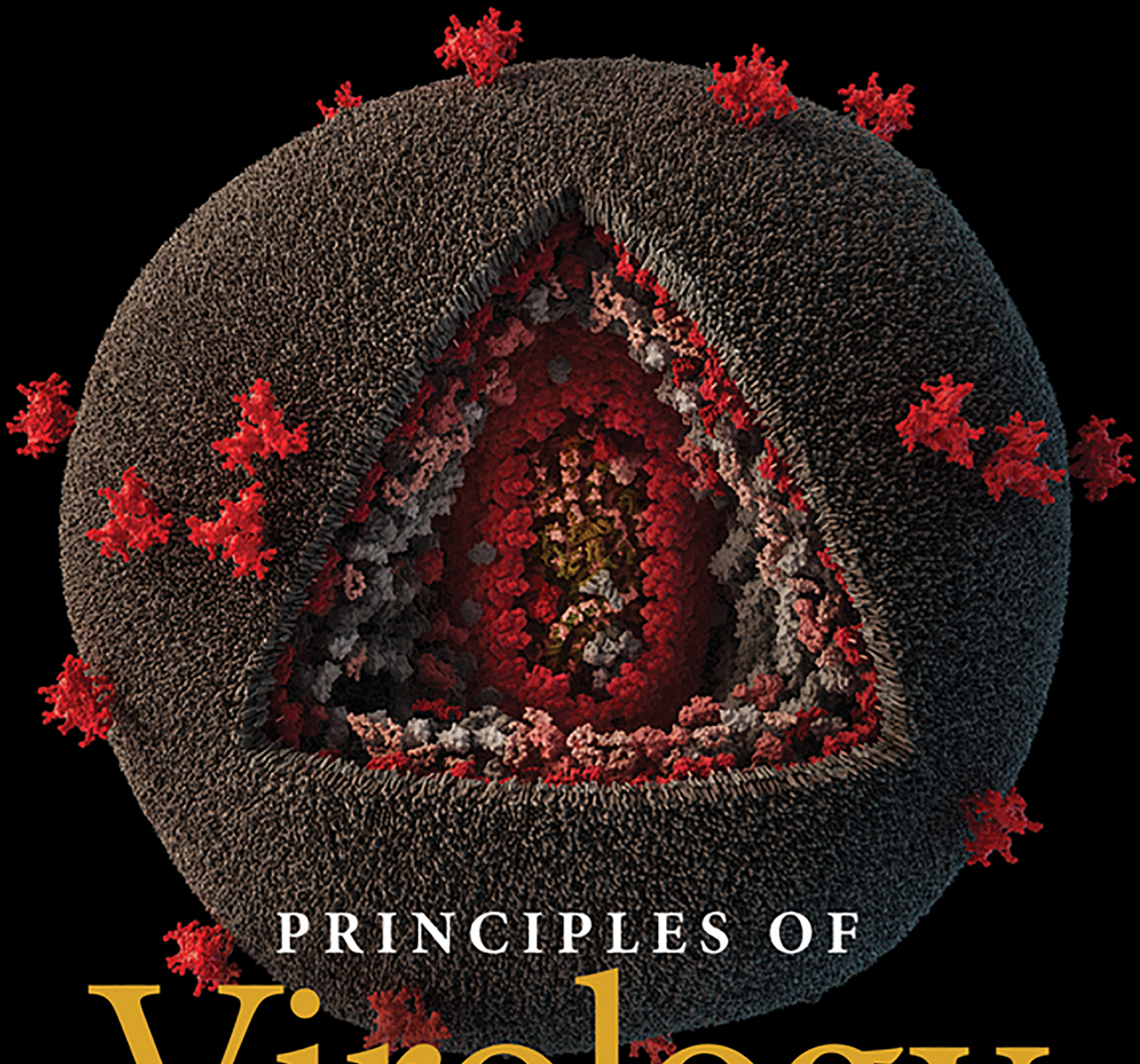
VOLUME I *Molecular Biology*



PRINCIPLES OF Virology

FIFTH EDITION

Jane Flint • Vincent R. Racaniello
Glenn F. Rall • Theodora Hatziioannou
Anna Marie Skalka

VOLUME I *Molecular Biology*

PRINCIPLES OF Virology

FIFTH EDITION

VOLUME I *Molecular Biology*

PRINCIPLES OF Virology

Fifth Edition

Jane Flint
Department of Molecular Biology
Princeton University
Princeton, New Jersey

Vincent R. Racaniello
Department of Microbiology & Immunology
Vagelos College of Physicians and Surgeons
Columbia University
New York, New York

Glenn F. Rall
Fox Chase Cancer Center
Philadelphia, Pennsylvania

Theodora Hatziioannou
The Rockefeller University
New York, New York

Anna Marie Skalka
Fox Chase Cancer Center
Philadelphia, Pennsylvania

ASM PRESS
WASHINGTON, DC

WILEY

Copyright © 2020 American Society for Microbiology. All rights reserved.

Copublication by the American Society for Microbiology and John Wiley & Sons, Inc.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means electronic, mechanical, photocopying, recording, scanning, or otherwise, except as permitted by law. Advice on how to reuse material from this title is available at http://wiley.com/go/permissions.

The right of Jane Flint, Vincent R. Racaniello, Glenn F. Rall, Theodora Hatziioannou, and Anna Marie Skalka to be identified as the author(s) of this work/the editorial material in this work has been asserted in accordance with law.

Limit of Liability/Disclaimer of Warranty
While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy of completeness of the contents of this book and specifically disclaim any implied warranties or merchantability of fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The publisher is not providing legal, medical, or other professional services. Any reference herein to any specific commercial products, procedures, or services by trade name, trademark, manufacturer, or otherwise does not constitute or imply endorsement, recommendation, or favored status by the American Society for Microbiology (ASM). The views and opinions of the author(s) expressed in this publication do not necessarily state or reflect those of ASM, and they shall not be used to advertise or endorse any product.

Editorial Correspondence:
ASM Press, 1752 N Street, NW, Washington, DC 20036-2904, USA

Registered Offices:
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products, visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Flint, S. Jane, author. | Racaniello, V. R. (Vincent R.), author. | Rall, Glenn F., author. | Hatziioannou, Theodora, author. | Skalka, Anna Marie, author.
Title: Principles of virology / Jane Flint, Department of Molecular Biology, Princeton University, Princeton, New Jersey, Vincent R. Racaniello, Department of Microbiology & Immunology, Vagelos College of Physicians and Surgeons, Columbia University, New York, New York, Glenn F. Rall, Fox Chase Cancer Center, Philadelphia, Pennsylvania, Theodora Hatziioannou, The Rockefeller University, New York, New York, Anna Marie Skalka, Fox Chase Cancer Center, Philadelphia, Pennsylvania.
Description: Fifth edition. | Washington, DC : American Society for Microbiology [2020] ; Hoboken, NJ : Wiley, [2020] | Includes bibliographical references and index. | Contents: volume 1. Molecular biology—volume 2. Pathogenesis and control.
Identifiers: LCCN 2020013722 (print) | LCCN 2020013723 (ebook) | ISBN 9781683670322 (set) | ISBN 9781683672845 (v. 1 ; paperback) | ISBN 9781683672852 (v. 2 ; paperback) | ISBN 9781683672821 (v. 1 ; adobe pdf) | ISBN 9781683673606 (v. 1 ; epub) | ISBN 9781683672838 (v. 2 ; adobe pdf) | ISBN 9781683673590 (v. 2 ; epub) | ISBN 9781683670339 (adobe pdf) | ISBN 9781683673583 (epub)
Subjects: LCSH: Virology.
Classification: LCC QR360 .P697 2020 (print) | LCC QR360 (ebook) | DDC 616.9/101—dc23
LC record available at https://lccn.loc.gov/2020013722
LC ebook record available at https://lccn.loc.gov/2020013723

Illustrations and illustration concepting: Patrick Lane, ScEYEnce Studios
Cover image: Visual Science
Cover and interior design: Susan Brown Schmidler

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

We dedicate this book to the students, current and future scientists, physicians, and all those with an interest in the field of virology, for whom it was written.
We kept them ever in mind.

