BMC
Bioinformatics

RESEARCH

Open Access

Interaction site prediction by structural similarity to neighboring clusters in protein-protein interaction networks

Hiroyuki Monji[1*], Satoshi Koizumi[2], Tomonobu Ozaki[3], Takenao Ohkawa[1*]

Abstract

Background: Recently, revealing the function of proteins with protein-protein interaction (PPI) networks is regarded as one of important issues in bioinformatics. With the development of experimental methods such as the yeast two-hybrid method, the data of protein interaction have been increasing extremely. Many databases dealing with these data comprehensively have been constructed and applied to analyzing PPI networks. However, few research on prediction interaction sites using both PPI networks and the 3D protein structures complementarily has explored.

Results: We propose a method of predicting interaction sites in proteins with unknown function by using both of PPI networks and protein structures. For a protein with unknown function as a target, several clusters are extracted from the neighboring proteins based on their structural similarity. Then, interaction sites are predicted by extracting similar sites from the group of a protein cluster and the target protein. Moreover, the proposed method can improve the prediction accuracy by introducing repetitive prediction process.

Conclusions: The proposed method has been applied to small scale dataset, then the effectiveness of the method has been confirmed. The challenge will now be to apply the method to large-scale datasets.

Background

The functional analysis of proteins is an important issue for elucidating the mechanism of living bodies. Since most of the functions of proteins are largely-related to their 3D structures, research on estimating the function of protein by revealing the relation between the 3D structure and the function is one of the main stream of the structural bioinformatics.

Most of proteins express their functions by interacting with other proteins or ligands. In many cases, interaction occurs at local portion of a protein, which is called an *interaction site*. The structural and physical characteristics on the interaction site often determine the function of the protein, which means that clarifying the

location of interaction site of the protein helps analyze the function of proteins.

Various methods for predicting interaction sites (or functionally significant sites) have been developed. Sacan et al. developed a tool for detecting family-specific local structural sites [1]. In their method, geometrically significant structural centers of the protein are detected, then features generated from the geometrical and biochemical environment around these centers are used to distinguish a family. Jones and Thornton proposed a method of predicting interaction sites by comparing the protein surface patches in terms of six properties [2]. In other approaches, interface residues in a protein are deduced by use of neural networks which have been trained with surface patches in protein structures and sequence profiles [3-6]. Support vector machines are also used in predicting interface residues [7-10]. Other methods involved in predicting interaction sites have been proposed [11].

* Correspondence: h.monji@cs25.scitec.kobe-u.ac.jp; ohkawa@kobe-u.ac.jp
[1]Graduate School of System Informatics, Kobe University, Rokkodai, Nada, Kobe 657–8501, Japan
Full list of author information is available at the end of the article

BioMed Central

Meanwhile, analyzing the function of proteins from the aspect of protein-protein interaction has gotten a lot of attention [12,13]. The development of experimental methods for observing interactions, such as the yeast two-hybrid, helps increase the data related to the protein-protein interaction, leading many databases [14-16], and is anticipated for understanding various biological phenomena [12,13,17]. Such data are mainly converted into the protein-protein interaction networks (PPI networks), which are often used as the protein function identification tools [18-21]. For example, six thousands of yeast genes library are used in creating protein-protein interaction map [18], several attempts analyzing over thousands kinds of protein-protein interactions have been addressed in full detail [19,20].

There is much research on identifying the function of proteins with PPI networks. Vazquez et al. have proposed the method predicting the function of protein nodes which are functionally unknown in PPI networks, and identifying the function of each node to optimize the function of whole nodes in the networks [12]. Also, they argue PPI networks are scalefree [22], which leads to many methods by probabilistic approach to complex networks. For example, Letovsky proposed a method for calculating probability of functional label given to nodes with propagation of the binomial model and Markov random field [13]. Deng et al. presented a protein function prediction method by assigning functions to all the unannotated proteins based on functions of the annotated proteins and the protein interaction network using Bayesian approaches [23].

In such PPI-based research, however, 3D protein structures are little considered. Since it is obvious that 3D protein structures make a strong contribution to the function of proteins, it must be significant to predict the interaction sites from the viewpoint of both the PPI networks and the 3D protein structures.

