Iranian Journal of

Biotechnology

Discovering Domains Mediating Protein Interactions

Changiz Eslahchi^{1,*}¹ Department of Computer Sciences, Shahid Beheshti University GC, Tehran, IR Iran

ARTICLE INFO

Article type:

Research Article

Article history:

Received: 10 Jun 2012

Revised: 11 Oct 2012

Accepted: 26 Nov 2012

Keywords:

Domain

Mediator Domain

Protein Interaction

ABSTRACT

Background: Protein-protein interactions do not provide any direct information regarding the domains within the proteins that mediate the interactions. The majority of proteins are multi domain proteins and the interaction between them is often defined by the pairs of their domains. Most of the former studies focus only on interacting domain pairs. However they do not consider the interactions that require the presence of a third domain.

Objectives: In this manuscript, we define the concept of necessary and sufficient triplets of domains and mediator domain.

Materials and Methods: We approximate these conditions by pragmatic statistical definitions on a set of gold-standard interacting protein pairs and a set of gold-standard non-interacting protein pairs.

Results: In this paper we introduce a new method for the prediction of the interaction between two domains using third domains as a mediator. we show that the mediator domain has an effective role in the interaction between proteins.

Conclusions: By using these concepts, we introduce a method for the prediction of the interaction between two domains. Subsequently by evaluating the performance of our method on the yeast protein interactions data set, we show that the mediator domain has an effective role in the interaction between proteins.

Published by Kowsar Corp, 2013. cc 3.0.

► Implication for health policy/practice/research/medical education:

This study has implication on further studies on bioinformatics and biological researches.

► Please cite this paper as:

Eslahchi C. Discovering Domains Mediating Protein Interactions. *Iran J Biotech.* 2013; **11**(1): 22-31. DOI: 10.5812/ijb.9309

1. Background

More than half of eukaryotic proteins are multi-domain proteins (1). It is often assumed that the interaction between two proteins involves binding two or more specific domains (2) or binding a domain in one protein to short regions (approximately three to eight residues) of the other protein (3). For example, multiple domains of Nkx3.1 are involved in contacting SRF (4). While more than two domains could be involved in mediating the

interaction of two proteins, most of the former works have been developed to identify interacting domain pairs either for the purpose of predicting or explaining the protein interactions (5-7). In particular, they have neglected the identification of interactions that require the presence of a third domain. Few exceptions include (8-10). However, while these works have considered domain combinations for predicting the protein interactions, they have not evaluated whether these domain combinations are required to mediate protein interactions.

* Corresponding author: Changiz Eslahchi, Department of Computer Sciences, Shahid Beheshti University GC, PO. Box: 1983963113, Tehran, IR Iran. Tel: +98-2129903005, Fax: +98-2122431653, E-mail: Ch-eslahchi@sbu.ac.ir

DOI: 10.5812/ijb.9309

Copyright © 2013, National Institute of Genetic Engineering and Biotechnology; Published by Kowsar Corp.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, Hou et al., use the concept of mediate domain in yeast proteome to construct a mediate protein-protein interaction network.

2. Objectives

In this study we find triplet of domains (A, B, C), such that the domain C has an effective role in the interaction of the proteins X and Y containing the domains A and B respectively. For this purpose, we emphasis on two related issues; firstly, identifying those domain pairs that occur frequently in interacting proteins and may not be necessary or sufficient for mediating the interactions of these proteins and secondly, we characterize those domain triplets that are necessary and sufficient for mediating the interactions of these proteins. A domain combination (a triplet of domains) is sufficient for mediating the interaction of a pair of co-localized proteins if they interact whenever the domain combination is observed in them. A domain combination is necessary for mediating a pair of interacted proteins if the deletion of any domains in the domain combination, stops the interaction of those proteins.

3. Materials and Methods

Characterizing domain combinations are necessary and sufficient for mediating protein interactions. The conditions of being “necessary” and “sufficient” cannot always be determined without additional laboratory experiments. For instance, to determine necessity, one has to delete a domain from a pair of interacting proteins and then test them in a laboratory whether the two proteins still interact or not. We approximate these conditions by pragmatic statistical definitions.

Let $D(X)$ be the set of domains of the protein X . For the proteins X and Y and the domains A, B and C , we have: $(A, B) \in (X, Y)$ if $A \in D(X) \& B \in D(Y)$ (Figure 1).

