13

Computational Approaches to Elucidating Transient Protein-Protein Interactions, Predicting Receptor-Ligand Pairings

Ernesto Iacucci[1], Samuel Xavier de Souza[2] and Yves Moreau[1]
[1]*K.U.Leuven*
[2]*Universidade Federal do Rio Grande do Norte*
[1]*Belgium*
[2]*Brazil*

1. Introduction

Protein-protein interactions (PPI) are one of the most important biological events which occur in the cell. As PPIs regulate almost all biological processes in the cell, aberrations in PPI may cause severe health problems. One specific area of PPI is receptor-ligand interactions. These interactions are transient yet account for a large part of cell-to-cell communication. As PPI is an important area of research, many groups have proposed methods to make computational predictions of PPI.

The basis of the majority of these methods rely largely on the phylogenetic profile analysis of candidate interactors. These methods determine the similarity of the phylogenetic history of a protein *A* and its putative protein partner *B*, examining the most accurate measure of similarity between the phylogenetic histories of *A* and *B* in order to predict interaction. As interacting proteins should co-adapt as they are under the same evolutionary pressures, it is self-evident that interacting receptors and ligands should be identifiable by application of the same methodology.

While several methods, described below, make use of phylogenetic information to predict protein-protein interaction (PPI), more contemporary work has been conducted in the area of data fusion and kernel learning. We describe one method [Iacucci et al. 2011] in detail which does both. In this work, the existing line of phylogenetic research is extended by using phylogenetic data to construct a kernel to train a least square support vector machines (LS-SVM) in order to classify candidate receptors and ligands as *interacting* or *non-interacting*.

In this chapter, we discuss the plethora of various methods for determining protein-protein interactions. In addition, we evaluate the application of LS-SVMs to the sub-problem of receptor-ligand interaction prediction.

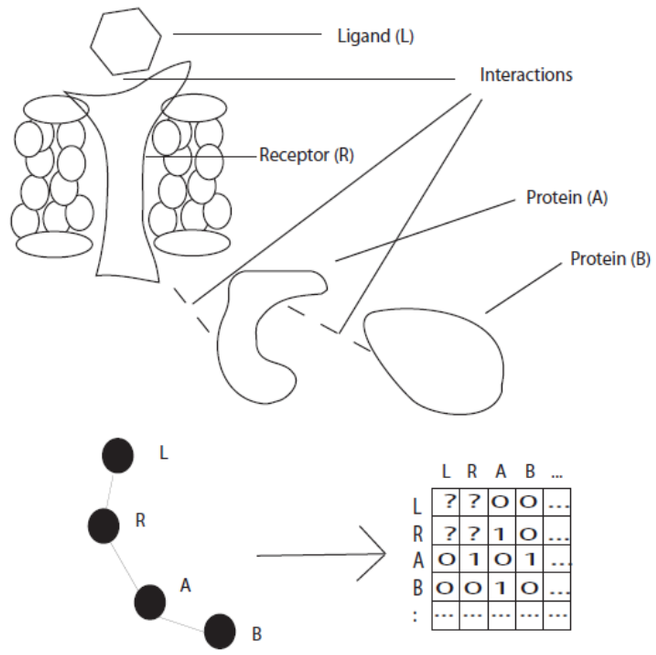


Fig. 1. The Receptor Ligand Schematic. Schematic of receptor-ligand and protein-protein interaction model. Top image is a representation of in-vivo interaction of proteins, receptors, and ligands while bottom image is the graph representation from which a PPI adjacency matrix may be derived. (Figure published in Iacucci *et al*. 2010)

2. Current computational approaches for predicting protein-protein interaction

During the past decade, many methods for prediction of interaction between proteins have been studied due to the crucial role that these interactions have in the understanding of the diverse cellular mechanisms of life forms. Many of these methods involve experimental analysis of specific protein pairs in a smaller scale or, in current high throughput methods [Uetz et al. 2000, Giot et al. 2003], a large amount of protein interactions. The later can be used to detect many interactions with reasonable sensitivity but rather low specificity. Another, relatively inexpensive, way to predict protein-protein interactions does not include wet lab analysis, using instead a variety of computational approaches. These approaches can complement experimental wet lab techniques and are often supported by either the hypothesis of protein co-evolution [Tan et al. 2004, Tillier et al. 2006, Izarzugaza et al. 2006], structural similarities [Gong et al. 2005, Ogmen et al. 2005] or amino-acids sequence conservation [Pitre et al. 2006].

While the entire genomes of many species are already completely sequenced, the interactone of these life forms is often many orders of magnitude larger and yet far from being fully mapped [Claverie et al. 2001, Rubin et al. 2001]. High throughput experimental techniques will certainly help to create this mapping and computational approaches can complement their results identifying false positive interactions, and therefore improving the specificity of these experimental techniques. Apart from the experimental techniques, computational methods are themselves a powerful and affordable alternative to contribute to interactome mapping.

Several computational approaches have been developed in recent years. Many of them are freely available as web tools offering a variety of services to biologists and bioinformatics that range from prediction of interactions between of proteins in pairs or in batch mode, through browsing of consolidated large scale analysis, up to visualization of binding sites and physical interactions in 3-dimensional images.

