**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, naïve 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/](http://code.google.com/p/simultaneous-pin/)

Contact: dshah@kaist.ac.kr

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 (Brochë and Helden, 2006; Han et al., 2004), one of which is known as the automatic protein complex prediction method. Protein complexes generally correspond to clusters in a PPIN because proteins in a complex are highly interactive with each other (Tong and Dreiss, 2002). Therefore, computational methods for protein complex predictions, such as MCODE (Molecular Complex Detection; Bader and Hogue, 2003), LCMA (Local Clique Merging Algorithm; Li et al., 2005), SPC (Super Para-magnetic Clustering; Blatt et al., 1997), RNSEC (Restricted Neighborhood Search Clustering; King et al., 2004) and DPClus (Altatif-Ul-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 et al., 2007). In a dynamic network, a protein complex is a group of interacting partners, some of which may cooperate or even compete with each other. As a result, computational methods for protein complex predictions, such as MCODE, LCMA, SPC, DPClus, etc., should correctly identify the protein complexes from the PPIN.
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.

Among a number of interaction partners, detecting the cooperative partners for a certain function is essential for an understanding of functional mechanisms of proteins. However, too few genes have been studied through experiments, which are typically accomplished only with great difficulty. Therefore, in this research, an understanding of the cooperation between a protein and its partners is approached by eliminating instances of interaction competition through computations.

This article, a network model is developed that incorporates interaction competition information drawn from the structural interface data of protein domains. A framework using the network model for graph-theoretic clustering methods is then proposed for protein complex predictions. A network model, simultaneous protein interaction network (SPIN), captures different sets of non-competitive interactions extracted from the original PPIN. Network clustering on non-competitive interactions excludes superfluous members in the formation of a protein complex.

This research seeks instances of interaction competition based on interaction interfaces. Many competitive interactions are mediated by the same interfacial surface. More than one protein cannot physically bind to the same or an overlapping surface on a protein at the same time; such interactions are identified as mutually exclusive interactions (MEIs; Hu et al., 2005; Valente et al., 2009).

A SPIN is a simple graph composed of nodes and edges which allows any naive graph-theoretic clustering algorithm to be applied to it to computationally predict protein complexes. In this article, MCODE and LCMA are applied to SPINs, as constructed from Saccharomyces cerevisiae (yeast) interactome and are then applied to plain yeast PPIN for comparison. The prediction results are compared with experimentally derived yeast protein complexes recorded in the MIPS complex database (Guldener et al., 2006).

According to the result analysis, SPIN-based clustering outperforms simple PPIN-based clustering. Our model results in a significantly improved F1-score when compared with PPIN-based methods. Moreover, it detects all of the complexes detected by PPIN-based clustering while also generating additional true positives in all thresholds. This result was possible because only superfluous members for a complex formation were removed by the SPIN-based method apart from a small number of cases.

2 METHOD

2.1 Competition between MEI partners

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; therefore, it is regarded as a subunit that mediates PPIs (Boxem et al., 2008).

Another consideration is that a pair of domains can interact through several different interfaces (Aragues et al., 2007a; 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., 2006) 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 identical DDI.

2.1.1 Generation of MEI data

Data mining for MEI information was performed from PDB (Bernstein et al., 1977). 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 complexes 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 the existence of a functional domain that can interact through its several interfaces (Aragues et al., 2007a; 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., 2006) 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 intra-molecular interaction (Gong et al., 2005).

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 XIntpi is used to represent interactions with
Protein complex prediction

Fig. 2. (a) 3D structures of proteins and complexes recorded in PDB. (b) PSIMAP detects interfacial residues between domains. (c) A DDI map including the information of mutually exclusive interfaces. (d) Two PPIs are mutually exclusive when their interaction structures correspond to mutually exclusive DDIs.

PPIN
- Proteins $P = \{ p_1, p_2, p_3 \}$
- Interactions $I = \{ i_1, i_2, i_3 \}$
- MEIs on $p_i$ $\text{mei}_{p_i} = \{ \langle i_1, i_2 \rangle \}$

Boolean expression for MEI:
- $\langle i_1 \otimes i_2 \rangle$ for $p_1$
- $i_1$ and $i_2$ are undetermined for $p_2$ and $p_3$

Boolean expression for interactions of each protein
- $x_{\text{intp}_1} = \langle i_1 \otimes i_2 \rangle$
- $x_{\text{intp}_2} = \{ i_1 \wedge i_2 \} \lor \{ \neg i_1 \land i_2 \} \lor i_2$
- $x_{\text{intp}_3} = \{ i_1 \land i_2 \} \lor \{ i_1 \land i_3 \} \lor \{ i_2 \land i_3 \}$

Boolean expression for the network with MEI information
- $x_{\text{spin}_1, p, \text{mei}} = \langle (i_1 \otimes i_2) \land i_2 \rangle$
- $x_{\text{spin}_2, p, \text{mei}} = (i_1 \land i_2 \land i_3) \lor (i_1 \land i_2 \land \neg i_3)$

Fig. 3. Boolean expression of MEI information for a simple network.

