Using gnome wide data for protein function prediction by exploiting gene ontology relationships

Using gnome wide data for protein function prediction by exploiting gene ontology relationships

Authors: Benso A., Di Carlo S., Politano G., Savino A., ur Rehman H.,

Published in the Proceedings of the IEEE International Conference on Automation, Quality and Testing Robotics (AQTR), 24-27 Nov. 2012, Cluj Napoca, RO.

DOI: 10.1109/AQTR.2012.6237762

Using Gnome Wide Data for Protein Function Prediction by Exploiting Gene Ontology Relationships

Alfredo Benso, Stefano Di Carlo, HafeezurRehman Hafeez, Gianfranco Politano and Alessandro Savino
Politecnico di Torino, I-10129, Torino, Italy
Department of Control and Computer Engineering
email: {alfredo.benso, stefano.dicarlo, hafeez.hafeezurrehman, gianfranco.politano, alessandro.savino}@polito.it

***Abstract*—Many new therapeutic techniques depend not only on the knowledge of the molecules participating in the biological phenomena but also their biochemical function. Advancements in prediction of new proteins are immense if compared with the annotation of functionally unknown proteins. To accelerate the personalized medicine effort, computational techniques should be used in a smart way to accurately predict protein function. In this paper, we propose and evaluate a technique that utilizes integrated biological data from different online databases. We use this information along-with Gene Ontology (GO) relationships of functional annotations in a wide-ranging way to accurately predict protein function. We integrate PPI (Protein Protein Interactions) data, protein motifs information, and protein homology data, with a semantic similarity measure based on Gene Ontology to infer functional information for unannotated proteins. Our method is applied to predict function of a subset of *homosapiens* species proteins. The integrated approach with GO relationships provides substantial improvement in precision and accuracy as compared to functional links without GO relationships. We provide a comprehensive assignment of annotated GO terms to many proteins that currently are not assigned any function.**

***Index Terms*—Function Prediction, Gene Ontology, PPI, Protein motifs**

[1]

I. Introduction

Targets for drug and vaccine design are almost always based on proteins, mainly those involving enzymatic functions. Unfortunately, since many proteins remain uncharacterized, they cannot be taken into account as potential protein targets in drug and vaccine manufacturing process. To make drugs more efficient and to widen the set of their possible targets, it has become necessary to devise effective automated tools for the precise annotation of uncharacterized proteins.

The existence of various recently available high throughput data sets, such as protein-protein interaction networks, microarray data and genome sequences offers a deep insight into the mechanisms related to a protein's function. Until recently, many approaches like [1],[2],[3] and [4] were developed to predict protein function using protein-protein interaction networks. Protein-protein networks, are graphs where each node represents a protein and edges between nodes represent different types of functional relationships. These methods are based on the idea that interacting proteins share common functions; therefore, these methods tend to assign functions to an unannotated protein based on the functions of its neighbors. But for precise and accurate function prediction, the context information of protein functions is necessary to be incorporated by encapsulating the relationship between them. A prominent standard that maintains a structural framework of protein functions is the Gene Ontology. Gene Ontology is a directed acyclic graph where each node represents a functional term with each term arranged in a parent-child relationship with others. The child term either IS A special case of the parent or is a PART OF the parent process i.e., a sub-process or component. Annotated proteins are linked to one or more functional terms of the GO structure and because of parent child relationship a protein is known to a child term it is also known to all of its parent terms. Some techniques, e.g., [5], tried to incorporate protein-protein interactions with Gene Ontology (GO) structural relationships to accurately predict protein function. One limitation of such methods is the increase in complexity to incorporate full functional coverage of a protein since they consider Gene Ontology terms of fixed size.