The methodologies of these many different approaches vary, but they all seem to be supported by the following findings: (a) evidences in favor of the hypothesis of protein co-evolution and the similarities observed in the phylogenetic trees of these proteins; and (b) datasets of already known protein-protein interactions verified by experimental techniques. Co-evolutionary methods find protein pairs with the highest co-evolutionary signal. This information is powerful to predict which members of interacting protein families are associated structurally or functionally although it is not specific enough to predict whether or not two protein families interact. On the other hand, methods supported by verified protein-protein interactions make use of the structural or amino-acid sequence similarities of interacting proteins partners to predict interaction between query protein pairs. This makes such methods more suitable to predict physical interactions rather than functional relationships.

We have reviewed 6 methods and their web tools for predicting protein-protein interactions. Three of them, supported by the protein co-evolution hypothesis, are: TSEMA [Izarzugaza et al. 2006], ADVICE [Tan et al. 2004], Codep [Tillier et al. 2006]. The other three, supported by datasets of verified interactions, are: PIPE [Pitre et al. 2006], PSIbase [Gong et al. 2005], and PRISM [Ogmen et al. 2005]. In the next Sections, we describe each one of these two types of methods.

2.1 Current co-evolutionary methods

Many studies of the problem of predicting protein-protein interactions investigate the similarity of the phylogenetic history of the interaction partners. Many examples of interaction between proteins have presented signs of co-evolution in such a way that members of different interacting protein families present similarity between their phylogenetic trees [Fryxell 1996, Goh et al. 2000, van Kesteren et al. 1996, Moyle et al. 1994, Pazos and Valencia 2001]. The core of co-evolutionary methods is based on measures of similarity for the phylogenetic trees of interacting protein partners.

There are several measures for similarity between phylogenetic trees. The trees can be compared directly [Goh et al. 2000], via distance matrices [Moyle et al. 1994, Goh and Cohen 2002, Ramani and Marcotte 2003, Gertz et al. 2003], or using multiple sequence alignments [Tillier et al. 2006]. In the following Sections, we present three co-evolutionary methods: TSEMA and ADVICE, which uses distance to compare the phylogeneitc trees, and Codep, which computes the correlation between co-evolving partners from their multiple sequence alignments.

2.1.1 Interactive prediction of protein pairing between interacting families TSEMA

TSEMA is a method and web tool to predict mappings between two families of homologous proteins. The probed protein families can either be inputted using the Newick format or in a format comparable with ClustalW, which is used to build the trees. The distances for all

pairs of proteins within both families are extracted from their phylogenetic trees by summing the length of the branches separating each pair of proteins in the trees. The algorithm of TSEMA finds the mapping between the two sets proteins which maximizes the matching between the sets of distances using a modified implementation of the Ramani and Marcotte's Monte Carlo Metropolis method [Ramani and Marcotte 2003].

Availability: http://tsema.bioinfo.cnio.es/

2.1.2 Automated Detection and Validation of Interaction by Co-Evolution – ADVICE

ADVICE predicts and validate protein-protein interactions using observed co-evolution between proteins. The web tool retrieves orthologous sequences of a list of input protein sequences and compute the similarities among the proteins evolutionary histories. The tool also provides visualization for the resulting network of co-evolved proteins.

The ADVICE algorithm infers interaction based on the correlation between distance matrices constructed from the evolutionary history using orthologous sequences of top 10 species. The tool uses BLAST [Altschul et al. 1990] to search the orthologous sequences from Swiss-Prot and TrEMBL databases [Boeckmann 2003]. The distance matrices are constructed using only pairs of orthologous sequences occurring together in the same species. By default, only the orthologous sequences of the top 10 species, based on the BLAST E-value, are used to construct the matrices, excluding those species where more than one orthologous sequence of the input sequence is found. The actual distance matrices are build from the respective multiple sequence alignments using ClustalW [Thompson et al. 1994]. The algorithm then calculates the correlation between pairs of matrices measuring the Pearson's correlation coefficients, which has values between -1, implying 100% anti-correlation, and 1,which representing 100% evolutionary history similarity, being velues above 0.8 good indicators of interaction and values below 0.3 a good cut-off value to detect potential spurious interaction.

Availability: http://advice.i2r.a-star.edu.sg

2.1.3 Maximizing co-evolutionary interdependencies to discover interacting proteins – Codep

Codep and the other co-evolutionary methods find proteins with the highest co-evolutionary signals, independent of physical or functional interaction. The main difference of Codep is that it uses multiple sequence alignments directly rather than distances obtained from the sequences. The user inputs two phylogenetic trees with orthologous sequences. The algorithm maximizes interdependency based on the maximal mutual information. It does this by fixing one of the multiple sequence alignments and varying the order of the other via exhaustive search or via simulated annealing.

The rationale to use directly multiple sequence alignments instead of the distance matrices, which provides a faster way to calculate correlation, is that character-state methods in the field of phylogenetic analysis are more powerful than distance method and some information can be lost in transforming character-state data into distance matrices.

Availability: http://www.uhnresearch.ca/labs/tillier/

2.2 Methods based on verified interactions

Another promising computational approach to predict new protein-protein interactions is to look at the physical structure and the conservation of amino-acid sequences in partners of interactions that are already reliably known to exist. Then, use the gathered information to find correlation with query protein partners of a probed interaction. Many methods apply this approach, which have delivered powerful tools for finding new interactions [Pitre et al. 2006] and even to corroborate with the protein co-evolution hypothesis [Kim et al. 2004]. In the next three Sections we describe three of these methods: PIPE, which compares amino-acid subsequences between probed protein partners and partners of verified protein interactions from a database; and PSIbase and PRISM, both which compare structural characteristics of probed and verified interactions.