We also dedicate it to our families:
Jonn, Gethyn, and Amy Leedham
Doris, Aidan, Devin, and Nadia
Eileen, Kelsey, and Abigail
Paul, Stefan, and Eve
Rudy, Jeannie, and Chris

Oh, be wiser thou!
Instructed that true knowledge leads to love.

William Wordsworth
Lines left upon a Seat in a Yew-tree
1888

About the Instructor Companion Website

This book is accompanied by a companion website for instructors:

www.wiley.com/go/flint/pov5 

The website includes:

- PowerPoints of figures
- Author podcasts
- Study Questions and Answers

Contents

Preface

The enduring goal of scientific endeavor, as of all human enterprise, I imagine, is to achieve an intelligible view of the universe. One of the great discoveries of modern science is that its goal cannot be achieved piecemeal, certainly not by the accumulation of facts. To understand a phenomenon is to understand a category of phenomena or it is nothing. Understanding is reached through creative acts.

A. D. HERSHEY
Carnegie Institution Yearbook 65

All five editions of this textbook have been written according to the authors' philosophy that the best approach to teaching introductory virology is by emphasizing shared principles. Studying the common steps of the viral reproductive cycle, illustrated with a set of representative viruses, and considering mechanisms by which these viruses can cause disease provides an integrated overview of the biology of these infectious agents. Such knowledge cannot be acquired by learning a collection of facts about individual viruses. Consequently, the major goal of this book is to define and illustrate the basic principles of virus biology.

In this information-rich age, the quantity of data describing any given virus can be overwhelming, if not indigestible, for student and expert alike. The urge to write more and more about less and less is the curse of reductionist science and the bane of those who write textbooks meant to be used by students. In the fifth edition, we continue to distill information with the intent of extracting essential principles, while providing descriptions of how the information was acquired and tools to encourage our readers' exploration of the primary literature. Boxes are used to emphasize major principles and to provide supplementary material of relevance, from explanations of terminology to descriptions of trailblazing experiments. Our goal is to illuminate process and strategy as opposed to listing facts and figures. In an effort to make the book readable, we have been selective in our choice of viruses that are used as examples. The encyclopedic *Fields' Virology* [Knipe DM, Howley PM (ed). 2020. *Fields Virology*, 7th ed. Lippincott Williams & Wilkins, Philadelphia, PA] is recommended as a resource for detailed reviews of specific virus families.

What's New

This edition is marked by a welcome addition to the author team. Our new member, Theodora Hatziioannou, brings expertise in retrovirology, entry, and intrinsic immunity, as well as authority regarding ancient Greek mythology and philosophy that the attentive reader will see is generously sprinkled throughout the text.

We have added an important new chapter in Volume II, "Therapeutic Viruses." While the majority of the chapters define how viruses reproduce and cause mayhem to both cell and host, this new chapter turns the tables to discuss how viruses can be beneficial to eliminate tumor cells, deliver therapeutic genes to specific cells, and expand our arsenal of vaccines for prevention of virus-mediated diseases.

The authors continually strive to make this text accessible and relevant to our readers, many of whom are undergraduates, graduate students, and postdoctoral fellows. Consequently, for this edition, we enlisted the aid of more than twenty of these trainees to provide guidance and commentary on our chapters and ensure that concepts are clearly explained and that the text is compelling to read. This unique group of editors has been invaluable in the design of all of our fully reworked and up-to-date chapters and appendices, and we extend a particular thank-you to them for sharing their perspectives.

A new feature is the inclusion of a set of study questions and/or, in some cases, puzzles, as aids to ensure that the key principles are evident within each chapter. This section complements the Principles that begin each chapter, focusing on unifying core concepts.

Finally, although the SARS-CoV-2 pandemic began as we were preparing to go to press, we have included additions to relevant chapters on the epidemiology, emergence, and replication of this global scourge, as well as some hopeful information concerning vaccine development. What is apparent is that, now more than ever, an appreciation of how viruses impact their hosts is not just an academic pursuit, but rather literally a matter of life and death. We extend our gratitude to all those who serve in patient care settings.

Principles Taught in Two Distinct, but Integrated Volumes

Volume I covers the molecular biology of viral reproduction, and Volume II focuses on viral pathogenesis, control of virus infections, and virus evolution. The organization into two volumes follows a natural break in pedagogy and provides considerable flexibility and utility for students and teachers alike. The two volumes differ in content but are integrated in style and presentation. In addition to updating the chapters and appendices for both volumes, we have organized the material more efficiently, and as noted above, added a new chapter that we believe reflects an exciting direction for the field. Links to Internet resources such as websites, podcasts, blog posts, and movies are provided within each chapter; the digital edition provides one-click access to these materials.

As in our previous editions, we have tested ideas for inclusion in the text in our own classes. We have also received constructive comments and suggestions from other virology instructors and their students. Feedback from our readers was particularly useful in finding typographical errors, clarifying confusing or complicated illustrations, and pointing out inconsistencies in content.

For purposes of readability, references are not included within the text; each chapter ends with an updated list of relevant books, review articles, and selected research papers for readers who wish to pursue specific topics. New to this edition are short descriptions of the key messages from each of the cited papers of special interest. Finally, each volume has a general glossary of essential terms.

These two volumes outline and illustrate the strategies by which all viruses reproduce, how infections spread within a host, and how they are maintained in populations. We have focused primarily on animal viruses, but have drawn insights from studies of viruses that reproduce in plants, bacteria, and archaea.

Volume I: The Science of Virology and the Molecular Biology of Viruses

This volume examines the molecular processes that take place in an infected host cell. Chapter 1 provides a general introduction and historical perspective, and includes descriptions of the unique properties of viruses. The unifying principles that are the foundations of virology,

including the concept of a common strategy for viral propagation, are then described. The principles of the infectious cycle, descriptions of the basic techniques for cultivating and assaying viruses, and the concept of the single-step growth cycle are presented in Chapter 2.