In this paper we propose a new approach to protein function prediction that overcomes this limitation by incorporating all the annotations of a protein present in the GO structure. The conceptual innovation of our method is to enhance annotated functional space of the GO terms by selecting flexible ontology structure size that represents all annotations of a protein. We build a computational model that integrates optimal potential information that give positive evidence to protein function. The integrated model is then used with GO structure annotations by calculating a semantic similarity measure between terms. Our method computes annotation for protein based upon its likely functional context i.e., set of annotation under an abstract function of the GO structure.

Multiple sources of information i.e., protein-protein interactions, inter-species homology information, and common protein motifs with appropriate similarity threshold values is used to define the functional potential of interacting proteins. This functional potential is a global indicator of functional similarity. We also compute a more specific similarity measure

[1]* **To whom correspondence should be addressed**

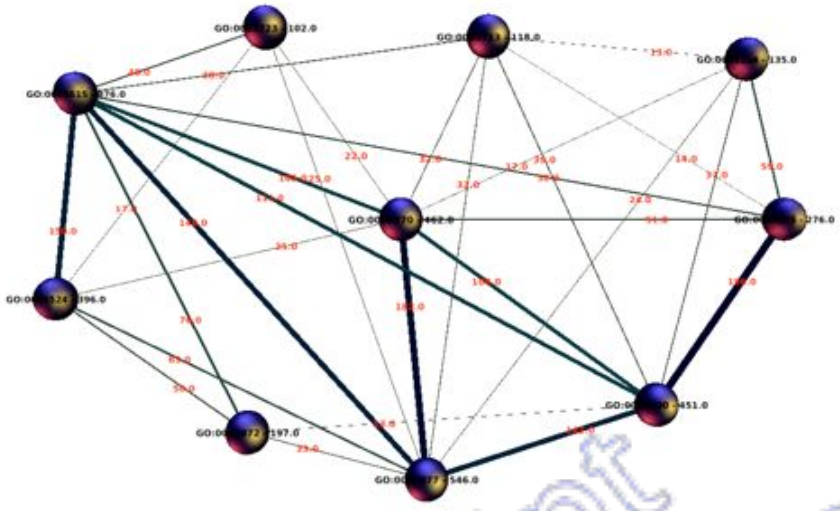

Figure 1. A set of human proteins that share common functions with an edge representing number of common proteins between them.

between the annotated neighbors by defining the functional context in which the interaction can occur. This two way integration improves both the precision and accuracy of the annotation.

The paper is organized as follows: In section II we give an overview of all the closely related protein interaction network approaches that utilize hybrid information with GO structural similarity to predict protein function. In section III we elaborate the proposed computational model that integrates protein-protein interactions, protein binding sites, inter-species homology, and GO functional similarity for protein function prediction. Section V details the effectiveness of the proposed method when applied to proteins of *homosapiens* species. In section V we present some preliminary results and conclude the paper with some future developments.

II. Related Work

Protein-protein interactions based approaches for protein function prediction can be classified into three major groups: *module-assisted, direct methods,* and *probabilistic methods* [6]. Module-assisted methods try to find modules in the network that perform a particular function. The annotation is assigned based upon the protein falling in a specific module. On the other hand direct methods consider the assumption that neighboring proteins in the network have similar functional annotations. Methods of this category, e.g., [7], predict the annotations based upon direct interactions among proteins. This approach is further extended to *indirect neighbor* [8], where the author distinguishes between direct and indirect functional associations by taking into account first and second level neighbors. The functional flow method [2] considers a network flow of annotations from functionally known proteins to unannotated proteins.

Earlier network based approaches assume that proteins with similar functions are always close to each other in the network, which is not true for all proteins[9]; so a third group of techniques considers probabilistic models based on Markov Random Fields [10], [11], [12]. The main hypothesis of these techniques is that a target protein annotation is independent of all other proteins, given the target neighbors [6]. These methods first estimate prior and conditional probabilities of annotations and then project the joint likelihood of unannotated protein to all target annotations.

