Journal Subline
LNBI 3380

Transactions on Computational Systems Biology I

Corrado Priami
Editor-in-Chief

Springer

Corrado Priami (Ed.)

Transactions on Computational Systems Biology I

Springer

Corrado Priami
Università di Trento
Dipartimento di Informatica e Telecomunicazioni
Via Sommarive, 14, 38050 Povo (TN), Italy
E-mail: priami@dit.unitn.it

Library of Congress Control Number: 2005922555

CR Subject Classification (1998): J.3, H.2.8, F.1

ISSN 0302-9743
ISBN 3-540-25422-6 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11410232 06/3142 5 4 3 2 1 0

Lecture Notes in Bioinformatics 3380

Edited by S. Istrail, P. Pevzner, and M. Waterman

Editorial Board: A. Apostolico S. Brunak M. Gelfand
T. Lengauer S. Miyano G. Myers M.-F. Sagot D. Sankoff
R. Shamir T. Speed M. Vingron W. Wong

Subseries of Lecture Notes in Computer Science

Preface

This is the first issue of a new journal of the LNCS journal subline. The aim of the journal is to encourage inter- and multidisciplinary research in the fields of computer science and life sciences. The recent paradigmatic shift in biology towards a system view of biological phenomena requires a corresponding paradigmatic shift in the techniques from computer science that can face the new challenges. Classical tools usually used in bioinformatics are no longer up to date and new ideas are needed.

The convergence of sciences and technologies we are experiencing these days is changing the classical terms of reference for research activities. In fact clear distinctions between disciplines no longer exist because advances in one field permit advances in others and vice versa, thus establishing a positive feedback loop between sciences. The potential impact of the convergence of sciences and technologies is so huge that we must consider how to control and correctly drive our future activities.

International and national funding agencies are looking at interdisciplinary research as a key issue for the coming years, especially in the intersection of life sciences and information technology. To speed up this process, we surely need to establish relationships between researchers of different communities and to define a common language that will allow them to exchange ideas and results. Furthermore, expectations of different communities can be merged only by running activities like common projects and experiences.

The Transactions on Computational Systems Biology could be a good forum to help life scientists and computer scientists to discuss together their common goals.

This first issue is made up of contributions by members of the Editorial Board to provide a smooth start-up of the journal. The first paper, by Gómez et al., surveys the new methods needed for acquiring data suitable to enable simulation of simple cellular systems. Then Shenhav et al. discuss how system biology can be of help also in studying very early organisms in the time evolution scale. The third contribution is by Roux-Rouquié and Soto and shows how useful is the notion of model and metamodel in the systemic approach to the study of biological systems. Feytmans et al. investigate the relationships between the complexity of a genome and the functional complexity arising from it. The next paper moves inside computer science. Priami and Quaglia show how calculi for describing concurrent systems can be used to model biological systems as well. The last invited contribution is by Uhrmacher et al. and copes with the problem of multilevel and multiscale simulation. Finally, Zobeley et al., in the regular paper of this issue, present a new complexity reduction method which is time-dependent and suited not only for steady states, but for all possible dynamics of a biochemical system.

Trento, January 10, 2005 Corrado Priami

LNCS Transactions on Computational Systems Biology – Editorial Board

Lecture Notes in Bioinformatics

Vol. 3388: J. Lagergren (Ed.), Comparative Genomics. VII, 133 pages. 2005.

Vol. 3380: C. Priami, Transactions on Computational Systems Biology I. IX, 111 pages. 2005.

Vol. 3370: A. Konagaya, K. Satou (Eds.), Grid Computing in Life Science. X, 188 pages. 2005.

Vol. 3318: E. Eskin, C. Workman (Eds.), Regulatory Genomics. VII, 115 pages. 2005.

Vol. 3240: I. Jonassen, J. Kim (Eds.), Algorithms in Bioinformatics. IX, 476 pages. 2004.

Vol. 3082: V. Danos, V. Schachter (Eds.), Computational Methods in Systems Biology. IX, 280 pages. 2005.

Vol. 2994: E. Rahm (Ed.), Data Integration in the Life Sciences. X, 221 pages. 2004.

Vol. 2983: S. Istrail, M.S. Waterman, A. Clark (Eds.), Computational Methods for SNPs and Haplotype Inference. IX, 153 pages. 2004.

