

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



New Perspectives in Predicting Membrane Protein-protein Interactions

Kelvin X. Zhang^{1,2} and B.F. Francis Ouellette²

¹University of British Columbia

²Ontario Institute for Cancer Research
Canada

1. Introduction

Cells are multi-molecular entities whose biological functions rely on stringent regulations both temporally and specially. These regulations are achieved through a variety of molecular interactions including protein-DNA interactions, protein-RNA interactions and protein-protein interactions (PPIs). PPIs are extremely important in a wide range of biological functions from enzyme catalysis, signal transduction and more structural functions. Owing to advanced large-scale techniques such as yeast two-hybrid and mass spectrometry, interactomes of several model organisms such as *Saccharomyces cerevisiae* (Gavin et al., 2006; Ho et al., 2002; Ito et al., 2001; Krogan et al., 2006; Uetz et al., 2000), *Drosophila melanogaster* (Formstecher et al., 2005; Giot et al., 2003) and *Caenorhabditis elegans* (Li et al., 2004) have recently been extensively studied. Such large-scale interaction networks have provided us with a good opportunity to explore and decipher new information from them. However, there are some limitations of these large-scale data sets: 1) the experimental techniques for detecting PPIs are time-consuming, costly and labor-intensive; 2) the quality of certain datasets is uneven; and 3) technical limitations such as the requirement to tag proteins of interest still exist. As a complementary alternative, computational approaches that identify PPIs have been studied intensively for years and have yielded some interesting results.

Proteins with at least one transmembrane domain constitute 20% to 35% of all known proteins, and therefore account for an important fraction of the proteins involved in biological mechanisms. However, for several reasons, the research on membrane protein interactions has been lagging behind. First, although the current available interactomes contain adequate interactions for analysis, the data sets still have a large amount of false positives. For example, compared to a gold-standard data set, identified protein-protein interactions from three frequently-used high-throughput methods (yeast two-hybrid (Uetz, et al., 2000), tandem affinity purification (TAP) (Gavin, et al., 2006) and high-throughput mass spectrometry protein complex identification (HMS-PCI)) (Ho, et al., 2002) yielded very low accuracy, coverage and overlap (von Mering et al., 2002). Second, some large-scale experimental techniques are biased against membrane proteins. For instance, in order to check whether proteins interact or not, they need to be expressed in the nucleus which may not be their native living environment.

Source: New Achievements in Evolutionary Computation, Book edited by: Peter Korosec,
ISBN 978-953-307-053-7, pp. 318, February 2010, INTECH, Croatia, downloaded from SCIYO.COM

The modified version of the yeast two-hybrid called the split-ubiquitin membrane yeast two-hybrid (MYTH) system (Stagljar et al., 1998) was developed for specially detecting the interactions between membrane proteins. However, it is still time-consuming and labor-intensive, making it infeasible to generate a complete picture of the interactome of membrane proteins at current stage. Several groups have tackled this problem using computational approaches. Miller and colleagues (Miller et al., 2005) worked on identifying interactions between integral membrane proteins in yeast using a modified split-ubiquitin technique. To address the challenges presented in experimental techniques, Xia and colleagues (Xia et al., 2006) developed a computational method to predict the interactions between helical membrane proteins in yeast by integrating 11 genomic features such as sequence, function, localization, abundance, regulation, and phenotype using logistic regression. It however suffers low prediction power and low verifiability with experimental results. In addition to utilizing genomic features to predict protein-protein interactions, graph theory based on the topology of network is an alternative approach to infer protein-protein relationship from protein interaction networks and showing interesting results (Nabieva et al., 2005; Valente & Cusick, 2006). Our group proposed a method to predict interactions between membrane proteins using a probabilistic model based on the topology of protein-protein interaction network and that of domain-domain interaction network in yeast (Zhang & Ouellette, 2008).

The objective of this chapter is to provide an overview focused on recent approaches in predicting membrane proteins by computational methods including a new approach to predict membrane protein-protein interactions developed in our own laboratory. We also discuss the applicability of each computational approach and also the strengths, weaknesses and challenges of all of them.

2. Experimental identification of PPIs between membrane proteins

Currently, the yeast two-hybrid (Y2H) and the tandem affinity purification (TAP) following by mass spectrometry are the two mainstream experimental techniques to identify protein-protein interactions on a large scale. In the yeast two-hybrid system, a bait protein containing a DNA binding domain hybridizes with a prey protein containing an activation domain. If the reporter gene is generated, it means that this pair of proteins interact with each other as the activation domain activates the transcription of the reporter gene. An alternative way is to tag a protein of interest and then express it in cells. The tagged protein and its interacting/binding proteins are purified as it binds to a column or bead. After purification, proteins interacted with the tagged protein are analyzed and identified through SDS-PAGE followed by mass spectrometry. These approaches have provided us with an important amount of valuable protein-protein interactions, which makes it possible to build a more robust interactome of cells.

Besides some intrinsic limitations of these approaches such as high false positives and the requirement to tag proteins of interest, both of them are biased against membrane proteins. In the yeast two-hybrid system, the generation of the reporter gene product indicates an interaction. As the activation of the transcription of the reporter gene takes place in the cell nucleus, participating proteins must be localized to the nucleus. However, membrane proteins usually locate at the cell membrane instead of in the cell nucleus, which makes them excluded from the results of the yeast two-hybrid system. Due to their chemical properties, membrane proteins are difficult to manipulate in protein purification, too.

Therefore, interactions between membrane proteins are less likely to be detected in such approaches.