We propose a method of predicting interaction sites of a protein (target protein) whose structure has been solved but whose interaction site is unknown using the information of 3D structures and PPI networks. As it is known that the function of a protein is often similar to the function of neighboring proteins on the PPI network, interaction sites may be predicted by extracting pockets from the surface of the target protein whose shape and physical properties are similar to those of the neighboring proteins. However, the functions of all of the neighboring proteins are not always similar to the function of the target protein. Hence, the neighboring proteins are classified into several non-disjoint groups, each of which shares the common features based on structural similarity. The interaction sites are predicted by extracting common pockets that appear both in one of these groups and in the target protein. In addition, information of the neighboring proteins whose interaction sites have been specified by this method itself may be effectively utilized. That is, we assume that the predicted interaction site of the target protein is considered as a known interaction site, then the prediction process is repeated for other target proteins.
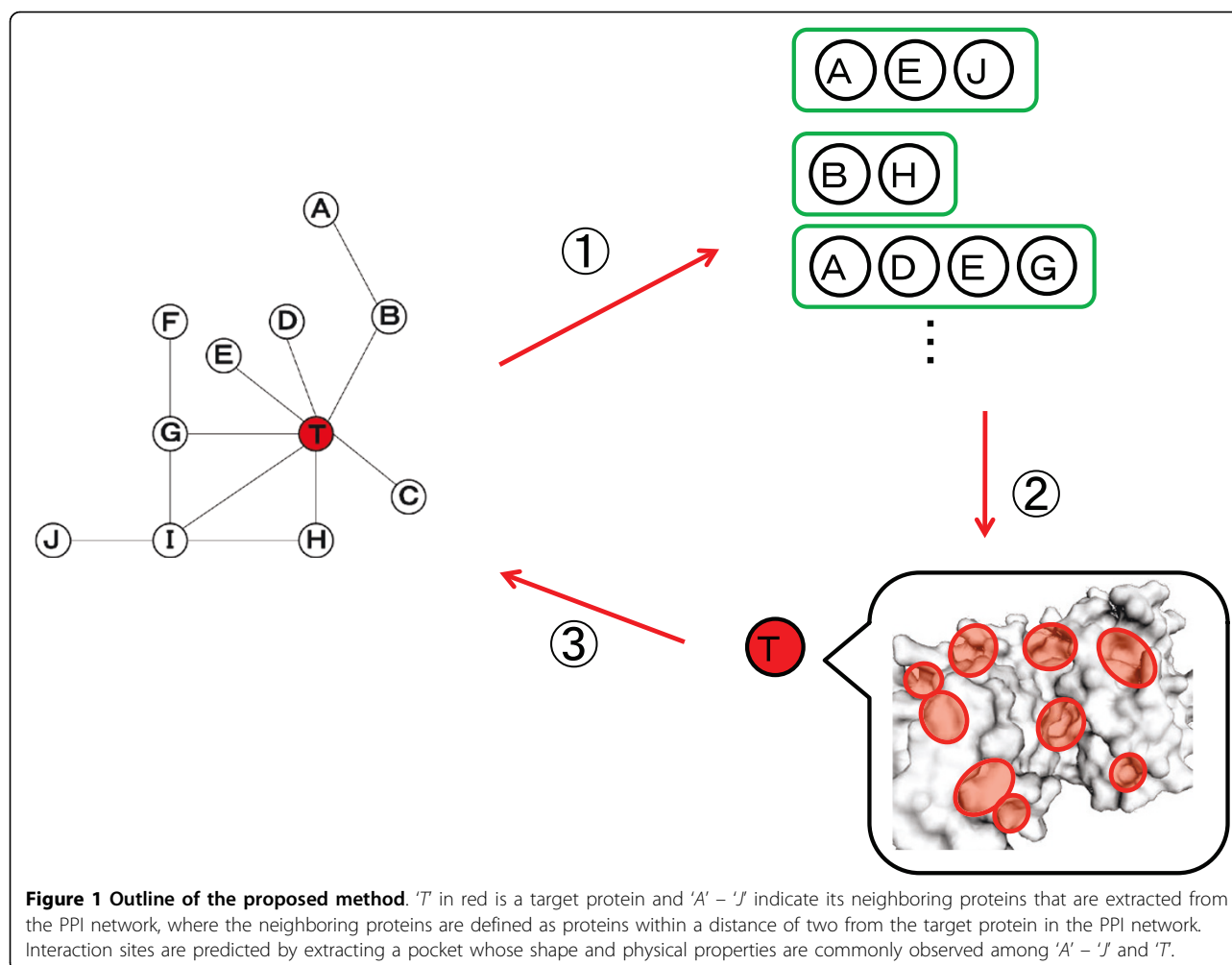
Method
Outline
Figure 1, in which '*T*' in red is a target protein and '*A*' – '*J*' indicate its neighboring proteins that are extracted from the PPI network, shows the outline of prediction of interaction sites, where the neighboring proteins are defined as proteins within a distance of two from the target protein in the PPI network. Since the interaction site often forms a concave structure, instead of the whole of molecular surface of the protein, only pockets are treated as candidates of the interaction sites. In other words, interaction sites are predicted by extracting a pocket whose shape and physical properties are commonly observed among '*A*' – '*J*' and '*T*'. In practical cases, however, all of neighboring proteins '*A*' – '*J*' do not always have similar functions. For this reason, the groups, called neighboring protein clusters, in which a similar pocket is commonly observed, are extracted from '*A*' – '*J*'. In our method, how to extract the cluster which shares discriminative pocket being similar in shape and physical properties is an important issue. If structurally similar groups are simply extracted from the neighboring proteins the cluster with similar structural features would be extracted, but the cluster which shares a "discriminative" pocket is not always obtained because the similarity of pockets which are observed in many proteins universally tend to be high. To cope with this problem, we introduce a restriction that each cluster must have at least one protein with known interaction sites. Next, the score is given for each pocket of the target protein which appears in all of extracted neighboring protein clusters commonly, and the top-ranked pockets are output as interaction sites.

Meanwhile, if the target protein '*T*' has no neighboring protein with a known interaction site, it is impossible to construct any neighboring proteins clusters. To handle this difficulty, the prediction process is repeated by considering the predicted interaction sites as known interaction sites. In addition, repetition of the prediction process increases the neighboring proteins having the predicted (i.e. known) interaction sites, reorganization of the clusters using them will improve the prediction accuracy.

Molecular surface data and pocket
In the proposed method, molecular surface data available from eF-site database [24] are used. A number of polygons represent the molecular surface, and every vertex composing polygons has the information of structure

Figure 1 Outline of the proposed method. '*T*' in red is a target protein and '*A*' – '*J*' indicate its neighboring proteins that are extracted from the PPI network, where the neighboring proteins are defined as proteins within a distance of two from the target protein in the PPI network. Interaction sites are predicted by extracting a pocket whose shape and physical properties are commonly observed among '*A*' – '*J*' and '*T*'.

(location, maximum curvature, and minimum curvature), the property values (electrostatic potential and hydrophobicity), and the connection information of vertices. Interaction sites are widely known having concave structures on surface because of binding stability, specificity, and reaction promotion. Much research on searching and extracting pockets from the protein surface as candidates of interaction sites has been conducted [25,26].

In fact, the number of vertices of molecular surface of some proteins is over 20,000, so it is impractical idea to handle the whole molecular surface for comparing protein structures. Thus focusing on only pockets extracted from the molecular surface has advantages. In our method, the LIGSITE [27] algorithm is utilized to extract pockets. About 30 pockets are extracted for each protein.

Representaion of pockets by histograms

It is known that proteins change their conformation in interacting, so comparing pockets by rigid superimposing of vertices which construct a pocket each other is inappropriate. So far, many methods for comparing surface patches have been proposed [28,29]. In order to compare molecular surfaces of the pockets from the viewpoint of mainly physical properties and roughly geometrical figures, we introduce a method of representing a molecular surface using histogram of structural and physical properties of the surface. Comparison of histogram is utilized in the area of such as image processing and it can compare pockets not definitely but roughly. As a pocket is constructed from vertex set of polygons, the pocket can be expressed with the four histograms, which are defined using three parameters, the range of rank d, the maximum value max, and the minimum value min, from four properties, namely, maximum curvature κ_{max}, minimum curvature κ_{min}, electrostatic potential C, and hydrophobicity H, of each vertex shown as follows. Values of the parameters max, min, and d are determined experimentally.

• Histogram of mean curvature:
$M = (\kappa_{max} + \kappa_{min})/2$

$max = 3.0$, $min = -3.0$, $d = 0.01$
• Histogram of Gaussian curvature:
$G = \kappa_{max} \cdot \kappa_{min}$
$max = 3.0$, $min = -3.0$, $d = 0.01$
• Histogram of electrostatic potential: C
$max = 0.6$, $min = -0.6$, $d = 0.01$
• Histogram of hydrophobicity: H
$max = 5.0$, $min = -5.0$, $d = 0.1$

Similarity among pockets

A pocket is expressed using four histograms of structural and physical properties. We define similarity among pockets by comparing the four histograms.