Vol. 2812: G. Benson, R.D.M. Page (Eds.), Algorithms in Bioinformatics. X, 528 pages. 2003.

Vol. 2666: C. Guerra, S. Istrail (Eds.), Mathematical Methods for Protein Structure Analysis and Design. XI, 157 pages. 2003.

Table of Contents

Accessible Protein Interaction Data for Network Modeling. Structure of the Information and Available Repositories

Manuel Gómez[1], Ramón Alonso-Allende[2], Florencio Pazos[3], Osvaldo Graña[4], David Juan[4], and Alfonso Valencia[4]

[1] Centro de Astrobiología (CSIC/INTA),
Instituto Nacional de Técnica Aeroespacial,
Ctra de Torrejón a Ajalvir, km 4,
28850 Torrejón de Ardoz, Madrid, Spain
[2] Bioalma, Ronda de Poniente, 4 - 2nd floor, Unit C-D,
28760 Tres Cantos, Madrid, Spain
[3] Structural Bioinformatics Group,
Biochemistry Building,
Department of Biological Sciences,
Imperial College, London SW7 2AZ, U.K
[4] Protein Design Group,
National Center for Biotechnology (C.N.B. - C.S.I.C.),
Cantoblanco, E-28049 Madrid, Spain
valencia@cnb.uam.es

Abstract. In recent years there has been an incredible explosion of computational studies of molecular biology systems, particularly those related to the analysis of the structure and organization of molecular networks, as the initial steps toward the possible simulation of the behavior of simple cellular systems. Needless to say, this task will not be possible without the availability of a new class of data derived from experimental proteomics. Large-scale application of the yeast two-hybrid system, affinity purification (TAPs-MS), and other methodologies are for the first time providing overviews of complete protein interaction networks. Interestingly a number of computational methods are also contributing substantially to the identification of protein interactions, by comparing genome organization and evolution. Other disciplines, such as structural biology and computational structural biology, are complementing the information on interaction networks by providing detailed molecular descriptions of the corresponding complexes, which will become essential for the direct manipulation of the networks using theoretical or experimental methods. The storage, manipulation and visualization of the huge volumes of information about protein interactions and networks pose similar problems, irrespective of the source of the information: experimental or computational. In this sense, a number of competing systems and emerging standards have appeared in parallel with the publication of the data. In this review, we will provide an overview of the main experimental, high-throughput methods for the study of protein interactions, the parallel developments of computational methods for the prediction of protein interactions based on genome and sequence information, and the development of databases and standards that facilitate the analysis of all this information.

C. Priami et al. (Eds.): Transactions on Computational Systems Biology, LNBI 3380, pp. 1 – 13, 2005.

1 Introduction

Proteins are involved in key cellular processes, including signal transduction, metabolism, cellular architecture and information transfer. To carry out these functions, proteins interact to form complexes of varying nature and stability, from stable interactions of structural proteins to transient contacts modulated by post-translational modifications, as is typical of signaling proteins.

During the last few years, proteomics has produced spectacular advances in the description of these complexes, utilizing high-throughput techniques such as systematic yeast 2-hybrid approaches [1-4], Tandem Affinity purification followed by Mass Spectrometry resolution of the isolated complexes [5], and various combinations of information obtained from peptide libraries [6, 7]. Other techniques, such as chromatin immunoprecipitation (ChIP), have systematically addressed the relationship between transcription factors and their specific DNA binding sites [8, 9]. Nevertheless, establishing the complete structure of the complexes and protein interactions in a living cell, including the modulation of the interactions in different cellular states (temporal) and compartments (spatial), is a formidably complex problem.

Despite its limited size, the public release of the first set of proteomic data has produced an avalanche of theoretical studies on the organization of protein interaction networks, the identification of the basic control and interaction motifs, and the comparison to other non-biological networks [10-18].

At the formal level, the structure of metabolic and protein interaction networks has been fitted to power law distributions similar to those of many other biological and non-biological systems [19, 20]. As in these other systems, the implication is that the protein interaction networks are in a meta-stable situation (or critical state), which makes it impossible to predict the future development of the network and the fate of individual interactions. Considerable effort has also been put into the search for well-defined regions of the interaction network associated with defined biological properties, such as metabolic pathways with distinctive patterns of interactions [15, 21-24].