Figure 1. $(A, B) \in (X, Y)$

We define $(A, B, C) \in (X, Y)$ if $(A, B) \in (X, Y)$ and $C \in D(X)$ or $C \in D(Y)$ (Figure 2).

Figure 2. $(A, B, C) \in (X, Y)$

if (a) or (b) holds.

We have (X, Y) contains (A, B) or (X, Y) contains (A, B, C) if $(A, B) \in (X, Y)$ or $(A, B, C) \in (X, Y)$ respectively.

Let I be a set of gold-standard interacting protein pairs and NI be a set of gold-standard non-interacting protein pairs. Let O be a domain combination (a triplet of domains). We define $I_0 = \{(X, Y) \in I \mid O \in (X, Y)\}$ and $NI_0 = \{(X, Y) \in NI \mid O \in (X, Y)\}$.

A domain combination that is observed in only one pair of interacting proteins is not easily justified as a real mediator of the interaction. Therefore it is reasonable to restrict our attention to domain combinations that are observed in at least k interacting protein pairs ($|I_0| \geq k$). In this manuscript we consider $k \in \{2, 3, 4\}$.

Let O be a domain combinations that is observed in at least k interacting protein pairs, we set equation (Equation 1), which can be thought as the probability that O mediating the interaction of the proteins containing it. Since $|NI|/|I|$ is independent of O , in our calculation we show it by m (in the results section it is shown that $m=797.09$).

$$\text{Odds}(O) = \frac{\frac{|I_0|}{|I|}}{\frac{|NI_0|}{|NI|}} = \frac{|I_0|}{|NI_0|} \cdot \frac{|NI|}{|I|}$$

Equation 1. The Probability that O Mediating the Interaction of the Containing Proteins

Let μ_P and σ_P be the mean and standard deviation of the odds of all domain combinations. We assume of μ_P as the odds expected of a random domain combination. Moreover, for a pair of proteins (X, Y) containing O , let:

$$\text{Pr ob}(X, Y|O) = \text{Pr ecision}(O) = |I_0|/(|I_0| \cup |NI_0|)$$

Which is the probability of the interaction of a pair of proteins (X, Y) containing O . Let μ_R and σ_R be the mean and standard deviation of the precision of all domain combinations. We would assume of μ_R as the precision expected of a random domain combination. Consider two thresholds t_1 and t_2 .

Let U be the set of domain combinations that:

(i) Are observed in at least k pairs of interacting proteins.

(ii) Have odds at least t_1 .

(iii) Have precision at least t_2 .

In the next section we would discuss how to obtain the

thresholds t_1 and t_2 . A domain combination O is defined to be “sufficient” for mediating the interaction of a pair of proteins (X, Y) if O is contained in (X, Y) and O is in U . We consider a domain combination O is “necessary” for mediating the interaction of a pair of proteins (X, Y) if there does not exist a domain combination O' that is “more sufficient” (i.e., has higher precision) for mediating (X, Y) than O . Note that O and O' need not have any domain in common.

Thus the domain combination O is necessary and sufficient for mediating the interaction of a pair of proteins (X, Y) if and only if:

$$O = \arg \max_{O' \in U, O' \in (X, Y)} \text{Prob}(X, Y | O').$$

Such a domain combination O is also the best explanation for the interaction of (X, Y) ; and we describe it as a necessary and sufficient domain combination, “ns domain combination”, of (X, Y) .

We denote the set of all ns domain combinations by U' .

3.1. Odds and Precision Threshold (t_1 and t_2 .)

In this section we do particular statistical approaches to obtain the threshold t_1 and t_2 . These statistic evaluations have been done in the data set which will be explained in the result section. It is obvious that all the above formulas depend on k . In the Figures 3, 4, 5, 6, 7 and 8 the distribution of *Odd* and precision of domain combinations for different $k \in \{2, 3, 4\}$ are presented.

Figure 3. Distribution of Odds for $k=2$

Figure 4. Distribution of Precision for $k=2$

Figure 5. Distribution of Odds for $k=3$

Table 1 reveals some of the statistical parameters of odds and precision distributions.