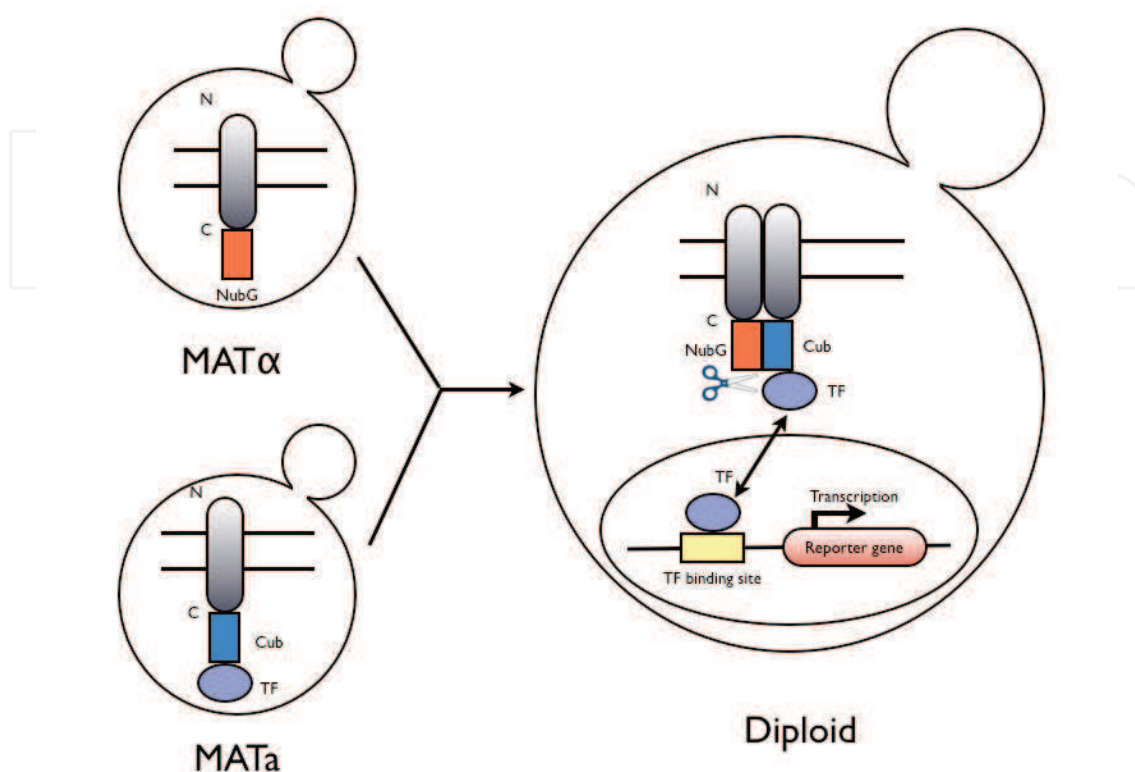


Fig. 1. The split-ubiquitin membrane yeast two-hybrid system. Two membrane proteins are fused to NubG and Cub-TF, respectively. They both are expressed in different mating type. If two membrane proteins interact with each other upon mating as a diploid, the two halves of ubiquitin reconstitute as a quasi-native ubiquitin, a target of ubiquitin-specific proteases that cleave the ubiquitin. The reporter gene is transcribed if two membrane proteins interact with each other as ubiquitin-specific proteases release TF into the nucleus and then activates the transcription of the reporter gene.

To overcome the drawback of the above methods, an approach called the split-ubiquitin membrane yeast two-hybrid (MYTH) system was first developed by Stagljar et al. (Stagljar, et al., 1998) and then was further modified in recent years. MYTH is a yeast-based genetic technology to detect detection of membrane protein interactions *in vivo*. This system is based on the split-ubiquitin approach, in which protein-protein interactions can direct the reconstitution of two ubiquitin halves. In such system (Figure 1), individual proteins are simultaneously introduced into the mutant yeast strain. The carboxy-terminal half of ubiquitin (Cub) and a LexA-VP16 transcription factor (TF) are fused onto the N- or C-terminus of a membrane protein while the amino-terminal half of ubiquitin bearing an Ile 13 Gly mutation (NubG-Prey or Prey-NubG) is fused onto the N- or C-terminus of another membrane protein. The protein fused to the Cub and TF can be referred to as the bait protein and is typically a known protein that the investigator is using to identify new binding partners. The protein fused to the NubG-Prey or Prey-NubG can be referred to as the prey protein and can be either a single known protein or a library of known or unknown proteins. If the bait protein interacts with the prey protein, quasi-native ubiquitin is

reconstituted. The resultant ubiquitin-specific proteases (UBPs) from the process of ubiquitin can cleave at the C-terminus of the Cub, which releases the TF, so some reporter genes such as HIS3, ADE2 and lacZ can be transcribed in the system.

The split-ubiquitin approach has been widely applied and has yielded interesting results. Thaminy et al. (Thaminy et al., 2003) identified the interacting partners of the mammalian ErbB3 receptor using the split-ubiquitin approach, which proved the effectiveness of such system. Miller et al. (Miller, et al., 2005) further applied this approach to construct an array of yeast expressing the fusion of membrane proteins of interest on a large scale. Recently, more applications of the split-ubiquitin approach have been proposed. For example, novel interactors of the yeast ABC transporter Ycf1p (Paumi et al., 2007) and the human Frizzled 1 receptor (Dirnberger et al., 2008) have been identified using such method.

3. Computational prediction of PPIs between membrane proteins

3.1 Multiple evidence-based

Thanks to current advanced techniques, the relationship between genes can be evaluated based on various types of biological data such as protein-protein interaction data, genetic interaction data, gene co-expression data and phylogenetic profiles. These data sets help us better understand gene functions in the context of specific pathways or biological networks and also enables us to discover gene relationships too weak to be detected in individual data type.