Here we review the sources of information available for protein interaction data, their organization in databases, and the potential of computational biology methods to complement the experimental information by inferring new interactions. Clearly, the availability of large-scale, well organized interaction data with the proper quality controls is essential for the success of theoretical studies of the properties of the molecular systems.

2 Large-Scale Studies of Protein Complexes: The Proteomes

2.1 Experimental Methods for the Large-Scale Detection of Protein Interactions

Several experimental methods are being applied for the large-scale detection of protein interactions. Some of these involve the implementation of standard techniques to study protein-protein interactions. One of the methods most often used is the yeast two-hybrid system (Y2H) [25, 26], based on the modular properties of the Gal4

protein of the yeast S. cerevisiae, as well as its modifications for application to membrane proteins [27]. A similar approach is based on beta-lactamase activity recovery [28]. Genome-wide studies involving variations of the Y2H protocol have been carried out in yeast, H. pylori, C. elegans and Drosophila [1-5, 29].

Ho et al., applied ultra-sensitive mass spectrometry to identify protein complexes in S. cerevisiae, covering 25% of the yeast proteome [30]. Tandem-affinity purification (TAP) and mass spectrometry was used by Gavin et al. to characterize multi-protein complexes in S. cerevisiae [5]. Yeast protein chips and microarrays have also been used to screen protein-protein interactions and protein-drug interactions [31]. Tong et al. applied a combination of computational prediction of interactions from phage-display ligand consensus sequences with large-scale two-hybrid physical interaction tests, to identify interaction partners of yeast SH3 domains [7].

Large-scale proteomics also implies some limitations, and the introduction of certain artefacts, such as those produced by the presence of promiscuous proteins with an artifactual preference to interact with many other proteins in Y2H assays or the over-representation of small proteins in complex purification strategies [32-36]. As in other high-throughput applications (e.g. DNA arrays), accuracy in the determination of individual properties is sacrificed in order to gain insight into the global properties of the system [37].

2.2 Extrapolating Experimental Information to Build Interaction Networks of Related Species

A number of attempts have been made to extrapolate the information on protein interactions obtained from model systems (S. cerevisae, C. elegans, H. pylori, D. melanogaster) to other genomes. In general, inferences have been made by assuming that orthologous sequences will participate in similar interactions. For example, the experimental interactions determined for H. pylori were extrapolated to E. coli by combining sequence similarity searches with a clustering strategy, based on interaction patterns and interaction domain information [38]. Lappe et al. developed an integration system to combine, compare and analyze interaction data from different sources and different organisms at a single level of abstraction [39]. Matthews et al. proposed a method to search for 'interologs' (potentially conserved interactions) in C. elegans using experimentally identified interacting protein partners of S. cerevisae [40-43].

These studies are very interesting, and certainly correspond to the most-simple assumption of conservation of interactions across different species. Nevertheless, the risk of extrapolating too far is considerable, even more so given that the principle of conservation of interactions across large evolutionary time has yet to be demonstrated and the combinatorial possibilities of protein domains complicates the situation significantly.

An interesting exploration of this problem has been published by Aloy and Russell [44] in which they calculated the degree of conservation of the interaction regions for pairs of proteins with different degrees of similarity. The conclusion of this study was that similar interaction sites can be predicted for proteins with sequence similarities as low as 30-40 %, even if the noise of the system is considerable. It is important to bear in mind that this study only implies that proteins that do interact tend to do so using

similar regions, and not that similar proteins will necessarily interact (see below for a discussion of the problem of predicting interaction specificity).

3 Computational Methods for the Prediction of Interaction Partners

A number of computational methods have recently appeared that use sequence information to predict physical or functional interactions between proteins. Five of them are described in Box 1 [45, 46], although others are likely to appear.

The possibility of using sequence and evolutionary information to identify potential interaction partners brings additional opportunities to enrich the collection of interactions available for modeling studies. However, a definitive evaluation of these methods is still incomplete since the collections of experimental data on interacting proteins that can be used as controls have their own limitations (see the section on experimental methods above) and the overlap between the sets of predicted or experimental interactions is currently limited. Nevertheless, taking all these limitations into account, the increasing availability of genomic sequence information and the improvement of the methods still makes it likely that computational methods for predicting protein-protein interactions could achieve coverage and accuracies similar to those of the high-throughput experimental methods [47, 48].

Not surprisingly, interaction networks predicted by the various experimental and computational methods that are based on similar principles tend to have similar organizations [17].