The fundamentals of viral genomes and genetics, and an overview of the surprisingly limited repertoire of viral strategies for genome replication and mRNA synthesis, are topics of Chapter 3. The architecture of extracellular virus particles in the context of providing both protection and delivery of the viral genome in a single vehicle is considered in Chapter 4. Chapters 5 to 13 address the broad spectrum of molecular processes that characterize the common steps of the reproductive cycle of viruses in a single cell, from decoding genetic information to genome replication and production of progeny virions. We describe how these common steps are accomplished in cells infected by diverse but representative viruses, while emphasizing common principles. Volume I concludes with a chapter that presents an integrated description of cellular responses to illustrate the marked, and generally irreversible, impact of virus infection on the host cell.

The appendix in Volume I provides concise illustrations of viral reproductive cycles for members of the main virus families discussed in the text. It is intended to be a reference resource when reading individual chapters and a convenient visual means by which specific topics may be related to the overall infectious cycles of the selected viruses.

Volume II: Pathogenesis, Control, and Evolution

This volume addresses the interplay between viruses and their host organisms. In Chapter 1, we introduce the discipline of epidemiology, and consider basic aspects that govern how the susceptibility of a population is controlled and measured. Physiological barriers to virus infections, and how viruses spread in a host, and to other hosts, are the topics of Chapter 2. The early host response to infection, comprising cell-autonomous (intrinsic) and innate immune responses, are the topics of Chapter 3, while the next chapter considers adaptive immune defenses, which are tailored to the pathogen, and immune memory. Chapter 5 focuses on the classical patterns of virus infection within cells and hosts, and the myriad ways that viruses cause illness. In Chapter 6, we discuss virus infections that transform cells in culture and promote oncogenesis (the formation of tumors) in animals. Next, we consider the principles underlying treatment and control of infection. Chapter 7 focuses on vaccines, and Chapter 8 discusses the approaches and challenges of antiviral drug discovery. In Chapter 9, the new chapter in this edition, we describe the rapidly expanding applications of viruses as therapeutic agents. The origin of viruses, the drivers of viral evolution, and host-virus conflicts are the subjects of Chapter 10. The principles of emerging virus infections, and humankind's experiences with epidemic and pandemic viral infections, are considered in Chapter 11. Chapter 12 is devoted entirely to the "AIDS virus," human immunodeficiency virus type 1, not only because it is the causative agent of the most serious current worldwide epidemic but also because of its unique and informative interactions with the human immune defenses. Volume II ends with a chapter on unusual infectious agents, viroids, satellites, and prions.

The Appendix of Volume II affords snapshots of the pathogenesis of common human viruses. This appendix has been completely re-envisioned in this edition, and now includes panels that define pathogenesis, vaccine and antiviral options, and the course of the infection through the human body. This consistent format should allow students to find information more easily, and compare properties of the selected viruses.

For some behind-the-scenes information about how the authors created the previous edition of *Principles of Virology*, see: http://bit.ly/Virology_MakingOf.

Acknowledgments

These two volumes of *Principles* could not have been composed and revised without help and contributions from many individuals. We are most grateful for the continuing encouragement from our colleagues in virology and the students who use the text. Our sincere thanks also go to colleagues who have taken considerable time and effort to review the text in its evolving manifestations. Their expert knowledge and advice on issues ranging from teaching virology to organization of individual chapters and style were invaluable and are inextricably woven into the final form of the book.

We also are grateful to those who gave so generously of their time to serve as expert reviewers of individual chapters or specific topics in these two volumes: Siddharth Balachandran (Fox Chase Cancer Center), Paul Bieniasz (Rockefeller University), Christoph Seeger (Fox Chase Cancer Center), and Laura Steel (Drexel University College of Medicine). Their rapid responses to our requests for details and checks on accuracy, as well as their assistance in simplifying complex concepts, were invaluable.

As noted in "What's New," we benefited from the efforts of the students and postdoctoral fellows who provided critiques on our chapters and helped to guide our revisions: Pradeep Morris Ambrose, Ruchita Balasubramanian, Mariana Nogueira Batista, Pierre Michel Jean Beltran, Marni S. Crow, Qiang Ding, Florian Douam, Jenna M. Gaska, Laura J. Halsey, Eliana Jacobson, Orkide O. Koyuncu, Robert LeDesma, Rebecca Markham, Alexa McIntyre, Katelynn A. Milora, Laura A. M. Nerger, Morgan Pantuck, Chen Peng, Katrien Poelaert, Daniel Poston, Anagha Prasanna, Pavithran T. Ravindran, Inna Ricardo-Lax, Fabian Schmidt, Andreas Solomos, Nikhila Shree Tanneti, Sharon M. Washio, Riley M. Williams, and Kai Wu.

Since the inception of this work, our belief has been that the illustrations must complement and enrich the text. The illustrations are an integral part of the text, and credit for their execution goes to the knowledge, insight, and artistic talent of Patrick Lane of ScEYEnce Studios. A key to common figure elements is provided following the "About the Authors" section. As noted in the figure legends, many could not have been completed without the help and generosity of numerous colleagues who provided original images. Special thanks go to those who crafted figures or videos tailored specifically to our needs, or provided multiple pieces in this latest edition: Jônatas Abrahão (Universidade Federal de Minas Gerais), Mark Andrake (Fox Chase Cancer Center), Irina Arkhipova (Marine Biological Laboratory, Woods Hole), Brian Baker (University of Notre Dame), Ben Beaden (Australia Zoo, Queensland), Paul Bieniasz (Rockefeller University), Kartik Chandran (Albert Einstein College of Medicine), Elliot Lefkowitz (University of Alabama), Joseph Pogliano (University of California,

San Diego), B.V. Venkatar Prasad and Liya Hu (Baylor College of Medicine), Bonnie Quigley (University of the Sunshine Coast, Australia), Jason Roberts (Victorian Infectious Diseases Reference Laboratory, Doherty Institute, Melbourne, Australia), Michael Rout (Rockefeller University), and Nuria Verdaguer (Molecular Biology Institute of Barcelona, CSIC).

The collaborative work undertaken to prepare the fifth edition was facilitated greatly by several authors' retreats. ASM Press generously provided financial support for these as well as for our many other meetings over the three years that this edition has been in preparation. We thank all those who guided and assisted in its production: Christine Charlip (Director, ASM Press) for her enduring support of our efforts; Megan Angelini (Managing Developmental Editor, ASM Press) for steering us through the complexities inherent in a team effort, and for keeping us on track during production; Susan Schmidler for her elegant and creative designs for the layout and cover; and Lindsay Williams (Editorial Rights Coordinator, ASM Press) for obtaining permissions for images and figures.

There is little doubt that in undertaking such a massive effort typographical errors and/or confusing statements still remain; we hope that the readership of this edition will help to remedy any mistakes. Even so, the three authors who have been part of this endeavor since it was first published in 1995, and the two who joined along the way, feel that with each new edition we get closer to our idealized vision of what this book would be. We aspire to convey more than information: we hope to educate, excite, and encourage future generations of science consumers. As Antoine de Saint-Exupéry, author of *The Little Prince*, once said: "If you want to build a ship, don't drum up the workers to gather wood, divide the labor, and give orders. Instead, teach them to yearn for the vast and endless sea."