The first attempt to predict interaction between membrane proteins on a large scale started from the work of Miller and colleagues (Miller, et al., 2005). They first generated a set of putative protein-protein interactions between membrane proteins through a modified split-ubiquitin technique. In order to test how reliable these putative protein-protein interactions are, they employed an artificial intelligent approach, support vector machine (SVM), to predict interactions at the different confidence levels. For training purposes, they compiled a positive training set containing 56 protein-protein interactions between membrane proteins from their experimental results and the literatures and a negative training set containing random protein pairs. Besides 10 features derived from experiments such as the number of interactions that the Cub-PLV participates, other 8 genomic features such as Gene Ontology term similarity and co-expression are included as input parameters to the SVM algorithm (Table 1). Finally, they tested 1,985 putative interactions from the experiment using the trained SVM and identified 131 highest confident interactions, 209 higher confident interactions, 468 medium confident interactions and 1,085 low confident interactions.

Xia et al. proposed a prediction method to identify 4,145 helical membrane protein interactions by optimally combining 14 genomic features (Table 1) (Xia, et al., 2006). After the fold enrichment analysis between interacting membrane protein pairs and all membrane protein pairs, they found 11 features are good indicators to predict interactions. Three features (relative protein abundance, relative mRNA expression and relative marginal essentiality) do not demonstrate statistically significant difference between interacting membrane protein pairs and all membrane protein pairs. The authors compiled a gold-standard positive set by selecting all membrane protein pairs in the same MIPS complex and a gold-standard negative set by paring all membrane proteins not in the MIPS complexes. They applied both the logistic regression classifier and the Naïve Bayes classifier on the gold-standard data sets using 11 genomic features. They demonstrated that the integration-

based classifier outperforms single evidence-based classifier. Also the logistic regression classifier has higher true positive rate than the Naïve Bayes classifier.

Features	Biological relevance	Ref1	Ref2
The number of interactions that the Cub-PLV participates	A membrane protein was proved to interact with other membrane proteins in the experiment.		*
The number of interactions that the NubG participates	A membrane protein was proved to interact with other membrane proteins in the experiment.		*
Whether both spots for a given NubG were found by the Cub-PLV in either repetition	A membrane protein was proved to interact with other membrane proteins in the experiment.		*
Whether repeated screens by using the same Cub-PLV found this NubG	A membrane protein was proved to interact with other membrane proteins in the experiment.		*
The total number of times that this interaction was observed in the screen	A membrane protein was proved to interact with other membrane proteins in the experiment.		*
Whether a reciprocal interaction is observed	A reciprocal interaction represents the more reliable interaction.		*
Whether the reciprocal interaction was tested	A reciprocal interaction represents the more reliable interaction.		*
The total number of times that this interaction was observed in this orientation or its reciprocal	A reciprocal interaction represents the more reliable interaction.		*
The strength of growth of the yeast in the positive colonies	Stronger interactions result in more growth of the yeast.		*
The relative strength of growth of the yeast in the positive colonies to the controls.	Stronger interactions result in more growth of the yeast.		*
The mutual clustering coefficients, the meet/ min coefficient, the geometric coefficient, and the hypergeometric coefficient	High coefficient score indicates interactions.		*
The difference in the codon enrichment correlation (CEC) between the two proteins	Interacting proteins might have comparable codon compositions.		*

GO functional similarity	A pair of membrane proteins tends to interact with each other if they share very similar Gene Ontology (GO) terms.	*	*
MIPS functional similarity	A pair of membrane proteins tends to interact with each other if they share very similar functional categories as defined in the MIPS database.	*	
Membrane co-localization	A pair of membrane proteins tends to interact with each other if they are assigned to the same cellular localization based on the SGD database.	*	*
Total protein abundance	A pair of membrane proteins tends to interact with each other if the sum of their protein abundance is high.	*	
Total mRNA expression	A pair of membrane proteins tends to interact with each other if the sum of their mRNA expression level is high.	*	*
Relative protein abundance	A pair of membrane proteins tends to interact with each other if the absolute difference between their protein abundance is low.	*	
Relative mRNA expression	A pair of membrane proteins tends to interact with each other if the absolute difference between their mRNA expression levels is low.	*	*
mRNA expression correlation	A pair of membrane proteins tends to interact with each other if the correlation of their mRNA expression profiles over time-course experiments is high.	*	*
Transcriptional co-regulation	A pair of membrane proteins tends to interact with each other if they are related by a same transcription factor.	*	
Co-essentiality	A pair of membrane proteins tends to interact with each other if they both are essential genes.	*	*

Total marginal essentiality	A pair of membrane proteins tends to interact with each other if the sum of their marginal essentiality is high.	*	
Relative marginal essentiality	A pair of membrane proteins tends to interact with each other if the absolute difference between their marginal essentiality is low.	*	
Genetic interaction	A pair of membrane proteins tends to interact with each other if they also genetically interact with each other.	*	
Gene fusion, phylogenetic profile, gene neighborhood, gene cluster	A pair of membrane proteins tends to interact with each other if they have high score in the Prolinks database representing functional relatedness.	*	

Table 1. A list of biological features indicating the interactions between membrane proteins. Ref1 represents the method proposed by Xia et al. and Ref2 represents the method proposed by Miller et al. A star sign means this feature has been applied to the corresponding approach.

3.2 Protein primary sequence and structure-based

Helix-helix interactions within a membrane protein or between membrane proteins play a critical role in protein folding and stabilization. Therefore, it has been of great importance to test if a pair of membrane proteins could interact with each other through helix-helix interactions.

