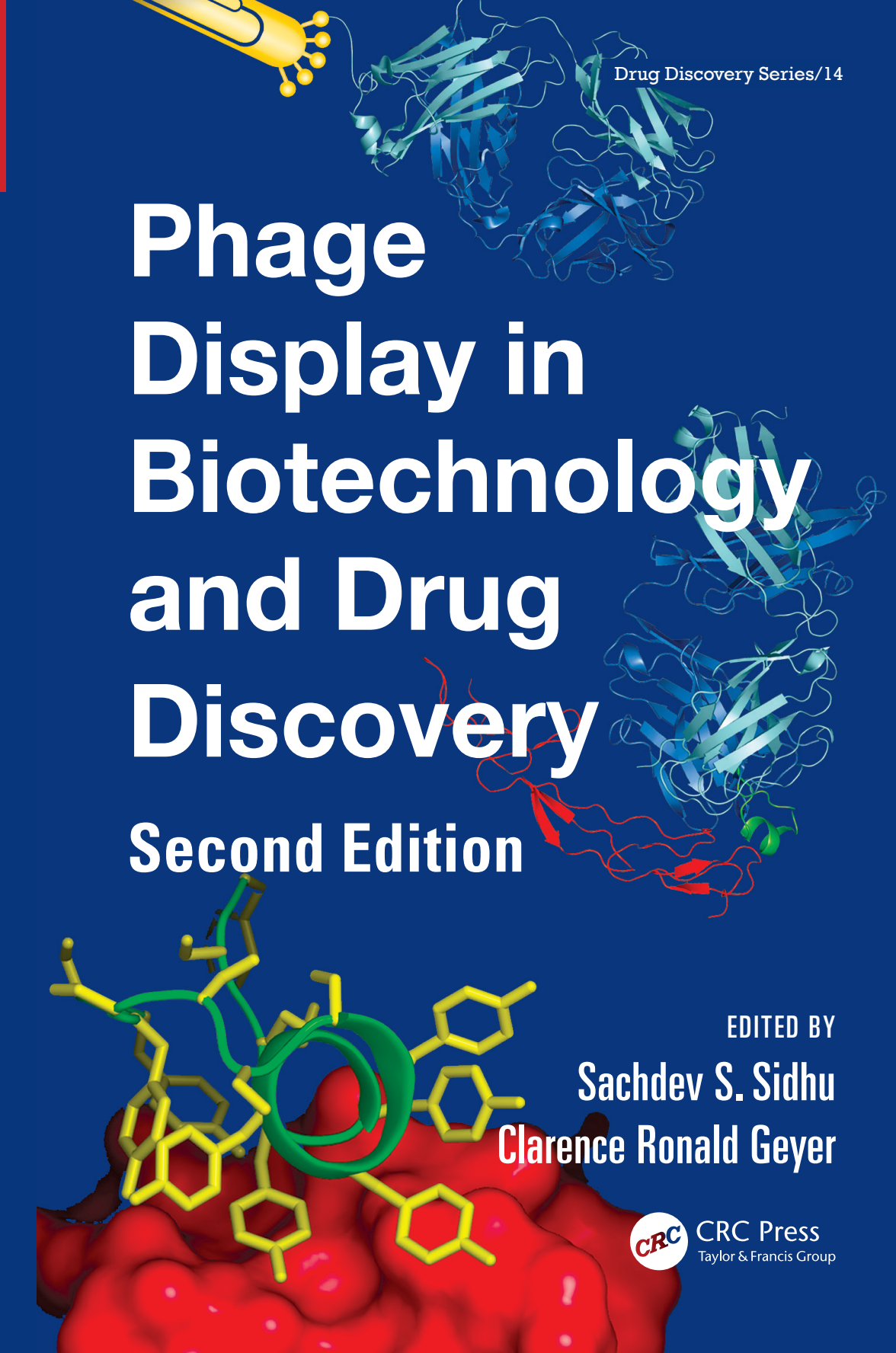
Drug Discovery Series/14
Phage Display in Biotechnology and Drug Discovery
Second Edition
EDITED BY
Sachdev S. Sidhu
Clarence Ronald Geyer
CRC Press
Taylor & Francis Group