3.1 Methods Based on Domain Composition

An alternative to the prediction of functional relationships between protein interactions is the study of the statistical association between proteins that share domains. The assumption in this case is that proteins that share a given domain will be functionally related by virtue of having this domain. Given the large number of multidomain proteins found in eukaryotes, it is easy to see that such a network will be highly complex and extremely dense. One approach attempts to elucidate which domains participate more often in protein interactions by considering the pairs of interacting yeast proteins recorded in the MIPS, MYGD and DIP databases, and the sequence domains included in the InterPro Database [49]. Another approach considers proteins as collections of conserved domains, where each domain is responsible for a specific interaction with another and a Markov chain Monte Carlo approach is used for the prediction of posterior probabilities of interaction between sets of proteins [50, 51].

3.2 Hybrid Methods Based on Sequence and Structure. Extrapolating from Interaction Partners to Interacting Regions

In order to manipulate molecular systems, by simulation or employing experimental methods, it is important to have information available not only about the general interaction networks, but also the details of the specific interaction at a molecular level. For example, the experimental manipulation of a signaling pathway with point

mutations requires specific knowledge of the amino acid residues involved in the interactions. In other words, it is important to develop methods for the discovery of interacting regions, as a way of channeling the capacity of molecular biology and simulation techniques for the exploration of interaction networks.

Computational methods for the prediction of interaction partners based on genome comparisons (phylogenetic profiles, conservation of gene neighborhood and gene fusion detection; see inset) do not provide information about the molecular details; the predictions remain at the level of functional relationships between sequences. In contrast, the predictions of the other two methods described here (mirror-trees and in-silico-two-hybrid) can be translated at the residue level for particular proteins.

Structural biology is also contributing substantially to the study of protein complexes, and perhaps the most important milestone in this area has been the determination of the structure of the ribosome [52]. Generally speaking, information about the structure of proteins is an essential component of the study of biological systems. From this type of experimental information we have learned about stable and transient protein complexes, about their interaction surfaces, and, to some extent, about the specificity of their interactions. A very interesting new avenue has been recently open by Aloy et al. [53] with the combination of experimental structure, protein models, and biochemical information to build the structure of new complexes whose general shape was solved by systematic electron microscopy studies of protein complexes purified by TAPs-MS.

From a computational point of view, major advances have been in the development of programs for the prediction of the structure of protein complexes (docking programs, [54, 55]), and a number of sequence-related analysis systems for the prediction of potential interaction regions.[56] In the near future, interesting progress is expected in the prediction of interaction regions by combining structural and sequence information.

Beyond the prediction of complex structure for interacting proteins of known structure, we still have to face the problem of distinguishing between potentially interacting proteins, e.g. all the pairs of proteins belonging to two protein families, versus the few protein pairs that are actually interacting. The specificity of those interactions is essential for the function of cellular systems in which members of the same protein family, using the same basic architecture, are able to trigger different signaling pathways. It is conceivable that a combination of protein modeling techniques and sequence information analysis will contribute to the search for the molecular basis of protein-protein recognition specificity. A few methods have been developed to this end. These methods make use of residue pair potentials obtained from interacting surfaces of known complexes. The information is then used to assess the extent to which the homologues of two interacting proteins of known structure will interact [57, 58]. Lu et al. have extended their protein structure prediction method to the prediction of the stability of protein complexes (Multiprospector). In this case, all combinations of protein sequences are tested for their compatibility in the framework of known protein complexes. The rationale is that proteins that will naturally form complexes will be more stable when associated with their partners than in isolation [59, 60]. The application of this method to complete genomes shows an impressive capacity for predicting potential interactions and an accuracy similar to other prediction methods [61]. Our group has studied the problem of molecular

specificity in various systems in which computational predictions have allowed us to manipulate the molecular basis of specific recognition in protein interactions [62-66]. However, in some cases accurate prediction of interactions is not possible due to the complexity of the conformational changes in the interaction surfaces.

4 Organization of the Information on Interactions in Specific Databases

In recent years, high-throughput methods have made molecular biology a data-intensive discipline. These data have to be stored in a structured way for data retrieval and analysis. A number of protein interaction results have been stored in this manner and made accessible via web services (see Table 1). All of these projects are still in an initial phase, which explains the current lack of differentiating characteristics that in the long run will determine their utility and survival in competition with other initiatives.