Eilers et. al proposed a method to calculate helix-helix packing values at the level of individual atoms, amino acids and entire proteins (Eilers et al., 2002). They found that packing values could be utilized to differentiate transmembrane proteins and soluble proteins as transmembrane helices pack more tightly. Besides packing values, they also demonstrated that helix contact plot, a method to calculate distances between all backbone atoms of each interacting helix pair, is another feature that can be used to classify transmembrane proteins and soluble proteins because the helix contact plot of transmembrane proteins display a broader distribution than that of soluble proteins. This study provides us with a good starting point to predict interactions between membrane proteins using helix packing and interhelical propensity.

Instead of using physical properties between residues, Fuchs et al. developed an approach to predict helical interactions based on the co-evolving mechanism of residues (Fuchs et al., 2007). The underlying hypothesis is that residues within the same particular protein structure tend to be mutated concurrently. They first generated a set of co-evolving residues from seven different prediction algorithms and the helix-helix interactions were then predicted by comparing helix pairs to their structural information in the Protein Data Bank (PDB) combined with this set of co-evolving residues. With this approach, interacting helices could be predicted at the specificity of 83% and the sensitivity of 42%. It is

demonstrated that evolutionarily conserved residues are a valuable feature to predict membrane protein interactions.

As more and more structural information related to residues becomes available, more sophisticated computational approaches are needed to improve prediction performance. In a recent publication, a two-level hierarchical method based on support vector machine (SVM) was proposed. In this study, they built two layers of SVMs (Lo et al., 2009). The first layer of SVM was to predict contact residues. Three input features were included at this level: residue contact propensity, evolutionary profile and relative solvent accessibility. The prediction of interactions between contact residues was implemented in the second layer of SVM in which contact residues were used as inputs. They selected five different features in this level: residue pair contact propensities, evolutionary profile, relative solvent accessibility, helix-helix interaction type and helical length. Tested on a set of 85 interacting helical pairs, 768 contact pairs and 939 contact residues, this method reaches to the sensitivity of 67% and specificity of 95%. This approach further proves the notion that the integration of diverse structural and sequence information with residue contact propensities is a good direction to predict helix-helix interactions and membrane protein interactions.

3.3 Network topology-based

A network topology-based approach was proposed by our group (Zhang & Ouellette, 2008). It is able to predict interactions between membrane proteins using a probabilistic model based on the topology of protein-protein interaction network and that of domain-domain interaction network in yeast. It has been demonstrated that the more likely a pair of proteins are functionally related to each other, the more likely they are to share interaction partners (Brun et al., 2003). Moreover, domain-domain interactions have also been shown as indicators of protein interactions due to the binding of modular domains or motifs (Jothi et al., 2006; Pawson & Nash, 2003). Therefore, we sought to examine the hypothesis that two proteins that share same interactors may interact with each other themselves. In order to address this question, we considered the internal protein-protein and domain-domain relationship of a pair of proteins and their protein-protein interaction partners.

Protein-protein interaction and domain-domain interaction data from disparate sources were integrated and then a log likelihood scoring method was applied on all putative integral membrane proteins in yeast to predict all putative integral membrane protein-protein interactions based on a cut-off threshold. It is shown that our approach improves on other predictive approaches when tested on a "gold-standard" data set and achieves 74.6% true positive rate at the expense of 0.43% false positive rate. Furthermore, it is also found that two integral membrane proteins are more likely to interact with each other if they share more common interaction partners. Recently, we proposed an improved approach to predict membrane PPIs by incorporating one more piece of evidence – gene ontology (GO) semantic similarity.

A scoring model can infer how closely a pair of genes is related in a protein-protein interaction network. According to previous research, if two proteins interact with a very similar group of proteins, they are likely to interact with each other (Ho, et al., 2002; Yu et al., 2006), thus, for a given pair of genes, we first mapped them to a pair of proteins, and then found a common set of interactors for this pair of genes and protein-protein interactions within the whole set of common interactors. A scoring method was employed to calculate the likelihood that a group of genes (a pair of query genes) and the whole set of

their common interactors are more densely connected (the number of PPIs within a group of proteins) than would be expected at random (Kelley & Ideker, 2005):

$$S_p(S, I) = \log \frac{P(S, I | N)}{P(S, I | N_{\text{control}})} = \log \frac{\prod_{(x,y) \in S \times S} \beta P_I(x, y) + (1 - \beta)(1 - P_I(x, y))}{\prod_{(x,y) \in S \times S} c_{x,y} P_I(x, y) + (1 - c_{x,y})(1 - P_I(x, y))} \quad (1)$$

where S is a set of common interactors plus a given pair of genes and I is a set of protein-protein interactions among those genes. $P_I(x, y)$ is an indicator function that equals 1 if and only if the interaction (x, y) occurs in I and otherwise 0. For network N , interactions are expected to occur with high probability for every pair of proteins in S . In our work, we followed the previous knowledge to estimate β and set β to 0.9 (Mewes et al., 2006). For network N_{control} , the probability of observing each interaction $c_{x,y}$ was determined by estimating the fraction of all control networks with randomly expected degree distribution which also contain that protein-protein interaction. Comparable control networks were randomly generated by rewiring interaction networks with same node number from the same gene set and same number of degrees, and by repeating the process 100 times.