Phage Display in Biotechnology and Drug Discovery

Drug Discovery Series

Series Editor
Daniel Levy

1. Virtual Screening in Drug Discovery, edited by Juan Alvarez and Brian Shoichet
2. Industrialization of Drug Discovery: From Target Selection Through Lead Optimization, edited by Jeffrey S. Handen, Ph.D.
3. Phage Display in Biotechnology and Drug Discovery, edited by Sachdev S. Sidhu
4. G Protein-Coupled Receptors in Drug Discovery, edited by Kenneth H. Lundstrom and Mark L. Chiu
5. Handbook of Assay Development in Drug Discovery, edited by Lisa K. Minor
6. *In Silico* Technologies in Drug Target Identification and Validation, edited by Darryl León and Scott Markel
7. Biochips as Pathways to Drug Discovery, edited by Andrew Carmen and Gary Hardiman
8. Functional Protein Microarrays in Drug Discovery, edited by Paul F. Predki
9. Functional Informatics in Drug Discovery, edited by Sergey Ilyin
10. Methods in Microarray Normalization, edited by Phillip Stafford
11. Microarray Innovations: Technology and Experimentation, edited by Gary Hardiman
12. Protein Discovery Technologies, edited by Renata Pasqualini and Wadih Arap
13. Pharmacoeconomics: From Theory to Practice, edited by Renée J. G. Arnold
14. Phage Display in Biotechnology and Drug Discovery, Second Edition, edited by Sachdev S. Sidhu, Clarence Ronald Geyer

Phage Display in Biotechnology and Drug Discovery

Second Edition

EDITED BY
Sachdev S. Sidhu
Clarence Ronald Geyer

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2015 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20150128

International Standard Book Number-13: 978-1-4398-3650-7 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Foreword

Science has always progressed by coupling insightful observations leading to testable hypotheses with innovative technologies that facilitate our ability to observe and test them. In the field of protein science, the technologies for protein display and in vitro selection have had an enormous impact on our ability to probe and manipulate protein functional properties.

The development of site-directed mutagenesis, which allowed one to systematically probe a gene sequence in the late 1970s, gave birth to the field of protein engineering in the early 1980s. Throughout the 1980s, most scientists in the protein engineering field would generate and purify one mutant protein at a time and characterize its functional properties. Some investigators had developed selections and screens that allowed one to test many variants simultaneously, but these tended to be highly specific for certain proteins (notably DNA binding proteins) and focused primarily on studying protein stability. Moreover, the selections were generally done in the context of a living cell, which limited the range of assays that could be performed. While replica plating screens were available to test variant proteins out of the cell, these tended to be quite labor intensive, thus limiting the number of variants that could be screened.

In 1985, George P. Smith published a paper showing that small peptides derived from EcoRI could be inserted into the gene III attachment protein in filamentous bacterial phage, which could then be captured using antibodies to the small peptide. This observation incubated several years, and then, in the late 1980s and early 1990s, other groups showed it was possible to display whole proteins on gene III that were folded and capable of binding their cognate ligands. Moreover, it was shown that by appropriate manipulation of the copy number on the phage, it was possible to select a range of binding affinities, from weak at a high copy number to strong at a low copy number. These selections could all be done in vitro and under a variety of selection conditions, limited only by binding to a support-bound ligand.

Throughout the 1990s up to today, huge improvements have been made to the display technology allowing massive increases in the library number (now routinely $>10^{10}$ variants per selection); recursive mutagenesis cycles allowing one to mutate as one selects; new display formats including other phage species, bacteria, yeast, and ribosomes; and automation to further simplify the process. As with any technology, there are limitations. For example, not all proteins can be readily displayed on phage, and expression effects can bias the outcome of the selection. Nonetheless, phage display has had a huge impact on probing, improving, and designing new functional properties into proteins and peptides including binding affinity, selectivity, catalysis, and chemical and thermal stability, among others. This book edited by Sachdev S. Sidhu provides an excellent review of the state-of-the-art in phage display technology now and in the near future.

James A. Wells
President and Chief Scientific Officer
Sunesis Pharmaceuticals
South San Francisco, California

Preface

Recent years have witnessed the sequencing of numerous genomes, including the all-important human genome itself. While genomic information offers considerable promise for drug discovery efforts, it must be remembered that we live in a protein world. The vast majority of biological processes are driven by proteins, and the full benefits of DNA databases will only be realized by the translation of genomic information into knowledge of protein function. Ultimately, drug discovery depends on the manipulation and modification of proteins, and thus, the genomic panacea comes with significant challenges for life scientists in the field of therapeutic biotechnology. Indeed, it has become clear that success in the modern era of biology will go to those who apply to protein analysis the high-throughput principles that made whole-genome sequencing a reality.

In this context, phage display is an established combinatorial technology that is likely to play an even greater role in the future of drug discovery. The power of the technology resides in its simplicity. Rapid molecular biology methods can be used to create vast libraries of proteins displayed on bacteriophage that also encapsulate the encoding DNA. Billions of different proteins can be screened en masse and individual protein sequences can be decoded rapidly from the cognate gene. In essence, the technology enables the engineering of proteins with simple molecular biology techniques that would otherwise only be applicable to DNA. In addition, the technology is very much suited to the methods currently used for high-throughput screening and thus can be readily adapted to the analysis of multiple targets and pathways.

This book comprises 19 chapters that provide a comprehensive view of the impact and promise of phage display in drug discovery and biotechnology. The chapters detail the theories, principles, and methods current in the field and demonstrate applications for peptide phage display, protein phage display, and the development of novel antibodies. The book as a whole is intended to give the reader an overview of the amazing breadth of the impact that phage display technology has had on the study of proteins in general and the development of protein therapeutics in particular. I hope that this work will serve as a comprehensive reference for researchers in the phage field and, perhaps more importantly, will serve to inspire newcomers to adapt the technology to their own needs in the ever-expanding world of therapeutic biology.

Sachdev S. Sidhu

Contributors

Sara K. Avrantinis
Department of Chemistry
University of California
Irvine, California

Jody D. Berry
Departments of Immunology and Medical Microbiology
University of Manitoba
Winnipeg, Manitoba, Canada

Julian Bertschinger
Institute of Pharmaceutical Sciences
Swiss Federal Institute of Technology
Zurich, Switzerland

Robert O. Carlson
Karyopharm Therapeutics
Natick, Massachusetts

Jacob Corn
Department of Protein Engineering
Genentech Inc.
South San Francisco, California

Reto Crameri
Swiss Institute of Allergy and Asthma Research (SIAF)
Davos, Switzerland

Mark S. Dennis
Department of Protein Engineering
Genentech, Inc.
South San Francisco, California

Kurt Deshayes
Department of Protein Engineering
Genentech Inc.
South San Francisco, California

Claire L. Dobson
MedImmune Ltd.
Cambridge, United Kingdom

Frederic A. Fellouse
Donnelly Centre
University of Toronto
Toronto, Ontario, Canada

Sabine Flückiger
BioVisioN Schweiz AG
Davos, Switzerland

Clarence Ronald Geyer
Department of Pathology
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

Paul T. Hamilton
Department of Plant and Microbial Biology
North Carolina State University
Raleigh, North Carolina

Zhaozhong Han
Alexion Pharmaceuticals, Inc.
Cheshire, Connecticut

Traver Hart
Donnelly Centre
University of Toronto
Toronto, Ontario, Canada

Celia P. Hart-Shorrock
MedImmune Ltd.
Cambridge, United Kingdom

Christian Heinis
Institute of Pharmaceutical Sciences
Swiss Federal Institute of Technology
Zurich, Switzerland

Robin Hyde-DeRuyscher
Biogen Idec
Research Triangle Park, North Carolina

Ece Karatan
Department of Biology
Appalachian State University
Boone, North Carolina

Brian K. Kay
Department of Biological Sciences
University of Illinois at Chicago
Chicago, Illinois

Malgorzata E. Kokoszka
Department of Biological Sciences
University of Illinois at Chicago
Chicago, Illinois

Zoltan Konthur
Max Planck Institute of Molecular Genetics
Berlin, Germany

Jianghai Liu
Department of Pathology
University of Saskatchewan
Saskatoon, Saskatchewan, Canada

Henry B. Lowman
Department of Antibody Engineering
Genentech, Inc.
South San Francisco, California

Lee Makowski
Combinatorial Biology Unit
Biosciences Division
Argonne National Laboratory
Argonne, Illinois