As the Figures 3, 4, 5, 6, 7 and 8 have shown the distribution of odds and precision are not distributed normally. To indicate the threshold t_1 and t_2 ; let H_0 (null hypothesis) be the assumption of the interaction between two proteins containing at least one ns domain combination and H_1 (alternative hypothesis) be the assumption that there is no interaction between two proteins contain-

Table 1. Statistical Parameters of Odds and Precision Distributions of R

k	μ_R	σ_R	Median of Precision	μ_p, m	σ_p, m	Median of Odds, m	Number of Domain Combinations
2	0.5599	0.3841	0.5000	1.3855	1.3084	1	10682
3	0.5749	0.3729	0.5714	1.9029	1.7782	1.3333	3717
4	0.5588	0.3688	0.5714	2.2986	2.2405	1.3333	1348

Table 2. Number of ns Domain Combinations Respect to Different Thresholds

k	γ	t_1	t_2	Number of Sufficient Domain Combination	Number of Sufficient and Necessary (ns) Domain Combinations	Number of Sufficient and Not Necessary Domain Combinations
2		0.55	0.01	5138	5080	58
3	0.2	0.5	0.01	2803	2783	20
4		1	0.1	813	811	2
2		2	0.2	3681	3643	38
3	0.1	3	0.3	1474	1472	2
4		4	0.5	731	729	2
2		2.2	0.4	1485	1483	2
3	0.05	3.01	0.5	637	635	2
4		5	0.59	249	247	2

Figure 6. Distribution of Precision for $k=3$ **Figure 7.** Distribution of Odds for $k=4$

ing at least one ns domain combination. We find t_1 and t_2 such that the related type-I error for the mentioned hypothesis, H_0 does not exceed γ . We obtain t_1 and t_2 by solving $P(\text{Precision}(O) \geq t_2 | H_0) = \gamma$ and $P(\text{Odds}(O) \geq t_1 | H_0) = \gamma$ for each data set. In the present study, we consider $\gamma = 0.05, 0.1$, and 0.2 . A hypothesis test is considered statistically significant if its P-value is less than or equal to a significance level. In this circumstance the null hypothesis is rejected.

Figure 8. Distribution of Precision for $k=4$

Typically the values of this significance level are considered to be 0.05 and 0.1. To test more domain combinations, we consider 0.2 as the significance level too. The number of ns domain combination respect to and are shown in Table 2.

In the method section, we defined three laws for prediction of interaction between proteins with respect to ns domain combinations.

3.2. Prediction

Let $O = (A, B, C)$ be ns domain combination and X, Y and Z be the three proteins. It is predicted that X and Y would merit an interaction by the following laws:

Law I) It is predicted X and Y have an interaction if:

1. X and Y have common localization
2. $O = (A, B, C) \in (X, Y)$ (See Figure 9)

Figure 9. Law I for prediction of interaction between X and Y

Law II) It is predicted X and have an interaction if:

1. X, Y and Z have common localization.
2. $(A, B) \in (X, Y)$ and $C \in O(Z)$
3. X and Z interact and Y and Z interact (see Figure 10)

Law III) It is predicted X and have an interaction if:

Law I or Law II holds.

4. Results

4.1. Dataset

The dataset DIP containing yeast protein interaction (<http://dip.doe-mbi.ucla.edu/dip/Download.cgi>) has been used. This dataset contains 4928 proteins and 17451 interactions and there are 3593 various domains in these proteins. In order to finding domains following address has been used: <http://dip.doe-mbi.ucla.edu/dip/services.cgi>.

In this data set there are two different tyos of domains:

1) Domains that are obtained experimentally. The number of these domains in this data set is 1077 (prefix of pfam codes of these domains is PF).

2) Domains that are obtained by automatic methods. The number of these domains in this data set is 2516 (prefix of pfam codes of these domains is PB).

We derive a reliable subset I from this dataset by including only those interactions that the two interacting proteins (i) have common localization; and (ii) a common partner. The localization of a protein is the location of the protein in the cell. This information can be obtained from Gen Ontology database which is available at www.genontology.org. Each of these conditions were highly associated with reliable interactions (12).

Figure 10. Law II for prediction of the interaction between X and Y

Therefore we consider the set I as the gold standard interaction protein pairs. The resulting subset I has 6955 interactions. Subsequently a set NI from those protein pairs that are assumed to be non-interacting with a high probability as follows had been constructed. A pair of interacting proteins that (i) do not have a common localization; and (ii) do not have a common partner, has been derived. As these protein pairs violate all the key conditions associated with reliable interactions (12), it is believed that NI is a gold standard of non-interacting proteins. The constructed set NI have 5543821 protein pairs.

Therefore in the above calculations $m = |NI|/|I| = 797.09$.