Should a given pair of proteins has a documented list of domain-domain interactions in iPfam, then we will have two sets of domains corresponding to two proteins. Hence, given a pair of proteins and their common interaction partners, a lot of domain-domain pairs among these sets of domains are possible. A modified model (2) implies dense domain-domain interactions existing in a group of common interactors of a given gene pair. A related log-odds score was used to evaluate the probability that the domain-domain interactions bridging between these two genes and their common interaction partners were denser than random based on the above scoring method:

$$S_d(S, I) = \log \frac{\prod_{(x,y) \in S \times S} \beta \frac{\sum_{(m,n) \in x \times y} D_I(m, n)}{D_x \times D_y} + (1 - \beta)(1 - \frac{\sum_{(m,n) \in x \times y} D_I(m, n)}{D_x \times D_y})}{\prod_{(x,y) \in S \times S} c_{x,y} \frac{\sum_{(m,n) \in x \times y} D_I(m, n)}{D_x \times D_y} + (1 - c_{x,y})(1 - \frac{\sum_{(m,n) \in x \times y} D_I(m, n)}{D_x \times D_y})} \quad (2)$$

Compared to the previous equation, $D_I(m, n)$ is an indicator function that equals 1 if and only if the domain-domain interaction (m, n) occurs in I and otherwise 0; D_x/D_y is the number of domains in each protein x and y ; for network N_{control} , the probability of observing each domain-domain interaction $c_{x,y}$ was determined by estimating the fraction of all control networks with randomly expected degree distribution that also contain that domain-domain interactions occurring between two proteins.

In order to measure the functional similarity between a pair of proteins, we developed a new scoring approach based on GO terms. Given two groups of GO terms (M , N) representing two proteins, the functional similarity between a pair of proteins was calculated by the following formula:

$$S_{GO}(M, N) = \frac{\sum_{i=1}^m \max_{j=1}^n (GO(i, j)) + \sum_{j=1}^n \max_{i=1}^m (GO(i, j))}{m + n} \quad (3)$$

where M is the set of unique GO terms of the protein x ; N is the set of unique GO terms of the protein y ; m is the number of GO terms in the set M ; n is the number of GO terms in the set N ; $GO(i,j)$ is the similarity score between GO term i and GO term j . The similarity scores between a pair of GO terms were computed based on the algorithm G-SESAME, a new advanced method to measure the semantic similarity of GO terms by considering the locations of their ancestor terms of the two specific terms (Wang et al., 2007).

To put the above three types of scores together, the final scoring function for a given pair of proteins was then:

$$S_{final} = S_p + S_d + S_{go} \tag{4}$$

For each possible interaction between integral membrane proteins, we calculated three different scores: PPI score, DDI score and a combined PPI/DDI/GO score according to (1)(2)(3)(4). This generated a table with 996,166 interacting pairs of proteins, each with three interaction probability scores. We compared the performance of our proposed approach by different types of scores: PPI score, DDI score, GO score and the combined score. A ROC curve was plotted by measuring sensitivity and specificity when tested against the gold-standard data set at different cut-off values (Fig. 2). The area under curve is 0.95 for combined score, 0.85 for PPI, 0.74 for DDI and 0.8 for GO terms, respectively, which indicates the good prediction performance of the proposed scoring method. Better performance can be achieved if we used combined scores rather than using PPI scores or DDI scores alone. It is estimated that there are around 5,000 interactions existing between

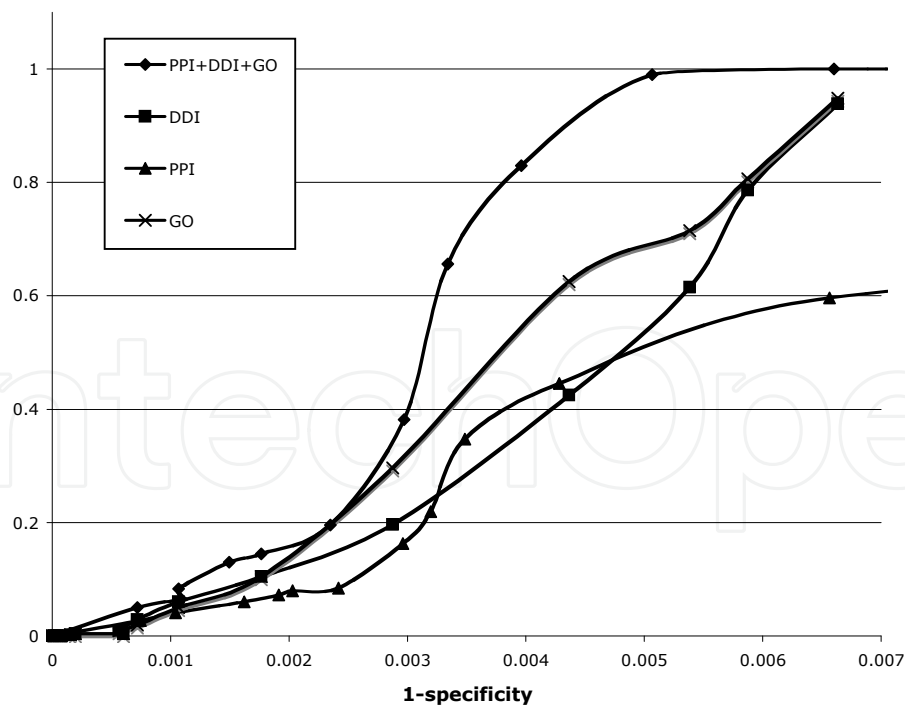


Fig. 2. Curve of receiver operating characteristics (ROC) plotted by the different cut-off values when tested against the gold-standard data set. The area under the curve plotted by PPIs combined with DDIs and GO terms is 0.95, 0.85 for PPI, 0.74 for DDI and 0.8 for GO, respectively.

membrane proteins [12]. Based on that number, we achieved 81.2% true positive rate (sensitivity) at the expense of 0.42% false positive rate ($1 - \text{specificity}$) for a cut-off score of 455, which predicted 4,531 interactions between integral membrane proteins, about 0.61% coverage of all possible interactions among integral membrane proteins.