More recently, research has shown that methods that incorporate GO structure into computational models by fully utilizing the semantic similarity offered by the Direct Acyclic Graph (DAG) architecture of Gene Ontology, are showing more promising results than those that do not utilize it. [13], [14],[15] and [16] use multifunctional GO terms to infer protein function. These methods exchange information within GO structure as well as between interacting proteins to infer protein function by calculating semantic similarity measures.

An integrated Markov Random Field (MRF) based approach is presented in [5], there the authors used GO structure with protein-protein networks and inter-species protein homologs information by constructing MRF based graphs among the GO terms of fixed size. In network based prediction all annotations/(GO terms) of a protein along with their functional relationships should be incorporated in order to accurately predict function. The GO structure has different hierarchy for each functional annotation of protein, e.g., *ADP-ribosylation factor 6* protein is annotated with GTP binding, GTPase activity, and thioesterase binding terms of the GO hierarchy. If the two functions are totally different the hierarchy structure is also very different, e.g., GTP binding, GTPase activity are related to binding and catalytic activity respectively and have different GO structural hierarchy. These different hierarchical information can be utilized by considering all the terms annotation and the semantic relationship between them. But there is always a tradeoff between choosing a network based prediction method and the size of the ontology structure. This bottleneck

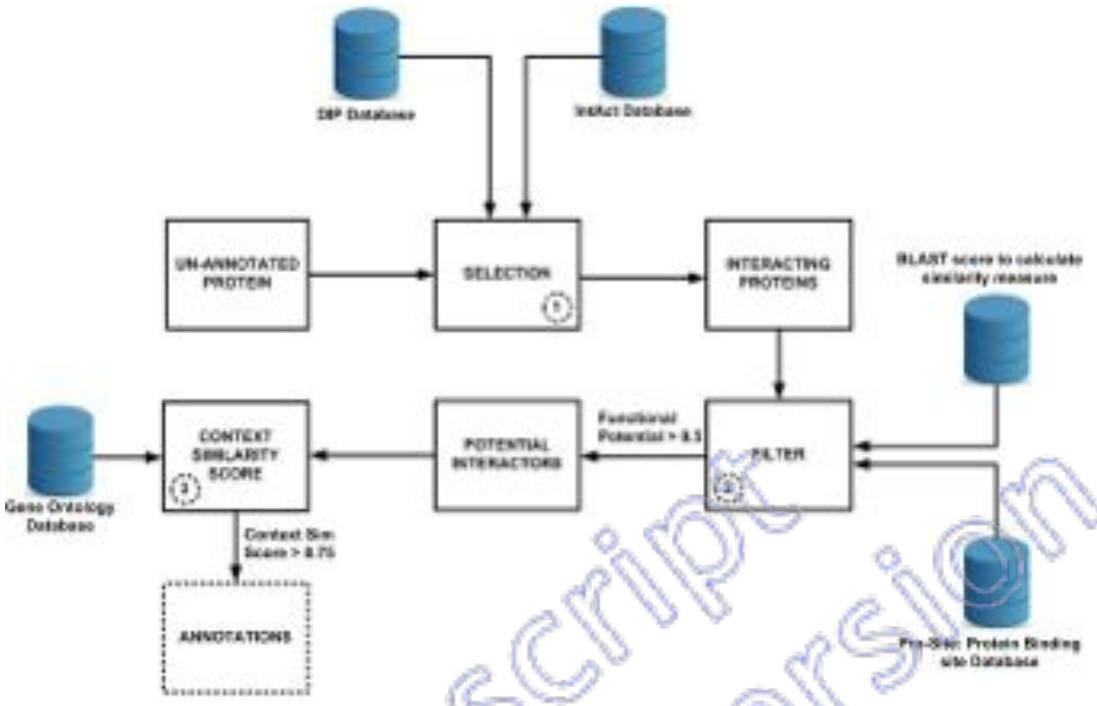

Figure 2. The general scheme of information integration for protein function prediction.

can be overcome by smartly integrating biological data with functional relationships and by taking into account network of potentially interacting proteins. In this way available data can be used to extract new knowledge, to accurately infer functions for unannotated proteins.