4.2. Evaluation of Prediction

To evaluate the performance of our prediction two measurements had been expressed; precision and recall which are defined by:

$$Recall = TP/|W| \quad Precision = TP/(TP+FP)$$

With respect to the refined data set (sets I and NI), it has been expressed that:

$$W = I$$

TP = the number of predicted edges that are in I.

FP = the number of predicted edges that are in NI.

And with respect to the primary data set, we define:

W = the primary data set.

TP = the number of predicted edges that exist in the data set.

FP = the number of predicted edges that do not exist in the data set.

In Tables 3 and 4 the results of prediction with respect to the refined and primary data sets using the three different laws had been described.

Table 3. Result of prediction respect to sets I and NI

	γ	k	TP	FP	Precision	Recall
Law I						
	0.2					
		2	1392	531	0.7239	0.2001
		3	716	373	0.6575	0.1029
		4	377	104	0.7838	0.0542
Law II						
	0.2					
		2	686	0	1	0.0986
		3	382	0	1	0.0549
		4	230	0	0.5287	0.0435
Law III						
	0.2					
		2	1483	531	0.7366	0.2135
		3	769	373	0.6734	0.1106
		4	403	104	0.7949	0.05579
Law I						
	0.1					
		2	1113	117	0.9049	0.1600
		3	500	35	0.9346	0.0719
		4	103	4	0.9626	0.0148
Law II						
	0.1					
		2	591	0	1	0.085
		3	281	0	1	0.0404
		4	82	0	1	0.0047
Law III						
	0.1					
		2	1210	117	0.9118	0.1740
		3	565	35	0.9417	0.0812
		4	129	4	0.9699	0.0158
Law I						
	0.05					
		2	506	41	0.9250	0.0728
		3	290	17	0.9446	0.041
		4	211	8	0.9635	0.0141
Law II						
	0.05					
		2	285	0	1	0.0410
		3	281	0	1	0.0279
		4	157	0	1	0.0226
Law III						
	0.05					
		2	569	41	0.9328	0.0818
		3	565	35	0.9417	0.0477
		4	251	8	0.9691	0.0361

On the other hand, by our laws, we predict interaction between a pair of proteins (X,Y) if they contain at least one pair of domain (A,B) which is contained in at least one ns domain combination. There are numerous interactions in the data set that do not contain any such pair of domains. Therefore it is expected that recall is not high. The best recall is obtained when we consider Law

III, $\gamma = 0.1$ and $k = 3$ with respect to both data sets. In the next section we reveal the effectiveness of the mediator domain in the interaction between two proteins.

4.3. Mediator Domains

We estimate the effectiveness of the mediator domain C in the interaction of the two proteins X and Y that contain

Table 4. Result of Prediction Respect to Primary Dataset

	γ	k	TP	FP	Precision	Recall
Law I						
	0.2					
		2	1599	5568	0.2231	0.0916
		3	836	2858	0.2263	0.0479
		4	446	962	0.3168	0.0256
Law II						
	0.2					
		2	686	1505	0.3131	0.0393
		3	382	653	0.369	0.021
		4	230	402	0.3639	0.0132
Law III						
	0.2					
		2	1692	6260	0.212	0.0970
		3	889	3104	0.2226	0.0509
		4	472	1147	0.2915	0.0270
Law I						
	0.1					
		2	1282	3169	0.288	0.0736
		3	588	1166	0.3352	0.0337
		4	112	195	0.3648	0.0064
Law II						
	0.1					
		2	591	1178	0.3341	0.033
		3	281	464	0.3772	0.0161
		4	82	118	0.41	0.0047
Law III						
	0.1					
		2	1210	117	0.9118	0.0792
		3	653	1393	0.3192	0.0374
		4	138	272	0.3366	0.0079
Law I						
	0.05					
		2	596	1201	0.3317	0.0342
		3	347	571	0.3780	0.0199
		4	246	346	0.4155	0.0141
Law II						
	0.05					
		2	285	474	0.3753	0.0163
		3	194	319	0.3782	0.0111
		4	157	209	0.4290	0.009
Law III						
	0.05					
		2	659	1428	0.3158	0.0378
		3	389	732	0.3470	0.0223
		4	286	470	0.3783	0.0164

(A, B) , in each ns domain combination $O = (A, B, C)$.