The map of the interactome of integral membrane protein was built based on 4,531 predicted protein-protein interactions between integral membrane proteins at the cutoff value of 455 (Fig. 3) by Cytoscape (Shannon et al., 2003). 53.4% (281/527) proteins in the interactome map contains at least one transmembrane helix according to the predictions by TMHMM. 80% (392/513) interactions within gold-standard data set overlaps with those within the interactome map but only accounts for 8.4% of the whole interactome of integral membrane proteins. By checking the topology properties of the interactome map, we found that most interactions in the gold-standard data set are in the same complex such as lipid biosynthesis, energy couple proton transport, protein biosynthesis, protein targeting to mitochondria and ATP synthesis coupled electron transport, which reflects the characteristics of performed experiments (detecting protein-protein interactions between same complexes). Our predicted interactions indicates some new members in some complexes such as transport, secretion, vesicle-mediated transport and intracellular transport, which is probably caused by some false negatives from experimental methods.

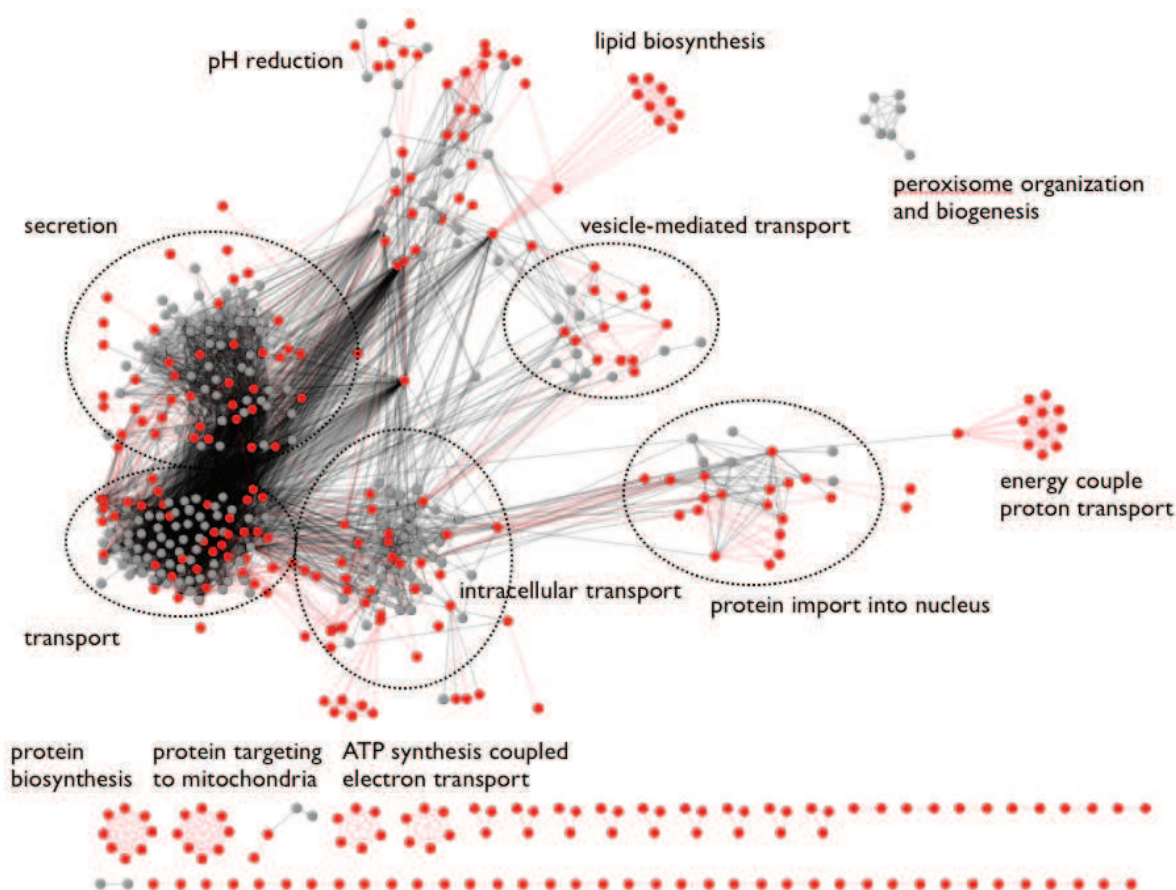


Fig. 3. The interactome map of membrane proteins in yeast. Nodes are represented as membrane proteins, and edges are represented as our predicted interactions between a pair of membrane proteins. Red nodes represent membrane proteins in the gold-standard data set and red edges represent interactions in the gold-standard data set.

One example is that in the group of protein import into nucleus, KAP95 and SSA1 do not interact with other proteins within the group according to the gold-standard data set, however they both play a critical role on nuclear localization signal (NLS)-directed nuclear transport by interacting other proteins to guide transport across the nuclear pore complex (Denning et al., 2001; Liu & Stewart, 2005). Furthermore, observed from the map, some interactions not within the gold-standard data set are found to bridge two complexes. For example, NUP116 and ATP14 are predicted to interact each other connecting two groups: protein import into nucleus and energy couple protein transport. Although there is no evidence demonstrating the direct interaction between NUP116 and ATP14, some research results indicate that ATP14 might be involved in ATP synthesis in the process of protein importing into nucleus (Dingwall & Laskey, 1986; Vargas et al., 2005). Interestingly, we found some new complexes such as peroxisome organization and biogenesis related to the functions of peroxisome membrane proteins such as peroxisome biogenesis and peroxisomal matrix protein import (Eckert & Erdmann, 2003; Heiland & Erdmann, 2005; Honscho et al., 2002).

4. Challenges in predicting membrane PPIs

Complemented by experimental methods, computational approaches provide us with a promising path to reveal a more complete picture of the membrane protein interactome. However, we should be aware of several challenges in predicting membrane PPIs.