Suneeta Mandava
Combinatorial Biology Unit
Biosciences Division
Argonne National Laboratory
Argonne, Illinois

Jonathan S. Marvin
Department of Antibody Engineering
Genentech, Inc.
South San Francisco, California

Megan McLaughlin
Donnelly Centre
University of Toronto
Toronto, Ontario, Canada

Ralph R. Minter
MedImmune Ltd.
Cambridge, United Kingdom

Jason Moffat
Donnelly Centre
and
Department of Molecular Genetics
University of Toronto
Toronto, Ontario, Canada

Dario Neri
Institute of Pharmaceutical Sciences
Swiss Federal Institute of Technology
Zurich, Switzerland

Shuichi Ohkubo
Tsukuba Research Center
Taiho Pharmaceutical Co., Ltd.
Tsukuba, Japan

Gabor Pal
Department of Biochemistry
Eötvös Loránd University
Budapest, Hungary

Valery A. Petrenko
Department of Pathobiology
College of Veterinary Medicine
Auburn University
Auburn, Alabama

Mikhail Popkov
Department of Molecular Biology
The Scripps Research Institute
La Jolla, California

Claudio Rhyner
Swiss Institute of Allergy and Asthma Research (SIAF)
Davos, Switzerland

Diane J. Rodi
Combinatorial Biology Unit
Biosciences Division
Argonne National Laboratory
Argonne, Illinois

Jamie K. Scott
Department of Molecular Biology and Biochemistry
and
Faculty of Health Sciences
Simon Fraser University
Burnaby, British Columbia, Canada

Sachdev S. Sidhu
Donnelly Centre
and
Department of Molecular Genetics
University of Toronto
Toronto, Ontario, Canada

George P. Smith
Division of Biological Sciences
University of Missouri
Columbia, Missouri

Jelena Tomic
Donnelly Centre
University of Toronto
Toronto, Ontario, Canada

Mihriban Tuna
Department of Biochemistry
School of Life Sciences
University of Sussex
Falmer, United Kingdom

Nienke E. van Houten
Faculty of Health Sciences
Simon Fraser University
Burnaby, British Columbia, Canada

Michael Weichel
Swiss Institute of Allergy and Asthma Research (SIAF)
Davos, Switzerland

Gregory A. Weiss
Department of Chemistry
University of California
Irvine, California

Derek N. Woolfson
Department of Biochemistry
School of Life Sciences
University of Sussex
Falmer, United Kingdom

CHAPTER 1

Filamentous Bacteriophage Structure and Biology

Diane J. Rodi, Suneeta Mandava, and Lee Makowski

CONTENTS

1.1 INTRODUCTION

Phage display technology provides a remarkably versatile tool for exploring the interactions between proteins, peptides, and small-molecule ligands. As such, it has become widely adapted for use in epitope mapping, identification of protein–peptide and protein–protein interactions, protein–small molecule interactions, humanization of antibodies, identification of tissue-targeting peptides, and many other applications as outlined throughout this book. However, it must be kept in mind that phage display is a combinatorial *biology* approach, not a combinatorial *chemistry* approach. The great strength of phage display over combinatorial methods that are strictly chemical is that the isolation of a single interacting protein or peptide attached to a phage particle is sufficient to allow the complete characterization of the isolate: the interacting virus can be grown up in bulk and the sequence of the displayed protein or peptide inferred from the DNA sequence carried within the viral particle. The other side of this coin is that phage display technology utilizes living systems, and is therefore constrained in its potential diversity by the molecular requirements of those systems.

The biological limitations that impact phage display technology are defined not simply by viral structure, but by the well-balanced phage–host system as a whole. The display of a protein or peptide on the surface of a bacteriophage particle involves insertion of the corresponding DNA into the gene of a structural protein and the expression of the foreign sequence as a fusion with the structural protein in such a way that it is exposed, at least in part, on the surface of the phage particle. This process perturbs the phage–host system and may result in anything from a negligibly small alteration in phage growth rate to a complete halt of phage production. Disruption of any step along the way between DNA cloning and production of virus, including protein synthesis, protein translocation, viral morphogenesis, viral stability, host cell binding, or subsequent steps in the infection process, can remove a particular display construct from the final phage population. Additionally, in the context of library screening methodology, it is also important to note that different inserts placed at the same site may have very different effects on the rate of viral production, resulting in biases that can seriously impact the diversity of a phage-displayed

library and, consequently, the results of affinity selection experiments. Some members of the libraries are present at much lower levels than others whereas others are absent. These biases must be well characterized in order to make optimal use of libraries in affinity selections or other experiments designed to take advantage of the unique properties of display libraries. Therefore, in order to understand the effect of biology on phage display, the way the phage interacts with its host must be considered in detail.

In this chapter, we outline the steps of phage–host interaction and discuss how those interactions may impact the diversity of phage-displayed libraries. Understanding these limitations in more detail should provide a starting point for engineering methods to minimize their effect on the use of phage display technology within the broad range of applications reviewed in this volume. Except for DNA replication, each step appears to have a detectable effect on the expression of some members of some libraries. Some effects appear significant whereas others are barely detectable. At the end, we briefly review identified bottlenecks in the viral life cycle and suggest simple strategies that can be implemented for minimizing the perturbations.

1.2 TAXONOMY AND GENETICS

The filamentous bacteriophages are a family of ssDNA-containing viruses (genus *Inovirus*) that infect a wide variety of gram-negative bacteria, including *Escherichia coli*, *Xanthomonas*, *Thermus*, *Pseudomonas*, *Salmonella*, and *Vibrio*. The best characterized of the filamentous phage are the Ff class of viruses, so named because of their method of host cell entry via the tip of the F conjugative pilus on the surface of male *E. coli* cells. The Ff viruses include M13, fd, and f1, all of which possess a 98% identity at the DNA sequence level. Ff virus particles are long, slender, and flexible rods with a diameter of about 65 Å. The wild-type Ff phages are between 0.8 and 0.9 μm long, giving the virus the proportions of a 4-foot-long pencil. Various engineered strains have somewhat longer genomes with the length of the particle increased proportionate to the length of the encapsulated DNA. Although there is considerable heterogeneity within the family, some similarities of sequence and genome organization are discernable among all group members. An electron micrograph shown in Figure 1.1 gives a rough idea of the proportions of the phage particles. The single-stranded, circular genome occupies the axis of the particle, stretched out for almost the entire length of the virion. Virus lengths are dependent upon both the size of the enclosed genome as stated above and on the physical distribution of the DNA within the capsid (axial distance per base), the latter of which has been demonstrated to be major coat protein charge dependent [1]. Little substructure is visible except at the end involved in host cell attachment. Each cross-section of the virion has an “up” strand and a “down” strand present, but these are not base paired because there is no complementary relationship between the sequences of the two strands except within the hairpin, which acts as the packaging signal that nucleates the initiation of viral assembly.

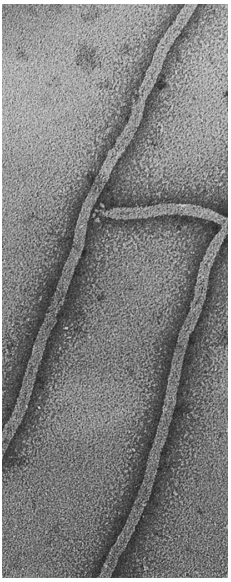

Figure 1.1 Electron micrograph of bacteriophage M13. This micrograph of M13 phage particles visually demonstrates the rationale for their designation as "filamentous" bacteriophage. The amino terminus of at least four copies of the gene *III* protein are visible at the end of the phage particle; the two subtilisin-cleavable N1 and N2 domains are seen as knobby structures at the proximal end of the virus. (Micrograph courtesy of Irene Davidovich.)

Ff viruses are not lytic but rather are parasitic. Productive infections result in viral release via extrusion or secretion across the inner and outer bacterial membranes in the absence of host cell death with the infected cells continuing to grow and divide (albeit at a significantly reduced rate). M13 produces anywhere from 200 to 2000 progeny phage per cell per doubling time [2,3]. This phage production represents a serious metabolic load for the infected *E. coli* with phage proteins making up 1%–5% of total protein synthesis and resulting in a reduction in cell growth of 30%–50% from uninfected cells. The nonlytic nature of Ff infection, along with the simultaneous presence of both single- and double-stranded forms of viral DNA, little size constraint on inserted DNA, and an exceptionally high viral titer capacity (typically 10^{11}–10^{12} particles per ml), has made the filamentous phage, primarily M13, a workhorse for molecular biology for the last 20 years.

1.3 VIRAL GENE PRODUCTS

The Ff phage genome encodes a total of 11 proteins (see Figure 1.2 for genome organization). There are five structural proteins, all of which are inserted into the inner host cell membrane prior to assembly (see Figure 1.3 for overall structural

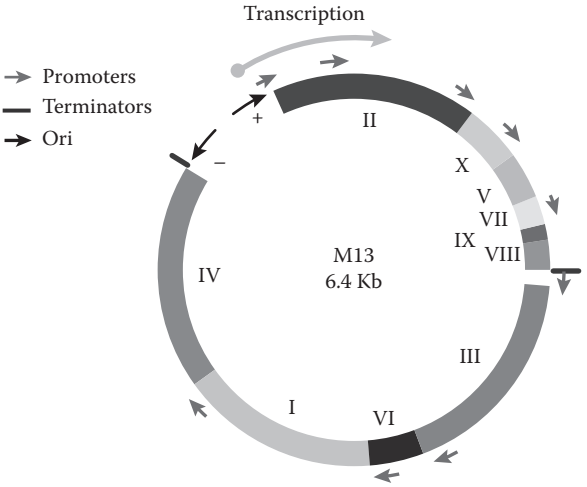

Figure 1.2 Ff phage genome. The location of each viral gene is indicated by number with the direction of transcription shown by arrow. The origin of replication lies within the intergenic region between the genes for pIV and pII. The packaging signal (PS) lies between the (–) strand origin of replication and gene *IV*.

organization of Ff phage). pVIII and pIII are synthesized with signal sequences that are removed subsequent to membrane insertion whereas pVI, pVII, and pIX are absent signal sequences. Three nonstructural phage proteins, pI, pXI, and pIV are required for phage morphogenesis but are not incorporated into the phage structure. pX and pXI are the result of in-frame internal translation initiation events in genes *II* and *I*, respectively, and are identical with the C-terminal portions of pII and pI in amino acid sequence, membrane localization, and topology [4]. In addition to the coding regions, there is an intergenic region, which contains the signals for the initiation of synthesis of both the plus (+) or viral-contained DNA strand, and the minus (–) strand; the initiation of capsid assembly signal (or packaging signal, PS), which lies between the (–) origin and the end of the pIV gene and the signal for termination of RNA synthesis. Parts of the intergenic region have been shown to be dispensable (reviewed in Ref. [3]), but all of the coding region products are necessary for the synthesis of the infectious progeny phage.

1.3.1 Replication Proteins (pII and pX)

pII is a 410 amino acid protein (MW = 46,137), which is required for all phage-specific DNA synthesis other than the formation of the complementary strand of the infecting ssDNA by host enzymes. pII has both endonuclease and topoisomerase activities required during the DNA replication phase of infection. pX is a 111 residue protein (MW = 12,672), which is encoded entirely within gene *II*, initiating at codon 300, an AUG that is in phase with the initiating AUG of gene *II*. Although pX has the same amino acid sequence as the carboxyl-terminal end of pII, it has been