Considering protein pairs that contain (A, B) , the recall and the precision is obtained by:

$$Recall = TP / |W| \quad Precision = TP / (TP + FP)$$

With respect to the refined data set (sets **I** and **NI**), we define:

$$W = \{(X, Y) \in I \mid \exists O = (A, B, C) \in U' \text{ s.t. } (A, B) \in (X, Y)\}$$

TP = the number of predicted edges that are in **I**.

FP = the number of predicted edges that are in **NI**.

$$NW = \{(X, Y) \in NI \mid \exists O = (A, B, C) \in U' \text{ s.t. } (A, B) \in (X, Y)\}$$

And with respect to the primary data set, we define:

$$NW = \{(X, Y) \mid \exists O = (A, B, C) \in U' \text{ s.t. } (A, B) \in (X, Y) \text{ and } (X, Y) \text{ are interacted}\}$$

TP = the number of predicted edges that are in data set.

FP = the number of predicted edges that are not in data set.

Table 5. Precision and Recall of Mediator Domain Respect to Data Sets I and NI

	γ	k	TP	FP	Precision	Recall	W	NW
Law I 0.2								
		2	1392	531	0.7239	0.8305	1676	26897
		3	716	373	0.6575	0.8149	879	11089
		4	377	104	0.7838	0.8667	435	3226
Law II 0.2								
		2	686	0	1	0.4093	1676	26897
		3	382	0	1	0.0549	879	11089
		4	230	0	0.5287		435	3226
Law III 0.2								
		2	1483	531	0.7366	0.8860	1676	26897
		3	769	373	0.6734	0.8749	879	11089
		4	403	104	0.7949	0.9264	435	3226
Law I 0.1								
		2	1113	117	0.9049	0.8202	1357	14297
		3	500	35	0.9346	0.7886	634	4510
		4	103	4	0.9626	0.2901	355	2108
Law II 0.1								
		2	591	0	1	0.43	1357	14297
		3	281	0	1	0.4432	634	4510
		4	82	0	1	0.2315	355	2108
Law III 0.1								
		2	1210	117	0.9118	0.8917	1357	14297
		3	565	35	0.9417	0.8912	634	4510
		4	129	4	0.9699	0.3634	355	2108
Law I 0.05								
		2	506	41	0.9250	0.7931	638	4524
		3	290	17	0.9446	0.8169	355	2108
		4	211	8	0.9635	0.7894	269	1805
Law II 0.05								
		2	285	0	1	0.4467	638	4524
		3	281	0	1	0.4432	355	2108
		4	157	0	1	0.5836	269	1805
Law III 0.05								
		2	569	41	0.9328	0.8918	638	4524
		3	565	35	0.9417	0.8912	355	2108
		4	251	8	0.9691	0.9331	269	1805

$NW = \{(X,Y) \mid \exists O = (A,B,C) \in U \text{ s.t. } (A,B) \in (X,Y) \text{ and } (X,Y) \text{ are not interacted}\}$

In Tables 5 and 6 it has been revealed that the mediator domain C has an appropriate effective role in the interaction between the two proteins X and Y that contain (A,B), in each ns domain combination $O = (A,B,C)$. For example

according to Table 5, under the case of Law III, $\gamma = 0.05$, $k = 4$, the precision is 0.9691 while the ratio $|W| / (|W| + |NW|) = 0.1297$. This means that the incorporation of C into Law III has improved the precision by $0.0901/0.1297 = 7.4719$ fold. That is, assuming the absence of errors in the data sets, a pair of proteins exhibiting a ns domain combina-

Table 6. Precision and Recall of Mediator Domain Respect to Primary Data Set

	γ	k	TP	FP	Precision	Recall	W	NW
Law I 0.2								
		2	1599	5568	0.2231	0.7647	2091	69674
		3	836	2858	0.2263	0.7741	1080	29863
		4	446	962	0.3168	0.8383	532	9881
Law II 0.2								
		2	686	1505	0.3131	0.3281	2091	69674
		3	382	653	0.369	0.3537	1080	29863
		4	230	402	0.3639	0.4323	532	9881
Law III 0.2								
		2	1692	6260	0.212	0.8092	2091	69674
		3	889	3104	0.2226	0.8231	1080	29863
		4	472	1147	0.2915	0.8872	532	9881
Law I 0.1								
		2	1282	3169	0.288	0.7653	1679	40458
		3	588	1166	0.3352	0.7626	771	13821
		4	112	195	0.3648	0.2587	433	6397
Law II 0.1								
		2	591	1178	0.3341	0.3520	1679	40458
		3	281	464	0.3772	0.3645	771	13821
		4	82	118	0.41	0.1894	433	6397
Law III 0.1								
		2	1210	117	0.9118	0.8917	1679	40458
		3	653	1393	0.3192	0.8470	771	13821
		4	138	272	0.3366	0.3187	433	6397
Law I 0.05								
		2	596	1201	0.3317	0.7680	776	13873
		3	347	571	0.3780	0.8014	433	6397
		4	246	346	0.4155	0.7616	323	5040
Law II 0.05								
		2	285	474	0.3753	0.3073	776	13873
		3	194	319	0.3782	0.448	433	6397
		4	157	209	0.4290	0.4861	323	5040
Law III 0.05								
		2	659	1428	0.3158	0.8492	776	13873
		3	389	732	0.3470	0.8984	433	6397
		4	286	470	0.3783	0.8854	323	5040