III. Methods

A single protein can be part of more than one biological process or molecular activities thus performs multiple functions at the same time. In Figure 1, a functional overlap graph between major molecular functions of *homosapiens* species is shown where each node represents a widely present GO term with the number of proteins annotated with this term. Edges between the terms represents the number of overlapping proteins that share those terms. The graph depicts high intersection of annotations for different GO terms which implies that a large number of proteins have more than one functional annotation. These functional annotations with diverse contexts add much complexity when used with network based prediction methods. We develop a technique that takes into account multi-function annotation with potential interactions to infer the function accurately.

Our technique is based on the fact that interacting proteins are likely to collaborate on a common purpose thus the function of an unannotated protein can be deduced when the function of its binding partners is known. For this purpose we use and integrate information from different biological databases to construct a network for an unannotated protein. Combining various types of information covers different aspects of a protein's activity and hence improves the overall predictive power of automated protein annotation.

We obtain our protein dataset from UniProt [17] database for *homosapiens* specie. We use a subset of these proteins to test our methodology. For unannotated proteins we consider related protein-protein network information which is passed as input to the proposed method. The general scheme of information integration for our method is shown in Figure 2. Our methodology consists of three major steps: selection, filtering and context similarity score.

A. Interacting Protein Selection

In the first step we construct a network for an unannotated protein based upon its interaction data. We select protein-protein interaction data from two databases: IntAct [18] and DIP (Database of Interacting Proteins) [19]. We only consider interactions for which there is an experimental evidence. Redundant and self interactions are removed to construct a protein interaction network for the protein under consideration.

B. Filtering of Potential Interactors

In the second step we compute a functional potential measure $FP_{(i,j)}$ to filter out proteins which have high potential of being functionally similar to the unannotated protein. The functional potential measure $FP_{(i,j)}$ is based upon two functional indicators: (1) protein motif information and (2) protein homolog information.

Motifs are structural elements that are conserved among different proteins. Proteins often have several motifs with distinct evolutionary histories. Patterns of evolutionarily conserved motifs in a protein-sequence reflect the tendency of biochemical functions of an annotated protein. These motifs

Algorithm 1 Function Prediction Algorithm

Input:
PPN(AP,UP,E): is the protein-protein interaction network where,
AP is the set of annotated proteins;
UP is the set of unannotated proteins;
E is the set of edges/inteactions among proteins.
Output:
AT; are the set of annotated terms for UP
Method: PredictFunction(PPN(AP,UP,E))

1: **for** each unannotated protein Pi in the set UP
2: calculate the value of functional potential $FP_{(i,j)}$ between protein P_i and interactor P_j defined in Eq. 2;
3: **if** $FP_{(i,j)} \geq threshold$
4: **for** each pair of interactors between protein P_i and interactor P_j
5: retrieve functional contexts terms F_c from the Gene Ontology graph annotations;
6: **end for**
7: **for** each functional context F_c among annotated neighbors of protein P_i
8: calculate a semantic similarity measure $Sim(P_j, P_k)$ among neighbor P_j and P_k of Protein P_i defined in Eq. 3.
9: **if** $Sim(P_j, P_k) \geq threshold$ **then**
10: add the Functional term to AT vector of protein Pi.
11: **end if**
12: **end for**
13: **end if**
14: **end for**
15: **end method**

can be conserved in unannotated proteins too, so the number of common motifs in two connected proteins can be a strong functional clue for functionally unknown proteins. We incorporate motif information from the ProSite database [20]. ProSite provides a number of conserved motifs for a query protein which are associated with a particular protein functional activity. Thus this information can be used to characterize the associated protein. We compute the functional relevance of proteins by calculating a similarity measure based on common motifs. This measure is normalized to $M_{i,j}$ and is calculated for same number of common motifs between two interacting proteins P_i and P_j as follow,