The Human Proteome Organization (HUPO) has launched an effort to establish standards for interaction databases that would be acceptable for all existing projects. These standards contain the minimum sufficient information to describe interactions, with the intention of facilitating information exchange between interaction databases. The consortium behind these initiatives has already designed the basic layer (XML) for the exchange, and a technical vocabulary for the description of the many experimental and theoretical techniques that produce data on protein interactions. Similar initiatives are taking place in related areas such as metabolic pathway databases[67]. The main databases of this kind have been running for years EMP [68, 69], WIT [70], KEGG [71], EcoCyc [72], and new ones are still appearing (aMAZE) [73, 74]. They are designed for storing information on enzymes, biochemical reactions and small molecules, and in some cases, the corresponding kinetic parameters. There are initiatives to create compatible standards between metabolic databases (see for example BioPAX-http://www.biopax.org/), which in the future may include protein interaction databases.

Alongside the data standardization structure, other projects have focused on a solution to another major database problem: data distribution. Many institutes and labs have relevant scientific information that is accessible through static web interfaces that are rarely visited. New technologies are now arising that are able to make all these data accessible through a single interface that can retrieve the information from its main source. An example of this technology is the PLANET project (see http://eu-plant-genome.net), where different data repositories are being made accessible through a single interface thanks to BioMoby technology [75].

The internet has offered a fast channel for information interchange. This has been particularly the case for the development of computational biology. Massive data exchange operations have made data reliability a major concern. Error propagation has proved to be a concern in areas with database annotation, making the link between annotation and the underlying experimental information an important issue. This need has increased the efforts in text mining research to recover the links between protein interaction databases and the corresponding sentences in the literature. During the last few years the technology in this area has developed rapidly [76-79]. Nevertheless, key

Table 1. Main databases on protein-protein interactions

Database	Site and Description
DIP [80-82]	Stores experimentally determined interactions between proteins. Currently, it includes 18,488 interactions for 7134 proteins in 104 organisms. http://dip.doe-mbi.ucla.edu/
MINT [98]	Designed to store functional interactions between biological molecules (proteins, RNA, DNA). It is now focusing on experimentally-verified direct and indirect protein-protein interactions. http://cbm.bio.uniroma2.it/mint/
BIND [99]	Contains full descriptions of interactions, molecular complexes and pathways http://www.bind.ca/
MIPS [100]	Large collection of diverse types of interactions. Commonly used as equivalent to 'hand-curated' sets of interactions. http://www.mips.biochem.mpg.de/
PathCalling Yeast Interaction Database [1]	Identifies protein-protein interactions on a genome-wide scale for functional assignment and drug target discovery http://portal.curagen.com/extpc/com.curagen.portal.servlet.Yeast
The GRID [101]	A database of genetic and physical interactions that contains interaction data from several sources, including MIPS and BIND http://biodata.mshri.on.ca/grid/servlet/Index
IntAct [67]	The project (funded by a European Commission grant, TEMBLOR) aims to represent and annotate protein-protein interactions, and to develop a public database of experimentally identified and predicted interactions. The database structure is designed to incorporate experimentally determined and predicted interactions, with special care in tracing the origin of the information. The interactions will be directly linked to original sentences in the literature describing them, for which text mining technology will be used. http//www.ebi.ac.uk/intact
STRING [46]	STRING is a database of known and predicted protein-protein interactions. http://string.embl.de/newstring_cgi/show_input_page.pl
HPID [42]	The human protein interaction database. Contains human protein interactions inferred by homology searches against experimental interaction data. http://www.hpid.org/
Prolinks [102]	A database of protein functional linkages derived from coevolution. Contains functional links predicted by several methods. http://169.232.137.207/cgi-dev/functionator/pronav
Predictome [103]	A database of putative functional links between proteins. Contains functional links establish by a variety of techniques, both experimental and computational. http://predictome.bu.edu/

problems remain in the field, such as the identification of protein and gene names. For example, in 2001 it was possible to link only 30% of the DIP database entries to the literature [80-82]. Only 20% of the missing links were explained by inaccuracies in the text mining system; the remaining 80% were produced because the protein names used in the database were not found in any of the available Medline entries, or because there was no information about the interactions in the literature. In the current status of the technology, the number of synonyms has grown, as well as the number of technical possibilities for detecting interactions[79]. Thus, this technology is maturing fast and may soon be able to facilitate the tasks of annotating databases, and to keep direct pointers between the interactions and the literature. (Very recently a collaborative effort has been launched to assess technologies in this area, see http://www.pdg.cnb.uam.es/BioLink).