Let $p_1,...,p_N$ be N pockets and each pocket is expressed with the histogram of mean curvature $M_i(1 \le i \le N)$, the histogram of Gaussian curvature G_i, the histogram of electrostatic potential C_i, and the histogram of hydrophobicity H_i. We simply define $S_{pkt}(p_1, ..., p_N)$, the similarity among pockets $p_1,...,p_N$, by

$S_{ptk}(p_1, ..., p_N) = J(M_1,...,M_N) \times J(G_1,...,G_N) \times J(C_1,...,C_N) \times J(H_1,...,H_N)$ (1)

where $J(A_1,...,A_N)$ represents the similarity among the histograms $A_1,...,A_N$, which is defined by

$$J\left(A_1,...,A_N\right) = \sum_{k=1}^{n} \frac{\min\left(\sum_{i=1}^{N} a_{i_k}\right)}{\max\left(\sum_{i=1}^{N} a_{i_k}\right)} \quad (2)$$

where $a_{i_k}(1 \le i \le N)$ represents frequency of k-th rank of i-th histogram, and n represents the maximum value of the rank. Equation (2) is based on the idea of Jaccard coefficient to comparing histograms. That is to say, the similarity among pockets S_{pkt} is defined as the product of the similarity of the four histograms expressing each pocket.

Extraction of neighboring proteins cluster

In our method, we define a neighboring proteins cluster as a subset of proteins sharing the pockets that are similar in shape and physical properties and are specific to the cluster, which are extracted from the set of the neighboring proteins. We introduce the similarity measure that shows how similar the pockets on each protein in the subset are. If each protein in the subset has the similar interaction site, they are likely to share common pockets, then the similarity of the pockets in proteins in the subset must be high. Therefore, the pockets of each protein in the subset are exhaustively compared by using the similarity among pockets given by equation (1), then the highest similarity is put to be the subset similarity. However, there is a possibility that this highest similarity is actually due to the non-specific pockets which appear universally in the several proteins.

To handle this matter, strong restriction is introduced, in which any subset must contain one or more proteins having a known interaction site. The following is an algorithm of extracting neighboring protein clusters.

1. Let P be a set of neighboring proteins, and S($\subseteq$P) be a set of proteins in P whose interaction sites are known. $PS(P, n)$ is a power set of P whose cardinality is $n(1 < n \le k)$, and ps is an element of $PS(P, n)$, namely $ps \in PS(P, n)$. Enumerate all of ps satisfying the following constraint.

$ps \cap S \ne \phi$ (3)

2. Let $P_{x_1}, ..., P_{x_n}$ be proteins in ps, namely $ps = \{P_{x_1},..., P_{x_n}\}$. Calculate $S_{set}(ps)$, the similarity among $\{P_{x_1},...,P_{x_n}\}$ by the following definition

$S_{set}(ps) = max\{S_{pkt}(p^1,...,p^n)|p^i \in pkt(P_{x_i}), 1 \le i \le n\}$ (4)

where $pkt(P_{x_i})$ denotes a set of pockets in protein P_{x_i}. $S_{set}(ps)$ means the similarity of the combination of the most similar pockets when exaustively comparing all the pockets of each protein.

3. Extract the elements ranked in the top Z of $PS(P, n)$ as clusters C as follows.

$C = \{x \in PS(P, n)|Rank(x, PS(P, n)) < Z\}$ (5)

where the function $Rank(x, PS(P, n))$ gives the ranking of x in $PS(P, n)$ in terms of the similarity, which is formally defined as

$Rank(x, PS(P, n)) = |\{y \in PS(P, n)|x \ne y, S_{set}(x) < S_{set}(y)\}|$ (6)
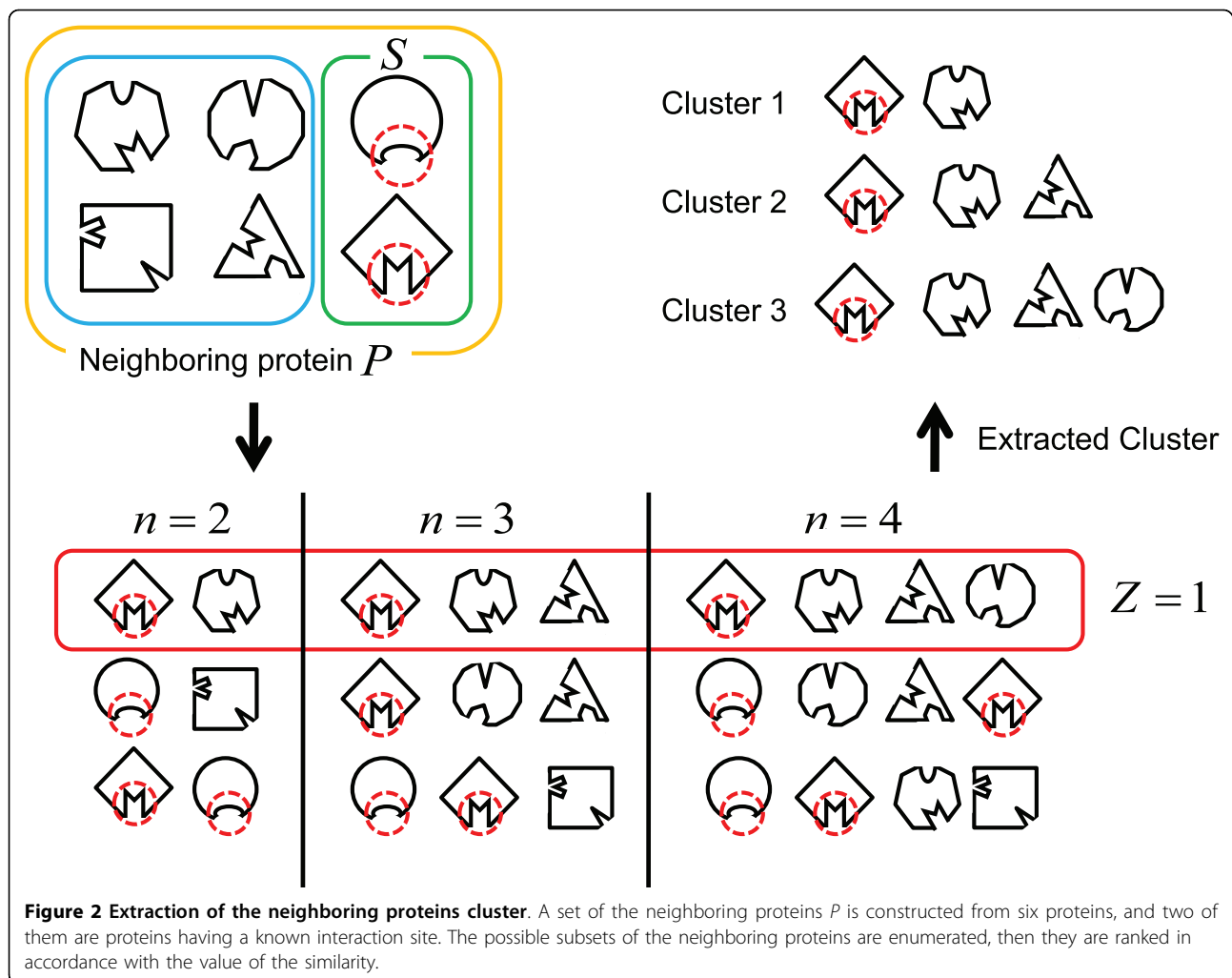
Figure 2 illustrates an example of extracting the neighboring protein clusters for $k = 4$ and $Z = 1$. In this example, a set of the neighboring proteins P is constructed from six proteins, and two of them are proteins having a known interaction site. $k = 4$ gives ps, the possible subset of the neighboring proteins set, whose size is 2, 3, or 4. First, ps which satisfies the constraint (3) is enumerated. Next, the enumerated ps is ranked in accordance with the value of the similarity. The constraint $Z = 1$ leads to extract ps with the highest similarity value as the clusters.