$$M_{i,j} = \frac{Common_{Motif}(P_i, P_j)}{Min_{Motif}(P_i, P_j)} \quad (1)$$

Where $Common_{Motif}(P_i, P_j)$ is the number of common motifs conserved between the two interacting proteins and $Min_{Motif}(P_i, P_j)$ is the minimum number of motifs conserved in one of the two proteins. If two proteins share the same functionally conserved motifs then there is a higher possibility that they share the same function which is computed using equation 1. The second measure that increases the functional potential of a protein is the homologs similarity between two proteins P_i and P_j of different species. We define a sequence similarity measure between protein P_i and P_j as $S_{(i,j)}$ a normalized pairwise BLAST score. A BLAST score is a numerical value that describes the overall quality of a sequence alignment. Higher numbers correspond to higher similarity. We use normalized BLAST scores, defined as the BLAST score (homolog) divided by self score of query (which is BLAST score of the protein against itself). The value of the normalized BLAST measure ranges from 0 to 1. We only consider score above 0.5 threshold value as in [5]. Normalized protein homology information $S_{(i,j)}$ of a protein adds to functional potential $FP_{(i,j)}$ a higher likelihood of sharing the same functions. Thus, if two proteins are homologs it increases the potential $FP_{(i,j)}$ for sharing the same functional information.

The overall functional linkage potential $FP_{(i,j)}$ between interacting protein P_i and its neighbor P_j is defined as follows,

$$\mathbf{FP_{(i,j)}} = \mathbf{M_{i,j}} + \mathbf{S_{i,j}} \quad (2)$$

where $S_{i,j}$ is the normalized pairwise sequence similarity score, and $M_{i,j}$ is the normalized score for common motifs between protein P_i and P_j, as defined in Eq. 1 and Eq. 2 respectively. The interacting nodes with high value of $FP_{(i,j)}$ are more likely to participate in common functions. After this step we have a network for unannotated protein with potential interactors.

C. Context based Similarity Measure

In the third step we define functional contexts and a similarity score among annotated neighboring proteins by utilizing Gene Ontology relationships. We use Gene Ontology structural data, downloaded from the Gene Ontology database[21]. The GO structure is organized in a DAG structure with three broader top hierarchies: (1) molecular function, (2) biological process and (3)cellular component. Each term of the GO structure refers to a protein function of one of the three hierarchies. The functional terms are organized into two fundamental assumptions: if a protein is positively annotated to a term, then it is also positively annotated to all of its parents or ancestors and if a protein is negatively annotated to a term, then it is also negatively annotated to all of its children or descendants. For our methodology we use the molecular function class of the GO hierarchy.

For proteins with multiple functions we define the functional context terms $F_1, F_2, ...F_n$ as the top most annotations of the Gene Ontology. A functional context is used to calculate the functional relevance of annotated child nodes with the unannotated protein under a given context. Functional contexts improve the predictive power of algorithm as the computations are more centered towards semantically related annotations. For each functional context we calculate a similarity measure among different annotations of neighboring proteins. For protein annotations under the same functional context we define a functional similarity measure between two annotated proteins as the measure of functional relevance as in [15]. The

functional similarity $Sim(P_i, P_j)$ between protein P_i and P_j is calculated as follows,

$$\mathbf{Sim(P_i, P_j)} = \frac{\mathbf{Sim_{TO}(P_i, P_j)}}{\mathbf{Min(annot_{P_i}, annot_{P_j})}} \quad (3)$$

Where $Sim_{\mathbf{TO}}(P_i, P_j)$ is the term overlap between two annotated proteins and $Min(annot_{P_i}, annot_{P_j})$ is the minimum number of annotations between the two proteins. We set a threshold of 0.75 as an adequate measure of similarity between two annotations. The set of annotations which cross this similarity threshold are considered as potential annotations for unannotated protein.