5 Concluding Remarks

Genomic sequencing, proteome characterization and structural genomics projects are providing a wealth of information about genes and proteins. Proteomics now offers the possibility of entering a new dimension of understanding, directly related to the organization of the basic components in protein networks and complexes. The experimental and computational approaches published in the last five years have provided the first wide ranging view of the properties, organization, evolution and complexity of protein interaction networks. Computational Biology is contributing to this collective effort with, firstly, new methods to identify protein interaction partners on a large scale, and secondly with new approaches able to provide detailed descriptions, and associated predictions, of protein interaction sites.

It is important to bear in mind that the characterization of protein interaction networks is only one initial step towards the understanding of cellular systems; a step for which high-throughput proteomics, bioinformatics and computational biology are inherently associated with the success of Computational Systems Biology.

Acknowledgements

This study was funded by the EC project TEMBLOR (EU grant QLRT-2001-00015). MJG is recipient of an I3P contract from the Spanish Research Council (CSIC).

Boxes

Box 1. Computational Methods for the Prediction of Interaction Partners

Phylogenetic profiles. This method is based on the identification of genes that have the same pattern of presence/absence in a number of genomes. A group of genes with the same phylogenetic profile is assumed to encode proteins that are functionally related (for example, they may be part of the same metabolic pathway) and that may or may not interact physically. The drawback of the method is that it can only be applied to complete genomes [83, 84].

Conservation of gene neighborhood. Especially in prokaryotes, the neighborhood of a gene has a tendency to be conserved, both in terms of identity and order of the genes. This is partly related to the fact that genes in prokaryotes are often organized in operons. Operons contain genes that need to be expressed in a coordinated fashion, usually because they are involved in related functions. The observed relationship between chromosome proximity and function [85] has been exploited to predict gene interactions, both in the physical and in the functional sense [86, 87].

Gene fusion. Two proteins, or protein domains, encoded by different genes are assumed to interact physically, or at least functionally, if in some species they are coded by a single gene, presumably originating from a gene fusion event [88, 89]. It has been shown that fusion events are particularly common in metabolic proteins [90].

Mirror trees. Interacting proteins are expected to co-evolve. Therefore, the corresponding phylogenetic trees should be more similar than those of non-interacting proteins. The first qualitative assessments of this concept were performed with the pairs composed of the insulin and their receptors [91], and dockerins and cohexins [92]. Later, linear correlation between the distance matrices used to construct the trees was proposed to measure tree similarity [93] and the approach was applied to large data sets [94]. Recently, a method based on this concept has been developed for predicting interaction specificity [95].

In silico two-hybrid. The co-evolution of interacting proteins can be studied by analysis of mutations in one of the partners that compensate mutations in the other. The detection of correlated mutations has been used to predict the tendency of pairs of residues to be in physical proximity [96]. This method has been applied to large data sets of proteins and domains [97].

References

1. Uetz, P., et al., A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature, 2000. 403(6770): p. 623-7.
2. Giot, L., et al., A protein interaction map of Drosophila melanogaster. Science, 2003. 302(5651): p. 1727-36.
3. Rain, J.C., et al., The protein-protein interaction map of Helicobacter pylori. Nature, 2001. 409(6817): p. 211-5.
4. Li, S., et al., A map of the interactome network of the metazoan C. elegans. Science, 2004. 303(5657): p. 540-3.
5. Gavin, A.C., et al., Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature, 2002. 415(6868): p. 141-7.
6. Landgraf, C., et al., Protein interaction networks by proteome Peptide scanning. PLoS Biol, 2004. 2(1): p. E14.
7. Tong, A.H., et al., A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. Science, 2002. 295(5553): p. 321-4.
8. Ren, B., et al., Genome-wide location and function of DNA binding proteins. Science, 2000. 290(5500): p. 2306-9.
9. Lee, T.I., et al., Transcriptional regulatory networks in Saccharomyces cerevisiae. Science, 2002. 298(5594): p. 799-804.
10. Milo, R., et al., Network motifs: simple building blocks of complex networks. Science, 2002. 298(5594): p. 824-7.