This often-consuming enterprise was made possible by the emotional, intellectual, and logistical support of our families, to whom the two volumes are dedicated.

About the Authors

L to R: Jane Flint, Vincent Racaniello, Theodora Hatziioannou, Ann Skalka, Glenn Rall

Jane Flint is a Professor Emerita of Molecular Biology at Princeton University. Dr. Flint's research focused on investigation of the molecular mechanisms by which viral gene products modulate host cell pathways and antiviral defenses to allow efficient reproduction in normal human cells of adenoviruses, viruses that are widely used in such therapeutic applications as gene transfer and cancer treatment. Her service to the scientific community includes membership on various editorial boards, several NIH study sections, and the NIH Recombinant DNA Advisory Committee.

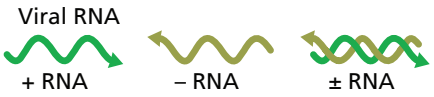
Vincent R. Racaniello is Higgins Professor of Microbiology & Immunology at Columbia University Vagelos College of Physicians & Surgeons. Dr. Racaniello has been studying viruses for over 40 years, including poliovirus, rhinovirus, enteroviruses, hepatitis C virus, and Zika virus. He teaches virology to undergraduate, graduate, medical, dental, and nursing students and uses social media to communicate the subject outside of the classroom. His Columbia University undergraduate virology lectures have been viewed by thousands at iTunes University, Coursera, and on YouTube. Vincent blogs about viruses at virology.ws and is host of the popular science program *This Week in Virology*, which, together with six other science podcasts, can be found at microbe.tv.

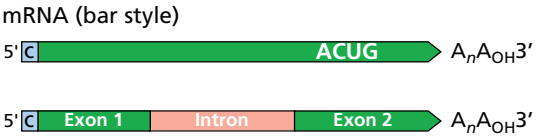
Glenn F. Rall is a Professor and the Chief Academic Officer at the Fox Chase Cancer Center in Philadelphia. He is an Adjunct Professor in the Microbiology and Immunology departments at the University of Pennsylvania and Thomas Jefferson, Drexel, and Temple Universities. Dr. Rall's laboratory studies viral infections of the brain and the immune responses to those infections, with the goal of defining how viruses contribute to disease in humans. His service to the scientific community includes former membership on the Autism Speaks Scientific Advisory Board, Editor of *PLoS Pathogens*, Career Development Chair and Program Chair of the American Society for Virology, and membership on multiple NIH grant review panels.

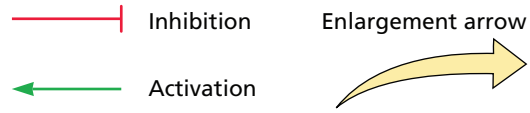
Theodora Hatziioannou is a Research Associate Professor at Rockefeller University in New York. Throughout her career, Dr. Hatziioannou has worked on multiple viruses, with a particular focus on retroviruses and the molecular mechanisms that govern virus tropism and on the improvement of animal models for human disease. She is actively involved in teaching programs at the Rockefeller University and the Albert Einstein College of Medicine, is an editor of *Journal of General Virology*, and serves as a reviewer for multiple scientific journals and NIH grant review panels.

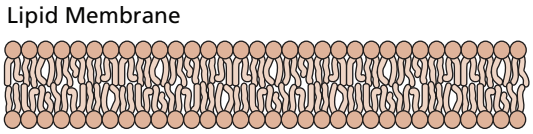
Anna Marie Skalka is a Professor Emerita and former Senior Vice President for Basic Research at the Fox Chase Cancer Center in Philadelphia. Dr. Skalka's major research interests are the molecular aspects of retrovirus biology. Dr. Skalka is internationally recognized for her contributions to the understanding of the biochemical mechanisms by which such viruses (including the AIDS virus) replicate and insert their genetic material into the host genome. Both an administrator and researcher, Dr. Skalka has been deeply involved in state, national, and international advisory groups concerned with the broader, societal implications of scientific research. She has also served on the editorial boards of peer-reviewed scientific journals and has been a member of scientific advisory boards including the National Cancer Institute Board of Scientific Counselors, the General Motors Cancer Research Foundation Awards Assembly, the Board of Governors of the American Academy of Microbiology, and the National Advisory Committee for the Pew Biomedical Scholars.