Scoring of pockets

If a pocket in the target protein is similar to the pockets that appear commonly in the high ranked neighboring protein cluster, it may be a candidate of the interaction site. To evaluate each candidate, we introduce the voting-based scoring scheme. In this scheme, a set of pockets consisting of one pocket from the target protein and the similar pockets from the neighboring protein cluster is evaluated from the viewpoint of similarity, and the pocket (from the target) that wins the highest similarity value is voted. Formally, for a target protein P_T and the proteins $P_1,...,P_n$ in the neighboring protein cluster, the pockets to be voted are enumerated as follows.

$p \in pkt(P_T)$ s.t.
$S_{pkt}(p^1,...p^n, p) = S_{set}(\{P_1,...,P_n, P_T\})$,

Figure 2 Extraction of the neighboring proteins cluster. A set of the neighboring proteins *P* is constructed from six proteins, and two of them are proteins having a known interaction site. The possible subsets of the neighboring proteins are enumerated, then they are ranked in accordance with the value of the similarity.

$$\forall i, \ 1 \le i \le n, \ p^i \in \ pkt(P_i) \ (7)$$

Complement and feedback by repetition

If the target protein has no neighboring proteins with the known interaction site, any cluster cannot be constructed because there is no *ps* satisfying the constraint (3). On the other hand, if the pockets of predicted protein are regarded as interaction sites we can get more proteins with known interaction sites on the PPI networks. That is, we can construct clusters using newly known (namely predicted) interaction sites, which enable repetitive prediction as shown in Figure 3. This repetitive process plays a complementary role for the protein which contains no interaction site in its neighboring proteins. Figure 3 shows that there are few proteins that have known interaction sites in a single cycle prediction, but the repetitive prediction increases the proteins having known interaction site.

Even if we deal in the target proteins whose neighboring proteins contain known interaction site, there is a possibility that the actual interaction site of the target protein may not be similar to any interaction sites of the neighboring proteins. The repetitive prediction plays a role of feedback for this problem. As the number of proteins whose neighboring proteins contain known interaction sites increases, the clusters of the neighboring proteins can be reconstructed. This feedback has the possibility to choose different pockets as interaction sites from the previous prediction process, and may get better result.

Results and discussion

To verify the effectiveness of the proposed method, we conducted experiments for prediction of interaction sites where we assumed that some of the known interaction sites are unknown. Figure 4 shows the PPI network used in these experiments. In Figure 4, *n* proteins are selected as target proteins, which mean that interaction sites are assumed to be unknown, from ten proteins whose interaction sites are actually known. Therefore,