Based upon the $FP_{(i,j)}$ and $Sim(P_i, P_j)$ measures we develop an algorithm (see Algorithm 1) to annotate the function of unannotated proteins. For each unannotated protein we build a network with related interactions. We calculate the value of $FP_{(i,j)}$ measure, for highly similar proteins showing similarity above defined potential; we include them for further analysis to accurately infer function. For set of annotations related to each protein, we calculate functional context terms. The protein annotations under each functional context are incorporated, for a set of highly similar annotations whose similarity score crosses the similarity threshold are considered as potential functions for protein under test.

IV. Experimental Setup and Results

We tested our methodology for a set of homosapeins proteins and have obtained positive results. The complete prototype of the method is under development and will be available in camera ready version after fixing some technical parameters. Here, we report only a single test which is performed on *Aurora Kinase A* protein that is annotated with *ATP binding, protein kinase binding, protein serine/threonine kinase activity, and protein serine/threonine/tyrosine kinase activity* functions. To test our methodology we consider this protein as an unannotated protein and try to infer its function using our methodology. We report all the values that are observed during the test, including all set of annotations.

A. Initial Interactions Dataset

In the the first step, for unannotated protein *Aurora Kinase A,* we construct an interaction network. In Table I, we report the set of interaction data for *Aurora Kinase A*(*O14965*) protein which is obtained from DIP and IntAct databases. Every protein in the table is represented by the relative UniProt Identifier.

B. Functional Potential

The interaction dataset is filtered by calculating a functional potential $FP_{(i,j)}$ measure for each interaction. In Table II, we report the values of homolog sequence similarity $S_{i,j}$, motif similarity measure $M_{i,j}$, and the overall functional potential $FP_{(i,j)}$. The set of protein interactors which attain functional potential $FP_{(i,j)}$ *>0.5* are considered as potential candidates to have functional information. The set of potential interactors is shown in Table III. The annotation information of these interactors is used to infer the function of unannotated protein under consideration.

C. Functional Contexts and Similarity scores

In the last step we define functional contexts for all neighboring interactors by using the Gene Ontology relationships between functions. We only report contexts whose child annotation crosses the similarity threshold i.e., Binding and Catalytic Activity contexts. A similarity score between different annotations of a protein is calculated. The annotation that crosses the similarity threshold i.e., $Sim(P_i, P_j)$ is greater than *0.75* are considered as potential annotations for protein under consideration. In Table IV, we report all functional contexts and related functional similarity values that cross the defined threshold.

Among all four annotations of *Aurora Kinase A* protein our method successfully predicted 3 out of 4 annotations accurately, for unpredicted annotation i.e., *serine/threonine/tyrosine kinase activity* our method predicted *IKappaB Kinase Activity* which is a sibling term of *serine/threonine/tyrosine kinase activity* function in Gene Ontology. Thus, semantically part of the same parent molecular activity.

V. Conclusion

In this paper we presented a new method that uses existing biological data with Gene Ontology functional dependencies to infer function of uncharacterized proteins. We combined three sources of information along with the incorporation of semantic relationships of the annotated functions. Incorporating Gene Ontology information enables simultaneous consideration of multiple but related functional categories. This relationship is utilized for defining functional context for each set of related functions. This context information improves the predictive ability by involving only related functions for the similarity measurements. This approach may easily be extended by integrating more sources of biological information to further increase the function prediction accuracy.

References

[1] S. Letovsky and S. Kasif, "Predicting protein function from protein-protein interaction data: A probabilistic approach," *Bioinformatics*, vol. 19, no. 1, pp. i197–i204, 2003.

[2] U. Karaoz, T. M. Murali, and et al., "Whole-genome annotation by using evidence integration in functional-linkage networks," *Proc. Nat'l Academy of Sciences USA*, vol. 101, pp. 2888–2893, 2004.