2.2.1 Protein-Protein Interaction Prediction Engine – PIPE

PIPE is a computational tool that can effectively identify protein-protein interactions among *S. cerevisiae* protein pairs. It relies on previously determined *S. cerevisiae* protein interactions compiled from the DIP [Salwinski et al. 2004] and MIPS [Mewes et al. 2002] databases to construct a graph where the nodes are proteins and the edges represent the relationship of interacting proteins.

The working principle of the PIPE algorithm to probe interaction between the pair of proteins A-B is to compare sliding subsequences of amino-acids of size w from A to subsequences of the same size of all proteins in the graph of known interactions; then compare sliding subsequences of B to the neighbors of all matches of A. If protein pair C-D are connected in the graph, representing a verified interaction, and if A has subsequence matches with C and B has matches with D, then the pair A-B is more likely to present interaction. The accumulation of all matches of subsequence comparisons presented in form of a matrix indicates a predicted interaction when the higher values in this matrix is above a given threshold of M matches.

The algorithm has three tuning parameters: w, M, and S_{PAM}, which is the threshold value that indicates a match between two subsequences of amino-acids. The author of PIPE chose to fix w in 20, and tune the other two parameter either by trial and error or by statistical evaluation.

PIPE is reported to have success rate comparable to biochemical techniques, with a sensitivity of 61% , specificity of 89%, and overall accuracy of 75%. The main disadvantages of PIPE is its heavy computational burden and its limitation to yeast proteins.

Availability: http://pipe.cgmlab.org

2.2.2 Protein Structural Interactome Map – PSIMAP

PSIMAP is a map that describes the information about domain-domain and protein-protein interactions known to exist in the Protein Data Bank of structures. It is based on the principle that interaction between protein structures is conserved as closely as protein structures themselves [Park et al. 2001, Aloy and Rossell, 2002; Aloy et al. 2003]. It that predicts if domains or proteins structures interact calculating if every possible pair of

structures has an Euclidean distance below a certain threshold. There are three different methods to do this: Full Atom Contact (FAC); Sample Atom Contact (SAC); and Bounding Box Contact (BBC). FAC is the most accurate, whereas SAC and BBC [Dafas et al. 2004] are faster methods.

PSIMAP extract the molecular interaction information of proteins from the PDB. It associates this information to domains using the Structural Classification of Proteins (SCOP) to assign the domains to the structures.

Availability: http://psimap.org and http://psibase.kaist.ac.kr/

2.2.3 Protein Interactions by Structural Matching – PRISM

The PRISM tool allows the user to explore protein interfaces and predict protein-protein interactions by comparing the structure of query proteins to those of a structurally and evolutionarily subset of biological and crystal interactions present in the Protein Data Bank (PDB) [Berman 2000]. Interfaces are defined as the set of residues forming the region of the structure through which two different protein chains bind to each other. This set consists the contacting residues between the chains and the neighboring residues up to a certain distance threshold.

The interfaces in PRISM were obtained from all higher complexes of proteins available in the PDB [Keskin et al. 2004]. From the 49512 interfaces extracted form the PDB, 8205 clusters were obtained using a sequence order-independent computer vision-based algorithm to structurally compare the interfaces. From these 8205 clusters, PRISM considers only 158 template interfaces (Oct/2011) that were found to have evolutionary hotspots [Keskin et al. 2005].

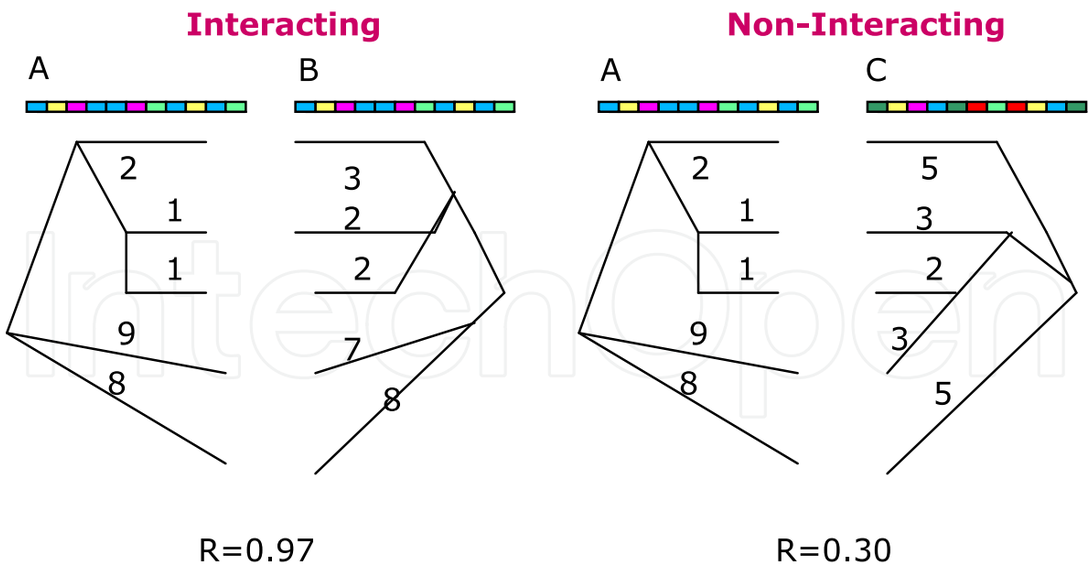


Fig. 2. Phylogenetic Analysis of Proteins

As proteins *A* and *B* are interacting proteins, they share a similar phylogenetic history and thus their phylogenetic profiles are highly correlated (R=0.97). Proteins A and C are non-interacting and are thus not strongly correlated (R=0.30).