Figure 3 Repetitive process of prediction. There are few proteins that have known interaction sites in a single cycle prediction, but the repetitive prediction increases the proteins having known interaction site.
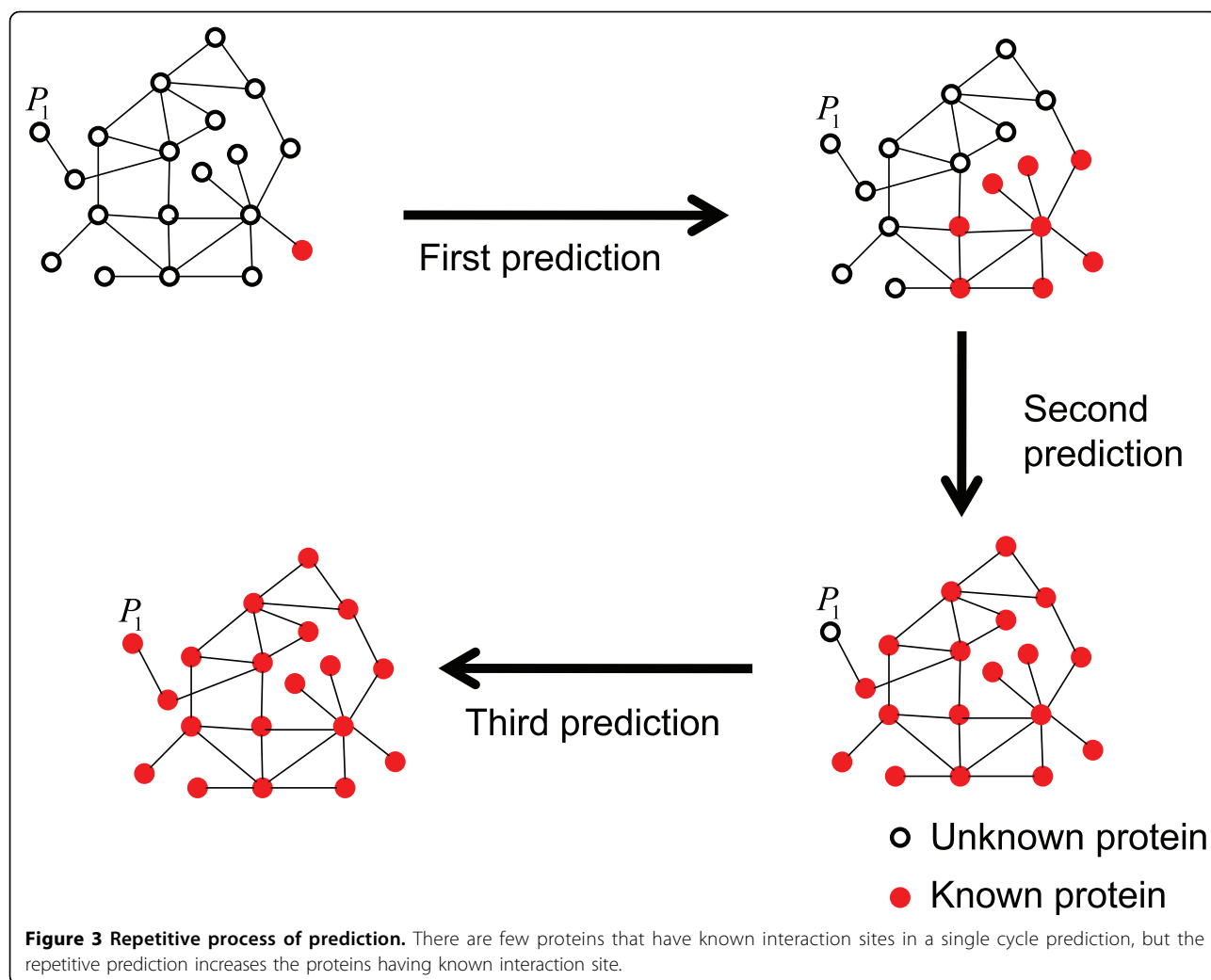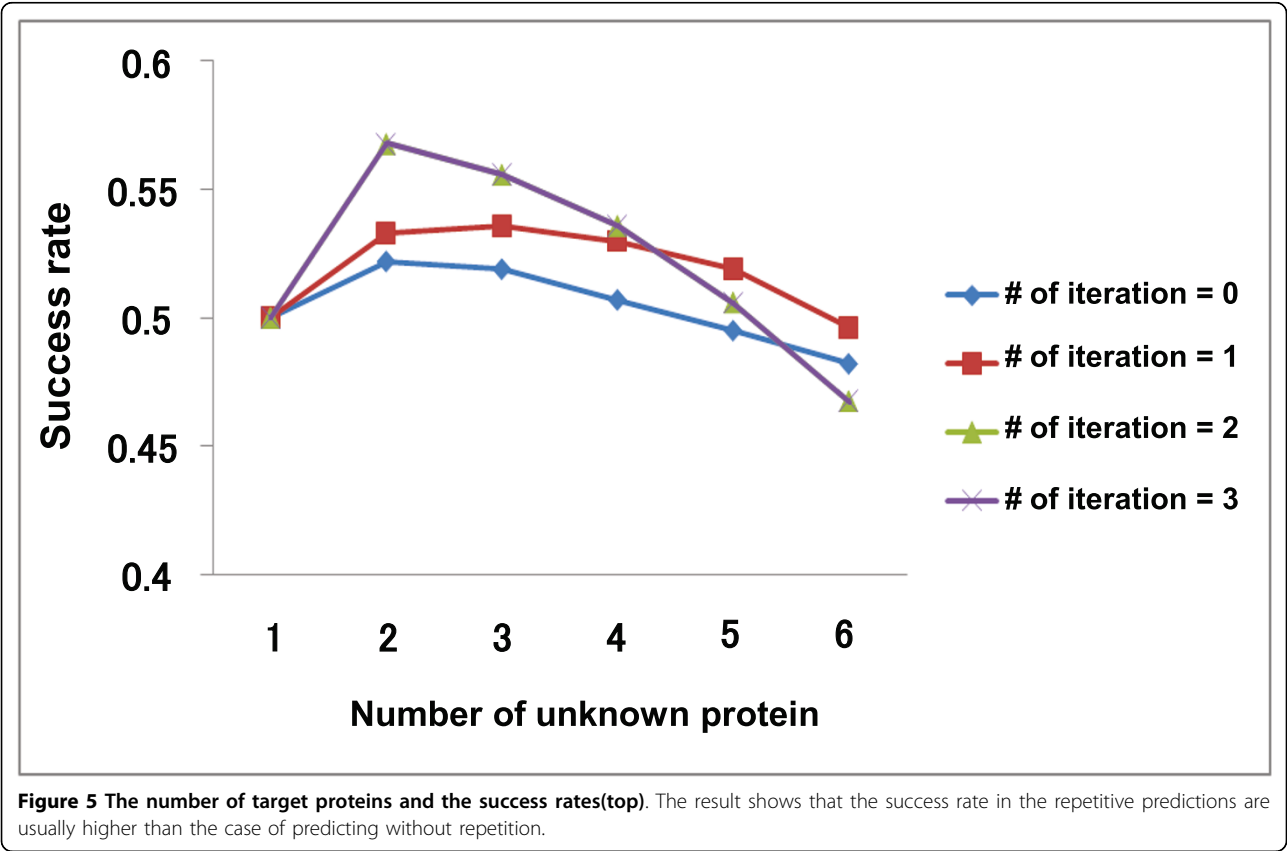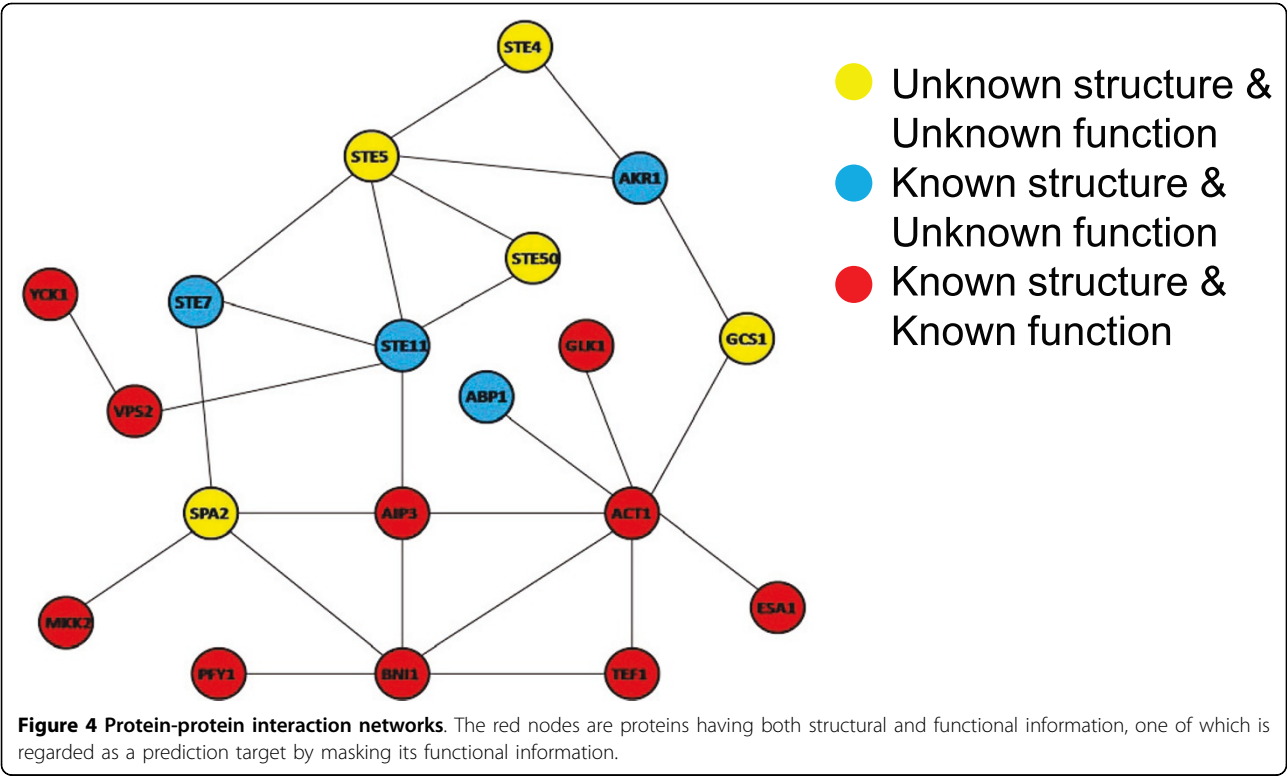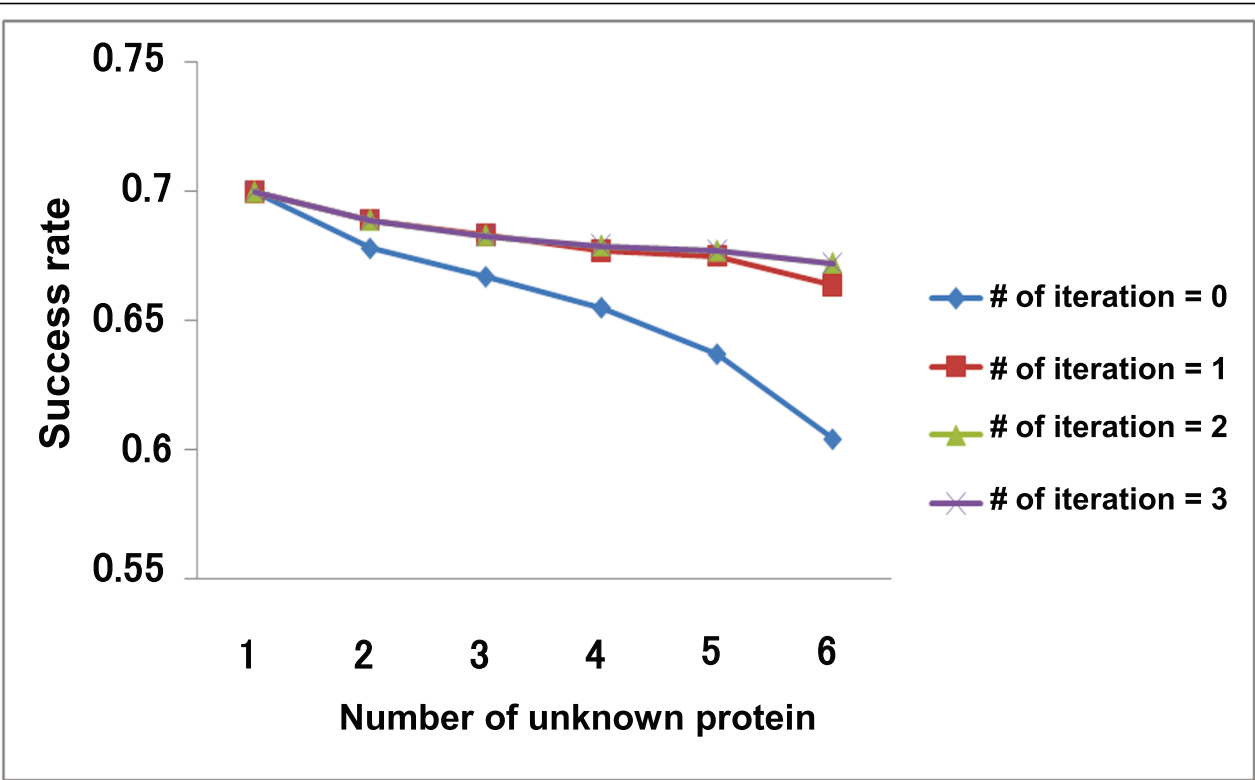
the interaction sites of n proteins are predicted at one cycle of the experiment. Furthermore, the n target proteins are combinatorially selected from the ten proteins (the number of combinations is $_{10}C_n$). We evaluate the results based on whether the pockets of the target proteins with the top-one score or the top-five score are true interaction sites. Figure 5 and 6 show the experimental results, in which the success rate in the case of top-one is about a half and in the case of top-five is about six or seven out of ten.