tion (A, B, C) is, on average, 7.4719 times more likely to interact than a pair of proteins exhibiting the domain pair (A, B) . It means that the domains X and Y ($(A, B) \in (X, Y)$) can be applicable for the interaction between proteins X and Y ($(A, B) \in (X, Y)$) if:

C is in $D(X) \cup D(Y)$

or

- There is a protein Z such that, $C \in D(Z)$, (Y, Z) and (X, Z) interact.

Therefore C has been named the “mediator domain” for A and B . According to the above results, we predicted the interaction between some pair of proteins, using ns domain combination and mediator domain in a good manner.

5. Discussion

In the present manuscript a method for the prediction of the protein interaction using ns domain combination and mediator domain is presented. It is revealed that the mediator domains have an effective role in the prediction protein interactions. Using ns domain combinations and mediator domains, we have predicted high reliable interactions. That is, a pair of proteins exhibiting a ns domain combination (*A, B, C*) is more likely to interact than a pair of proteins exhibiting the domain pair (*A, B*).

Acknowledgements

I would like to thank the department of research affairs of Shahid Beheshti University.

Authors' Contribution

The whole manuscript has been conducted by C. Eslahchi.

Financial Disclosure

None declared.

Funding/ Support

University of Shahid Beheshti and IPM.

References

- Ekman D, Bjorklund AK, Frey-Skott J, Elofsson A. Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *J Mol Biol.* 2005;**348**(1):231-43.
- Costa S, Cesareni G. Domains mediate protein-protein interactions and nucleate protein assemblies. *Handb Exp Pharmacol.* 2008;**186**:383-405.
- Neduva V, Linding R, Su-Angrand I, Stark A, de Masi F, Gibson TJ, et al. Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.* 2005;**3**(12):e405.
- Zhang Y, Fillmore RA, Zimmer WE. Structural and functional analysis of domains mediating interaction between the bagpipe homologue, Nkx3.1 and serum response factor. *Exp Biol Med (Maywood).* 2008;**233**(3):297-309.
- Guimaraes KS, Przytycka TM. Interrogating domain-domain interactions with parsimony based approaches. *BMC Bioinformatics.* 2008;**9**:171.
- Raghavachari B, Tasneem A, Przytycka TM, Jothi R. DOMINE: a database of protein domain interactions. *Nucleic Acids Res.* 2008;**36**(Database issue):D656-61.
- Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol.* 2007;**3**(4):e43.
- Han DS, Kim HS, Jang WH, Lee SD, Suh JK. PreSPI: a domain combination based prediction system for protein-protein interaction. *Nucleic Acids Res.* 2004;**32**(21):6312-20.
- Chua HN, Hugo W, Liu G, Li X, Wong L, Ng SK. A probabilistic graph-theoretic approach to integrate multiple predictions for the protein-protein subnetwork prediction challenge. *Ann N Y Acad Sci.* 2009;**1158**:224-33.
- Li L, Zhao B, Du J, Zhang K, Ling CX, Li SS. DomPep—a general method for predicting modular domain-mediated protein-protein interactions. *PLoS One.* 2011;**6**(10):e25528.
- Hou T, Li N, Li Y, Wang W. Characterization of domain-peptide interaction interface: prediction of SH3 domain-mediated protein-protein interaction network in yeast by generic structure-based models. *J Proteome Res.* 2012;**11**(5):2982-95.
- Chua HN, Wong L. Increasing the reliability of protein interactions. *Drug Discov Today.* 2008;**13**(15-16):652-8.