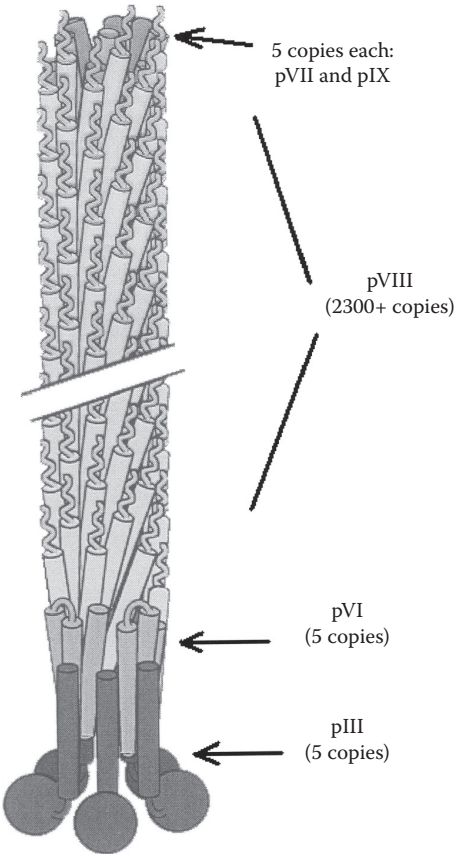

Figure 1.3 Schematic diagram of the Ff bacteriophage. This diagram depicts the structural organization of M13 as a representative of the Ff viral family. At the top of the diagram lies the distal end of the particle at which viral assembly initiates. At the bottom of the figure is the proximal or infectious end of the virus with five copies of the pIII anchored to the particle by five copies of the pVI.

shown to possess unique functions within the viral life cycle, such as inhibition of pII function [5].

1.3.2 Single-Stranded DNA Binding Protein (pV)

Gene *V* codes for an 87 amino acid protein (MW = 9682) that exists as a stable dimer even at a concentration as low as 1 nM [6]. The crystal structure of the protein has been solved to 1.8 Å resolution using multiwavelength anomalous dispersion on a selenomethionine-containing protein and is shown in Figure 1.4 [7]. Each monomer is largely β-structure, with 58 out of 87 residues arranged in a five-stranded antiparallel β-sheet; two antiparallel β-ladder loops protrude from this sheet. The remainder of the molecule is arranged into 3_{10} helices (residues 7–11 and 65–67),

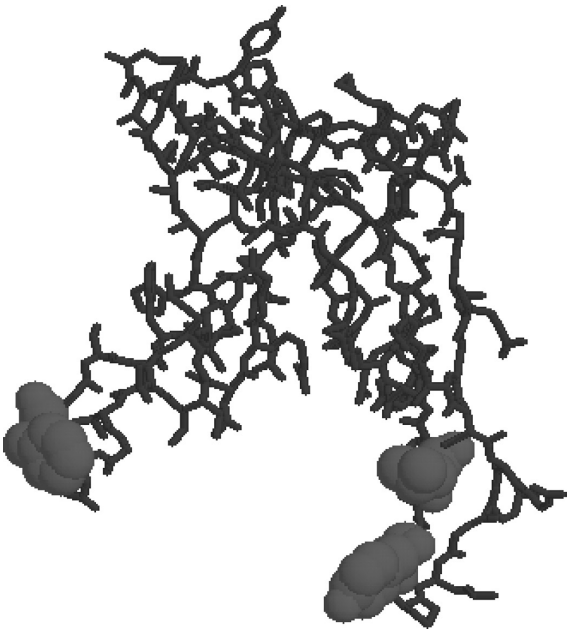

Figure 1.4 Crystal structure of the gene *V* ssDNA binding protein. The crystal structure of pV has been solved to 1.8 Å resolution using multiwavelength anomalous dispersion on a selenomethionine derivative and is shown here in a backbone format [7] (PDB accession code 2GN5). The protein normally exists as a dimer and wraps around the single-stranded form of the viral DNA within the host cell cytoplasm. Residues Tyr26, Leu28, and Phe73 have been shown to be critical for DNA binding [7] (residues shown in spacefill format).

β-bends (residues 21–24, 50–53, and 71–74), and one five-residue loop (residues 38–42). Nuclear magnetic resonance (NMR) analysis of the gene *V* protein [8] suggests that the DNA binding loop (residues 16–28; see Figure 1.4) is flexible in solution. This protein serves the dual functions of sequestering the intracellular ssDNA viral genomes [3] and modulating the translation of gene *II* mRNA [9].

1.3.3 Major Structural Protein (pVIII)

Gene *VIII* codes for the major coat protein of the virus. The major coat proteins of all filamentous phages are short, ranging from 44 to 55 amino acids, with most being encoded with a signal sequence (*Pseudomonas aeruginosa*-infecting phage Pf3 is an example of a pVIII absent a signal sequence). In the Ff group of phage, the major coat protein is 50 amino acids long (MW = 5235) with a 23 amino acid–long signal sequence. Approximately 2800 copies of pVIII are required to coat one full-length wild-type Ff virion. The concentration of pVIII in the inner cell membrane is very high—at least 5×10^5 molecules of pVIII are exported as virions per infected cell per doubling, making it one of the most abundant proteins in the infected cell [10].