In addition, the result shows that the success rate in the repetitive predictions are usually higher than the case of predicting without repetition. This means the complement and the feedback by repetition work well. As for the effectiveness of the repetition process, Figure 6 shows the success rate gets better as the repetition count increases, and is converged at two times repetition regardless of the number of the target proteins. Figure 5 indicates that when the number of the target protein is large (i.e. $n \geq 5$), the appropriate repetitive

process contributes to rising the success rate, but repeating too much leads the decline of the success rate. It is considered that the first repetitive prediction improves the success rate by the effect of the complement. The feedback after the second repetitive prediction for a few of target proteins has a fine effect on success rate because of many neighboring proteins whose interaction sites have been known actually. In the case of a lot of target proteins are considered, however, the repetitive process may work worse because of the side effect of the feedback which are brought by the neighboring proteins having not actual but assumed interaction sites.

Figure 7 shows the success rates of prediction for each protein with $n = 5$ fixed. This figure means the success rate varies considerably depending on the protein. The success rate for proteins TEF1 and GLK1 is 0, so the interaction site cannot be specified at all. Figure 8 illustrates the molecular surfaces of TEF1 and GLK1. The molecular surfaces of these proteins have a lot of local

Figure 4 Protein-protein interaction networks. The red nodes are proteins having both structural and functional information, one of which is regarded as a prediction target by masking its functional information.

Figure 5 The number of target proteins and the success rates(top). The result shows that the success rate in the repetitive predictions are usually higher than the case of predicting without repetition.