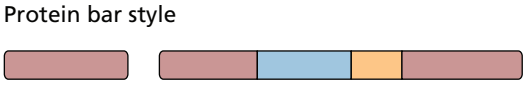
Key of Repetitive Elements

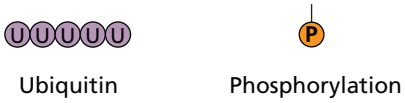


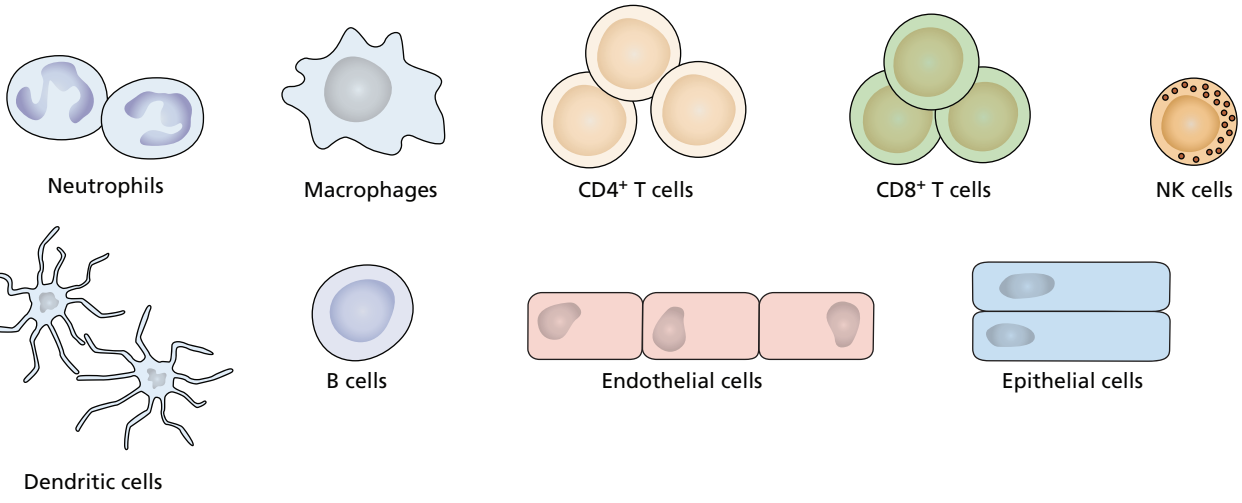






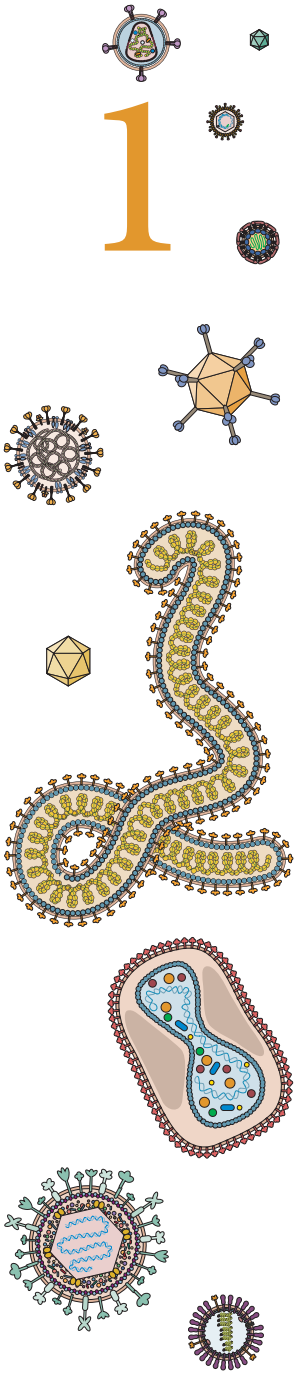






PART I

The Science of Virology



1 Foundations

LINKS FOR CHAPTER 1

- *Video: Interview with Dr. Donald Henderson*
  http://bit.ly/Virology_Henderson
- *This Week in Virology (TWIV): A weekly podcast about viruses featuring informal yet informative discussions and interviews with guests about the latest topics in the field*
  http://www.microbe.tv/twiv
- *Marine viruses and insect defense*
  http://bit.ly/Virology_Twiv301
- *Giants among viruses*
  http://bit.ly/Virology_Twiv261
- *Whiter reefs, fresh breath*
  http://www.microbe.tv/twiv/twiv-391/
- *Latest update of virus classification from the ICTV*
  https://talk.ictvonline.org/taxonomy/
- *The abundant and diverse viruses of the seas*
  http://bit.ly/Virology_3-20-09
- *How many viruses on Earth?*
  http://bit.ly/Virology_9-6-13

Thus, we cannot reject the assumption that the effect of the filtered lymph is not due to toxicity, but rather to the ability of the agent to replicate.

F. Loeffler, 1898

Luria's Credo

"There is an intrinsic simplicity of nature and the ultimate contribution of science resides in the discovery of unifying and simplifying generalizations, rather than in the description of isolated situations—in the visualization of simple, overall patterns rather than in the analysis of patchworks." More than half a century has passed since Salvador Luria wrote this credo in the introduction to the classic textbook *General Virology*.

Despite an explosion of information in biology since Luria wrote these words, his vision of unity in diversity is as relevant now as it was then. That such unifying principles exist may not be obvious considering the bewildering array of viruses, genes, and proteins recognized in modern virology. Indeed, new viruses are being described regularly, and viral diseases such as acquired immunodeficiency syndrome (AIDS), hepatitis, and influenza continue to challenge our efforts to control them. Yet Luria's credo still stands: even as our knowledge of viruses continues to increase, it is clear that their reproduction and survival depend on similar pathways. This insight has been hard-won over many years of observation, research, and debate; the history of virology is rich and instructive.

Viruses Defined

Viruses are microscopic infectious agents that can reproduce only inside a cell that they infect: they are **obligate parasites** of their host cells. Viruses spread from cell to cell via infectious particles called **virions**, which contain genomes comprising RNA or DNA surrounded by a protective protein coat. Upon particle entry and disassociation in a host cell, the viral genome directs synthesis of viral components by cellular systems. Progeny virus particles are formed in the infected cell by *de novo* self-assembly from the newly synthesized components.

As will be discussed in the following chapters, advances in knowledge of the structure of virus particles and the mechanisms by which they are produced in their host cells have been accompanied by increasingly accurate definitions of these unique agents. The earliest pathogenic viruses, distinguished by their small size and dependence on a host organism for reproduction, emphasized the importance of viruses as agents of disease. But there are many other important reasons to study viruses.

Why We Study Viruses

Viruses Are Everywhere

Viruses are all around us, comprising an enormous proportion of our environment, in both number and total mass (Box 1.1). All living things encounter billions of virus particles every day. For example, they enter our lungs in the 6 liters of air each of us inhales every minute; they enter our digestive systems with the food we eat; and they are transferred to our eyes, mouths, and other points of entry from the surfaces we touch and the people with whom we interact. Viral nucleic acids (the **virome**) can be found in the respiratory, gastrointestinal, and urogenital tracts even of normal, healthy individuals (Fig. 1.1). Our bloodstreams harbor up to 100,000 virus particles per milliliter. In addition to viruses that can infect us, our intestinal tracts are loaded with myriad plant and insect viruses, as well as many hundreds of bacterial species that harbor their own constellations of viruses.

PRINCIPLES *Foundations*

- Viruses are obligate intracellular parasites and depend on their host cell for all aspects of their reproduction.
- The field of virology encompasses viral discovery; the study of virus structure and reproduction; and the importance of viruses in biology, ecology, and disease.
- This text focuses primarily on viruses that infect vertebrates, especially humans, but it is important to keep in mind that viruses infect **all** living things including insects, plants, and bacteria.
- Viruses are not solely pathogenic nuisances; they can be beneficial. Viruses contribute to ecological homeostasis, keep our immune responses activated and alert, and can be used as molecular flashlights to illuminate cellular processes.
- Viruses have been part of all of human history: they were present long before *Homo sapiens* evolved, and the majority of human infections were likely acquired from other animals (zoonoses).
- While Koch's postulates were essential for defining many agents of disease, not all pathogenic viruses can be shown to fulfill these criteria.
- Viruses can be described based on their appearance, the hosts they infect, or the nature of their nucleic acid genome.
- All viruses must produce mRNA that can be translated by cellular ribosomes. The Baltimore classification allows relationships among viruses with RNA or DNA genomes to be determined based on the pathway required for mRNA production.
- A common program underlies the propagation of all viruses. This textbook describes that strategy and the similarities and differences in the manner in which different viruses are reproduced, spread, and cause disease.

Viruses Infect All Living Things

While most of this textbook focuses on viral infections of humans, it is important to bear in mind that viruses also infect pets, domestic and wild animals, plants, and insects throughout the world. They infect microbes such as algae, fungi, and bacteria, and some even interfere with the reproduction of other viruses. Viral infection of agricultural plants and animals can have enormous economic and societal impact. Outbreaks of infection by foot-and-mouth disease and avian influenza viruses have led to the destruction (**culling**) of millions of cattle, sheep, and poultry, including healthy animals, to prevent further spread. Losses in the United Kingdom during the 2001 outbreak of foot-and-mouth disease ran into billions of dollars, and caused havoc for both farmers and the government (Box 1.2). More recent outbreaks of the avian influenza virus H5N1 and other strains in Asia have resulted in similar disruption and economic loss. Viruses that infect crops such as potatoes and fruit trees are common, and can lead to serious food shortages as well as financial devastation.

