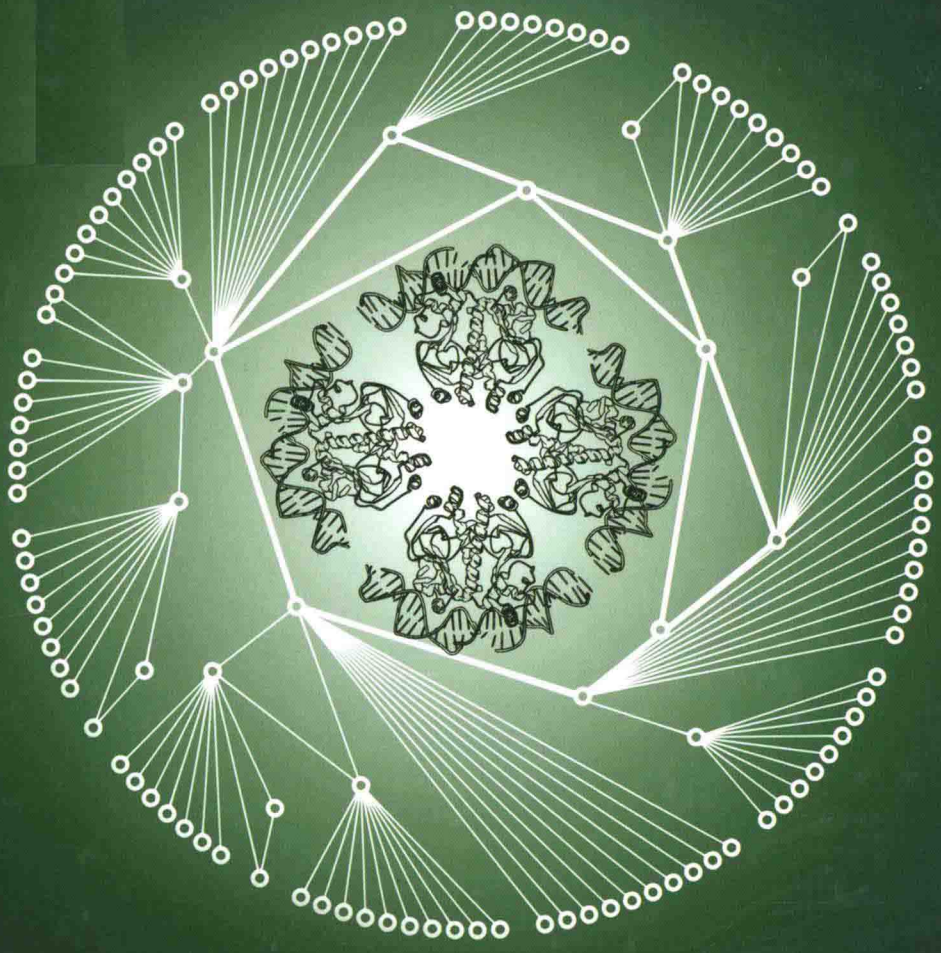

Networks in Cell Biology

EDITED BY
Mark Buchanan, Guido Caldarelli,
Paolo De Los Rios, Francesco Rao
and Michele Vendruscolo

CAMBRIDGE

Networks in Cell Biology

Edited by
M. BUCHANAN

G. CALDARELLI
CNR-INFM

P. DE LOS RIOS
Ecole Polytechnique Federale de Lausanne

F. RAO
University of Strasbourg

M. VENDRUSCOLO
University of Cambridge

常州大学山书馆
藏书章

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521882736

© Cambridge University Press 2010

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2010

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Modelling cell biology with networks / edited by M. Buchanan . . . [et al.].
p. cm.
ISBN 978-0-521-88273-6 (hardback)
1. Cellular control mechanisms. 2. Biological systems. 3. System analysis.
I. Buchanan, M. (Mark), 1961– II. Title.
QH604.M63 2010
571.6–dc22
2010000127

ISBN 978-0-521-88273-6 Hardback

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party Internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

NETWORKS IN CELL BIOLOGY

The science of complex biological networks is transforming research in areas ranging from evolutionary biology to medicine. This is the first book on the subject, providing a comprehensive introduction to complex network science and its biological applications.

With contributions from key leaders in both network theory and modern cell biology, this book discusses the network science that is increasingly foundational for systems biology and for the quantitative understanding of living systems. It surveys studies in the quantitative structure and dynamics of genetic regulatory networks, molecular networks underlying cellular metabolism, and other fundamental biological processes. The book balances empirical studies and theory to give a unified overview of this interdisciplinary science. It is a key introductory text for graduate students and researchers in physics, biology and biochemistry, and presents ideas and techniques from fields outside the reader's own area of specialization.

MARK BUCHANAN is a physicist and independent author. He writes a monthly column for the journal *Nature Physics*.

GUIDO CALDARELLI is Associate Professor in the Centre for Statistical Mechanics, CNR-INFM, Rome, and collaborates with the Research Centre 'E. Fermi', the Laboratory LINKALAB and the London Institute for Mathematical Sciences.

PAOLO DE LOS RIOS is a Professor in the Laboratory of Statistical Biophysics, Institute of Theoretical Physics, Ecole Polytechnique Federale de Lausanne (EPFL).

FRANCESCO RAO researches in the Laboratoire de Chimie Biophysique at the University of Strasbourg and at the Research Centre 'E. Fermi', Rome.

MICHELE VENDRUSCOLO is a Reader in Theoretical Chemical Biology in the Department of Chemistry at the University of Cambridge.

Contributors

S. Ahnert
TCM Group Cavendish Laboratory
University of Cambridge
Cambridge
CB3 0HB
UK

M. Madan Babu
MRC-Laboratory of Molecular Biology
University of Cambridge
Hills Road
Cambridge
CB2 0QH
UK

Matteo Brilli
Biométrie et Biologie Evolutive
UMR CNRS 5558
Université Lyon 1
43, Bvd du 11 Novembre
69622 Villeurbanne Cedex
France

Mark Buchanan

Gerard Cagney
Room F067
Conway Institute
University College Dublin
Belfield
Dublin 4
Ireland

G. Caldarelli
Centro SMC, INFM-CNR
Dipartimento di Fisica
Università 'Sapienza'
Piazzale Aldo Moro 5
Rome
00185
Italy

Elisa Calistri
Centro Interdipartimentale per lo studio
delle Dinamiche Complesse (CSDC)
Università di Firenze
Via G. Sansone 1
1-50022 Sesto Fiorentino
Firenze
Italy

试读结束：需要全本请在线购买： www.ertongbook.com

P. De Los Rios
Laboratoire de Biophysique
Statistique, ITP SB EPFL
Lausanne
1015
Switzerland

Thomas M. A. Fink
Institute M. Curie
5 Rue d'Ulm 75005
Paris
France

A. Gabrielli
Centro SMC, INFM-CNR
Dipartimento di Fisica
Università 'Sapienza'
Piazzale Aldo Moro 5
Rome
00185
Italy

D. Garlaschelli
Dipartimento di Fisica
Università di Siena
Via Roma 43 Siena
Italy

S. C. Janga
MRC-Laboratory of Molecular Biology
University of Cambridge
Hills Road
Cambridge
CB2 0QH
UK

Pietro Lió
Computer Laboratory
University of Cambridge
Cambridge
UK

Seesandra V. Rajagopala
J Craig Venter Institute (JCVI),
formerly The Institute of Genomic
Research (TIGR)
9704 Medical Center Drive
Rockville
MD 20850
USA

Francesco Rao
Centro Studi e Ricerche e Museo della
Fisica 'E. Fermi'
Compendio Viminale
Rome
Italy

Erzsébet Ravasz Regan
Beth Israel Deaconess Medical Center
Harvard University
Boston
USA

Gian Paolo Rossini
Dipartimento di Scienze Biomediche
Università di Modena e Reggio Emilia
Via G. Campi 287
Modena
I-41100
Italy

Daniel Segrè
Bioinformatics program
Boston University
44 Cummington St.
Boston, MA 02215
USA

Björn Titz
Crump Institute for Molecular Imaging
Graeber Lab
University of California, Los Angeles
570 Westwood Plaza
Los Angeles
CA 90095-1770
USA

Peter Uetz
The J Craig Venter Institute (JCVI),
formerly The Institute of Genomic
Research (TIGR)
9712 Medical Center Drive
Rockville
MD 20850
USA

Michele Vendruscolo
Department of Chemistry
University of Cambridge
Lensfield Road
Cambridge
CB2 1EW
UK

Contents

List of contributors *page* viii

Introduction 1
M. Buchanan, G. Caldarelli, P. De Los Rios, F. Rao and M. Vendruscolo

1 Network views of the cell 4
Paolo De Los Rios and Michele Vendruscolo
- 1.1 The network hypothesis 4
- 1.2 The central dogma and gene regulatory networks 5
- 1.3 Protein–protein interaction networks 7
- 1.4 Metabolic networks 9
- 1.5 Signaling networks 11
- 1.6 Networked networks and cell functionality 12
- 1.7 Concluding remarks 13

2 Transcriptional regulatory networks 14
Sarath Chandra Janga and M. Madan Babu
- 2.1 Introduction 14
- 2.2 Transcriptional regulation in prokaryotes and eukaryotes 19
- 2.3 Structure of transcriptional regulatory networks 20
- 2.4 Evolution of transcriptional regulatory networks 25
- 2.5 Dynamics of transcriptional regulatory networks 29
- 2.6 Conclusions 34

3 Transcription factors and gene regulatory networks 36
Matteo Brilli, Elisa Calistri and Pietro Lió
- 3.1 Introduction 36
- 3.2 Promoters' complexity/eukaryotic gene promoters 37
- 3.3 Transcription factors 41
- 3.4 Bioinformatics of regulatory networks 46

4 Experimental methods for protein interaction identification 53
Peter Uetz, Björn Titz, Seesandra V. Rajagopala and Gerard Cagney
4.1 Introduction 53
4.2 Protein complementation techniques 57
4.3 Affinity purification methods 62
4.4 Protein complex purification and mass spectrometry 65
4.5 Protein and peptide chips 73
4.6 Other methods for interaction detection and functional analysis 74
4.7 Quality of large-scale interaction data 75
4.8 Comparison of methods 77
4.9 Conclusions 80
5 Modeling protein interaction networks 83
Francesco Rao
5.1 Introduction 83
5.2 Scaling laws and network topology 85
5.3 Predicting protein interactions 87
5.4 Towards models at an atomic level of resolution 88
5.5 Concluding remarks 90
6 Dynamics and evolution of metabolic networks 93
Daniel Segrè
6.1 Introduction 93
6.2 Cellular metabolism and its regulation 93
6.3 Metabolism across disciplines 96
6.4 Dynamics of a metabolic system 98
6.5 Stoichiometric analysis 100
6.6 Constraint-based modeling: feasible states and optimality 103
6.7 Predicting genetic perturbations 108
6.8 Double perturbations and epistatic interactions 109
6.9 The ancient history of metabolism: from cell-scale to biosphere-scale 112
6.10 Conclusions 116
7 Hierarchical modularity in biological networks: the case of metabolic networks 117
Erzsébet Ravasz Regan
7.1 Introduction 117
7.2 Modularity and hubs in biological networks 119
7.3 Scaling of the clustering coefficient: a signature of hierarchy 121
7.4 Method for finding network modules 123
7.5 A case study: the *E. coli* metabolic network 126

7.6 Hierarchy, fractality and the small world of networks 131
7.7 Conclusions 134
8 Signalling networks 135
Gian Paolo Rossini
8.1 Introduction 135
8.2 Chemical signalling: many pathways following a few general themes 136
8.3 Cross-talks among signal transduction pathways 159
8.4 Signalling networks, system organization and modelling 166
8.5 Conclusions and outlook 169
Appendix A Complex networks: from local to global properties 170
D. Garlaschelli and G. Caldarelli
Appendix B Modelling the local structure of networks 188
D. Garlaschelli and G. Caldarelli
Appendix C Higher-order topological properties 210
S. Ahnert, T. Fink and G. Caldarelli
Appendix D Elementary mathematical concepts 219
A. Gabrielli and G. Caldarelli
References 235
Index 269

Introduction

M. BUCHANAN, G. CALDARELLI, P. DE LOS RIOS, F. RAO
AND M. VENDRUSCOLO

Biologists now have access to a virtually complete map of all the genes in the human genome, and in the genomes of many other species. They are aggressively assembling a similarly detailed knowledge of the proteome, the full collection of proteins encoded by those genes, and the transcriptome, the diverse set of mRNA molecules that serve as templates for protein manufacture. We increasingly know the "parts list" of molecular biology. Yet we still lack a deep understanding of how all these parts work together to support the complex and coherent activity of the living cell; how cells and organisms manage the concurrent tasks of production and re-production, signalling and regulation, in fluctuating and often hostile environments.

Building a more holistic understanding of cell biology is the aim of the new discipline of systems biology, which views the living cell as a network of interacting processes and gives concrete form to the vision of François Jacob, one of the pioneers in the study of genetic regulatory mechanisms, who spoke in the 1960s of the "logic of life." Put simply, systems biologists regard the cell as a vastly complex biological "circuit board," which orchestrates diverse components and modules to achieve robust, reliable and predictable operation. Systems biology suggests that the mechanisms of cell biology can be related to the information sciences, to ideas about information flow and processing in de-centralized networks.

This view, of course, has long been implicit in the study of cell signalling and other key pathways of molecular bio-chemistry. It has long been clear that such pathways do not act through simple sequential action in chain-like reaction paths, but as a rule exhibit a much richer dynamics involving multiple pathways working in parallel, with interactions passing between, and feeding backward as well as forward. Yet with rapidly advancing technology, and new theoretical tools coming from physics, engineering and mathematics, systems biology is beginning to reach beyond the recognition that such systems exist to elucidate specific quantitative

Networks in Cell Biology, ed. M. Buchanan, G. Caldarelli, P. De Los Rios, F. Rao and M. Vendruscolo.
Published by Cambridge University Press. © Cambridge University Press 2010.

试读结束：需要全本请在线购买： www.ertongbook.com

regularities in biological information flows, often involving operations such as feedback, synchronization, amplification and error correction that are familiar to engineers.

Significantly, the potential power of this perspective is being multiplied by an explosion of recent work in physics and mathematics, both theoretical and empirical, showing that many of the world's complex networks have hidden structural regularities of great importance. These networks – ranging from social networks and food webs to the Internet, and including genetic regulatory or metabolic networks within the cell – have a surprisingly universal character, and can be fruitfully described with a unified conceptual language. This work has infused science with new analytical measures such as betweenness centrality, network dimension, degree distribution and so on, which reflect local or global network properties. These measures have been found to have direct relevance to a network's stability or resilience, information processing efficiency, or dynamical richness, offering hope that the network perspective will prove invaluable for understanding networks in cellular biology.

Networks in Cell Biology aims to fill a yawning gap in the modern literature of systems biology. While a number of excellent texts at the graduate level cover recent developments in cell biology, and others offer timely introductions to recent advances in complex network science, no single volume yet focuses explicitly on advances in our understanding of cellular networks. The present book offers an up-to-date snapshot of such work, aiming for a balance between empirical studies and theory, which are natural complements in furthering knowledge in almost any area of modern biology. What do we know about the qualitative and quantitative structure and dynamics of genetic regulatory networks, or the network biochemistry underlying cellular metabolism? What do empirical studies tell us about the large-scale architecture of protein–protein interactions, and what are the key weaknesses and gaps in such data? What are the prospects for building realistic dynamical models for metabolic or regulatory processes, possibly with predictive capability?

This volume covers these and other topics of current research in systems biology, with an emphasis on recent advances in complex networks science, and on how the massive data now available in biology can help test and inform such theories in their application. With contributions from key leaders in both network theory and modern cell biology, this volume makes available in one publication the naturally diverse range of studies supporting a unified view of biological networks that is likely to be increasingly important in the future.

The contributions also illustrate a timeless truism of science – that scientific progress is not linear or predictable, and that it is often a new idea emerging from

an unexpected direction – a technique for gathering data of a new kind, or a small shift in perspective – that transforms and renews a field by framing old problems in new ways. The new idea in this case is an old one, the basic perspective of network science, but updated and vastly furthered by modern data and powerful concepts and techniques coming from statistical physics.

1

Network views of the cell

PAOLO DE LOS RIOS AND MICHELE VENDRUSCOLO

1.1 The network hypothesis

A cell is an enormously complex entity made up by myriad interacting molecular components that perform the biochemical reactions that maintain life. This book is about the network hypothesis, according to which it is possible to describe a cell through the set of interconnections between its component molecules. Hence, it becomes convenient to focus on these interactions rather than on the molecules themselves to describe the functioning of the cell.

The central dogma in molecular biology describes the way in which a cell processes the information required to produce the molecules necessary to sustain its existence and reproduction. It is also becoming increasingly clear, however, that in order to establish a more complete description of the manner in which a cell works we require a deeper understanding of the manner in which the sets of interconnections between these molecules are defining the identity of the cell itself. It is therefore important to investigate whether the genetic makeup of an organism does not only specify the rules for generating proteins, but also the way in which these proteins interact among themselves and with the other molecules in a cell.

Complex networks Networks are a way to represent an ensemble of objects together with their relations. Objects are described by means of vertices (sometimes also called nodes) and their relations by edges (sometimes also called links or connections) connecting them, which can be weighted to reflect their strength. A network is thus entirely characterized by the set of its connections, not by the way in which it is drawn. The actual distance between two vertices is given by the minimum number of edges that can be found between them. A key quantity to characterize networks is the degree: it is defined as the number of edges per vertex. Most networks present in nature have large fluctuations in the degree value. This feature has profound consequences both on the stability of the sytems represented by the network and on the dynamics of the processes defined on this structure.

Networks in Cell Biology, ed. M. Buchanan, G. Caldarelli, P. De Los Rios, F. Rao and M. Vendruscolo.
Published by Cambridge University Press. © Cambridge University Press 2010.

Networks provide a way to organize and regulate efficiently complex systems [12, 52, 88, 89, 98, 153, 436, 583, 636]. In an effective network different parts are linked by reducing at a minimum the number of the interconnections. This feat can be achieved at the cost of a fragility of the structure of connections. If one passage goes wrong, the whole structure falls down. A bit of redundancy is a good compromise for a fail-safe system without too much effort. A trade off between redundancy and cost is probably at the basis of the statistical properties of such objects and explains in part their ubiquitous presence in the various cell activities. From the point of view of the researcher a network is also a powerful method to represent the data in one object, and to enable the quantitative assessment of the fragility or robustness of the system. The first step is to establish the topology, the second is to establish the dynamics – that is how the topology changes with time. The study of these two aspects is at the heart of current research about biological networks and it constitutes the core of the following chapters in this book.

1.2 The central dogma and gene regulatory networks

The central dogma The central dogma, which was introduced in 1958 by F. Crick [131] and later restated in 1970 [132], defines the way in which information flows between DNA, RNA and proteins. At the most fundamental level, DNA passes information to itself (through replication) and to RNA (through transcription), and RNA passes it to proteins (through translation). We also now know, however, that additional information also follows other routes that taken together link DNA, RNA and proteins in a much interconnected network. This view of the central dogma defines our current core understanding of molecular biology.

The central dogma of molecular biology defines the manner in which the genetic information is exploited to generate proteins. In this process, DNA is first transcribed into messenger RNA (mRNA), which is then translated into proteins (Fig. 1.1a). This description, originally proposed by Francis Crick [131], is at the foundation of molecular biology.

Our understanding of the more subtle aspects and of the fundamental implications of the central dogma is becoming increasingly detailed. While initially the information was thought to flow in one direction from genes to proteins, it is now well established that there are much more articulate interactions between DNA, RNA and proteins. Gene transcription is a highly regulated process, in which a major role is played by transcription factors, which are proteins that bind

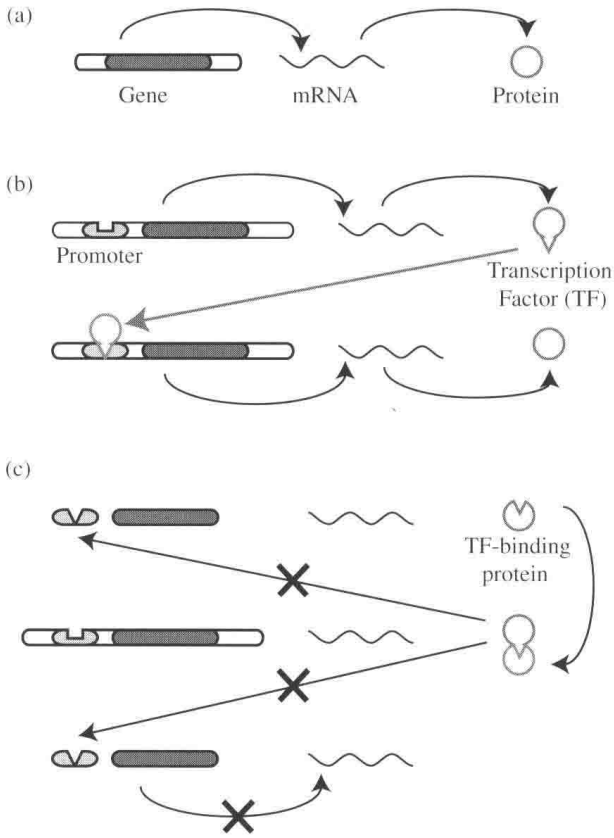

Fig. 1.1. The flow of information between genes and proteins is highly regulated in the cell by a network of interacting molecules. (a) In the simplest view, genes are transcribed into mRNA molecules, which are then translated into proteins. (b) The transcription of genes is controlled by transcription factors that bind to specific promoter regions that precede the genes themselves. (c) The activity of transcription factors can be regulated in a number of manners, including by binding with co-regulatory proteins that block their ability to bind to their corresponding promoters.

to specific promoter regions and control the transcription of the corresponding genes. As transcription factors are gene products themselves a backward flow of information from proteins to DNA is indeed present in the form of a gene-to-protein-to-gene control (Fig. 1.1b). For example, some genes encode proteins that bind to transcription factors, inactivating them. In this case the regulation implies a gene-to-protein-to-protein-to-gene control (Fig. 1.1c).

Furthermore, the cross-specificity between transcription factors and promoter regions is not extreme, allowing for some promiscuity. Indeed the same transcription factor can bind to several promoters and the same promoter can bind to several transcription factors. Moreover, promoter regions are often duplicated so

that several genes might have the same promoters and are regulated by the same transcription factors (Fig. 1.1c).

Our understanding of the mechanism of regulation of gene expression is rapidly increasing. It has been clarified that RNA molecules have a variety of roles in the regulation of protein expression. For example, the recently discovered *small interfering RNA* (siRNA) and *microRNA* (miRNA) [513] molecules bind to the mRNA products of specific genes and inhibit or even enhance their further translation into proteins. This is a gene-to-RNA-to-gene form of control adding further layers of complexity to the way in which gene expression is regulated. It is thus impossible to represent such a rich pattern of relations as simple unidirectional flows. Rather, a network representation is particularly well suited to capture the feedback and feedforward regulation mechanisms that modulate gene and protein expression. Moreover, the network of gene regulation is highly dynamic in order to respond to the changing environment of a cell and to the different requirements through the cell cycle, and should therefore be considered in a time-dependent manner. Chapters 2 and 3 provide a description of the modes of regulation of gene expression, of their actuators and of the theoretical methods in use to reconstruct gene regulatory networks.

1.3 Protein–protein interaction networks

Protein–protein interaction network A protein–protein interaction network is a graph whose vertices are proteins and edges represent interactions between them, including for example those required to form macromolecular complexes or to establish signaling processes. Given this rather general definition, it is easy to understand why protein–protein interaction networks are becoming a very popular way to represent a variety of functions into the cell. Because of the disparate nature of these interactions, a variety of experimental techniques have been developed to detect them. These methods can be divided into physical (e.g. concentration, immunoprecipitation), library-based (e.g. protein probing, two-hybrid systems) and genetic methods (e.g. synthetic lethal effects) [474].

It is increasingly clear that the majority of proteins do not carry out their functions in isolation but by interacting together in complex manners [506]. There are two main ways to look at protein–protein interaction networks – physical and logical.

Physical interaction networks define interacting protein pairs according to physico-chemical principles. The characterization of this type of network is promoting strong links between structural biology and systems biology, and it is

试读结束：需要全本请在线购买： www.ertongbook.com