Figure 6 The number of target proteins and the success rates(top five). The success rate gets better as the repetition count increases, and is converged at two times repetition regardless of the number of the target proteins.
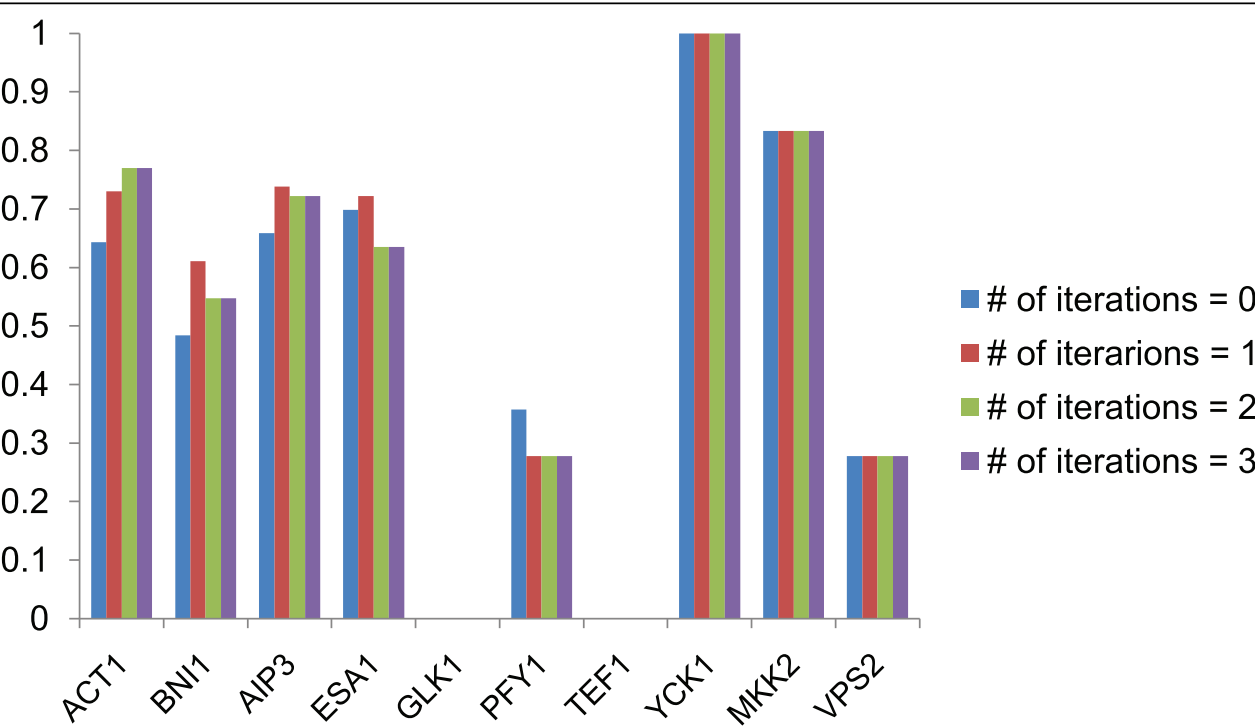
Figure 7 Accuracy rates of predicting per proteins. The success rate varies considerably depending on the protein. The success rate for proteins TEF1 and GLK1 is 0, so the interaction site cannot be specified at all.
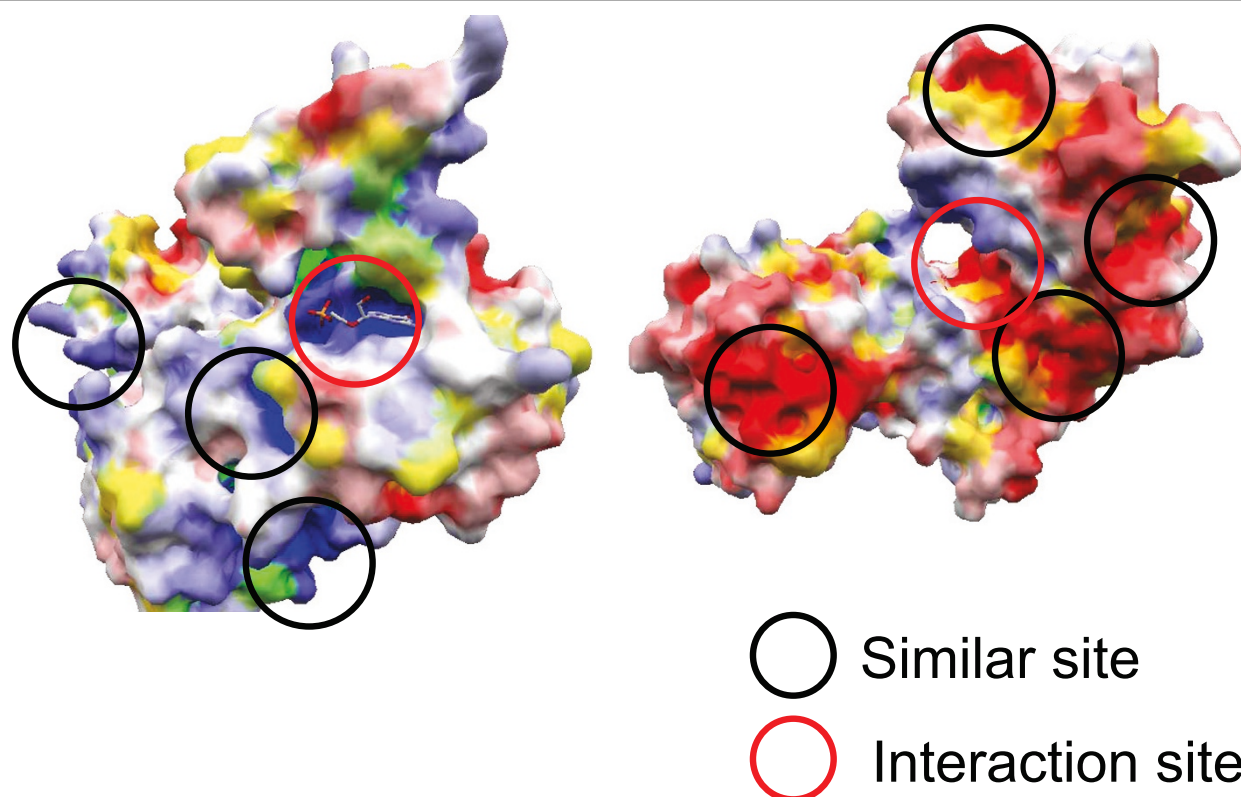
Figure 8 Molecular surfaces of proteins (TEF1 and GLK1). The molecular surfaces of proteins TEF1 and GLK1 have a lot of local structures whose properties are similar to the true interaction sites, which might cause the prediction failure.

structures whose properties are similar to the true interaction sites, which might cause the prediction failure. Our method assumes that the interaction sites have common structural features and share distinctive properties. This suggests that it is difficult to predict the interaction sites which have universal structures among proteins.

Finally we conducted comparative experiment, in which only structural information (i.e. without PPI networks) was utilized. In the basic situation, in which the number of target proteins is limited to one and no repetitive process is done, the success rate is 0.4, which is lower than the result by the proposed method (0.5). This result is under the limited situation only, but shows the effectiveness of the complementary use of the PPI network along with the protein structural information. We will perform comparative analysis for various experimental setting.