First, we are still in lack of reliable membrane PPIs, which results in the difficulty of compiling the gold-standard data set. Currently, positive interaction data is collected from protein pairs in the same protein complex and negative interaction data is derived from those pairs not in the same protein complex. The data quality problem arises as the complex data itself is limited by experimental approaches and contains false positive PPIs. On the other hand, the complex data is biased against membrane proteins, therefore, making it difficult to access the prediction performance of various approaches due to the scarcity of membrane PPIs in the gold-standard data set and the small coverage of membrane interactome. Furthermore, another concern is that large amount of negative data may bring false negatives during the training.

Moreover, it is challenging to interpret the prediction results from different approaches. Inconsistency of predicted membrane proteins has been observed. For example, Miller and colleagues (Miller, et al., 2005) identified 1,949 putative non-self interactions among 705 integral membrane proteins. Xia and colleagues (Xia, et al., 2006) predicted 4,145 helical membrane protein interactions among 516 proteins. Our group recently predicted 4,660 PPIs between integral membrane proteins using the PPIs network and the DDIs data (Zhang & Ouellette, 2008). Interestingly, only 79 protein-protein interactions are overlapped between the results from all three approaches (Figure 4). The reason for these differences among three large-scale sets of membrane protein interactions may be that each approach focuses on different aspects. The experimental result from Miller *et al.* is reliable but probably contains false positives and false negatives due to the intrinsic limitation of experimental techniques they employed. The approach proposed by Xia *et al.* is more focused on the interactions between complexes instead of on binary protein-protein interactions, so the result from Xia *et al.* is prone to predict interactions in the complex. Our approach emphasizes the interactions through the topological properties of PPI and DDI networks and appears to improve on the above methods because these interactions are probably

important features for membrane protein interactions. The better prediction accuracy may be achieved by more sophisticated approaches by incorporating various biologically meaningful evidence such as network topological features, protein primary sequences and structures. Currently, computational membrane protein interaction prediction is intensively studied but focuses only on yeast. Theoretically, methodologies can be applicable to a variety of organisms. However, even with the unprecedented increase of heterogeneous biological data, the data of some organisms such as *Mus musculus*, *Drosophila melanogaster* and especially *Homo sapiens* is far from complete. Therefore, prediction approaches based on multiple lines of evidence undertake the challenge caused by data incompleteness.



Fig. 4. Comparison of the prediction results from three large-scale methods. There are 438 predicted protein-protein interactions overlapping between data sets from Miller et al. and Zhang and Ouellette, 79 between Miller et al. and Xia et al., 372 between Xia et al. and Zhang and Ouellette, respectively.

5. Conclusions

In this chapter, we reviewed various computational approaches to predict protein-protein interactions between membrane proteins. In spite of some limitations caused by incompleteness of existing experimental data, computational methods have demonstrated reasonable prediction accuracy, which make them to be good resources to provide testable hypotheses for experimental validation. With an emergence of different types of high-throughput data at the systematic level, it prompts us to develop and propose computational methods to identify PPIs between membrane proteins by integrating these data sets. Therefore, complemented with various prediction methods and experimental approaches, such studies lead us to elucidate a cell’s interactome.

6. Acknowledgment

The authors are grateful to thank Michelle Brazas and Li Zhang for comments on the manuscripts.

This work was conducted with the support of the Ontario Institute for Cancer Research through funding provided by the government of Ontario. KXZ is supported by the CIHR/MSFHR Strategic Training Program in Bioinformatics. KXZ is also supported by the CIHR Canada Graduate Scholarships Doctoral Award.

7. References

- Brun, C., Chevenet, F., Martin, D., Wojcik, J., Guenoche, A., and Jacq, B. (2003). *Genome Biol* 5, R6.
- Denning, D., Mykytka, B., Allen, N. P., Huang, L., Al, B., and Rexach, M. (2001). *J Cell Biol* 154, 937-50.
- Dingwall, C., and Laskey, R. A. (1986). *Annu Rev Cell Biol* 2, 367-90.
- Dirnberger, D., Messerschmid, M., and Baumeister, R. (2008). *Nucleic Acids Res* 36, e37.
- Eckert, J. H., and Erdmann, R. (2003). *Rev Physiol Biochem Pharmacol* 147, 75-121.
- Eilers, M., Patel, A. B., Liu, W., and Smith, S. O. (2002). *Biophys J* 82, 2720-36.
- Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., Jacq, B., Arpin, M., Bellaiche, Y., Bellusci, S., Benaroch, P., Bornens, M., Chanet, R., Chavrier, P., Delattre, O., Doye, V., Fehon, R., Faye, G., Galli, T., Girault, J. A., Goud, B., de Gunzburg, J., Johannes, L., Junier, M. P., Mirouse, V., Mukherjee, A., Papadopoulo, D., Perez, F., Plessis, A., Rosse, C., Saule, S., Stoppa-Lyonnet, D., Vincent, A., White, M., Legrain, P., Wojcik, J., Camonis, J., and Daviet, L. (2005). *Genome Res* 15, 376-84.
- Fuchs, A., Martin-Galiano, A. J., Kalman, M., Fleishman, S., Ben-Tal, N., and Frishman, D. (2007). *Bioinformatics* 23, 3312-9.
- Gavin, A. C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L. J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M. A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A. M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J. M., Kuster, B., Bork, P., Russell, R. B., and Superti-Furga, G. (2006). *Nature* 440, 631-6.
- Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C. A., Finley, R. L., Jr., White, K. P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R. A., McKenna, M. P., Chant, J., and Rothberg, J. M. (2003). *Science* 302, 1727-36.
- Heiland, I., and Erdmann, R. (2005). *Febs J* 272, 2362-72.
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A. R., Sassi, H., Nielsen, P. A., Rasmussen, K. J., Andersen, J. R., Johansen, L. E., Hansen, L. H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B. D., Matthiesen, J., Hendrickson, R. C., Gleeson, F., Pawson, T., Moran, M. F., Durocher, D., Mann, M., Hogue, C. W., Figgeys, D., and Tyers, M. (2002). *Nature* 415, 180-3.