PRISM algorithm compares the 158 template interfaces to a target dataset of 18698 structures obtained from passing the structures extracted form the PDB through a 50% sequence identity filter, splitting multimeric proteins into constituent chains, and counting homologous chains only once. The user can also probe a protein structure that is originally not present in the target dataset. To compare target proteins to template interfaces PRISM algorithm do as follow: (a) extract target protein surfaces; (b) compare the target surface with all interface complementary partners from the template dataset using MULTIPROT [Shatsky et al. 2004] in order to detect common geometrical cores in a sequence-order-independent way; (c) check for the presence of hotspots in the target structure. The final prediction score is calculated weighting the structural match ratio and the hotspot match ratio.

Availability: http://prism.ccbb.ku.edu.tr/prism/

3. Phylogenetics and beyond, how multiple kernel learning can improve predictions of receptor-ligand pairings

As seen in the sections above, there are several groups which have used phylogenetic analysis to predict PPI. Here we examine the use of multiple kernel learning in the task of PPI prediction. Kernel learning provides the ability to utilize directly and indirectly related data (such as expression measures, domain content, etc.) and perform classification in high dimensional space. When different data sources are used, separate kernel classifiers can be built and the combined output used to provide a final result.

One of the first groups to look at predicting PPI using multiple data sources was Bhardwaj et al. (2003). They use both phylogenetic information as well as expression data to make their predictions. The use of both data sources were proved, in their work, to provide results with greater accuracy than with using phylogenetic analysis alone. Co-expression is a logical source of information for use in this setting as proteins which interact for the purpose of performing a common function are likely to be co-expressed as they will need to be present at the same time in the cell [Bhardwaj et al. 2003, Grigoriev et al. 2001].

The idea of combining expression and phylogenetic information to predict PPI is clearly a step on a path which leads one to consider a wider variety of data integration. Other data sources include domain information as domains are known to interact and it is clear that this data would provide additional insight into the task of protein-protein interaction. Combining the above mentioned data sources can be carried out by using multiple kernel learning.

To examine the utility of multiple kernel learning with respect to this task, it is necessary to cite an example in which it performs better than other settings. One such example exists when one looks at the work of Gertz et al. (2003) and compare it with the work presented in Iacucci et al. (2011). Both groups look at the receptor-ligand prediction task and apply computational methods to the same dataset. The datasets consist of members of the chemokine and tgfβ ligand families with their respective receptor families. In the case of Gertz et al (2003), distances matrices are created for the families and are matched according to their similarity. Using a Metropolis Monte Carlo optimization algorithm, the Gertz et al. (2003) group explored and scored possible matches between the two matrices, until they reached optimal solutions. A limitation of this approach is that it relied on phylogenetic distance information alone.

Contrary to the work of Gertz et al (2003), the work presented by Iacucci et al 2011 proposes that the integration of multiple data sources results in more accurate matches. This work involved the creating of a combined kernel classifier to carry out the learning task. While other kernel-based works have been applied to the PPI task [Kim et al. 2010, Miwa et al. 2009], the work of Iacucci et al (2011) is unique as they apply multiple kernel learning to the receptor-ligand problem. More specifically, they apply the least-squares support vector machines (LS-SVM) method based on the conclusions by Suykens et al. (2001) which shows this implementation to be robust.

The ability of Iacucci et al. (2011) to predict candidate receptor-ligand pairs has been show to outpace that of Gertz et al. (2003) on the same dataset. This work involves using multiple data sources (expression, phylogenetic, and protein-domain content information), computing separate kernels for each data type, creating LS-SVM classifiers and combining the results to predict receptor-ligand pairs. The specifics of these steps will be discussed below.

3.1 Data sources

Several choices for data sources can be considered when addressing the PPI prediction task. While the studies, mentioned above, which use phylogenetic information rely on sequence data, other sources are available. Such sources include domain content data and expression data.

The phylogenetic data used in the Iacucci et al. (2011) study was derived through several steps. First, candidate receptor and ligand sequences were retrieved for seven species (*Rattus norvegicus*, *Mus* musculus, *Homo sapiens*, *Pan troglodytes*, *Canis familiaris*, *Cavia porcellus*, and *Bos taurus*) from ensemble build 51 [Hubbard et al. 2009]. Following this, the sequences were aligned using ClustalW [Thompson et al. 1994] Once aligned, the sequences were edited so as to eliminate the positions which were not conserved across the seven orthologous sequences. Finally, the pair-wise alignment score was then taken for each possible species to species comparison between the edited orthologous sequences (as seven species are used, a total of 21 pair-wise comparisons for each candidate are created). The distance scores form a phylogenetic vector which was then used to create the phylogenetic kernel.

The expression data used in the Iacucci et al. (2011) work was taken from the well-known GNF human expression atlas (79 tissues) [Su et al. 2004], the data was normalized (values were mean-zeroed and the standard deviation was set to one) and was further transformed into the expression kernel.

For the Iacucci et al. (2011) work, the domain content of each candidate protein (receptor or ligand) was taken from the Interpro Database [Hunter et al. 2009]. A vector for each candidate protein was created where the presence of a protein domain was indicated with a ’1’ and the absence of a domain was indicated by a ‘0’. This data was then transformed to create the domain content kernel.

The “Golden Standard” for the verification of the Gertz et al (2003) and the Iacucci (2011) et al. work is based on the Database of Ligand-Receptor Partners (DLRP) [Graeber et al. 2001]. This dataset is an experimentally derived dataset where known receptor-ligand pairs are stored. The information found here was used to train the LS-SVM described below. In addition, it was also used as the “Golden Standard” to determine which predictions, by both

groups, were true positives and false positive as well as false negatives and true negatives. These values were then used to calculate specificity and sensitivity of each groups' predictions to ultimately determine which approach provided better results.

3.2 Kernel creation and the LS-SVM

The creation of the kernels and the training of the least-squares support vector machine (LS-SVM) in the work presented by Iacucci et al. (2011) required multiples steps. First, the data sources, discussed above, were used to create data matrices (phylogenetic, expression, and domain content) which were then used to create three kernels for each receptor-ligand family. Following this, the LS-SVMs were trained using the three kernels to predict outcomes for receptor-ligand pairs known from the DLRP "Golden Standard".

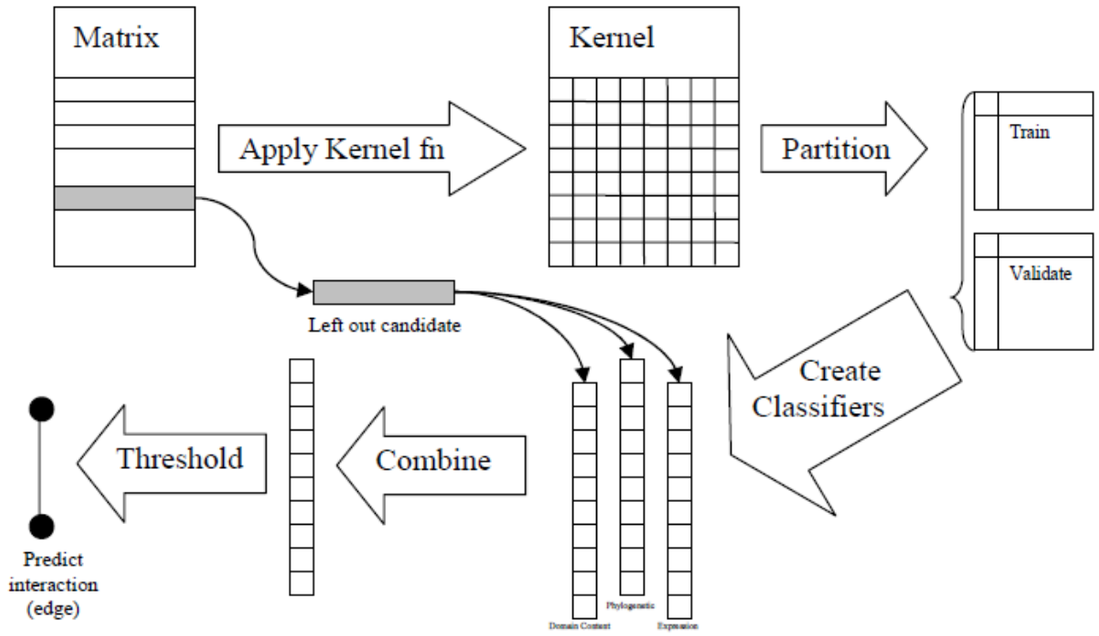


Fig. 3. Work flow of the combined kernel classifier

Data was partitioned into training and validation sets and parameters were tuned using a five fold validation strategy. The final output of the classifiers was achieved by a leave one out strategy. The classifier values were combined for a final result and a threshold was applied to determine which values are predicted edges (Figure published in Iacucci *et al.* 2011).

The kernel function used by Iacucci et al. (2011) measures the similarity between two proteins *A* and *B* (K(*A*,*B*)), one a candidate receptor *A* and the other a candidate ligand *B*. The LS-SVM classifier produced by Iacucci et al. (2011) is a binary predictor which assigns new examples in "interacting" or "non-interacting" classes. Creating the kernels from the various data matrices involved trials with different kernel functions, with linear functions ultimately being found to give the best performance in all cases. Data was partitioned into training and validation sets and parameters were tuned using a five fold validation strategy The final output of the classifiers was achieved by a leave-one-out strategy. The classifier values were scaled (minimum set to zero, maximum set to one). The values were then

combined, as defined in (1), for a final result. Figure 3 provides an overview of the workflow as described above.

$$g_{comb}(x) = \frac{g_{phylo}(x) + g_{\exp}(x) + g_{dom}(x)}{3} \tag{1}$$

3.3 Results and discussion

The comparison of the phylgenetic based method of Gertz et al. (2003) and the combined kernel classifer method of Iacucci et al. (2011) provides a clear perspective on the advantages of multiple kernel learning in the PPI prediction task. As both groups use the same dataset and have results which can be summarized and contrasted using recall, precision, and the *F*-measures.