[3] B. Schwikowski, P. Uetz, and S. Fields, "A network of protein-protein interactions in yeast," *Nature Biotechnology*, vol. 18, pp. 1257–1261, 2000.

[4] N. Yosef, R. Sharan, and N. Stafford, "Improved network-based identification of protein orthologs," *Bioinformatics*, vol. 24 no. 16, pp. i200–i206, 2008.

[5] A. Mitrofanova, V. Pavlovic, and B. Mishra, "Prediction of protein functions with gene ontology and interspecies protein homology data," *IEEE/ACM Transactions on*

Computational Biology and Bioinformatics, vol. 8 no. 3, pp. 775–784, 2011.

[6] R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function," *Molecular Systems Biology*, vol. 3, pp. 1–13, 2007.

[7] B. Schwikowski, P. Uetz, and S. Fields, "A network of protein-protein interactions in yeast," *Nature Biotechnology*, vol. 18, pp. 1257–1261, 2000.

[8] H. Chua, W. Sung, and L. Wong, "Exploiting indirect neighbors and topological weight to predict protein function from protein-protein interactions," *Bioinformatics*, vol. 19, pp. i197–i204, 2006.

[9] P. Bogdanov and A. K. Singh, "Molecular function prediction using neighborhood features," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 7 no.2, pp. 208–217, April-June 2010.

[10] M. Deng, Z. Tu, F. Sun, and T. Chen, "Mapping gene ontology to proteins based on protein-protein interaction data," *Bioinformatics*, vol. 20, pp. 895–902, 2004.

[11] M. Deng, K. Zhang, S. Mehta, T. Chen, and F. Sun, "Prediction of protein function using protein-protein interaction data," *J. Computational Biology*, vol. 10, pp. 947–960, 2003.

[12] S. Letovsky and S. Kasif, "Predicting protein function from protein/protein interaction data: A probabilistic approach," *Bioinformatics*, vol. 19, pp. i197–i204, 2003.

[13] S. Carroll and V. Pavlovic, "Protein classification using probabilistic chain graphs and th gene ontology structure," *Bioinformatics*, vol. 22 no. 15, pp. 1871–1878, 2006.

[14] A. del Pozo, F. Pazos, and A. Valencia, "Defining functional distances over gene ontology," *BMC Bioinformatics*, 25 Jan 2008.

[15] M. Mistry and P. Pavlidis, "Gene ontology term overlap as a measure of gene functional similarity," *BMC Bioinformatics*, 04 Aug 2008.

[16] G. Pandey, C. L. Myers, and V. Kumar, "Incorporating functional inter-relationships into protein function prediction algorithms," *BMC Bioinformatics*, 12 May 2009.

[17] "The uniprot onsortium: Reorganizing the protein space at the universal protein resource (uniprot)," *Nucleic Acids Res. 40: D71-D75 (2012).*

[18] S. Kerrien and et al., "The intact molecular interaction database in 2012. [pmid: 22121220]," *Nucl. Acids Res.*, doi: 10.1093/nar/gkr1088. [Online]. Available: http://www.ebi.ac.uk/intact

[19] L. Salwinski and et al., "The database of interacting proteins," *Nucl. Acids Res.*, pp. 449–51, 2004. [Online]. Available: http://dip.doe-mbi.ucla.edu

[20] N. Hulo, A. Bairoch, V. Bulliard, L. Cerutti, and et al., "The prosite database," *Nucl. Acids Res.*, pp. D227–230, 2006. [Online]. Available: http://prosite.expasy.org/

[21] "The gene ontology database." Jan 2012. [Online]. Available: http://www.geneontology.org/