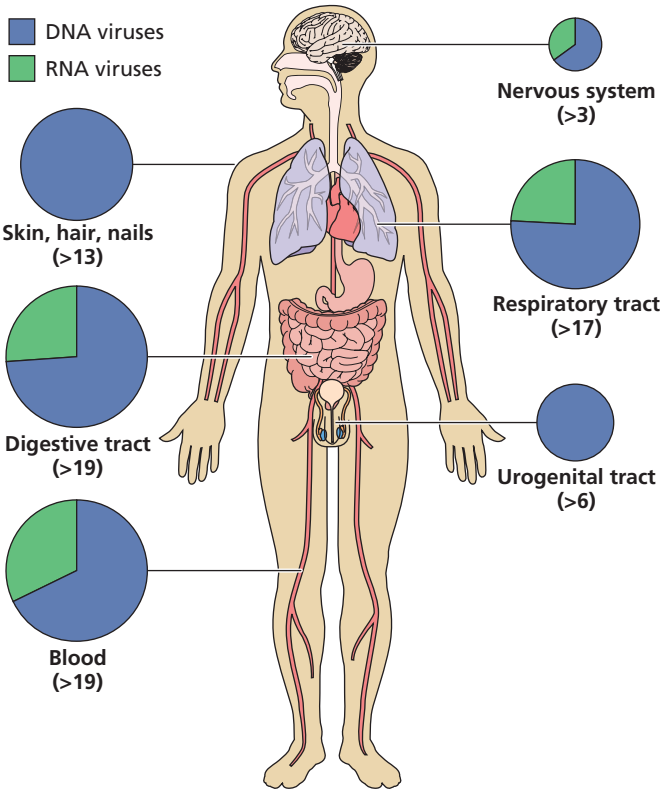


Figure 1.1 The human virome. Our knowledge of the diversity of viruses that can be present in or on a normal human (including some potential pathogens) has increased greatly with the development of high-throughput sequencing techniques and new bioinformatic tools. Current estimates of the numbers of distinct viral families with DNA or RNA genomes in various sites are in parentheses; the > symbol signifies the presence of additional viruses not yet assigned to known families. The numbers may increase as diagnostic tools improve and new viral families are identified. Data from Popgeorgiev N et al. 2013. *Intervirology* 56:395-412; see also http://www.virology.ws/2017/03/23/the-viruses-in-your-blood/.

BOX 1.1

BACKGROUND

Some astounding numbers

- Viruses are the most abundant entities in the biosphere. The biomass on our planet of bacterial viruses *alone* exceeds that of all of Earth's elephants by more than 1,000-fold. There are more than 10^{30} particles of bacterial viruses in the world's oceans, enough to extend out into space for 200 million light-years if arranged head to tail (http://www.virology.ws/2009/03/20/the-abundant-and-diverse-viruses-of-the-seas/; http://www.phagehunter.org/2008/09/how-far-do-those-phages-stretch.html).
- Whales are commonly infected with a member of the virus family *Caliciviridae* that causes rashes, blisters, intestinal problems, and diarrhea, and that can also infect humans. Infected whales excrete more than 10^{13} calicivirus particles daily.
- The average human body contains approximately 10^{13} cells, but almost an equal number of bacteria, and as many as 100-fold more virus particles.
- With about 10^{16} human immunodeficiency virus type 1 (HIV-1) genomes on the planet today, it is highly probable that somewhere there exist HIV-1 genomes that are resistant to every one of the antiviral drugs that we have now or are likely to have in the future.

Viruses reside in Earth's vast oceans and everywhere else on our planet. Courtesy of NASA's Earth Observatory, Suomi NPP satellite image courtesy of NASA/GSFC.

BOX 1.2

DISCUSSION

The first animal virus discovered remains a scourge today

Foot-and-mouth disease virus infects domestic cattle, pigs, and sheep, as well as many species of wild animals. Although mortality is low, morbidity (illness) is high and infected farm animals lose their commercial value. The virus is highly contagious, and the most common and effective method of control is by the slaughter of entire herds in affected areas.

Outbreaks of foot-and-mouth disease were widely reported in Europe, Asia, Africa, and South and North America in the 1800s. The largest epidemic ever recorded in the United States occurred in 1914. After entry into the Chicago stockyards, the virus spread to more than 3,500 herds in 22 states. This calamity accelerated epidemiological and disease control programs, eventually leading to the field- and laboratory-based systems maintained by the U.S. Department of Agriculture to protect domestic livestock from foreign animal and plant diseases. Similar control systems have been established in other Western countries, but this virus still presents a formidable challenge throughout the world. A 1997 outbreak of foot-and-mouth disease among pigs in Taiwan resulted in economic losses of greater than $10 billion.

In 2001, an epidemic outbreak in the United Kingdom spread to other countries in Europe and led to the slaughter of more than 6 million infected and uninfected farm animals. The associated economic, societal, and political costs jolted the British government. Images of mass graves and horrific pyres consuming the corpses of dead animals (see figure) sensitized the public as never before. Minor outbreaks that occurred later in the United Kingdom and parts of Asia were also controlled by culling. But in 2011, South Korea was reported to have destroyed 1.5 million pigs, roughly 12% of its population, to curb a more serious outbreak spread of the virus.

Hunt J.3 January 2013. Foot-and-mouth is knocking on Europe's door. *Farmers Weekly*. http://www.fwi.co.uk/articles/03/01/2013/136943/foot-and-mouth-is-knocking-on-europe39s-door.htm.

Murphy FA, Gibbs EPJ, Horzinek MC, Studdert MJ. 1999. *Veterinary Virology*, 3rd ed. Academic Press, Inc, San Diego, CA.

Mass burning of cattle carcasses during the 2001 foot-and-mouth disease outbreak in the United Kingdom. Courtesy of Dr. Pamela Hullinger, California Department of Food and Agriculture.

Viruses Can Cause Human Disease

With such constant exposure, it is nothing short of amazing that the vast majority of viruses that infect us have little or no impact on our health or well-being. As described in Volume II, we owe such relative safety to our elaborate immune defense systems, which have evolved under the selective pressure imposed by microbial infection. When these defenses are compromised, even the most common infection can be lethal. Despite such defenses, some of the most devastating human diseases have been or still are caused by viruses; these diseases include smallpox, yellow fever, poliomyelitis, influenza, measles, and AIDS. Viral infections can lead to life-threatening diseases that impact virtually all organs, including the lungs, liver, central nervous system, and intestines. Viruses are responsible for approximately 15% of the human cancer burden, and viral infections of the respiratory and gastrointestinal tracts kill millions of children in the developing world each year. As summarized in Volume II, Appendix, there is no question about the biomedical importance of these agents.

Viruses Can Be Beneficial

Despite the appalling statistics from human and agricultural epidemics, it is important to realize that viruses can also be beneficial. Such benefit can be seen most clearly in the marine ecosystem, where virus particles are the most abundant biological entities (Box 1.1). Indeed, they comprise 94% of all nucleic acid-containing particles in the oceans and are 15 times more abundant than *Bacteria* and *Archaea*. Viral infections in the ocean kill 20 to 40% of marine microbes daily, converting these living organisms into particulate matter. In so doing they release essential nutrients that supply phytoplankton

at the bottom of the ocean's food chain, as well as carbon dioxide and other gases that affect the climate of the earth. Pathogens can also influence one another: infection by one virus can have an ameliorating effect on the pathogenesis of a second virus or even bacteria. For example, mice latently infected with some murine herpesviruses are resistant to infection with the bacterial pathogens *Listeria monocytogenes* and *Yersinia pestis*. The idea that viruses are solely agents of disease is giving way to an appreciation of their positive, even necessary, effects, and a realization that their unique properties can actually be harnessed for human benefit (Volume II, Chapter 9).

