Advances in Experimental Medicine and Biology 1053

Theam Soon Lim *Editor*

Recombinant Antibodies for Infectious Diseases

Springer

Advances in Experimental Medicine and Biology

Volume 1053

Editorial Board:

IRUN R. COHEN, *The Weizmann Institute of Science, Rehovot, Israel*
ABEL LAJTHA, *N.S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA*
JOHN D. LAMBRIS, *University of Pennsylvania, Philadelphia, PA, USA*
RODOLFO PAOLETTI, *University of Milan, Milan, Italy*
NIMA REZAEI, *Tehran University of Medical Sciences Children's Medical Center, Children's Medical Center Hospital, Tehran, Iran*

More information about this series at http://www.springer.com/series/5584

Theam Soon Lim
Editor

Recombinant Antibodies for Infectious Diseases

Springer

Editor
Theam Soon Lim
Institute for Research in Molecular Medicine
Universiti Sains Malaysia
Penang, Malaysia

ISSN 0065-2598 ISSN 2214-8019 (electronic)
Advances in Experimental Medicine and Biology
ISBN 978-3-319-72076-0 ISBN 978-3-319-72077-7 (eBook)
https://doi.org/10.1007/978-3-319-72077-7

Library of Congress Control Number: 2018931166

© Springer International Publishing AG, part of Springer Nature 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The fate of antibodies and infectious diseases has been entwined since the early days when Emil von Behring and Shibasaburo Kitasato first witnessed the unique toxin neutralization ability of antibodies in sera from immunized rabbits. The hybridoma technology by Köhler and Milstein in the mid-1970s revolutionized the way antibodies can be applied in the future. Their contributions set about the first monospecific antibodies to be generated against a specific target. This brought about major revolutions in the way we carry out basic research, medical diagnostics, and therapeutics. However, the ability to produce murine monospecific antibodies was not the finish article due to major side reactions associated with the use of animal-derived antibodies in humans.

The quest for further improvements in the way antibodies were being generated meant there was a need to generate monoclonal antibodies at a more rapid pace with more human-like characteristics. The biotechnology boom at the turn of the century resulted in rapid advancements in recombinant DNA technology, molecular biology, and DNA sequencing technologies. The technological developments allowed the floodgates of genetic information to open making molecular-based technologies accessible to many laboratories around the world. The knowledge, information, and technological advancements worked symbiotically to fuel the advancements of complementary technologies. Recombinant antibody technology also benefited from this evolution with the improved understanding of genes and mechanisms associated with *in vivo* antibody production.

In the quest to make antibodies more human, researchers sought after new alternative methods to generate recombinant versions of human antibodies. This possibility was only realized with the introduction of phage display technology by George Smith, which allowed the presentation of peptides on the surface of bacteriophages. This allowed for the evolution of the technology to present antibodies on the surface, which catalyzed the growth of human monoclonal antibody technology. Since then, numerous versions of human antibody formats have been developed with astonishing success. As a result, new antibody libraries have been developed and are now an important tool for monoclonal human antibody development work. The technology has allowed for different antibodies to be developed against many different kinds of targets which was otherwise impossible with conventional approaches. The ability to generate fully human monoclonal antibodies at a rapid pace has shaped the pharmaceutical landscape in recent years. This signaled a turning

point for many medical approaches applied at that time and has helped shaped the way modern immunotherapeutics are designed.

The ongoing challenges associated with infectious diseases like antibiotic resistance and the cost for drug discovery meant that an alternative treatment was required. This book provides an in-depth introduction to bacteriophage biology as well as its application for antibody phage display. The book also includes examples of different forms of antibody libraries that are used to tackle the issue of infectious diseases. It also provides a comprehensive list of antibody phage display technologies and the application of antibodies against different infectious agents. In addition to that, the book also includes concepts of computational-based antibody design, antibody engineering strategies, and considerations in the application of antibody-based therapy for infectious diseases.

On a personal note, the treatment of infectious disease is a topic close at heart due to the constant threat it poses around Southeast Asia. This is a key focus area for me personally as part of the antibody technology initiative at the Institute for Research in Molecular Medicine (INFORMM), which is the brainchild of the Malaysian Ministry of Higher Education under the Higher Institutions' Centre of Excellence (HICoE) program together with Universiti Sains Malaysia (USM). It is our aim that this book can provide technical assistance to new start-up laboratories and researchers looking to apply antibody phage display for infectious diseases. We also hope this book will help spur interest and ideas in the field while at the same time expand research focusing on antibody-based therapy for infectious diseases.

I would like to thank the authors whose contributions to this book have allowed it to be a comprehensive guide for antibody phage display in infectious diseases. I would also like to thank Prof. Michael Hust for his guidance and advice throughout the preparation of this book. My scientific career would not have been possible without the influence of great mentors like Zoltàn Konthur and Jörn Glökler. On a personal note, I would like to thank Poi Hong, Hayley, Hayden, and my parents for their support while preparing this book and throughout my scientific career.