The Iacucci et al. (2011) predictions for the tgfβ family accurately reconstructed over 76% of the supported edges (0.76 recall and 0.67 precision) of the know DLRP receptor-ligand pairs. In this case, the combined kernel classifier was able to relatively improved upon the Gertz et al. (2003) work by a factor of approximately two as the Gertz et al. (2003) work reconstructs 44% of the supported edges (0.44 recall and 0.53 precision) of the know DLRP receptor-ligand pairs. Comparing *F*-measures, we see that the combined kernel classifer method improved upon that of Gertz et al. (2003) significantly as the Iacucci et al. (2011) method has an *F*-measure of 0.71 while that of Gertz et al. (2003) has a value of 0.48.

The Iacucci et al. (20011) predictions for the chemokine family accurately reconstructed over 65% of the supported edges (0.65 recall and 0.23 precision) of the know DLRP receptor-ligand pairs. In this case, the combined kernel classifier was able to relatively improved upon the Gertz et al. (2003) work by a factor of approximately three as the Gertz et al. (2003) work reconstructs 22% of the supported edges (0.22 recall and 0.37 precision) of the know DLRP receptor-ligand pairs. Comparing *F*-measures, we see that the combined kernel classifer method improved upon that of Gertz et al. (2003) significantly as the Iacucci et al (2011) method has an *F*-measure of 0.33 while that of Gertz et al. (2003) has a value of 0.27.

Qualitatively, the performance of the Iacucci et al (2011) method also seems to be matching the performance of Gertz et al. (2003), as the novel interaction of CCR1 with SCY11 [Gao et al. 1996] reported in their work is also discovered using Iacucci et al (2011) method.

The comparison of the results of the two methods discussed here support the notion that kernel learning presents a useful methodology for elucidating receptor-ligand pairings. The benefits of the combined kernel classifier method over the Gertz et al. (2003) method are clear. Foremost in the advantages are the ability to predict multiple ligands for one receptor, which represents an necessary feature for receptor-ligand research. Also, as the classifier output is continuous, the results can be considered to be prioritized, this presents a major convenience to researchers as often the set of candidate ligands are large and financial and time resources to validate few.

4. Conclusion

The task of PPI prediction is a difficult and important area of bioinformatics research. As the number of possible interacting protein pairs in the cell is huge, wet-lab experimentation

validation of all of them is essentially impossible. In addition to being time consuming, in-vivo validation costs are also a consideration. Having a computational method for predicting PPI is therefore a necessary tool for researchers.

Several groups have addressed the PPI prediction task. While several have used phylogenetics to solve the problem, others have used physical protein structures and amino-acid sequence information to assist in making the predictions. We have reviewed these methods and discussed the key differences among them.

Methods, which rely on the physical structure and the conservation of amino-acid sequences in partners of interactions that are already reliably known to exist, also give researchers additional insight to function prediction as the methods are based on known examples. The drawback of these methods is that one has to have a known example for a comparison, which is not always the case when researching candidate receptor-ligand pairs.

Methods which rely on phylogenetic histories to determine PPI are based on a well-establish rational which holds that as interacting proteins co-evolve, there phylogenetic histories should be similar. This explains why the methods which rely on phylogenetic information are largely based on measures of similarity for the phylogenetic trees of interacting protein partners.

The advantage of using multiple kernel learning to predict PPI is apparent when using multiple sources of data. Many of the methods, mentioned above, rely on an ever growing amount of publicly available data. The ever expanding amount of high throughput data which continues to become available to the bioinformatics community represents an excellent opportunity to enhance the kernel classifier method presented in Iacucci et al. (2011).

A practical advantage of using multiple data sources allows one to extend the method as new and higher quality sources become available. For example, if better micro-array dataset becomes available in the future, it is an advantage to be able remove the existing expression-based kernel with one derived from the new dataset without having to the retrain a global classifier. Likewise, if additional data sources become available, adding an additional sub-classifier based on the new data source would take less time to train than adding the data source and retraining the global classifier.

Looking forward many exciting challenges remain to be addressed in this field. While the task of PPI is daunting and complex, the work reviewed above demonstrates that it is also rich with opportunities for improvement and further development.

5. Acknowledgments

Funding: The authors would like to acknowledge support from:

- Research Council KUL:
  ProMeta, GOA Ambiorics, GOA MaNet, CoE EF/05/007 SymBioSys en KUL PFV/10/016 SymBioSys , START 1, several PhD/postdoc & fellow grants
- Flemish Government:
  FWO: PhD/postdoc grants, projects , G.0318.05 (subfunctionalization), G.0553.06 (VitamineD), G.0302.07 (SVM/Kernel), research communities (ICCoS, ANMMM, MLDM); G.0733.09 (3UTR); G.082409 (EGFR)
  IWT: PhD Grants, Silicos; SBO-BioFrame, SBO-MoKa, TBM-IOTA3

FOD:Cancer plans
IBBT
- Belgian Federal Science Policy Office: IUAP P6/25 (BioMaGNet, Bioinformatics and Modeling: from Genomes to Networks, 2007-2011) ;
-EU-RTD: ERNSI: European Research Network on System Identification; FP7-HEALTH CHeartED

6. References

Aloy, P., & Russell, R. B. (2002). Interrogating protein interaction networks through structural biology. Proceedings of the National Academy of Sciences of the United States of America, 99(9), 5896-901. doi:10.1073/pnas.092147999

Aloy, P., Ceulemans, H., Stark, A., & Russell, R. B. (2003). The relationship between sequence and interaction divergence in proteins. Journal of molecular biology, 332(5), 989-98. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14499603

Berman, H. M. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235-242. doi:10.1093/nar/28.1.235

Bhardwaj N, Lu H: Correlation between gene expression profiles and protein-protein interactions within and across genomes. Bioinformatics 2005, 21:2730-2738.