Conclusions
We proposed a method of predicting interaction sites by comparing the pockets of proteins whose interaction sites are unknown to pockets of the neighboring proteins in the PPI networks. The challenge will be to apply this method to large-scale protein-protein interaction networks. In this paper, experimental results for the PPI networks consisting of only a few dozen of proteins have been presented. The nodes of actual PPI networks are, however, so massive that there is a need for experiments in larger networks. Now, therefore, it is desirable to apply the proposed method by dividing large-scale PPI networks into subgraphs comprising a few dozen of nodes. Currently, we put the strong restriction on our method, that is, we assume that the structural information of all prediction target proteins or neighboring proteins has been known. However, the method should be modified so that not only proteins with structural information but proteins having no structural information can be treated for practicality improvement. In addition, comparing accuracy of the proposed method to that of existing methods is a crucial remaining work in the near future.

This article has been published as part of *BMC Bioinformatics* Volume 12 Supplement 1, 2011: Selected articles from the Ninth Asia Pacific Bioinformatics Conference (APBC 2011). The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2105/12?issue=S1.

Author details
¹Graduate School of System Informatics, Kobe University, Rokkodai, Nada, Kobe 657–8501, Japan. ²Graduate School of Engineering, Kobe University, Rokkodai, Nada, Kobe 657–8501, Japan. ³Cybermedia Center, Osaka University, Yamadaoka, Suita 565–0871, Japan.

Authors' contributions
HM carried out the arrangement of the data set and the experimental results and drafted the manuscript. SK carried out the implementation of the algorithm and performed experiment. TO participated in the algorithm development and the design of the study. TO conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Published: 15 February 2011

References
1. Sacan A, Ozturk O, Ferhatosmanoglu H, Wang Y: **LFM-Pro: A Tool for Detecting Significant Local Structural Sites in Proteins.** *Bioinformatics* 2007, **23**:709-716.
2. Jones S, Thornton JM: **Prediction of protein-protein interaction sites using patch analysis.** *J. Mol. Biol* 1997, **272**:133-143.
3. Zhou HX, Shan Y: **Prediction of protein interaction sites from sequence profile and residue neighbor list.** *Proteins* 2001, **44**:336-343.
4. Fariselli P, Pazos F, Valencia A, Casadio R: **Prediction of protein-protein interaction sites in heterocomplexes with neural networks.** *European Journal of Biochemistry* 2002, **269**(5):1356-1361.
5. Ofran Y, Rost B: **Predicted protein-protein interaction sites from local sequence information.** *FEBS Letters* 2003, **544**:236-239.
6. Porollo A, Meller J: **Prediction-based fingerprints of protein-protein interactions.** *Proteins* 2007, **66**:630-645.
7. Bradford JR, Westhead DR: **Improved prediction of protein.protein binding sites using a support vector machines approach.** *Bioinformatics* 2005, **21**(8):1487-1494.
8. Koike A, Takagi T: **Prediction of protein-protein interaction sites using support vector machines.** *Protein Eng Des Sel* 2004, **17**(2):165-173.
9. Bordner AJ, Abagyan R: **Statistical analysis and prediction of protein-protein interfaces.** *Proteins* 2005, **60**:353-366.
10. Res I, Mihalek I, Lichtarge O: **An evolution based classifier for prediction of protein interfaces without using protein structures.** *Bioinformatics* 2005, **21**:2496-2501.
11. Zhou HX, Qin SB: **Interaction-site prediction for protein complexes: a critical assessment.** *Bioinformatics* 2007, **23**:2203-2209.
12. Vazquez A, Flammini A, Maritan A, Vespignani A: **Global protein function prediction in protein-protein interaction networks.** *Nature Biotechnology* 2003, **21**:697-700.
13. Letovsky S, Kasif S: **Predicting protein function from protein-protein interaction data:a probabilistic approach.** *Bioinformatics* 2003, **19**(Suppl.1):i197-i204.
14. Xenarios I, Salwinski L, Duan XJ, Higney P, M KS, Eisenberg D: **DIP, the Database of Interacting Proteins:a research tool for studying cellular networks of protein interactions.** *Bioinformatics* 2002, **30**(1):303-305.
15. Bader GD, Donaldson I, Wolting C, Ouellette BF, Pawson T, Hogue CW: **BIND-The Biomolecular Interaction Network Database.** *Nucleic Acids Research* 2001, **29**(1):242-245.
16. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TK, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao Z, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Chandran S, Krishna S, Joy M, Anand SK, Madavan V, Joseph A, Wong GW, Schiemann WP, Constantinescu SN, Huang L, Khosravi-Far R, Steen H,
Tewari M, Ghaffari S, Blobe GC, Dang CV, Garcia JG, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A: **Development of human protein reference database as an initial platform for approaching systems biology in humans.** *Genome Res* 2003, **13**(10):2363-2371.
17. Salwinski L, Eisenberg D: **Computational methods of analysis of protein-protein interactions.** *Current Opinion in Structural Biology* 2003, **13**:377-382.
18. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM: **A comprehensive analysis of proteinprotein interactions in Saccharomyces cerevisiae.** *Nature* 2000, **403**:623-627.
19. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y: **Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins.** *Proc Natl Acad Sci U S A* 2000, **97**(3):1143-1147.
20. Schwikowski B, Uetz P, Fields S: **A network of protein-protein interactions in yeast.** *Nat Biotechnol* 2000, **18**:1257-1261.
21. Hishigaki H, Nakai K, Ono T, Tanagami A, Takagi T: **Assessment of prediction accuracy of protein function from protein-protein interaction data.** *Yeast* 2001, **18**(6):523-531.
22. Thomas A, Cannings R, Monk NAM, Cannings C: **On the structure of protein-protein interaction networks.** *Biochemical Society Transactions* 2003, **31**:1491-1496.
23. Deng M, Zhang K, Mehta S, Chen T, Sun F: **Prediction of protein function using protein-protein interaction data.** *J Comput Biol* 2002, **10**(6):947-960.
24. Kinoshita K, Nakamura H: **eF-site and PDBjViewer: database and viewer for protein functional sites.** *Bioinformatics* 2004, **20**:1329-1330.
25. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J: **CASTp, computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.** *Nucleic Acids Res* 2006, **34**(Web Server issue):W116-W118.
26. Laskowski RA: **SURFNET, a program for visualizing molecular surfaces, cavities, and intermolecular interactions.** *Journal of molecular graphics* 1995, **13**(5):323-330.
27. Huang B, Schroeder M: **LIGSITE: predicting protein binding sites using the Connolly surface and degree of conservation.** *BMC structural Biology* 2006, **6**(19).
28. Kaneta Y, Shoji N, Ohkawa T, Nakamura H: **A method of comparing protein molecular surface based on normal vectors with attributes and its application to function identification.** *Information Sciences* 2002, **146**(1-4):41-54.
29. Ritchie DW, Kemp GJL: **Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces.** *Jornal of Computational Chemistry* 1999, **20**(4):383-395.