Protein Name	IntAct Database Interactors	DIP Database Interactors
O14965	P42771	Q9Y6K9
	P04198	Q9Y297
	O75410	O14920
	Q14008	O15111
	Q8K1R7	Q9ULWO
	O76095	Q9NQS7
	Q6P2K8	O15392
	B3KPG9	
	P61026	
	Q9WU62	
	P04179	
	Q8TEP8	
	AOAUL9	
	Q8VDQ8	
	Q96RR5	
	Q9NWT8	
	P68036	
	Q01469	
	Q9ULW0	

Table I
STEP-1 INTERACTION SET FOR *Aurora Kinase A* PROTEIN

Unannotated Protein P_i	Interactor P_j	Homolog Similarity Score $S_{i,j}$	Motif Similarity $M_{i,j}$	Functional Potential $FP_{(i,j)}$
O14965	P42771	0.507	NA	0.507
	P04198	0.8731	NA	0.8731
	O75410	0.4097	NA	0.4097
	Q14008	0.1311	NA	0.1311
	Q8K1R7	0.7917	1	1.7917
	O76095	0.2279	NA	0.2279
	Q6P2K8	0.3456	NA	0.3456
	B3KPG9	NA	0	NA
	P61026	0.6132	0	0.6132
	Q9WU62	0.2018	NA	0.2018
	P0417	NA	0	NA
	Q8TEP8	0.2072	NA	0.2072
	AOAUL9	NA	NA	NA
	Q8VDQ8	0.5762	NA	0.5762
	Q96RR5	0.5831	NA	0.5831
	Q9NWT8	0.8332	NA	0.8332
	P68036	0.5798	0	0.5798
	Q01469	0.7143	0	0.7143
	Q9ULW0	NA	NA	NA
	Q9Y6K9	0.2083	NA	0.2083
	Q9Y297	0.2166	0	0.2166
	O14920	0.8013	NA	0.8013
	O15111	0.504	1	1.504
	Q9ULWO	NA	NA	NA
	Q9NQS7	0.271	NA	0.271
	O15392	NA	NA	NA

Table II
FUNCTIONAL POTENTIAL BETWEEN *Aurora Kinase A* PROTEIN AND ITS INTERACTORS

Unannotated Protein P_i	Interactor P_j	Homolog Similarity Score $S_{i,j}$	Motif Smilarity $M_{i,j}$	Functional Potential $FP_{(i,j)}$
O14965	P42771	0.507	NA	0.507
	P04198	0.8731	NA	0.8731
	Q8K1R7	0.7917	1	1.7917
	P61026	0.6132	0	0.6132
	Q8VDQ8	0.5762	NA	0.5762
	Q96RR5	0.5831	NA	0.5831
	Q9NWT8	0.8332	NA	0.8332
	P68036	0.5798	0	0.5798
	Q01469	0.7143	0	0.7143
	O14920	0.8013	NA	0.8013
	O15111	0.504	1	1.504

Table III
POTENTIAL INTERACTORS OF *Aurora Kinase A* PROTEIN WITH $FP_{(i,j)}$>0.5

Functional Context	Interactor P_j	$Sim(P_i, P_j)$	Predicted Function	Prediction Result
Binding	O14920	1	ATP Binding	True
	O15111			
	Q8k1r7			
	P68036			
	Q8k1r7	1	Protein Kinase Binding	True
	P42771			
	O14920	1	Protein Binding	True(because Protein Binding is a parent of ATP binding)
	O15111			
	P42771			
	P04198			
	Q8K1R7			
	98VDQ8			
	Q9NWT8			
	P68036			
	P01469			
	P8Q1R7	1	GTP Binding	This term is more specific or detailed annotation of ATP binding.
	P61026			
Catalytic Activity	O14920	1	IKappaB Kinase Activity	This is more detailed annotation of Protein S/T Kinase Activity
	O15111			
	O14920	1	Protein Serine Therine Kinase Activity	True
	O15111			
	P8Q1R7			

Table IV
FUNCTIONAL CONTEXT AND ANNOTATIONS FOR *Aurora Kinase A* PROTEIN WITH $Sim(P_i, P_j)$ >*0.75*