Penang, Malaysia Theam Soon Lim

Contents

Contributors

Arnaud Avril Département des maladies infectieuses, Unité biothérapies anti-infectieuses et immunité, Institut de Recherche Biomédical des Armées, Brétigny-sur-Orge, France

Adriana Badarau Arsanis Biosciences GmbH/Arsanis, Inc, Vienna, Austria

Felix Breden Department of Biological Sciences, Simon Fraser University, Burnaby, BC, Canada

Sam J. Brooke Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Soo Khim Chan Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Gang Chen Donnelly Centre for Cellular and Biomolecular Research, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada

Yee Siew Choong Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Julia M. Costa-Cruz Laboratory of Parasitology, Institute of Biomedical Sciences, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil

Gaowei Fan Department of Clinical Laboratory, Beijing Chaoyang Hospital, Capital Medical University, Beijing, China

Luiz R. Goulart Laboratory of Nanobiotechnology, Institute of Genetics and Biochemistry, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil

Department of Medical Microbiology and Immunology, University of California-Davis, Genome and Biomedical Sciences Facility, Davis, CA, USA

Adam Seidelin Hey Novartis Institutes for BioMedical Research, Basel, Switzerland

Sofia Khanum Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Jing Yi Lai Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Cheh Tat Law Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Jinming Li National Center for Clinical Laboratories, Beijing Hospital, National Center of Gerontology, Beijing, China

Theam Soon Lim Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Analytical Biochemistry Research Centre, Universiti Sains Malaysia, Minden, Penang, Malaysia

Yee Ying Lim Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Yie Vern Lee Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Eszter Nagy Arsanis Biosciences GmbH/Arsanis, Inc, Vienna, Austria

Gábor Nagy Arsanis Biosciences GmbH/Arsanis, Inc, Vienna, Austria

Johan Nilvebrant Donnelly Centre for Cellular and Biomolecular Research, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada

Division of Protein Technology, School of Biotechnology, Royal Institute of Technology, AlbaNova University Center, Stockholm, Sweden

Soumya Palliyil Scottish Biologics Facility, University of Aberdeen, Aberdeen, UK

Christine A. Power Arsanis Biosciences GmbH/Arsanis, Inc, Vienna, Austria

Anizah Rahumatullah Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Marina Rajič Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland, New Zealand

Jasna Rakonjac Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland, New Zealand

Vanessa da S. Ribeiro Federal University of Goias, Catalão, Goias, Brazil

Marjorie Russel The Rockefeller University, New York, NY, USA

Hyunbo Shim Departments of Bioinspired Science and Life Science, Ewha Womans University, Seoul, South Korea

Sachdev S. Sidhu Donnelly Centre for Cellular and Biomolecular Research, Banting and Best Department of Medical Research, University of Toronto, Toronto, ON, Canada

Jia Xin Soong Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Minden, Penang, Malaysia

Valéria Szijártó Arsanis Biosciences GmbH/Arsanis, Inc, Vienna, Austria

Obinna Ubah Scottish Biologics Facility, Elasmogen Ltd, Aberdeen, UK

Corey T. Watson Department of Biochemistry and Molecular Genetics, University of Louisville School of Medicine, Louisville, KY, USA

Abbreviations

%	Percent
ΔG_{Bind}	Binding free energy
T3SS	Type three secretion system
AAC	Antibody–antibiotic conjugate
Ab	Antibody
ACSs	Antibody-secreting cells
ActIIR	Myostatin/activin type II receptor
ADAs	Antidrug antibodies
ADC	Antibody–drug conjugate
ADCC	Antibody-dependent cell-mediated cytotoxic
ADCP	Antibody-dependent cell phagocytosis
ADCVI	Antibody-dependent cell-mediated virus inhibition
ADE	Antibody-dependent enhancement
ADIN	Antibody-dependent intracellular neutralization
AID	Activation-induced cytidine deaminase
AIDS	Acquired immunodeficiency syndrome
AIGIV	Anthrax immunoglobulin intravenous
AMA	Antibody modelling assessment
AMA1	Apical membrane antigen 1
AMBER	Assisted Model Building with Energy Refinement
AMF	*Aspergillus fumigatus* membrane fraction
ANN	Artificial neural network method
APC	Antigen-presenting cells
ART	Antiretroviral therapy
ARTs	ADP-ribosyltransferases
ASCs	Antigen-specific
ASPD	Artificially selected proteins/peptides database
B. anthracis	*Bacillus anthracis*
BCRs	B cell receptors
BDB	Biopanning data bank
BER	Base excision repair
BLAST	Basic local alignment search tool
bnAbs	Broadly neutralizing antibodies
BoNT	Botulinum neurotoxin
bp	Base pairs
BSA	Bovine serum albumin
BsAbs	Bispecific antibodies

BSL4	Biosafety level 4
C. albicans	*Candida albicans*
CASP	Critical Assessment of Techniques for Protein Structure Prediction
CD4bs	CD4 binding site
CDC	Centers for Disease Control and Prevention
CDC	Complement-dependent cytotoxicity
CDCs	Cholesterol-dependent cytolysins
CDI	*C. difficile* infection
cDNA	Complementary deoxyribonucleic acid
CDR	Complementary determining region
CDT	*Clostridium difficile* transferase
CF	Compactness factor
cfu	Colony-forming unit
C_H	Constant region genes
CH1	First heavy chain constant domain
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CHESS	Centocor: HA-1A Efficacy in Septic Shock
ClfA	Clumping factor A
COGs	Cost of goods
CP	Cysteine proteinases
cryo-EM	Cryo-electron microscopy
CsCl	Cesium chloride
CSF	Cerebrospinal fluid
CSR	Class switch recombination
Dabs	Domain antibodies
DALYs	Disability-adjusted life years
DBP	Duffy binding protein
DENV	Dengue virus
DNA	Deoxyribonucleotide acid
dPNAG	Deacetylated form of poly-β-1,6-N-acetylglucosamine
DSB	Double-strand break
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E	Envelope
E. coli	*Escherichia coli*
e.g.	For example
EBC	Epstein–Barr virus
EBOV	Ebola virus
EbpC	Major pilus component of *Enterococcus faecalis*
EBV	Epstein–Barr virus
ECLIA	Electrochemiluminescent immunoassay
EF	Edema factor
EGFR	Epidermal growth factor receptor
EHEC	Enterohemorrhagic *E. coli*
ELISA	Enzyme-linked immunosorbent assay
EMV	Extracellular enveloped virions
EPPND	Enhanced pre- and postnatal development

ET	Edema toxin
Fab	Fragment antigen binding regions
FACS	Fluorescence-activated cell sorting
FBI	Federal Bureau of Investigation
Fc	Fragment crystallizable
FcRn	Fc receptor neonate
FDA	US Food and Drug Administration
fHbp	Factor H binding protein of *N. meningitidis*
FML	Fucose–mannose ligand
FR	Framework
FRs	Framework regions
Gb3	Globotriaosylceramide
GC	Germinal center
GD2	Disialoganglioside
GENESIS	Generalized-ensemble simulation system
GIPL	Glycosylinositolphospholipid
GP	Glycoprotein
GP_{CL}	Cleaved glycoprotein
GPI-APs	Glycosylphosphatidylinositol-anchored proteins
GP_{UNCL}	Uncleaved glycoprotein
GRA	Granule antigens
GROMOS	Groningen molecular simulation
GS	Glycine–serine
GVHD	Graft versus host disease
HA	Hemagglutinin
HAMA	Human anti-mouse antibody
HAP	Hospital-acquired pneumonia
HBcAg	Hepatitis B core antigen
HBsAg	Hepatitis B surface antigen
HBV	Hepatitis B virus
HC	Heavy chain
HCAbs	Heavy-chain-only antibodies
HeV	*Hendra* virus
HIV	Human immunodeficiency virus
Hla	Alpha-hemolysin
hLD_{50}	Mean lethal dose
HSC	Human string content
HSLs	Homoserine lactones
HTS	High-throughput sequencing
HUS	Hemolytic uremic syndrome
IBS	Inflammatory bowel syndrome
IEDB	Immune epitope database
IFA	Immunofluorescence assay
Ig	Immunoglobulin
IgG	Immunoglobulin G
IHC	Immunohistochemical
IMV	Intracellular mature virions
IND	Investigational new drug

IP	Intellectual property
IPP	Ileal Peyer's patches
IRIS	Immune reconstitution inflammatory syndrome
ISAAC	Immunospot array assay on a chip
JCV	Polyomavirus JC
JEV	Japanese encephalitis virus
K_D	Dissociation constant
kDa	Kilodalton
KGB	Komitet Gosudarstvennoy Bezopasnosti
L. acidophilus	*Lactobacillus acidophilus*
LC	Light chain
LcrV	Low-calcium response V antigen
LD	Linear dichroism
LF	Lethal factor
LF	Lymphatic filariasis
LFIA	Lateral flow immunoassay
LPS	Lipopolysaccharide
LT	Lethal toxin
LVS	Live vaccine strain
mAb	Monoclonal antibody
MAC	Membrane attack complex
MD	Molecular dynamic
MDR	Multidrug-resistant
MFP	Membrane protein fraction
MHC	Major histocompatibility complex
MIC	Microneme proteins
MM-GBSA	Molecular mechanics/generalized born solvent area
MM-PBSA	Molecular mechanics Poisson–Boltzmann surface area
MMR	Mismatch repair
MoMp	Mitochondrial outer membrane permeabilization
MPER	Membrane-proximal external region
MPFIA	Magnetic particle fluorogenic immunoassay
MrkA	Type 3 fimbrial shaft subunit of *Klebsiella*
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Multiple sclerosis
MSP	Merozoite surface protein
MSSA	Methicillin-sensitive *Staphylococcus aureus*
NA	Neuraminidase
NAbs	Neutralizing antibodies
NC	Neurocysticercosis
NHEJ	Nonhomologous end joining
nM	Nanomolar
NMR	Nuclear magnetic resonance
NPC1	Niemann–Pick C1
NS	Nonstructural
NSG	Next-generation sequencing
nt	Nucleotide
OPK	Opsonophagocytic uptake and killing

PA	Protective antigen
PACE	Phage-assisted continual evolution of protein
PBL	Peripheral blood lymphocyte
PBMCs	Peripheral blood mononuclear cells
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PD	Phage display
PDA	Protein design automation
PDB	Protein data bank
PEG	Polyethylene glycol
PfHRP2	*P. falciparum* histidine-rich protein 2
PFTs	Beta-barrel pore-forming toxins
pI	Isoelectric point
PIGS	Prediction of immunoglobulin structure
Ply	Pneumolysin
pM	Picomolar
PML	Progressive multifocal leukoencephalopathy
PNAG	Poly-β-1,6-N-acetylglucosamine
PNH	Paroxysmal nocturnal hemoglobinuria
PS	Packaging signal
Psa/Ph6	Pathogenesis
Psp	Phage shock protein
PTM	Posttranslational modification
PTx	Pertussis toxin
PVL	Panton–Valentine leukocidin
QS	Quorum sensing
RA	Rheumatoid arthritis
RAG	Recombination-activating genes
RF	Random forest
RF	Replicative form
RMSD	Root-mean-square deviation
RNA	Ribonucleic acid
ROA	Route of administration
ROP	Rhoptry proteins
RSS	Recombination signal sequences
RSV	Respiratory syncytial virus
RT-PCR	Reverse transcription polymerase chain reaction
S. aureus	*Staphylococcus aureus*
SAbDab	Structural antibody database
SAR	Structure–activity relationship
SARS-CoV	Severe acute respiratory syndrome coronavirus
scFab	Single-chain Fab
scFv	Single-chain variable fragment
sdAb	Single-domain antibody
SDRs	Specificity-determining residues
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEB	Staphylococcal enterotoxin B
SHM	Somatic hypermutation

sIBM	Inclusion body myositis
SLE	Systemic lupus erythematosus
SlpA	Surface layer A
SoC	Standard-of-care
SPA	Sequence prediction algorithm
SpA	Staphylococcal surface protein a
SPR	Surface plasmon resonance
spvB	*Salmonella* virulence plasmid factor B
ssDNA	Single-stranded DNA
STEC	Shiga toxin-producing *E. coli*
Stxs	Shiga toxins
SUDV	Sudan virus
SurA	Survival protein A
TB	Tuberculosis
Tc	Transchromosomic
TcdA	Enterotoxin A
TcdB	Cytotoxin B
TCR	Tissue cross-reactivity
TCR	T-cell receptor
TeNT	Tetanus neurotoxin
TG	Thyroglobulin
TLS	Tumor lysis syndrome
TNF	Tumor necrosis factor
TRIM	Trinucleotide-directed mutagenesis
TRIM21	Tripartite motif-containing 21
TSST	Toxic shock syndrome toxin
USAMRIID	US Army Medical Research Institute of Infectious Disease
USFDA	US Food and Drug Administration
VAP	Ventilator-associated pneumonia
VEEV	Venezuelan equine encephalitis virus
VEGF	Vascular endothelial growth factor
V_H	Variable heavy chains
VH	Heavy chain variable domain
VHF	Viral hemorrhagic fevers
VIG	Vaccinia immune globulin
V_L	Variable light chains
VL	Visceral leishmaniasis
VLA1	Human integrin domain I
VP1	Viral protein 1
VSG	Variant-specific surface glycoprotein
Vκ	Kappa variable domain
WEEV	Western equine encephalitis virus
WHO	World Health Organization
WNV	West Nile virus
WNV	West Nile
YFV	Yellow fever virus
ZIKV	Zika virus

Filamentous Phage: Structure and Biology

1

Jasna Rakonjac, Marjorie Russel, Sofia Khanum, Sam J. Brooke, and Marina Rajič

Abstract

Ff filamentous phage (fd, M13 and f1) of *Escherichia coli* have been the workhorse of phage display technology for the past 30 years. Dominance of Ff over other bacteriophage in display technology stems from the titres that are about 100-fold higher than any other known phage, efficacious transformation ensuring large library size and superior stability of the virion at high temperatures, detergents and pH extremes, allowing broad range of biopanning conditions in screening phage display libraries. Due to the excellent understanding of infection and assembly requirements, Ff phage have also been at the core of phage-assisted continual protein evolution strategies (PACE). This chapter will give an overview of the Ff filamentous phage structure and biology, emphasizing those properties of the Ff phage life cycle and virion that are pertinent to phage display applications.

Keywords

Bacteriophage · Filamentous phage · Phage display · Ff structure · Infection mechanism · Ff life cycle

1.1 Introduction

The appearance and mode of assembly of filamentous phage are oddities in the world of bacteriophage, as they do not fit the archetypal head and tail image, and they do not lyse (kill) the host in order to reproduce. Most of what is known about filamentous phage assembly comes from work on *E. coli* bacteriophage Ff (f1, M13 and fd) which are 98% identical at the DNA sequence level and have been studied interchangeably [97, 98, 106, 121, 134]. Like fila-

This article is dedicated to the memory of Peter Model, a pioneer of filamentous bacteriophage research and a greatly admired mentor to students and junior faculty at the Rockefeller University.

J. Rakonjac (✉) · M. Rajič
Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland, New Zealand
e-mail: j.rakonjac@massey.ac.nz; m.rajic@massey.ac.nz

M. Russel
The Rockefeller University, New York, NY, USA
e-mail: russelm@mail.rockefeller.edu

S. Khanum · S. J. Brooke
Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand
e-mail: s.khanum@massey.ac.nz; s.j.brooke@massey.ac.nz

© Springer International Publishing AG, part of Springer Nature 2017

T. S. Lim (ed.), *Recombinant Antibodies for Infectious Diseases*, Advances in Experimental Medicine and Biology 1053, https://doi.org/10.1007/978-3-319-72077-7_1

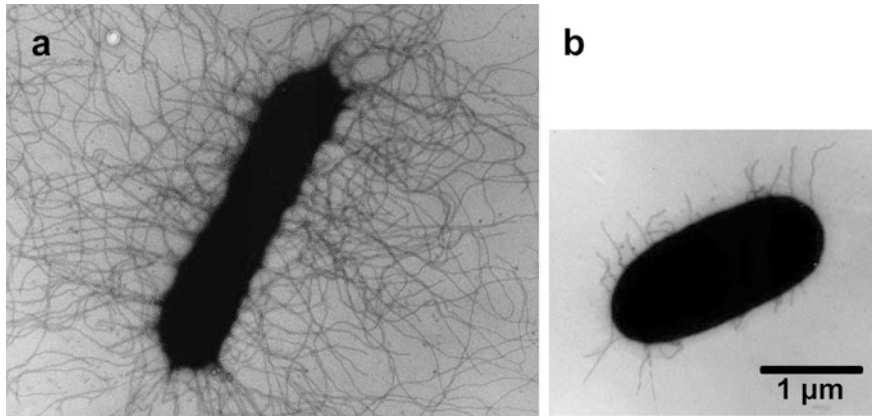

Fig. 1.1 *E. coli* cells assembling the Ff phage. (**a**) Cell infected with a ΔgIII Ff phage. (**b**) Cell infected with a wild-type Ff phage (Reproduced from Ref. [120])

mentous bacterial surface appendages, pili, Ff are assembled by a secretion-like process, aided by ATP and proton motive force [48, 130]. Simultaneous assembly and secretion are executed by a dedicated membrane-embedded assembly machinery, composed of an outer membrane channel and an inner-membrane-embedded, phage-encoded ATPase [48]. Deletion of the phage gene encoding the Ff phage release protein (pIII) results in permanent assembly of multi-length filaments, morphologically converting the Ff into DNA-containing pili (Fig. 1.1; [123]). The factory-like production of Ff phage (1000 phage per cell per hour at the peak of assembly) necessitates production and membrane targeting of millions of major coat protein copies and 300–400 assembly complexes traversing the envelope. Both of these seemingly unsustainable tasks are achieved and endured by *E. coli*, which remains viable. Ff assembly and secretion do have a consequence on cell physiology, causing a doubling of generation time and inducing the phage shock protein (Psp) stress response [14, 82]. A consequence of the extended generation time is a thinner bacterial lawn in comparison to uninfected cells, allowing formation of turbid "plaques". Infected cells form colonies that are small and transparent in comparison to the uninfected cells. Protein and gene nomenclature used for the Ff phage is numerical, expressed in Arabic or Roman numerals. This review will use the Roman numerals.

The virion of filamentous phage is composed of a circular single-stranded DNA (ssDNA) genome in the form of a two-stranded helix, surrounded by a tube formed by thousands of major coat protein (pVIII) subunits (Fig. 1.2; [98]). The tube is composed of the helically arrayed 50-residue α-helical major coat protein pVIII, and is capped at both ends by two different pairs of proteins (pVII/pIX and pIII/pVI).

Ff bacteriophage infect *Escherichia coli* by binding to the tip of the F-pilus, which then retracts [10, 83]. Like Ff, other filamentous bacteriophage appear to bind the retractable pili [8, 69, 103]. The secondary receptor in Ff and several distantly related phage is the TolQRA complex [28, 63]; hence this complex is likely a universal filamentous phage secondary receptor. Gram-negative bacteria are predominant hosts of filamentous phage. Among nearly 100 different filamentous phage that have been identified to date, only two were found to infect Gram-positive hosts [35].

Ff phage replicate as plasmid-like extrachromosomal replicons (episomes), by a rolling circle mechanism [106]. Whereas many other filamentous phage replicate exclusively as episomes, there is also a large number that integrate into the host chromosome [89]. Regulation of replication, integration and excision in these "temperate" filamentous phage follows a few different strategies, but the common characteristic is that the host survives both the integrated

(inactive) and induced states [78, 103, 125]. Many temperate filamentous phage replicate at extremely low rates, even in the induced state, producing less than one phage per cell per generation. These phage cannot make plaques and are being discovered through bacterial genome and microbial communities' metagenome sequencing. This review will focus on the Ff phage; for a detailed recent review on other filamentous phage please refer to [89].

Ff filamentous phage of *E. coli* have been used extensively in phage display technology and, recently, nanotechnology, due to their resistance to pH extremes, detergents and high temperature [12, 13]. Phage display is a powerful combinatorial technology; it allows identification, amongst billions of peptides/antibodies/proteins, of rare variants that bind to a ligand (or bait) of interest [11, 102, 141]. Ff is the phage of choice for antibody display due to the fact that displayed proteins are folded in the oxidative environment of the *E. coli* periplasm, allowing formation of S-S bridges in the immunoglobulin fold of antibodies [49]. Ff are also suitable for display of bacterial and archaeal membrane, cell-surface and secreted proteins, at genomic and metagenomics scale [24, 52, 108, 109]. Tailed phage, T4, T7 and λ, have also been used in phage display [4, 53, 160]. These phage are more suitable for display of cytosolic and nuclear proteins, given that their virion proteins fold in the reducing environment of the *E. coli* cytoplasm.

Ff phage display technology branched into nanotechnology at the start of the twenty-first century, through selection and display of nanocrystal-nucleating peptides for a number of inorganic materials [16, 66]. The liquid crystalline character of filamentous bacteriophage further expands nanotechnology applications of these viruses [23, 91, 115].

In basic research, Ff-phage-encoded coat and assembly/secretion proteins have served as models for studies of membrane targeting and secretion of proteins in Gram-negative bacteria [92, 136], whereas the phage of pathogenic bacteria have been investigated from the standpoint of bacterial virulence, physiology and effect on biofilm properties [89]. Ff have been used most recently in the phage-assisted continual evolution of proteins (PACE), resulting in e.g. novel polymerases, proteases or transcription factors with novel specificities [15, 47, 116].

1.2 Ff Structure

The genome of filamentous bacteriophage is circular single stranded DNA (ssDNA) which forms an anti-parallel two-stranded helix (similar to A-DNA or B-DNA). Due to the circular nature of the DNA, the helix ends with two loops [36]. In Ff phage one of the two loops is a 32-nucleotide hairpin called the packaging signal (PS), which targets the genome for packaging and initiates filamentous phage assembly [133]. Phosphates along the DNA helix interact with positively charged residues of the major coat protein, which forms a helically symmetrical tube that gives the filamentous appearance to the virion [57, 98, 99]. Due to the absence of complementarity, in the sequences beyond the packaging signal, the Watson-Crick type of pairing is only maintained for about 25% nucleotide pairs. Some filamentous bacteriophage (e.g. Pf1 of *Pseudomonas aeruginosa*) form a helix in which phosphates are in the centre, with the bases on the outside, interacting with the capsid [84].

The virions of filamentous phage are long, flexible filaments 6–7 nm in diameter and of varied length, which depends on the length of the packaged DNA (Fig. 1.2b). If initiation or termination of filament assembly is impaired, multiple genomes can be assembled into long filaments that can extend to up to 20 μm [122]. An additional factor determining length is the rise per nucleotide, or h, that ranges between 0.28 (in Ff phage) and 0.61 nm (in Pf1, in which the DNA helix is in the extended P-DNA-like form) [35]. The virion tube is composed of a helical array of the α-helical major coat protein pVIII (Fig. 1.2). There are two types of virions based on the symmetry of pVIII arrays: class I (5-start helix, C_2S_5) or class II (1-start helix, $C_1S_{5.4}$) [36]. Structures of the major coat protein have been determined for both classes of virions (i.e. in Ff and Pf1 phage) [57, 99]. However, very little is known about the structure of the virion "caps" at the two ends of the filament.

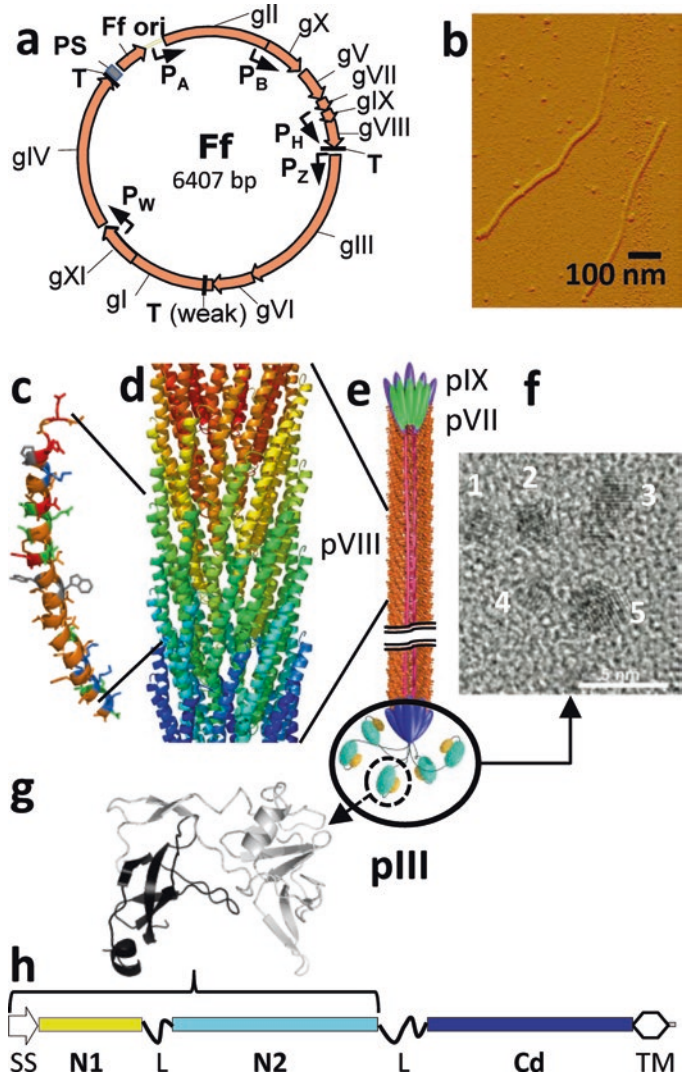

Fig. 1.2 The Ff genome and virion. (**a**) The Ff genome; *gI* to *gXI*, genes (orange block arrows); *PS*, packaging signal; *Ff ori*; origin of replication; $P_{A,}$ $P_{B,}$ $P_{H,}$ $P_{w,}$ P_{z}; promoters; *T*, terminators. (**b**) Atomic force microscope image of two Ff virions, one (longer) that has encapsidated the helper phage R408 genome, and the smaller that contains a phagemid vector whose genome is smaller (Rakonjac, Russel and Model, unpublished). (**c**) Structure of a pVIII (major coat protein) monomer. (**d**) pVIII arrangement within the filamentous phage capsid. (**e**) Schematic representation of the virion. (**f**) High-resolution TEM lattice fringe images of five ZnS nanocrystals (labelled *1–5*) at the pIII-end of a single Ff virion (From Ref. [81], reprinted with permission from AAAS). (**g**) Ribbon representation of the N1 domain (*dark gray*) and the N2 domain (*light gray*) of pIII. (**h**) Domain organization of pIII preprotein. *SS*, signal sequence, *N1*, *N2*, *Cd*, domains of pIII; G, glycine-rich linkers; TM, transmembrane helix. The images of the pVIII subunit (**c**), the capsid (**d**), and the N1-N2 domains of pIII (**g**) were derived from coordinates of the RCSB PDB database accession numbers 2cOw [99] and 1g3p [87], respectively, using PyMOL [37]

The Ff virions (including the receptor-binding protein pIII) remain intact and infectious after exposure to a broad range of pH and to high temperature [12, 13]. This is of great importance to affinity-screening (biopanning) of phage display libraries which regularly include e.g. pH extremes used for elution of the binders from the cognate 'bait'. Ff are also resistant to ionic detergents below the critical micellar concentration [149]. Ff phage are sensitive to chloroform, to which the tailed phage are generally resistant. However, a peptide displayed on all copies of the major coat protein was identified that rendered the virion resistant to this organic solvent [118]. Filamentous bacteriophage virions contain no lipids [35], a surprising fact, given that all virion proteins are integral inner membrane proteins prior to assembly [44]. At high concentration, filamentous bacteriophage behave as a liquid crystal and can be aligned in a strong magnetic field. These properties of Pf1 phage have been utilized in structural studies of other proteins by nuclear magnetic resonance (NMR) by promoting alignment of the investigated proteins in the magnetic field [163].

The liquid-crystalline properties of filamentous phage and ability to align them by shear flow [25] have seen them become a basis for diagnostics devices based on linear dichroism (LD), where a signal from aligned filaments, displaying antibodies against e.g. surface antigen of a pathogenic bacterium, changes upon recognition of the cognate bacterial cell [115]. A similar principle was used in design of "phage litmus", where the spectral changes are used to monitor detection of an analyte [111]. Chemical and enzymatic modifications have been applied to Ff phage to further expand the range of their applications as bionanoparticles [please refer to the recent reviews [7, 22, 65]]. The liquid crystalline Ff virions form membrane-like sheets when exposed to high-osmolarity solutions (dextran, PEG) that assume several macroscopic topologies, including disc, sphere, ribbon or more complex star-like disc-ribbon intermediate structure [40, 142].

1.2.1 Minor Virion Proteins

An atomic-resolution structure of the Ff virion caps has not been determined, in contrast to the detailed knowledge of pVIII structure and packing along the filament. Minor proteins pVII and pIX are small, containing only 32 (pVII) and 33 (pIX) amino acids; both proteins are hydrophobic. Like all other virion proteins, they are integrated into the inner membrane prior to assembly; however they have no signal sequence and are thought to spontaneously insert into the membrane [44]. These two proteins are incorporated into the virion at the initiation step of assembly (Figs. 1.2 and 1.5); [44, 59]. Genetic analysis suggests that the packaging signal interacts with the C termini of these two proteins to initiate assembly [133]. Both pVII and pIX have been used as display platforms for scFv's, however in those cases, SecA-dependent signal sequences must precede the scFv in order to target the fusions to the inner membrane and into the virion [54, 55, 71].

Proteins pIII and pVI terminate filament assembly and release the virion from the cell [122, 123]. The largest protein in the virion, pIII, also mediates receptor binding and entry of the phage into the host cell (Figs. 1.2e–h and 1.3). Both pIII and pVI are integral membrane proteins prior to assembly into the filamentous phage virion [9, 44].

The pVI is a 112-residue that is mostly hydrophobic. It is predicted by the TMHMM2.0 [77] to contain three transmembrane α helices, with the N terminus in the periplasm and the C terminus in the cytoplasm. This protein is not exposed on the virion surface [44] and is assumed to serve as a "base" or "adaptor" for attachment of pIII in forming the virion cap. The C-, but not the N-terminus of pVI can be used, however, as a point of fusion for display of peptides [72]. The C-terminus therefore appears to be near the surface of the virion.

The length of the pIII pre-protein (including signal sequence) and mature protein is 424 and

406 amino acids, respectively. It is composed of three domains (N1, N2 and C) separated by long glycine-rich linkers (Fig. 1.2h). Prior to assembly into the virion, pIII is targeted to the inner membrane by its N-terminal signal sequence and anchored in the phospholipid bilayer by a hydrophobic transmembrane α helix near the C-terminus, in a SecYEG and SecA-dependent manner [9]. Given the position of the membrane anchor, most of pIII is localized in the periplasm prior to assembly into the virion; only five C-terminal residues are located in the cytoplasm [33, 34].

The pIII copy number per virion was indicated as a by-product of a nanotechnology application, in which ZnS nanocrystals were nucleated by N-terminally displayed peptides [81]. The TEM of the nanocrystals associated with the virion tip shows five ZnS nanocrystals, corresponding to five copies of pIII (Fig. 1.2f). Since pVI and pIII are equimolar in the virion cap [59], if there are five copies of pIII, there are also five copies of pVI subunits per virion. Thus the distal cap appears to maintain the fivefold axial symmetry of the major coat protein in the filament shaft.

The N1 and N2 domains of pIII interact with the host receptors; the structure of these two domains has been determined using X-ray crystallography and NMR (Fig. 1.1g; [70, 87]). The three-dimensional structure of the C-domain, which is required for assembly of the virion end-cap, termination of phage assembly, formation of a detergent-resistant virion cap and for late steps in phage infection [6, 122], is yet to be determined.

Display on full-length pIII results in up to five fusion copies per virion and is used preferentially for screening of the naïve antibody libraries, where low-affinity binders are expected [60]. However, display at the N-terminus of the mature full-length pIII often leads to proteolysis of the displayed moiety, reducing the copy number of displayed protein per virion, and increasing infectivity of the particles, providing advantage to these variants in amplification steps during library screening, without binding to the bait [101]. It was recently suggested that this problem might be overcome by interdomain display, where the coding sequence for the displayed peptide (protein) is inserted between the pIII N1 and N2 or N2 and C domains [150]. This strategy eliminates clones containing inserts that are susceptible to proteolysis, by rendering the virion non-infectious due to the loss of the N1, or both N1 and N2 domains, depending on the site of the insert.

Not only full-length, but also the truncated pIII, containing solely the C-terminal domain, are used as a platform for display of antibody variable domains. The fusions in this case are inserted between a signal sequence and the pIII C domain, and have to be co-expressed with wild-type pIII in the same cell and co-assembled into the virion [3] in order to allow amplification of the virions i*n E. coli*. The insert in these vectors does not affect infectivity of the virion; however the C-terminal fusions are displayed only in one or two copies per virion, resulting in low avidity. This mode of display is used for screening the libraries derived from immunized individuals, which results in selection of high-affinity recombinant antibodies [2].

An antibody specific for the C-terminal 10 residues of pIII cannot bind to pIII when it is in the virion (Rakonjac, unpublished). Therefore, the C-terminus must be buried within the virion cap. Nevertheless, it was reported that, in conjunction with a long flexible linker, it was possible to display proteins at the C-terminus of pIII, albeit in combination with the wild-type pIII copies [50].

1.3 Ff Gene Organisation, Transcription and Translation

The genome of Ff phage is 6407 nt in length (Fig. 1.2a). The genes required for infection, replication and virion assembly are clustered along the genome in three groups: (i) replication (gII/gX and gV); (ii) virion structure (gVII, gIX, gVIII, gIII and gVI); (iii) assembly/secretion (gI/gXI and gIV), and are organised in two operons [106]. gVII, gIX and gVIII (encoding two minor proteins that initiate assembly and the major coat

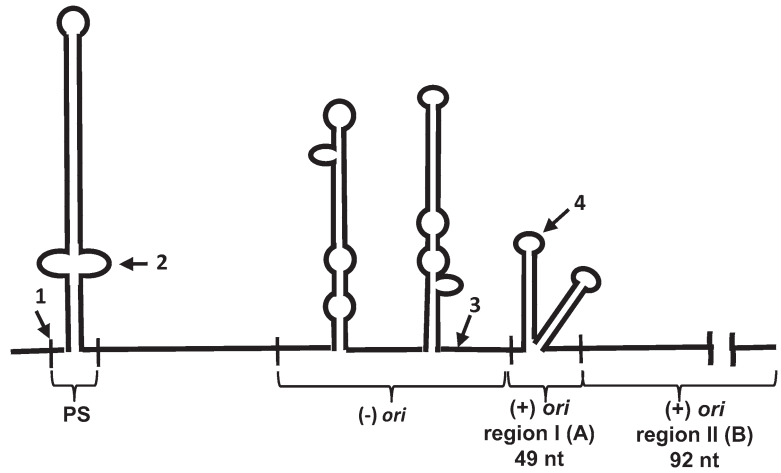

Fig. 1.3 Ff intergenic sequence (IG) containing origin of replication. Packaging signal (*PS*); (−) *ori*, negative strand origin of replication; *(+) ori region I (A)*, the region I of the positive strand origin of replication; *(+) ori region II (B)*, the region II (B) of the positive strand origin of replication; *(1)* stop codon of gIV; *(2)* terminator of gIV mRNA; *(3)* initiation site of (−) strand primer synthesis; *(4)* initiation site of (+) strand synthesis

protein, respectively) are transcribed within an operon that also contains genes gII/gX and gV (encoding replication proteins), whereas gIII and gVI (genes encoding two minor virion proteins that terminate assembly), are in the same operon with downstream gI/gXI and gIV (encoding the secretion/assembly machinery). Two genes, gX and gXI, encode truncated translational products of gII and gI, respectively, and these shorter proteins are required, respectively, for regulation of phage replication and for assembly [62, 106]. Upstream of the replication module, there is an intergenic region (IG; Fig. 1.3) that contains origins for positive (+) and negative (−) strand replication and a packaging signal [162].

The large body of published work on Ff transcription, translation and replication has been used recently for developing a mathematical model of the Ff life cycle and performing simulations of single- and multi-generation infection cycles [143, 144]. A number of early publications on Ff phage were devoted to transcription, resulting in a good qualitative and quantitative picture of transcription and translation (Reviewed by [106]) Both operons, gII-gV-gVII-gIX-gVIII and gIII-gVI-gI-gIV end with strong terminators. Multiple nested promoters, however, result in a number of overlapping transcripts [43, 107]. These multiple transcripts undergo differential turnover and are translated at different rates [58, 79, 104, 106]. In the absence of transcriptional regulators controlling Ff gene expression, these modes of regulation are sufficient to ensure that the requirements for extremely high levels of phage assembly are met. In particular, high level production (10^5–10^6 per cell per generation) is ensured for the ssDNA-binding protein pV and major coat protein pVIII [82, 143, 144].

In contrast to Ff phage that contain no specific regulators of transcription, in lysogenic filamentous phage, dedicated phage- and host-encoded transcriptional regulators determine the level of phage proteins, which is generally much lower than those of the Ff phage even in the induced state (reviewed in [89, 121]).

1.4 Ff Infection

The primary receptor for the Ff filamentous phage is the tip of the conjugative pilus (F); these phage were isolated originally as “male”-specific bacteriophage [68, 85, 96]. The secondary receptor is the TolQRA complex of inner membrane proteins, highly conserved in Gram-negative bacteria [27]. This secondary receptor appears to