Bleakley K, Yamanishi Y: Supervised prediction of drug-target interactions using bipartite local models. Bioinformatics 2009, 25:2397-2403.

Boeckmann, B. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Research, 31(1), 365-370. doi:10.1093/nar/gkg095

Claverie, J. M. (2001). Gene number. What if there are only 30,000 human genes? Science (New York, N.Y.), 291(5507), 1255-7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11233450

Dafas, P., Bolser, D., Gomoluch, J., Park, J., & Schroeder, M. (2004). Using convex hulls to extract interaction interfaces from known structures. Bioinformatics (Oxford, England), 20(10), 1486-90. doi:10.1093/bioinformatics/bth106

Fryxell, K. J. (1996). The coevolution of gene family trees. Trends in Genetics, 12(9), 364-369. doi:10.1016/S0168-9525(96)80020-5

Gao JL, Sen AI, Kitaura M, Yoshie O, Rothenberg ME, Murphy PM, Luster AD: Identification of a mouse eosinophil receptor for the CC chemokine eotaxin. Biochem Biophys Res Commun 1996, 223:679-684.

Ge H, Liu Z, Church GM, Vidal M: Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. Nat Genet 2001, 29:482-486.

Gertz J, Elfond G, Shustrova A, Weisinger M, Pellegrini M, Cokus S, Rothschild B: Inferring protein interactions from phylogenetic distance matrices. Bioinformatics 2003, 19:2039-2045.

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., et al. (2003). A protein interaction map of Drosophila melanogaster. Science (New York, N.Y.), 302(5651), 1727-36. doi:10.1126/science.1090289

Goh, C. S., Bogan, A. A., Joachimiak, M., Walther, D., & Cohen, F. E. (2000). Co-evolution of proteins with their interaction partners. Journal of molecular biology, 299(2), 283-93. doi:10.1006/jmbi.2000.3732

Gong, S., Yoon, G., Jang, I., Bolser, D., Dafas, P., Schroeder, M., Choi, H. H., et al. (2005). PSIbase: a database of Protein Structural Interactome map (PSIMAP). Bioinformatics (Oxford, England), 21(10), 2541-3.

doi:10.1093/bioinformatics/bti366

Graeber TG, Eisenberg D: Bioinformatic identification of potential autocrine signaling loops in cancers from gene expression profiles. Nat Genet 2001, 29:295-300.

Grigoriev A: A relationship between gene expression and protein interactions on the proteome scale: analysis of the bacteriophage T7 and the yeast Saccharomyces cerevisiae. Nucleic Acids Res 2001, 29:3513-3519.

Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L et al.: Ensembl 2009. Nucleic Acids Res 2009, 37:D690-D697.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L et al.: InterPro: the integrative protein signature database. Nucleic Acids Res 2009, 37:D211-D215.

Izarzugaza, J. M. G., Juan, D., Pons, C., Ranea, J. a G., Valencia, A., & Pazos, F. (2006). TSEMA: interactive prediction of protein pairings between interacting families. Nucleic acids research, 34(Web Server issue), W315-9. doi:10.1093/nar/gkl112

Jacob L, Vert JP: Protein-ligand interaction prediction: an improved chemogenomics approach. Bioinformatics 2008, 24:2149-2156.

Keskin, O., Ma, B., & Nussinov, R. (2005). Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. Journal of molecular biology, 345(5), 1281-94. doi:10.1016/j.jmb.2004.10.077

Keskin, O., Tsai, C.-J., Wolfson, H., & Nussinov, R. (2004). A new, structurally nonredundant, diverse data set of protein-protein interfaces and its implications. Protein science : a publication of the Protein Society, 13(4), 1043-55. doi:10.1110/ps.03484604

Kim S, Yoon J, Yang J, Park S: Walk-weighted subsequence kernels for protein-protein interaction extraction. BMC Bioinformatics 2010, 11:107.

Kim, W. K., Bolser, D. M., & Park, J. H. (2004). Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). Bioinformatics (Oxford, England), 20(7), 1138-50. doi:10.1093/bioinformatics/bth053

Mewes, H. W., Frishman, D., Güldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., et al. (2002). MIPS: a database for genomes and protein sequences. Nucleic acids research, 30(1), 31-4. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=99165&tool=pmcentrez&rendertype=abstract

Miwa M, Saetre R, Miyao Y, Tsujii J: Protein-protein interaction extraction by leveraging multiple kernels and parsers. Int J Med Inform 2009, 78:e39-e46.

Moyle, W. R., Campbell, R. K., Myers, R. V., Bernard, M. P., Han, Y., & Wang, X. (1994). Co-evolution of ligand-receptor pairs. Nature, 368(6468), 251-5. doi:10.1038/368251a0

Nagamine N, Sakakibara Y: Statistical prediction of protein chemical interactions based on chemical structure and mass spectrometry data. Bioinformatics 2007, 23:2004-2012.

