tr>
<tr>
<td>OS ≥ 0.2</td>
<td>57 (8, 0)</td>
<td>65 (8, 0)</td>
<td>63 (4, 8)</td>
</tr>
<tr>
<td>OS ≥ 0.3</td>
<td>43 (1, 0)</td>
<td>51 (8, 0)</td>
<td>40 (1, 4)</td>
</tr>
<tr>
<td>OS ≥ 0.4</td>
<td>33 (5, 0)</td>
<td>38 (5, 0)</td>
<td>30 (0, 3)</td>
</tr>
<tr>
<td>OS ≥ 0.5</td>
<td>28 (4, 0)</td>
<td>32 (4, 0)</td>
<td>28 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.6</td>
<td>17 (6, 0)</td>
<td>23 (6, 0)</td>
<td>17 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.7</td>
<td>11 (5, 0)</td>
<td>16 (5, 0)</td>
<td>11 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.8</td>
<td>10 (4, 0)</td>
<td>14 (4, 0)</td>
<td>10 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 0.9</td>
<td>7 (2, 0)</td>
<td>9 (2, 0)</td>
<td>7 (0, 0)</td>
</tr>
<tr>
<td>OS ≥ 1.0</td>
<td>0 (0, 0)</td>
<td>2 (0, 0)</td>
<td>2 (0, 0)</td>
</tr>
</tbody>
</table>

The values of gain in parentheses are the number of true positives found in addition to the result of the PPIN-based method.

The values of loss were all zero for the SPIN-based methods. This finding indicates that the results of the SPIN-based methods perfectly covered all the true positive matches from the PPIN-based methods with the thresholds listed in the table while also generating additional true positives. Unlike SPIN constructed using structural MEIs, alternating with random MEIs results in some loss of known complexes as well as additional gains. This result was possible because randomizing the MEI information may remove true and false positive proteins all together as results.

Our model may incorrectly remove true positive protein members, although it generates no loss of matched complexes. In this experiment, SPIN_MCORE removed only false positive proteins, whereas SPIN_LCMA showed two cases of protein loss as it removed two true positive members for matching with known complexes. (See MIPS complexes 410.20 and 160 in Supplementary Table 4.) However, in these cases, the known complexes had a higher OS with the SPIN cluster compared with those that used the PPIN cluster, as the SPIN framework removed many superfluous proteins. This result indicates that the proposed network model using structural MEI information can be successfully applied to graph-theoretic clustering methods for complex predictions with few faults.

3.5 The effect of the SPIN construction

The proposed network model refines a PPIN by excluding interaction competitions and it generates several subnetworks that represent possible coactive states in a process of interaction dynamics.
could not differentiate these overlapping complexes but predicted a massive cluster matched by these three and another complex 270.10.10 with several superfluous proteins. On the other hand, given the structural interface data, our SPIN construction process specified competitions among exchangeable proteins YFL099w, YJR090c and YIL056w for the interaction with the core protein YDR328c. Consequently, for the network region in which the PPIN YJR090c and YIL046w for the interaction with the core protein This study introduces a network refinement model based on the all of the complexes that the PPIN-based method found as well framework outperforms the plain PPIN-based method. It found PPIN for comparison. The comparison showed that the SPIN-based two graph-theoretic clustering algorithms on SPIN and on a simple construction reserves sets of non-competitive interactions by entity, includes competitive interactions that cannot participate in complex formations together. In the proposed framework, a predictions. A simple PPIN, which is represented as a static structural interface data of protein pairs for protein complex entity, includes competitive interactions that cannot participate in complex formations together. In the proposed framework, a predictions. A simple PPIN, which is represented as a static structural interface data of protein pairs for protein complex entity, includes competitive interactions that cannot participate in complex formations together. In the proposed framework, a predictions. A simple PPIN, which is represented as a static structural interface data of protein pairs for protein complex

4 CONCLUSIONS

This study introduces a network refinement model based on the structural interface data of protein pairs for protein complex predictions. A simple PPIN, which is represented as a static entity, includes competitive interactions that cannot participate in complex formations together. In the proposed framework, a SPIN construction reserves sets of non-competitive interactions by

Supplementary Table 3 lists MIPS complexes matched by MCCODEs based on PPIN and SPIN along with their overlapping scores and matched proteins, and Supplementary Table 4 lists those for LCMAs.

**REFERENCES**