MEI information for protein $p_1$ has an MEI pair, $\langle i_1, i_2 \rangle$; thus, its interactions are represented as $\text{intp}_1 = \langle i_1 \otimes i_2 \rangle$. Another consideration for representing MEI information is that the interaction of an MEI that will occur cannot be determined. Therefore, interactions of the protein $p_2$ are represented as $\text{intp}_2 = \{ i_1 \wedge i_2 \} \lor \{ \neg i_1 \land i_2 \} \lor i_2$, and consequently $i_2$, which ignores $i_1$, which participates in the MEI process on counterpart protein $p_1$. Using $\text{intp}_2$ annotated to each protein, the Boolean expression $x_{\text{spin}_1,p,\text{mei}}$ is generated to represent the MEI information in a PPIN, where $P$ is a protein set. $I$ is an interaction set and MEI is a set of MEIs in the network. $x_{\text{spin}_1,p,\text{mei}}$, is reserved by the conjunction of $\text{intp}_2$ for all proteins in the network. Accordingly, it represents all interactions and mutually exclusive relationships in the network. In the disjunctive normal form (DNF) of $x_{\text{spin}_1,p,\text{mei}}$, each conjunctive clause represents a set of non-competitive interactions in the PPIN.

2.2 SPIN

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 $x_{\text{spin}_1,p,\text{mei}}$ computed from the MEIs, SPINS are extracted from the PPIN based on each conjunctive clause in $x_{\text{spin}_1,p,\text{mei}}$. In Figure 4, the PPIN has two MEIs $\langle i_3, i_5 \rangle$ and $\langle i_4, i_6 \rangle$. Two SPINS are generated from each MEI:

- $\langle i_3 \rangle$ and $\langle i_5 \rangle$
- $\langle i_4 \rangle$ and $\langle i_6 \rangle$

Fig. 4. An example of a SPIN construction: from a PPIN with two MEIs in (a), the SPIN construction process generates two SPINS in (b and c).

< $i_3, i_5$ >; therefore, its set of interactions is represented by $\langle i_3 \land i_5 \land \neg i_4 \land i_6 \rangle \lor \{ i_3 \land i_5 \land i_4 \land i_6 \} \lor \{ i_3 \land i_5 \land i_4 \land \neg i_6 \} \lor \{ i_3 \land i_5 \land i_4 \land \neg i_6 \}$. PSIMAP detects interfacial residues between domains. (d) A DDI map including the information of mutually exclusive interfaces. (d) Two PPIs are mutually exclusive when their interaction structures correspond to mutually exclusive DDIs.

2.3 SPIN-based framework for protein complex prediction

SPINS are not necessarily generated from the whole interactome because the SPIN is constructed only to find a set of possibly coactivated interactions cooperating for a function. In addition, the computation cost of the SPIN construction process is high because the number of SPINS is at most $2^n$ with $n$ MEIs based on the two choices of including one interaction or the other for each MEI. Therefore, its set of interactions is represented by $\langle i_3 \land i_5 \land \neg i_4 \land i_6 \rangle \lor \{ i_3 \land i_5 \land i_4 \land i_6 \} \lor \{ i_3 \land i_5 \land i_4 \land \neg i_6 \} \lor \{ i_3 \land i_5 \land i_4 \land \neg i_6 \}$. In each conjunctive clause represents a non-competitive interaction set, two SPINS are generated from each clause.

Fig. 5 illustrates SPIN framework for protein complex prediction consisting of three phases. In the proposed framework, subnetwork preparation precedes the SPIN construction process. A subnetwork is finally generated on the generated SPINS to predict the protein complexes. The subnetwork preparation adopts a naive clustering algorithm which is used in post-clustering as well. Adopting the same clustering algorithm dramatically reduces the computation cost for SPIN construction but does not change the prediction results because the generated subnetworks cover all of the complexes that can be predicted by the post-clustering algorithm. Although a subnetwork is a cluster, a SPIN generated from the network may not be a cluster as it will lose some interactions. Therefore, clustering is performed on generated SPINS in the post-clustering phase.

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.
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 entities (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 extracted 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 distinct proteins in SPIN clusters is fewer than that in PPIN clusters.