Ogmen, U., Keskin, O., Aytuna, a S., Nussinov, R., & Gursoy, a. (2005). PRISM: protein interactions by structural matching. Nucleic Acids Research, 33(Web Server), W331-W336. doi:10.1093/nar/gki585

Park, J., Lappe, M., & Teichmann, S. A. (2001). Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. Journal of molecular biology, 307(3), 929-38. doi:10.1006/jmbi.2001.4526

Pazos, F., & Valencia, A. (2001). Similarity of phylogenetic trees as indicator of protein-protein interaction. Protein engineering, 14(9), 609-14. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11707606

Pitre, S., Dehne, F., Chan, A., Cheetham, J., Duong, A., Emili, A., Gebbia, M., et al. (2006). PIPE: a protein-protein interaction prediction engine based on the re-occurring short polypeptide sequences between known interacting protein pairs. BMC bioinformatics, 7, 365. doi:10.1186/1471-2105-7-365

Rubin, G. M. (2001). The draft sequences. Comparing species. Nature, 409(6822), 820-1. doi:10.1038/35057277

Salwinski, L., Miller, C. S., Smith, A. J., Pettit, F. K., Bowie, J. U., & Eisenberg, D. (2004). The Database of Interacting Proteins: 2004 update. Nucleic acids research, 32(Database issue), D449-51. doi:10.1093/nar/gkh086

Sato T, Yamanishi Y, Kanehisa M, Toh H: The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. Bioinformatics 2005, 21:3482-3489.

Shatsky, M., Nussinov, R., Wolfson, H., Guigó, R., & Gusfield, D. (2002). Algorithms in Bioinformatics. (R. Guigó & D. Gusfield, Eds.) (Vol. 2452, pp. 235-250). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/3-540-45784-4

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke,M.P., Walker,J.R., Hogenesch,J.B: A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 2004, 101:6062-6067.

Suykens JA, Vandewalle J, De MB: Optimal control by least squares support vector machines. Neural Netw 2001, 14:23-35.

Tan, S.-H., Zhang, Z., & Ng, S.-K. (2004). ADVICE: Automated Detection and Validation of Interaction by Co-Evolution. Nucleic acids research, 32(Web Server issue), W69-72. doi:10.1093/nar/gkh471

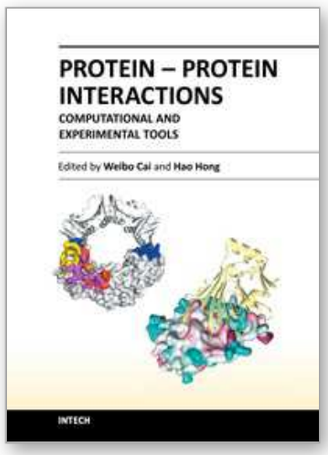
Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994, 22:4673-4680.

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research, 22(22), 4673-80. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=308517&tool=pmcentrez&rendertype=abstract

Tillier, E. R. M., Biro, L., Li, G., & Tillo, D. (2006). Codep : Maximizing Co-Evolutionary Interdependencies to Discover Interacting Proteins, 831(December 2005), 822- 831. doi:10.1002/prot

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature, 403(6770), 623-7. Macmillian Magazines Ltd. doi:10.1038/35001009

van Kesteren, R. E., Tensen, C. P., Smit, A. B., van Minnen, J., Kolakowski, L. F., Meyerhof, W., Richter, D., et al. (1996). Co-evolution of ligand-receptor pairs in the vasopressin/oxytocin superfamily of bioactive peptides. The Journal of biological chemistry, 271(7), 3619-26. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8631971



Protein-Protein Interactions - Computational and Experimental Tools

Edited by Dr. Weibo Cai

ISBN 978-953-51-0397-4
Hard cover, 472 pages
Publisher InTech
Published online 30, March, 2012
Published in print edition March, 2012

Proteins are indispensable players in virtually all biological events. The functions of proteins are coordinated through intricate regulatory networks of transient protein-protein interactions (PPIs). To predict and/or study PPIs, a wide variety of techniques have been developed over the last several decades. Many in vitro and in vivo assays have been implemented to explore the mechanism of these ubiquitous interactions. However, despite significant advances in these experimental approaches, many limitations exist such as false-positives/false-negatives, difficulty in obtaining crystal structures of proteins, challenges in the detection of transient PPI, among others. To overcome these limitations, many computational approaches have been developed which are becoming increasingly widely used to facilitate the investigation of PPIs. This book has gathered an ensemble of experts in the field, in 22 chapters, which have been broadly categorized into Computational Approaches, Experimental Approaches, and Others.

How to reference

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

Ernesto Iacucci, Samuel Xavier De Souza and Yves Moreau (2012). Computational Approaches to Elucidating Transient Protein-Protein Interactions, Predicting Receptor-Ligand Pairings, Protein-Protein Interactions - Computational and Experimental Tools, Dr. Weibo Cai (Ed.), ISBN: 978-953-51-0397-4, InTech, Available from: http://www.intechopen.com/books/protein-protein-interactions-computational-and-experimental-tools/computational-approaches-to-elucidating-transient-protein-protein-interactions-predicting-receptor-l

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

IntechOpen