- Honsho, M., Hiroshige, T., and Fujiki, Y. (2002). *J Biol Chem* 277, 44513-24.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). *Proc Natl Acad Sci U S A* 98, 4569-74.
- Jothi, R., Cherukuri, P. F., Tasneem, A., and Przytycka, T. M. (2006). *J Mol Biol* **362**, 861-75.
- Kelley, R., and Ideker, T. (2005). *Nat Biotechnol* 23, 561-6.
- Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrin-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., Beattie, B., Richards, D. P., Canadian, V., Lalev, A., Mena, F., Wong, P., Starostine, A., Canete, M. M., Vlasblom, J., Wu, S., Orsi, C., Collins, S. R., Chandran, S., Haw, R., Rilstone, J. J., Gandi, K., Thompson, N. J., Musso, G., St Onge, P., Ghanny, S., Lam, M. H., Butland, G., Altaf-Ul, A. M., Kanaya, S., Shilatifard, A., O'Shea, E., Weissman, J. S., Ingles, C. J., Hughes, T. R., Parkinson, J., Gerstein, M., Wodak, S. J., Emili, A., and Greenblatt, J. F. (2006). *Nature* 440, 637-43.
- Li, S., Armstrong, C. M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P. O., Han, J. D., Chesneau, A., Hao, T., Goldberg, D. S., Li, N., Martinez, M., Rual, J. F., Lamesch, P., Xu, L., Tewari, M., Wong, S. L., Zhang, L. V., Berriz, G. F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H. W., Elewa, A., Baumgartner, B., Rose, D. J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S. E., Saxton, W. M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K. C., Harper, J. W., Cusick, M. E., Roth, F. P., Hill, D. E., and Vidal, M. (2004). *Science* 303, 540-3.
- Liu, S. M., and Stewart, M. (2005). *J Mol Biol* 349, 515-25.
- Lo, A., Chiu, Y. Y., Rodland, E. A., Lyu, P. C., Sung, T. Y., and Hsu, W. L. (2009). *Bioinformatics* 25, 996-1003.
- Mewes, H. W., Frishman, D., Mayer, K. F., Munsterkotter, M., Noubibou, O., Pagel, P., Rattei, T., Oesterheld, M., Ruepp, A., and Stumpflen, V. (2006). *Nucleic Acids Res* 34, D169-72.
- Miller, J. P., Lo, R. S., Ben-Hur, A., Desmarais, C., Stagljar, I., Noble, W. S., and Fields, S. (2005). *Proc Natl Acad Sci U S A* 102, 12123-8.
- Nabieva, E., Jim, K., Agarwal, A., Chazelle, B., and Singh, M. (2005). *Bioinformatics* 21 Suppl 1, i302-10.
- Paumi, C. M., Menendez, J., Arnoldo, A., Engels, K., Iyer, K. R., Thaminy, S., Georgiev, O., Barral, Y., Michaelis, S., and Stagljar, I. (2007). *Mol Cell* 26, 15-25.
- Pawson, T., and Nash, P. (2003). *Science* 300, 445-52.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). *Genome Res* 13, 2498-504.
- Stagljar, I., Korostensky, C., Johnsson, N., and te Heesen, S. (1998). *Proc Natl Acad Sci U S A* 95, 5187-92.
- Thaminy, S., Auerbach, D., Arnoldo, A., and Stagljar, I. (2003). *Genome Res* **13**, 1744-53.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., 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., and Rothberg, J. M. (2000). *Nature* 403, 623-7.
- Valente, A. X., and Cusick, M. E. (2006). *Nucleic Acids Res* 34, 2812-9.

- Vargas, D. Y., Raj, A., Marras, S. A., Kramer, F. R., and Tyagi, S. (2005). *Proc Natl Acad Sci U S A* 102, 17008-13.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S. G., Fields, S., and Bork, P. (2002). *Nature* 417, 399-403.
- Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., and Chen, C. F. (2007). *Bioinformatics* 23, 1274-81.
- Xia, Y., Lu, L. J., and Gerstein, M. (2006). *J Mol Biol* 357, 339-49.
- Yu, H., Paccanaro, A., Trifonov, V., and Gerstein, M. (2006). *Bioinformatics* 22, 823-9.
- Zhang, K. X., and Ouellette, B. F. F. (2008). *IEEE Congress on Evolutionary Computation*, 1801-1806.



New Achievements in Evolutionary Computation

Edited by Peter Korosec

ISBN 978-953-307-053-7

Hard cover, 318 pages

Publisher InTech

Published online 01, February, 2010

Published in print edition February, 2010

Evolutionary computation has been widely used in computer science for decades. Even though it started as far back as the 1960s with simulated evolution, the subject is still evolving. During this time, new metaheuristic optimization approaches, like evolutionary algorithms, genetic algorithms, swarm intelligence, etc., were being developed and new fields of usage in artificial intelligence, machine learning, combinatorial and numerical optimization, etc., were being explored. However, even with so much work done, novel research into new techniques and new areas of usage is far from over. This book presents some new theoretical as well as practical aspects of evolutionary computation. This book will be of great value to undergraduates, graduate students, researchers in computer science, and anyone else with an interest in learning about the latest developments in evolutionary computation.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

X. Zhang and B.F. Francis Ouellette (2010). New Perspectives in Predicting Membrane Protein-Protein Interactions, New Achievements in Evolutionary Computation, Peter Korosec (Ed.), ISBN: 978-953-307-053-7, InTech, Available from: <http://www.intechopen.com/books/new-achievements-in-evolutionary-computation/new-perspectives-in-predicting-membrane-protein-protein-interactions>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen