Protein Protein Interactions And Networks Identification Computer Analysis And

Protein Interactions And Networks Identification Computational Analysis And Prediction Computational Biology

The rapidly evolving field of protein science has now come to realize the ubiquity and importance of protein-protein

interactions. It had been known for some time that proteins may interact with each other to form functional complexes, but it was thought to be the property of only a handful of key proteins. However, with the

advent of high throughput proteomics to monitor protein-protein interactions at an organism level, we can now safely state that protein-protein interactions are the norm and not the exception.

Thus, protein function must be understood in the larger context of the various binding complexes that each protein may form with interacting partners at a given time in the life cycle of a cell.

Proteins are now seen as forming sophisticated interaction networks subject to remarkable regulation. The study of these interaction networks and regulatory mechanism, which I would like

to term "systems proteomics," is one of the thriving fields of proteomics. The bird-eye view that systems proteomics offers should not however mask the fact that proteins are each characterized by

a unique set of physical and chemical properties. In other words, no protein looks and behaves like another. This complicates enormously the design of high-throughput proteomics

methods. Unlike genes, which, by and large, display similar physico-chemical behaviors and thus can be easily used in a high throughput mode, proteins are not easily amenable to the same treatment. It is

thus important to remind researchers active in the proteomics field the fundamental basis of protein chemistry. This book attempts to bridge the two extreme ends of protein science: on one end,

systems proteomics, which describes, at a system level, the intricate connection network that proteins form in a cell, and on the other end, protein chemistry and biophysics, which describe the

molecular properties of individual proteins and the structural and thermodynamic basis of their interactions within the network. Bridging the two ends of the spectrum is bioinformatics

and computational chemistry. Large data sets created by systems proteomics need to be mined for meaningful information, methods need to be designed and implemented to improve

experimental designs, extract signal over noise, and reject artifacts, and predictive methods need to be worked out and put to the test. Computational chemistry faces similar

challenges. The prediction of binding thermodynamics of protein-protein interaction is still in its infancy. Proteins are large objects, and simplifying assumptions and shortcuts still need to be

applied to make simulations manageable, and this despite exponential progress in computer technology. Finally, the study of proteins impacts directly on human health. It is an obvious

statement to say that, for decades, enzymes, receptors, and key regulator proteins have been targeted for drug discovery. However, a recent and exciting development is the exploitation

of our knowledge of protein-protein interaction for the design of new pharmaceuticals. This presents particular challenges because protein-protein interfaces are generally shallow and interactions are

weak. However, progress is clearly being made and the book seeks to provide examples of successes in this area. This volume presents a timely and comprehensive overview of biological

networks at all organization levels in the spirit of the complex system approach. It discusses the transversal issues and fundamental principles as well as the overall structure, dynamics, and modeling of a

wide array of biological networks at the molecular, cellular, and population levels. Anchored in both empirical data and a strong theoretical background, the book therefore lends valuable

credence to the complex systems approach. Protein-protein interactions (PPI) are at the heart of the majority of cellular processes, and are frequently dysregulated or usurped in disease. Given

this central role, the inhibition of PPIs has been of significant interest as a means of treating a wide variety of diseases. However, there are inherent challenges in developing molecules

capable of disrupting the relatively featureless and large interfacial areas involved. Despite this, there have been a number of successes in this field in recent years using both traditional drug

discovery approaches and innovative, interdisciplinary strategies using novel chemical scaffolds. This book comprehensively covers the various aspects of PPI inhibition, encompassing

small molecules, peptidomimetics, cyclic peptides, stapled peptides and macrocycles. Illustrated throughout with successful case studies, this book provides a holistic, cutting-edge view of the subject area and

is ideal for chemical biologists and medicinal chemists interested in developing PPI inhibitors.

The term 'protein-protein interaction (PPI)' refers to the study of

associations between proteins as manifested through biochemical processes such as formation of structures, signal transduction, transport, and phosphorylation. PPI play an important role in

the study of biological processes. Many PPI have been discovered over the years and several databases have been created to store the information about these interactions. von Mering (2002)

states that about 80,000 interactions between yeast proteins are currently available from various high-throughput interaction detection methods. Determining PPI

using high-throughput methods is not only expensive and time-consuming, but also generates a high number of false positives and false negatives. Therefore, there is a need for

computational approaches that can help in the process of identifying real protein interactions. Several methods have been designed to address the task of predicting protein-protein

interactions using machine learning. Most of them use features extracted from protein sequences (e.g., amino acids composition) or associated with protein sequences directly (e.g., GO

annotation). Others use relational and structural features extracted from the PPI network, along with the features related to the protein sequence. When using the PPI network to design

features, several node and topological features can be extracted directly from the associated graph. In this thesis, important graph features of a protein interaction network that help

in predicting protein interactions are identified. Two previously published datasets are used in this study. A third dataset has been created by combining three PPI databases. Several classifiers

are applied on the graph attributes extracted from protein interaction networks of these three datasets. A detailed study has been performed in this present work to determine if

graph attributes extracted from a protein interaction network are more predictive than biological features of protein interactions. The results indicate that the performance

criteria (such as Sensitivity, Specificity and AUC score) improve when graph features are combined with biological features. Networks in Nature Understanding Cellular Function

Through the Analysis of Protein Interaction Networks Protein-protein Interactions and Networks Computational Methods for Discovering Functional Modules from

Protein Interaction Networks Intelligent Computing Theories and Application ***Proteins perform every function in a cell. With the advent of genome***

sequencing projects for different organisms, large amounts of DNA and protein sequence data is available, whereas their biological function is still unknown

in the most of the cases. Predicting protein function is the most challenging problem in post-genomic era. Using sequence homology, phylogenetic profiles, gene

expression data, and function of unknown protein can be predicted. Recently, the large interaction networks constructed from high throughput

techniques like Yeast2Hybrid experiments are also used in prediction of protein function. As experimental techniques for detection and validation of protein interactions

are time consuming, there is a need for computational methods for this task. Based on the concept that a protein performs similar function like

its neighbor in protein interaction network, a method is proposed to predict protein function using protein-protein interaction data.This analysis should enlighten the

path for predicting unannotated protein function hence identifying diseases and inventing methods of it's cureness. This volume explores techniques that

study interactions between proteins in different species, and combines them with context-specific data, analysis of omics datasets, and assembles individual

interactions into higher-order semantic units, i.e., protein complexes and functional modules. The chapters in this book cover computational methods that solve diverse

tasks such as the prediction of functional protein-protein interactions; the alignment-based comparison of interaction networks by SANA; using the RaptorX-ComplexContact

webserver to predict inter-protein residue-residue contacts; the docking of alternative confirmations of proteins participating in binary interactions and the

visually-guided selection of a docking model using COZOID; the detection of novel functional units by KeyPathwayMiner and how PathClass can use such de novo pathways

to classify breast cancer subtypes. Written in the highly successful Methods in Molecular Biology series format, chapters include introductions

to their respective topics, lists of the necessary hardware- and software, step-by-step, readily reproducible computational protocols, and tips on

troubleshooting and avoiding known pitfalls. Cutting-edge and comprehensive, Protein-Protein Interaction Networks: Methods and Protocols is a valuable resource for

both novice and expert researchers who are interested in learning more about this evolving field. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics

combines elements of computer science, information technology, mathematics, statistics and biotechnology, providing the methodology and in silico solutions to

mine biological data and processes. The book covers Theory, Topics and Applications, with a special focus on Integrative –omics and Systems Biology. The

theoretical, methodological underpinnings of BCB, including phylogeny are covered, as are more current areas of focus, such as translational bioinformatics, cheminformatics

, and environmental informatics. Finally, Applications provide guidance for commonly asked questions. This major reference work spans basic and cutting-edge

methodologies authored by leaders in the field, providing an invaluable resource for students, scientists, professionals in research institutes, and a broad swath

of researchers in biotechnology and the biomedical and pharmaceutical industries. Brings together information from computer science, information technology,

mathematics, statistics and biotechnology Written and reviewed by leading experts in the field, providing a unique and authoritative resource Focuses on the main

theoretical and methodological concepts before expanding on specific topics and applications Includes interactive images, multimedia tools and crosslinking to

further resources and databases Recent studies have suggested that molecular interaction networks within cells could be decomposed into different subnetworks of molecules that

are involved in common biological processes. Such subnetworks are known as pathways, protein complexes or, in general, as functional modules. Many computational

methods have been developed to discover functional modules based on various hypotheses. For example, network motifs are abundant subnetworks in natural networks but

not random networks with similar global properties. Networks motifs have been utilized for comparing protein-protein interaction (PPI) networks of various organisms and

for assessing the random models in terms of capturing the global and local properties of PPI networks. In another example, subnetwork markers are connected

subnetworks from PPI networks in which member gene expressions correlate with labels of the samples. Such subnetwork markers could be used as predictors for

phenotype of the samples such as the disease statuses of the patients. In this dissertation, I first present novel computational methods for discovering

network motifs that use the confidence scores from protein interactions. Since there are many false positives and false negatives in the current binary PPI networks,

utilizing confidence scores could result in better network motifs. I have used this algorithm to compare PPI networks of prokaryotic unicellular, eukaryotic

unicellular and multicellular organisms. Later, I present two efficient and optimal computational approaches for identifying subnetwork markers. The first one

utilizes confidence scores from PPIs. And the second one is a randomized algorithm for discovering the subnetworks markers with the best predicting performance. I

have applied these algorithms to predict disease statuses of colon cancer and breast cancer patients and treatment outcomes of a combinatory therapy for a breast cancer

study. Protein-Protein Interaction Assays Protein-Protein Interactions in Human Disease Protein Interactions: Computational Methods, Analysis And Applications

Data Management of Protein Interaction Networks Biological Data Mining in Protein Interaction Networks

In this thesis we propose some new approaches to the study of

complex networks, and apply them to multiple domains, focusing in particular on protein-protein interaction networks. We begin by examining the roles of

individual proteins; specifically, the influential idea of 'date' and 'party' hubs. It was proposed that party hubs are local coordinators whereas date hubs are global

connectors. We show that the observations underlying this proposal appear to have been largely illusory, and that topological properties of hubs do not in general

correlate with interactor co-expression, thus undermining the primary basis for the categorisation. However, we find significant correlations between

interaction centrality and the functional similarity of the interacting proteins, indicating that it might be useful to conceive of roles for protein-protein interactions,

as opposed to individual proteins. The observation that examining just one or a few network properties can be misleading motivates us to attempt to develop a more holistic

methodology for network investigation. A wide variety of diagnostics of network structure exist, but studies typically employ only small, largely arbitrarily

selected subsets of these. Here we simultaneously investigate many networks using many diagnostics in a data-driven fashion, and demonstrate how this approach serves to

organise both networks and diagnostics, as well as to relate network structure to functionally relevant characteristics in a variety of settings. These include finding fast estimators

for the solution of hard graph problems, discovering evolutionarily significant aspects of metabolic networks, detecting structural constraints on

particular network types, and constructing summary statistics for efficient model-fitting to networks. We use the last mentioned to suggest that du plication-

divergence is a feasible mechanism for protein-protein interaction evolution, and that interactions may rewire faster in yeast than in larger genomes like human and fruit

fly. Our results help to illuminate protein-protein interaction networks in multiple ways, as well as providing some insight into st ructure-function relationships

in other types of networks. We believe the methodology outlined here can serve as a general-purpose, data-driven approach to aid in the understanding of networked systems.

The biological interactions of living organisms, and protein-protein interactions in particular, are astonishingly diverse. This comprehensive book provides a broad, thorough and multidiscip

linary coverage of its field. It integrates different approaches from bioinformatics, biochemistry, computational analysis and systems biology to offer the reader a comprehensive

global view of the diverse data on protein-protein interactions and protein interaction networks. Proteins participate in the majority of cellular processes. To

determine the function of a protein it is not sufficient to solely know its sequence, its structure in isolation, or how it works individually. Additionally, we need to know how the protein

interacts with other proteins in biological networks. This is because most of the proteins perform their main function through interactions. This thesis sets out to improve the

understanding of protein-protein interaction networks (PPINs). For this, we propose three approaches: (1) Studying measures and methods used in social and

complex networks. The methods, measures, and properties of social networks allow us to gain an understanding of PPINs via the comparison of different types of

network families. We studied models that describe social networks to see which models are useful in describing biological networks. We investigate the similarities

and differences in terms of the network community profile and centrality measures. (2) Studying PPINs and their role in evolution. We are interested in the

relationship of PPINs and the evolutionary changes between species. We investigate whether the centrality measures are correlated with the variability and similarity in orthologous

proteins. (3) Studying protein features that are important to evaluate, classify, and predict interactions. Interactions can be classified according to

different chara cteristics. One characteristic is the energy (that is the attraction or repulsion of the molecules) that occurs in interacting proteins. We identify which type of energy

values contributes better to predicting PPIs. We argue that the number of energetic features and their contribution to the interactions can be a key

factor in predicting transient and permanent interactions. Proteins are some of the building blocks of organisms. They usually perform their functions by interacting

with each other and forming protein complexes. A protein -protein interaction network is a graph that consists of proteins as vertices and their

interactions as edges. Protein-protein interaction network alignment is very important in identifying protein complexes and predicting protein functions. Many

algorithms based on graph theory have been developed to improve the accuracy of alignment, but due to the sparsity of protein-protein interactions, the result is far from

satisfactory. We propose to improve the network alignment through adding protein interactions to existing PPI networks. In order to assess the improvement, we

devise four groups of experiments and compare their results. The quality of PPI network alignment is assessed through the number of known protein complexes that

are discovered. Significant improvement is obtained, up to 70% additional complexes being discovered after adding interactions. Other consequences are observed as

well. Out of the two programs we compare, AlignMCL and MaWISH, the former performs significantly better whereas the latter is more stable. Further, adding predicted PPIs

is not as efficient as adding PPIs from existing databases. Finally, we show that smaller but more reliable sets of interactions perform better than larger PPI

sets.

Proteomics for Biological Discovery

Role of Protein-Protein Interactions in Metabolism: Genetics, Structure, Function, 2nd Edition

On the

Structure and Evolution of Protein Interaction Networks A Two Pass Neighborhood Approach Computational Analysis

New Approaches of Protein Function Prediction from Protein

Interaction Networks contains the critical aspects of PPI network based protein function prediction, including semantically assessing the reliability of PPI data, measuring the functional similarity between proteins, dynamically selecting prediction domains, predicting functions, and establishing

corresponding prediction frameworks. Functional annotation of proteins is vital to biological and clinical research and other applications due to the important roles proteins play in various biological processes. Although the functions of some proteins have been annotated via biological experiments,

there are still many proteins whose functions are yet to be annotated due to the limitations of existing methods and the high cost of experiments. To overcome experimental limitations, this book helps users understand the computational approaches that have been rapidly developed for protein function

prediction. Provides innovative approaches and new developments targeting key issues in protein function prediction Presents heuristic ideas for further research in this challenging area Knowledge of protein sequences has exploded, but knowledge of protein function is needed to make use of

sequence information, and this lags behind. A protein's function must be understood in context and part of this is the network of interactions between proteins. What are the relationships between protein function and the structure of the interaction network? In the first part of my thesis, I investigate the

functional relevance of clusters, or communities, of proteins in the yeast protein interaction network. Communities are candidates for biological modules. The work I present is the first to systematically investigate this structure at multiple scales in such networks. I develop novel tests to

assess whether communities are functionally homogeneous, and demonstrate that almost every protein is found in a functionally homogeneous community at some scale. The evolution of protein sequences IS well-studied, but comparatively little is known about the

evolution of protein function. Such knowledge is needed to un- derstand when it is appropriate to annotate newly sequenced proteins by transferring functional information from homologs-i.e. evolutionarily related proteins. In the sec- ond part of my thesis, I assess the success of transferring protein-

protein interactions across species and use this to estimate the rate at which interactions are lost in evolu- tion. At levels of sequence similarity associated with functional annotation transfer, I demonstrate that protein-protein interaction transfer is unreliable. The relevance of community structure for

understanding protein function and the low conservation of individual interactions, suggests a possible role for communities in the evolution of cellular function. I discuss this possibility in my conclusions.

A major challenge of post-genomic biology is understanding the complex networks of

interacting genes, proteins and small molecules that give rise to biological form and function. Advances in whole-genome approaches are now enabling us to characterize these networks systematically, using procedures such as the two-hybrid assay and protein co-immunoprecipitation to

screen for protein-protein interactions (PPI). Large protein networks are now available for many species like the baker's yeast, worm, fruit fly and the malaria parasite P. falciparum. These data also introduce a number of technical challenges: how to separate true protein-protein interactions

from false positives; how to annotate interactions with functional roles; and, ultimately, how to organize large-scale interaction data into models of cellular signaling and machinery. Further, as protein interactions form the backbone of cellular function, they can potentially be used

in conjunction with other large-scale data types to get more insights into the functioning of the cell. In this dissertation, I try to address some the above questions that arise during the analysis of protein networks. First, I describe a new method to assign confidence scores to protein interactions

derived from large-scale studies. Subsequently, I perform a benchmarking analysis to compare its performance with other existing methods. Next, I extend the network comparison algorithm, NetworkBLAST, to compare protein networks across multiple species. In particular, to elucidate cellular machinery on a

global scale, I performed a multiple comparison of the protein-protein interaction networks of m>C. elegans, D. melanogaster and S. cerevisiae. This comparison integrated protein interaction and sequence information to reveal 71 network regions that were conserved across all

three species and many exclusive to the metazoans. I then applied this technique to the analysis of the protein network of the malaria pathogen Plasmodium falciparum and showed that its patterns of interaction, like its genome sequence, set it apart from other species. Finally, I integrated the

PPI network data with expression Quantitative Loci (eQTL) data in yeast to efficiently interpret them. I present an efficient method, called 'eQTL Electrical Diagrams' (eQED), that integrates eQTLs with protein interaction networks by modeling the two data sets as a wiring diagram of current sources and

resistors. eQED achieved a 79% accuracy in recovering a reference set of regulator-target pairs in yeast, which is significantly higher performance than three competing methods. eQED also annotates 368 protein-protein interactions with their directionality of information flow with

an accuracy of approximately 75%. "The goal of this book is to disseminate research results and best practices from cross-disciplinary researchers and practitioners interested in, and working on bioinformatics, data mining, and proteomics"--Provided by publisher.

Dynamics, Evolution, and Modularity
Evidence of Probabilistic Behaviour in Protein Interaction Networks
Protein-Protein Interaction Networks
Methods and Protocols
Proteomics and Protein-Protein Interactions
The study of protein interactions from the networks point of view

has yielded new insights into systems biology [Bar03, MA03, RSM+02, WS98]. In particular, "network motifs" become apparent as a useful and systematic tool for describing and exploring networks [BP06, MKFV06, MSOI+02, SOMMA02, SV06]. Finding motifs has

involved either exact counting (e.g. [MSOI+02]) or subgraph sampling (e.g. [BP06, KIMA04a, MZW05]). In this thesis we develop an algorithm to count all instances of a particular subgraph, which can be used to query whether a given subgraph is a significant motif. This

method can be used to perform exact counting of network motifs faster and with less memory than previous methods, and can also be combined with subgraph sampling to find larger motifs than ever before -- we have found motifs with up to 15 nodes and explored subgraphs up to 20

nodes. Unlike previous methods, this method can also be used to explore motif clustering and can be combined with network alignment techniques [FNS+06, KSK+03]. We also present new methods of estimating parameters for models of biological network growth, and present a

new model based on these parameters and underlying binding domains. Finally, we propose an experiment to explore the effect of the whole genome duplication [KBL04] on the protein-protein interaction network of S. cerevisiae, allowing us to distinguish between cases of subfunctionalization

and neofunctionalization. This thesis describes the development of PINALOG, a protein interaction network alignment method, and its application to the area of protein function prediction and protein complex detection. Protein-protein interactions (PPI) play an

important role in the function of biological processes. Advances in high-throughput technology have produced a large amount of protein-protein interaction data, enabling analyses at the system level. Although protein-protein interaction networks (PPINs) vary between species, there

are components of them that perform similar biological functions and these are likely to be conserved across species. Comparison of the protein interaction networks from different species yields understanding of the evolution of species, as well as a means to predict protein

function and conserved components. An alignment method, PINALOG, has been developed which globally aligns the similar parts of the networks using information from protein sequences, protein functions and network topology in a seed-and-extend framework. The

results on human and yeast network alignment revealed conserved subnetworks that are components of similar biological processes such as the proteasome or transcription related processes. The alignments of several pairs of species confirm the superior performance of

PINALOG over commonly used methods such as Graemlin and IsoRank in terms of finding a large conserved network as well as detecting biologically meaningful mappings of the proteins in the two aligned species. The alignment method also suggested an approach to perform

protein complex prediction by knowledge transfer from one species to another. In addition the implications for function prediction of proteins in the "twilight" zone where there is little or no sequence similarity were explored. A web server for PINALOG was developed to

provide users access to the alignment method. In this thesis, we present two computational platforms for future biological research. The first, FNAC, is a flexible programmatic Framework for Network Analysis and Comparison that simplifies many common operations on

biological networks. As a demonstration of FNAC, we investigate the properties of several prominent protein function and protein-protein interaction networks. In doing so, we uncover evidence suggesting that a recently-developed technique for annotating proteins

may also have substantial value in the computational prediction of protein-protein interactions. Our second computational platform, the Coiled-Coil Database (CCDB), serves as a central and easily queryable repository for information about the coiled coil protein

structural motif in a variety of organisms. The first full survey of statistical, topological, data-mining, and ontology-based methods for analyzing protein-protein interaction networks. Protein Interaction Networks and Their Applications to Protein Characterization and Cancer Genes

Prediction
Communities and Homology in Protein-protein Interactions
Understanding Protein-protein Interaction Networks
Graph-based Analysis of Protein-protein Interaction Data Sets
Generation and Application of Human Protein-protein Interaction Networks

in Systems Biology

The chapters in this book are written by a team of well-reputed international researchers. The objective is to provide advanced and updated information related to protein-protein interactions. I hope the methods, resources and approaches described

here will enhance the available knowledge of the reader significantly. Genetic variations may change the structure and function of individual proteins as well as affect their interactions with other proteins and thereby impact metabolic processes dependent on protein-protein

interactions. For example, cytochrome P450 proteins, which metabolize a vast array of drugs, steroids and other xenobiotics, are dependent on interactions with redox and allosteric partner proteins for their localization, stability, (catalytic) function and

metabolic diversity (reactions). Genetic variations may impact such interactions by changing the splicing and/or amino acid sequence which in turn may impact protein topology, localization, post translational modifications and three dimensional structure. More

generally, research on single gene defects and their role in disease, as well as recent large scale sequencing studies suggest that a large number of genetic variations may contribute to disease not only by affecting gene function or expression but also by modulating

complex protein interaction networks. The aim of this research topic is to bring together researchers working in the area of drug, steroid and xenobiotic metabolism who are studying protein-protein interactions, to describe their recent advances in the field. We are aiming for a

comprehensive analysis of the subject from different approaches including genetics, proteomics, transcriptomics, structural biology, biochemistry and pharmacology. Of particular interest are papers dealing with translational research describing the role of novel genetic

variations altering protein-protein interaction. Authors may submit original articles, reviews and opinion or hypothesis papers dealing with the role of protein-protein interactions in health and disease. Potential topics include, but are not limited to: Role of protein-protein

interactions in xenobiotic metabolism by cytochrome P450s and other drug metabolism enzymes. Role of classical and novel interaction partners for cytochrome P450-dependent metabolism which may include interactions with redox partners,

interactions with other P450 enzymes to form P450 dimers/multimers, P450-UGT interactions and proteins involved in posttranslational modification of P450s. Effect of genetic variations (mutations and polymorphisms) on metabolism affected by protein-

protein interactions. Structural implications of mutations and polymorphisms on protein-protein interactions. Functional characterization of protein-protein interactions. Analysis of protein-protein interaction networks in health and disease.

Regulatory mechanisms governing metabolic processes based on protein-protein interactions. Experimental approaches for identification of new protein-protein interactions including changes caused by mutations and polymorphisms.

As the mysteries stored in our DNA have been more completely revealed, scientists have begun to face the extraordinary challenge of unraveling the int- cate network of protein□protein interactions established by that DNA fra- work. It is

increasingly clear that proteins continuously interact with one another in a highly regulated fashion to determine cell fate, such as proliferation, diff- entiation, or death. These protein□protein interactions enable and exert str- gent control over DNA replication, RNA

transcription, protein translation, macromolecular assembly and degradation, and signal transduction; essentially all cellular functions involve protein□protein interactions. Thus, protein□p- tein interactions are fundamental for normal physiology in

all organisms. Alt-ation of critical protein□protein interactions is thought to be involved in the development of many diseases, such as neurodegenerative disorders, cancers, and infectious diseases. Therefore, examination of when and how protein□p-tein interactions occur

and how they are controlled is essential for understa- ing diverse biological processes as well as for elucidating the molecular basis of diseases and identifying potential targets for therapeutic interventions. Over the years, many innovative biochemical,

biophysical, genetic, and computational approaches have been developed to detect and analyze p-tein□protein interactions. This multitude of techniques is mandated by the diversity of physical and chemical properties of proteins and the sensitivity of

protein□protein interactions to cellular conditions.

This two-volume set of LNCS 12463 and LNCS 12464 constitutes - in conjunction with the volume LNAI 12465 - the refereed proceedings of the 16th International Conference on Intelligent Computing,

ICIC 2020, held in Bari, Italy, in October 2020. The 162 full papers of the three proceedings volumes were carefully reviewed and selected from 457 submissions. The ICIC theme unifies the picture of contemporary intelligent computing techniques as an

integral concept that highlights the trends in advanced computational intelligence and bridges theoretical research with applications. The theme for this conference is "Advanced Intelligent Computing Methodologies and Applications." Papers

related to this theme are especially solicited, addressing theories, methodologies, and applications in science and technology.

Advances in Protein Chemistry and Structural Biology

16th International Conference, ICIC 2020, Bari, Italy,

October 2–5, 2020, Proceedings, Part II
Identification, Computer Analysis, and Prediction
Protein Function Prediction
Prediction from Protein Interaction Network
Protein-Protein Interaction Networks: Structures, Evolution, and Application to Drug Design

Data from high-throughput experiments of protein-protein interactions are commonly used to probe the nature of biological organization and extract functional

relationships between sets of proteins. What has not been appreciated is that the underlying mechanisms involved in assembling these networks may exhibit

considerable probabilistic behaviour. We find that the probability of an interaction between two proteins is generally proportional to the numerical product of their

individual interacting partners, or degrees. The degree-weighted behaviour is manifested throughout the protein-protein interaction networks

studied here, except for the high-degree, or hub, interaction areas. However, we find that the probabilities of interaction between the hubs are still high. Further evidence is

provided by path length analyses, which show that these hubs are separated by very few links. The results suggest that protein-protein interaction networks

incorporate probabilistic elements that lead to scale-rich hierarchical architectures. These observations seem to be at odds with a biol ogically-guided organization.

One interpretation of the findings is that we are witnessing the ability of proteins to indiscriminately bind rather than the protein-protein interactions that

are actually utilized by the cell in biological processes. Therefore, the topological study of a degree-weighted network requires a more refined methodology to

extract biological information about pathways, modules, or other inferred relationships among proteins. Resumen La importancia de comprender los procesos

biológicos ha estimulado el desarrollo de métodos para la detección de interacciones pr oteína-proteína. Esta tesis presenta PIANA (Protein Interactions And Network

Analysis), un programa informático para la integración y el análisis de redes de interacción proteicas. Además, describimos un método que identifica

motivos de interacción basándose en que las proteínas con parejas de interacción comunes tienden a interaccionar con esas parejas a través del

mismo motivo de interacción. Encontramos que las proteínas altamente conectadas (i.e., hubs) con múltiples motivos tienen mayor probabilidad de

ser esenciales para la viabilidad de la célula que los hubs con uno o dos motivos. Finalmente, presentamos un método que predice genes relacionados con cáncer mediante

la integración de redes de interacción proteicas, datos de expresión diferenciada y propiedades estructurales, funcionales y evolutivas. El valor de predicción

positiva es 71% con sensitividad del 1%, superando a otros métodos usados indepen dientemente. Abstract The importance of understanding cellular processes

prompted the development of experimental approaches that detect protein-protein interactions. Here, we describe a software platform called PIANA (Protein

Interactions And Network Analysis) that integrates interaction data from multiple sources and automates the analysis of protein interaction networks.

Moreover, we describe a method that delineates interacting motifs by relying on the observation that proteins with common interaction partners tend to

interact with these partners through the same interacting motif. We find that highly connected proteins (i.e., hubs) with multiple interacting motifs are more

likely to be essential for cellular viability than hubs with one or two interacting motifs. Furthermore, we present a method that predicts cancer genes by

integrating protein interaction networks, differential expression studies and structural, functiona. Written by recognized experts in the

study of proteins, Proteomics for Biological Discovery begins by discussing the emergence of proteomics from genome sequencing projects and a

summary of potential answers to be gained from proteome-level research. The tools of proteomics, from conventional to novel techniques, are

then dealt with in terms of underlying concepts, limitations and future directions. An invaluable source of information, this title also provides a

thorough overview of the current developments in post-translational modification studies, structural proteomics, biochemical proteomics,

microfabrication, applied proteomics, and bioinformatics relevant to proteomics. Presents a comprehensive and coherent review of the major issues faced in terms

of technology development, bioinformatics, strategic approaches, and applications Chapters offer a rigorous overview with summary of limitations, emerging

approaches, questions, and realistic future industry and basic science applications Discusses higher level integrative aspects, including technical challenges and

applications for drug discovery Accessible to the novice while providing experienced investigators essential information Proteomics for Biological Discovery is an

essential resource for students, postdoctoral fellows, and researchers across all fields of biomedical research, including biochemistry, protein

chemistry, molecular genetics, cell/de velopmental biology, and bioinformatics. Current PPI databases do not offer sophisticated querying interfaces and

especially do not integrate existing information about proteins. Current algorithms for PIN analysis use only topological information, while emerging approaches

attempt to exploit the biological knowledge related to proteins and kinds of interaction, e.g. protein function, localization, structure, described in

Gene Ontology or PDB. The book discusses technologies, standards and databases for, respectively, generating, representing and storing PPI data. It also describes main

algorithms and tools for the analysis, comparison and knowledge extraction from PINs. Moreover, some case studies and applications of PINs are also discussed.

Biological Networks Encyclopedia of Bioinformatics and Computational Biology Development and Application of a Computational Approach to

Align Protein Interaction Networks Network Tools Forthe Analysis and Prediction of Protein-protein Interactions Towards a Better Understanding of Protein-

Protein Interaction Networks

The identification and mapping of protein-protein interactions (PPIs) is a major goal in systems biology. Experimental data are currently produced in large

scale using a variety of high-throughput assays in yeast or mammalian systems. Analysis of these data using computational tools leads to the construction of large protein interaction networks, which help researchers identify novel

protein functions. However, our current view of protein interaction networks is still limited and there is an active field of research trying to further develop this concept to include important processes: the topology of

interactions and their changes in real time, the effects of competition for binding to the same protein region, PPI variation due to alternative splicing or post-translational modifications, etc. In particular, a clinically relevant topic for

development of the concept of protein interactions networks is the consideration of mutant isoforms, which may be responsible for a pathological condition. Mutations in proteins may result in loss of normal interactions

and appearance of novel abnormal interactions that may affect a protein's function and biological cycle. This Research Topic presents novel findings and recent achievements in the field of protein interaction networks

with a focus on disease. Authors describe methods for the identification and quantification of PPIs, the annotation and analysis of networks, considering PPIs and protein complexes formed by mutant proteins associated with

pathological conditions or genetic diseases. Protein-Protein Interactions in Human Disease, Volume 111, Part B, promotes further research and development in the protein interaction network in order to identify critical

proteins involved in the etiology of human diseases and locate new protein targets for drug development. Thus, this volume is of considerable interest to protein chemists, pharmacologists, cell biologists, immunologists,

structural biologists, computational biochemists and other researchers working in the field. In addition, these articles would be of great benefit to medical, biology and pharmacology students who specialize in this field. Describes

advances in the application of powerful techniques in studying and analyzing protein-protein interactions Ideal for a wide audience of researchers, specialists and students Written by authorities in their field Contains a

number of high quality illustrations, figures and tables that support the presented information Understanding the cell as a system has become one of the foremost challenges in the post-genomic era. As a result of advances in high-

throughput (HTP) methodologies, we have seen a rapid growth in new types of data at the whole-genome scale. Over the last decade, HTP experimental techniques such as yeast two-hybrid assays and co-affinity purification couple with mass

spectrometry have generated large amounts of data on protein-protein interactions (PPI) for many organisms. We focus on the sub-domain of systems biology related to understanding the interactions between proteins that ultimately drive

all cellular processes. Representing PPIs as a protein interaction network has proved to be a powerful tool for understanding PPIs at the systems level. In this representation, each node represents a protein and each

edge between two nodes represents a physical interaction between the corresponding two proteins. With this abstraction, we present algorithms for the prediction and analysis of such PPI networks as well as web servers and databases for

their public availability: 1. In many organisms, the coverage of experimental determined PPI data remains relatively noisy and limited. Given two protein sequences, we describe an algorithm, called Struct2Net, to

predict if two proteins physically interact, using insights from structural biology and logistic regression. Furthermore, we create a community-wide web-resource that predicts interactions between any protein

sequence pair and provides proteome-wide pre-computed PPI predictions for Homo sapiens, Drosophila melanogaster, and Saccharomyces cerevisiae. 2. Comparative analysis of PPI networks across organisms can

provide valuable insights into evolutionary conservation. We describe an algorithm, called IsoRank, for global alignment of multiple PPI networks. The algorithm first constructs an eigenvalue problem

that models the network and sequence similarity constraints. The solution of the problem describes a k partite graph that is further processed to find the alignments. Furthermore, we create a communitywide web

database, called IsoBase, that provides network alignments and orthology mappings for the most commonly studied eukaryotic model organisms: Homo sapiens, Mus musculus, Drosophila melanogaster,

Caenorhabditis elegans, and Saccharomyces cerevisiae. High-throughput methods for detecting protein-protein interactions (PPI) have recently gained popularity. These rapid advances in technology have

given researchers an initial global picture of protein interactions on a genomic scale. The usefulness of this understanding is, however, typically compromised by noisy data and intrinsic complexity of the biological system. In this

dissertation, we attempt to solve some problems in effectively analyzing the data. Firstly, since there are lots of false positives in experimentally detected interactions, we propose a novel topological measurement to

select reliable interactions from the noisy data. Our method is based on the small-world network property of the protein interaction network and generalizes purely local measures adopted previously. Based on our observation

that the true positive interactions in protein complexes and tightly coupled networks demonstrate dense interactions, we propose to measure the significance of two proteins' co-existence in a dense network as an index of interaction

reliability. Our topological measure also integrates the prior confidence of each data set. The experiments demonstrate that our measure can be used to identify reliable interactions and to predict potential interactions with

improved performance. Meanwhile, we discovered two additional properties: namely, the short alternative path property and the local clustering of network property of the protein interaction network, which are

generalizations of previously known protein interaction network properties. Secondly, we address the problem of effectively incorporating domain knowledge into the protein clustering process. Based on our analysis of the

relationship of network topology and biological relevance, we propose a novel semi-supervised clustering algorithm suitable for the noisy protein interaction network. We choose to estimate the pairwise similarity

between each protein pair and use this similarity as input to clustering algorithms. Therefore, it is not bounded to any specific clustering methods. We select topological features in the network and define a model to map these features

to pairwise similarities. The known protein annotations are used to train the model. Using this model, we can estimate the pairwise similarity between each pair of proteins. Finally, normal unsupervised

clustering algorithms can be applied using the similarity matrix. Since our similarity measure has already incorporated prior protein annotations, our algorithm can detect clusters with improved performance. Also,

the unsupervised clustering algorithms we adopt maintain the explorative nature and therefore are capable of detecting new protein functional groups. Thirdly, we investigate the problem of protein complex detection.

Protein complexes can be roughly considered as densely connected subgraphs in the network. The difficulties in this problem are caused by the fact that protein complexes may overlap with each other, i.e. containing shared

proteins, and the protein interaction network contains a lot of noise. To overcome these difficulties, we propose a novel subgraph quality measure, and based on the measure, we propose a novel "seed-refine" algorithm. Our

subgraph quality measure achieves two goals: (1) it provides a statistically meaningful combination of inside links, outside links and the size of the subgraph and, (2) it provides a statistically meaningful

combination of the quality contribution of each vertex in the subgraph. Our "seed-refine" algorithm consists of a two-layer seeding heuristic to find good seeds and a novel subgraph refinement method that controls the overlap between

subgraphs. Our algorithm allows to output overlapping subgraphs but methodologically makes it possible only when there is strong evidence to do so. Experiments confirm the effectiveness of our method.

Graph-based

Protein-protein Interaction Prediction in Saccharomyces Cerevisiae
ABC of Bioinformatics
New Approaches of Protein Function Prediction from Protein Interaction Networks
Protein Interaction

Networks Methods and Applications

"Protein-protein interactions represent a crucial source of information for the understanding of the biological

mechanisms of the cell. In order to be useful, high quality protein-protein interactions must be computationally extracted from the noisy datasets produced by

high-throughput experiments such as affinity purification. Even when filtered protein-protein interaction datasets are obtained, the task of analyzing the

network formed by these numerous interactions remains tremendous. Protein-protein interaction networks are large, intricate, and require computational

approaches to provide meaningful biological insights. The overall objective of this thesis is to explore algorithms assessing the quality of protein-protein

interactions and facilitating the analysis of their networks. This work is divided into four results: 1) a novel Bayesian approach to model contaminants

originating from affinity purifications, 2) a new method to identify and evaluate the quality of protein-protein interactions independently in different cell compartments,

3) an algorithm computing the statistical significance of clusterings of proteins sharing the same functional annotation in protein-protein interaction networks, and 4) a

computational tool performing sequence motif discovery in 5' untranslated regions as well as evaluating the clustering of such motifs in protein-protein interaction networks." --

At the outset of this project, computationally accessible human protein-protein interaction (PPI) data was limited to that available in a single hand-curated

database, the Database of Interacting Proteins (DIP). The sum total of these interactions was 621, a mere drop in the bucket of the potential interactome, considering

that the International Protein Index (IPI) database contains nearly 70,000 human proteins. Therefore, this research project involved three aims: (1) to create a

resource that expanded our knowledge of the human interactome by predicting novel interactions in silico; (2) to aid in the expansion of the human interactome

through experimental means; and (3) to gain new knowledge about the inner workings of the human cell through the networks of interacting proteins. To address the

first aim, we created the Online Predicted Human Interaction Database (OPHID) by transferring PPI from model organisms to human proteins, generating 23,889

predicted interactions. For the second aim, we collaborated on the analysis of a novel PPI detection technology, Luminescence-based Mammalian Interactome Mapping

(LUMIER), establishing protocols to evaluate and integrate experimental findings. This resulted in the elucidation of 947 PPI involved in TGFbeta signaling, and

1,484 PPI involved in Wnt signaling. Finally, for the third aim, through examining the evolutionary conservation of the human PPI network, we explored the extent to which

PPIs could be transferred between organisms. This revealed a preferential conservation of PPI involved in protein complexes when interologs are used to predict PPI networks.

While our work and that of many others has greatly expanded the knowledge of multiple interactomes, there is a considerable amount of knowledge yet to be gained

exploring the dynamic aspects of PPI networks. The Advances in Protein Chemistry and Structural Biology series is an essential resource for protein chemists. Each

volume brings forth new information about protocols and analysis of proteins, with each thematically organized volume guest edited by leading experts in a broad

range of protein-related topics. Provides cutting-edge developments in protein chemistry and structural biology Chapters are written by authorities in

their field Targeted to a wide audience of researchers, specialists, and students Protein'Protein Interactions Protein-protein Interaction Network Alignment Protein-protein

Interaction Confidence Assessment and Network Clustering Computational Analysis Inhibitors of Protein-Protein Interactions Biology, Chemistry, Bioinformatics,

and Drug Design