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 results were assessed using an evaluation metric used in earlier known complexes. Precision of the prediction set quantifies the extent to which a prediction set captures the Recall of the predicted cluster and the known complex. In Equation (1), \( |V_p \cap V_m| \) 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 match if \( \text{OS}(p, m) \geq 0.2 \) (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.
- **\(F_1\)**: the \(F_1\) 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 \(F_1\) 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 criterions 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 most 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 \(F_1\), 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 \(F_1\) 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 \(F_1\)-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 superfused 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 (a–c) 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 1. Summary of the prediction results from the four methods

<table>
<thead>
<tr>
<th></th>
<th>MCODE</th>
<th>LCMA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PPIN_MCODE</td>
<td>SPIN_MCODE</td>
</tr>
<tr>
<td>(a) Predicted clusters</td>
<td>140</td>
<td>171</td>
</tr>
<tr>
<td>(b) Proteins in clusters</td>
<td>620</td>
<td>543</td>
</tr>
<tr>
<td>(c) Identical clusters</td>
<td>131</td>
<td>151</td>
</tr>
<tr>
<td>(d) Unique clusters</td>
<td>9</td>
<td>40</td>
</tr>
<tr>
<td>(e) MEIs included</td>
<td>147</td>
<td>0</td>
</tr>
</tbody>
</table>

### 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>(F_1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCODE</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>LCMA</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>
</tbody>
</table>

The performance comparison is presented in Table 2. For each method, we report the precision, recall, and \(F_1\), 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 \(F_1\) 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 \(F_1\)-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 superfused 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 (a–c) 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.
but missed after our modification, and gain denotes the number of true positives found in addition to the result of the PPIN-based method.

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 PPIN 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 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 additional true positives.

Our model may incorrectly remove true positive protein members, although it generates no loss of matched complexes. In this experiment, SPIN_MCODE removed only false positive proteins, whereas SPIN_MCSC 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. This refinement consequently removes superfluous proteins and identifies overlapping complexes in a network clustering.

Figure 6 is an example illustrating a refinement effect by contrasting the known complexes and clusters predicted by LCMA based on the PPIN and the SPIN. The gray oval represents known complexes from MIPS, the quadrangle is a PPIN cluster, and the dotted quadrangles are SPIN clusters. A protein that appears in several complexes is underlined, the grey ovals represent known complexes from PPIN and SPIN, and the dotted quadrangles are SPIN clusters. A protein that appears in several complexes is underlined.

Fig. 6. Comparisons among the known complexes and clusters predicted by LCMA based on PPIN and SPIN. The gray oval represents known complexes from MIPS, the quadrangle is a PPIN cluster, and the dotted quadrangles are SPIN clusters. A protein that appears in several complexes is underlined.
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 YIL06w for the interaction with the core protein YDR328c. Consequently, for the network region in which the PPIN complex formations. As additional true positives by removing superfluous proteins for 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 network construction reserves sets of non-competitive interactions by considering mutual exclusions among the interactions in a network. This allows network-clustering algorithms to identify stable clusters that may possibly be matched by to actual protein complexes. An evaluation of the proposed framework involved the testing of two graph-theoretic clustering algorithms on SPIN and on a simple PPIN for comparison. The comparison showed that the SPIN-based framework outperforms the plain PPIN-based method. It found a massive cluster that the PPIN-based method found as well as additional true positives by removing superfluous proteins for complex formations. From the evaluation, it is concluded that considering MEBs is worthwhile for complex predictions. Information on mutual exclusiveness is drawn from structural interface data, which remains insufficient. This indicates that SPIN-based methods will become more useful as the additional interface data becomes available. The authors are planning to extend the concept of SPIN so that Supplementary Table 3 lists MIPS complexes matched by MCODEs based on PPIN and SPIN along with their overlapping scores and matched proteins, and Supplementary Table 4 lists those for LCMAs.

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 model is defined that the PPIN-based method found as well as additional proteins share an interface. Consequently, for the network region in which the PPIN complex formations. As additional true positives by removing superfluous proteins for 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 network construction reserves sets of non-competitive interactions by considering mutual exclusions among the interactions in a network. This allows network-clustering algorithms to identify stable clusters that may possibly be matched by to actual protein complexes. An evaluation of the proposed framework involved the testing of two graph-theoretic clustering algorithms on SPIN and on a simple PPIN for comparison. The comparison showed that the SPIN-based framework outperforms the plain PPIN-based method. It found a massive cluster that the PPIN-based method found as well as additional true positives by removing superfluous proteins for complex formations. From the evaluation, it is concluded that considering MEBs is worthwhile for complex predictions. Information on mutual exclusiveness is drawn from structural interface data, which remains insufficient. This indicates that SPIN-based methods will become more useful as the additional interface data becomes available. The authors are planning to extend the concept of SPIN so that considering mutual exclusiveness is drawn from structural interface data, which remains insufficient. This indicates that SPIN-based methods will become more useful as the additional interface data becomes available. The authors are planning to extend the concept of SPIN so that considering mutual exclusiveness is drawn from structural interface data, which remains insufficient. This indicates that SPIN-based methods will become more useful as the additional interface data becomes available. The authors are planning to extend the concept of SPIN so that considering mutual exclusiveness is drawn from structural interface data, which remains insufficient. This indicates that SPIN-based methods will become more useful as the additional interface data becomes available.

Conflict of Interest: none declared.

REFERENCES