Viruses "R" Us

Every cell in our body contains viral DNA. Human endogenous retroviruses, and elements thereof, make up about 8% of our genome. Most are inactive, fossil remnants from infections of germ cells that occurred over millions of years during our evolution. Some of them are suspected to be associated with specific diseases, but the regulatory sequences and protein products of other endogenous retroviruses have been coopted during our evolution for their unique functions. For example, retroviral gene products may play a role in the regulation of pluripotency in germ cells, in transmission of signals at neuronal synapses, and clearly in the way that we give birth. The development of the human placenta depends on cell fusion promoted by a retroviral protein. If not for these endogenous retroviruses, we might be producing our young in eggs, like birds and reptiles.

Recent genomic studies have revealed that our viral "heritage" is not limited to retroviruses. Human and other vertebrate genomes harbor sequences derived from several other RNA and DNA viruses. As many of these insertions are estimated to have occurred some 40 million to 90 million years ago, this knowledge has provided unique insight into the ages and evolution of their currently circulating relatives. The conservation of some of these viral sequences in vertebrate genomes suggests that they may have been selected for beneficial properties over evolutionary time.

Viruses Can Cross Species Boundaries

Although viruses generally have a limited host range, they can and do spread across species barriers. As the world's human population continues to expand and impinge on the wilderness, cross-species (**zoonotic**) infections of humans are occurring with increasing frequency. In addition to the AIDS pandemic, the highly fatal Ebola hemorrhagic fever, severe acute respiratory syndrome (SARS), and Middle East respiratory syndrome (MERS) are recent examples of viral diseases to emerge from zoonotic infections. The influenza virus H5N1 continues to spread among poultry and wild birds in areas of the Middle East and Asia. The virus is deadly to humans who catch it from infected birds. The frightening possibility that it could gain the ability to spread among humans is a major incentive for monitoring for person-to-person transmission in case of infection by this and other pathogenic avian influenza viruses. Given the eons over which viruses have had the opportunity to interact with various species, today's "natural" host may simply be a way station in viral evolution.

Viruses Are Unique Tools To Study Biology

Because viruses are dependent on their hosts for propagation, studies that focus on viral reprogramming of cellular mechanisms have provided unique insights into genetics, cellular biology, and functioning of host defenses. Groundbreaking studies of viruses that infect bacteria (called **bacteriophages**) in the mid-20th century established the molecular basis of genetic inheritance. Through development and use of stringent, quantitative methods with these relatively simple biological entities, this research confirmed that DNA encodes genes and genes encode proteins. General mechanisms of genetic recombination, repair, and control of gene expression were also elucidated, laying the foundations of modern molecular biology and recombinant DNA technology. Subsequent studies of animal viruses established many fundamental principles of cellular function, including the presence of intervening sequences in eukaryotic genes. The study of cancer (transforming) viruses established the genetic basis of this disease.

With the development of recombinant DNA technology and our increased understanding of viral systems, it has become possible to use viral genomes as vehicles for the delivery of genes to cells and organisms for both scientific and therapeutic purposes. The use of viral vectors to introduce genes into various cells and organisms to study their function has become a standard method in biology. Viral vectors are also being used to treat human disease, for example, via "gene therapy," in which functional genes delivered by viral vectors compensate for faulty genes in the host cells (Volume II, Chapter 9).

The study of viruses has contributed in a unique way to the field of anthropology. As ancient humans moved from one geographic area to another, the viral strains unique to their original locations came along with them. The presence of such strains can be detected by analysis of viral nucleic acids, proteins, and antibodies from ancient human specimens and in modern populations. Together with archeological information, identification of these virological markers has been used to trace the pathways by which humans came to inhabit various regions of our planet (Fig. 1.2).

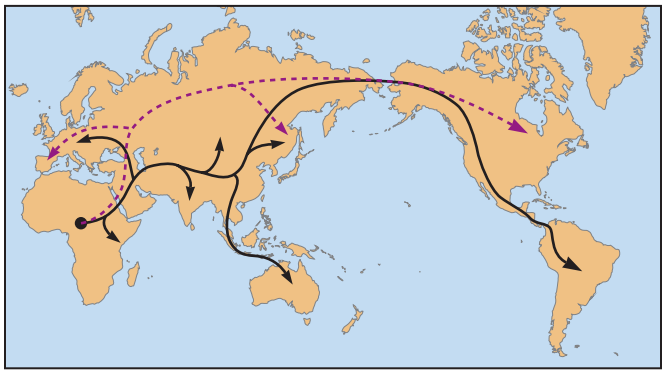


Figure 1.2 Tracking ancient human migrations by the viruses they carried. The polyomavirus known as JC virus is transmitted among families and populations and has coevolved with humans since the time of their origin in Africa. This virus produces no disease in normal, healthy people. Most individuals are infected in childhood, after which the virus establishes a persistent infection in the gastrointestinal tract and is shed in urine. Analysis of the genomes of JC virus in human populations from different geographic locations has suggested an expansion of ancient humans from Africa via two distinct migrations, each carrying a different lineage of the virus. Results from these studies are consistent with analyses of human DNAs (shown by the solid line). They also suggest an additional route that was undetectable in the human DNA analyses (indicated by the dashed line). Data from Pavesi A. 2005. *J Gen Virol* 86:1315–1326.

Virus Prehistory

Although viruses have been known as distinct biological entities for only about 120 years, evidence of viral infection can be found among the earliest recordings of human activity, and methods for combating viral disease were practiced long before the first virus was recognized. Consequently, efforts to understand and control these important agents of disease began only in the last century.

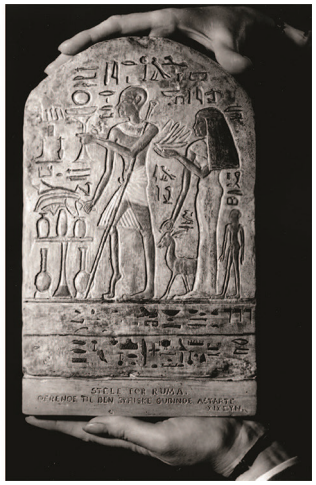
Viral Infections in Antiquity

Reconstruction of the prehistoric past to provide a plausible account of when or how viruses established themselves in human populations is challenging. However, extrapolating from current knowledge, we can deduce that some modern viruses were undoubtedly associated with the earliest precursors of mammals and coevolved with humans. Other viruses entered human populations only recently. The last 10,000 years of history was a time of radical change for humans and our viruses: animals were domesticated, the human population increased dramatically, large population centers appeared, and commerce and technology drove worldwide travel and interactions among unprecedented numbers of people.

Viruses that established themselves in human populations were undoubtedly transmitted from animals, much as still happens today. Early human groups that domesticated and lived with their animals were almost certainly exposed to different viruses than were nomadic hunter/gatherer societies. Similarly, as many different viruses are **endemic** in the tropics, human societies in that environment must have been exposed to a greater variety of viruses than societies established in temperate climates. When nomadic groups met others with domesticated animals, human-to-human contact could have provided new avenues for virus spread. Even so, it seems unlikely that viruses such as those that cause measles or smallpox could have entered a permanent relationship with small groups of early humans. Such highly virulent viruses, as we now know them to be, either kill their hosts or induce lifelong immunity. Consequently, they can survive only when large, interacting host populations offer a sufficient number of naive and permissive hosts for their continued propagation. Such viruses could not have been established in human populations until large, settled communities appeared. Less virulent viruses that enter into a long-term relationship with their hosts were therefore more likely to be the first to become adapted to reproduction in the earliest human populations. These viruses include the modern retroviruses, herpesviruses, and papillomaviruses.

Evidence for knowledge of several diseases that we now know to be caused by viruses can be found in ancient records. The Greek poet Homer characterizes Hector as "rabid" in *The Iliad* (Fig. 1.3A), and Mesopotamian laws that outline the responsibilities of the owners of rabid dogs date from before 1000 B.C.E. Their existence indicates that the communicable nature of this viral disease was already well-known by that time. Egyptian hieroglyphs illustrate what appear to be the consequences of poliovirus infection (a withered leg typical of poliomyelitis [Fig. 1.3B]). Pustular lesions characteristic of smallpox have also been found on Egyptian mummies. The smallpox virus was probably endemic in the Ganges River basin by the fifth century B.C.E. and subsequently spread to other parts of Asia and Europe. This viral pathogen has played an important part in human history. Its introduction into the previously unexposed native populations of Central and South America by colonists in the 16th century led to lethal **epidemics**, which are considered an important factor in the conquests achieved by a small number of European soldiers. Other viral diseases known in ancient times include mumps and, perhaps, influenza. Europeans have described yellow fever since they discovered Africa, and it has been suggested that this scourge of the tropical trade was the basis for legends about ghost ships, such as the *Flying Dutchman*, in which an entire ship's crew perished mysteriously.

Humans have not only been subject to viral disease throughout much of their history but have also manipulated these agents, albeit unknowingly, for much longer than might be imagined. One classic example is the cultivation of marvelously patterned tulips, which were of enormous value in



Here this firebrand, rabid Hector, leads the charge.
Homer, *The Iliad*,
translated by Robert Fagels
(Viking Penguin)

Figure 1.3 References to viral diseases from the ancient literature. (A) An image of Hector from an ancient Greek vase. Courtesy of the Penn Museum, object 30-44-4. **(B)** An Egyptian stele, or stone tablet, from the 18th dynasty (1580–1350 B.C.E.) depicting a man with a withered leg and the "drop foot" syndrome characteristic of poliomyelitis. Image courtesy of SPL/Science Source.

17th-century Holland. Such efforts included deliberate spread of a virus (tulip breaking virus or tulip mosaic virus) that we now know causes the striping of tulip petals so highly prized at that time (Fig. 1.4). Attempts to control viral disease have an even more venerable history.

The First Vaccines

Measures to control one viral disease have been used for the last millennium. The disease is smallpox (Fig. 1.5), and the practice is called **variolation**. The process entails taking material directly from the smallpox lesions of an infected individual and scratching it onto the skin of healthy individuals with a lancet. Widespread in China and India by the 11th century, variolation was based on the recognition that smallpox survivors were protected against subsequent bouts of the disease. Variolation later spread to Asia Minor, where its value was recognized by Lady Mary Wortley Montagu, wife of the British ambassador to the Ottoman Empire. She introduced this practice into England in 1721, where it became quite widespread following the successful inoculation of children of the royal family. George Washington is said to have introduced the practice among Continental Army soldiers in 1776. However, the consequences of variolation were unpredictable and never pleasant: serious skin lesions invariably developed at the site of inoculation and were often accompanied by more generalized rash and disease, with a fatality rate of 1 to 2%. From the comfortable viewpoint of

Figure 1.4 *Three Broken Tulips*. A painting by Nicolas Robert (1624–1685), now in the collection of the Fitzwilliam Museum, Cambridge, United Kingdom. Striping patterns (color breaking) in tulips were described in 1576 in western Europe and were caused by a viral infection. This beautiful image depicts the remarkable consequences of infection with the tulip mosaic virus. © Fitzwilliam Museum, Cambridge.

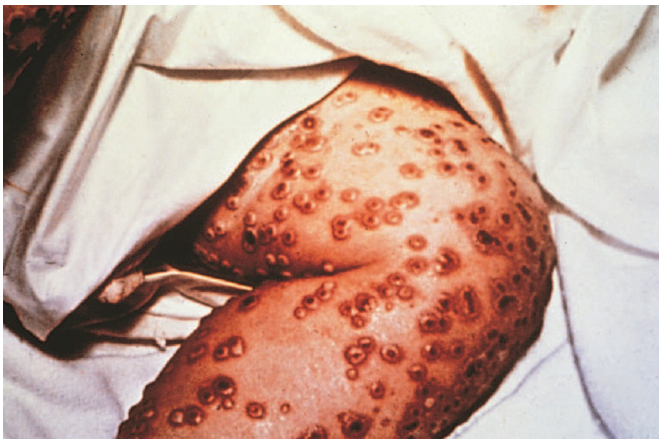


Figure 1.5 Characteristic smallpox lesions in a young victim. Illustrations like these were used as examples to track down individuals infected with the smallpox virus (variola virus) during the World Health Organization campaign to eradicate the disease. Courtesy of CDC/Dr. Robinson (CDC PHIL ID#10398). See also the interview with Dr. Donald Henderson: http://bit.ly/Virology_Henderson.

an affluent country in the 21st century, such a death rate seems unacceptably high. However, in the 18th century, variolation was perceived as a much better alternative than naturally contracting natural smallpox, a disease with a fatality rate of 25% in the whole population and 40% in babies and young children.

In the 1790s, Edward Jenner, an English country physician, established the principle on which modern methods of viral immunization are based, even though viruses themselves were not to be identified for another 100 years. Jenner himself was variolated with smallpox as a young boy and was undoubtedly familiar with its effects and risks. Perhaps this experience spurred his abiding interest in this method. Although it is commonly asserted that Jenner's development of the smallpox vaccine was inspired by his observations of milkmaids, the reality is more prosaic. As a physician's apprentice at age 13, Jenner learned about a curious observation of local practitioners who had been variolating farmers with smallpox. No expected skin rash or disease appeared in farmers who had previously suffered a bout with cowpox. This lack of response was typical of individuals who had survived earlier infection with smallpox and were known to be immune to the disease. It was supposed therefore that, like smallpox survivors, these nonresponding farmers must somehow be immune to smallpox. Although the phenomenon was first observed and later reported by others, Jenner was the first to appreciate its significance fully and to follow up with direct experiments. From 1794 to 1796, he demonstrated that inoculation with material from cowpox lesions induced only mild symptoms in the recipient but protected against the far more dangerous disease. It is from these experiments that we derive the term **vaccination** (*vacca* = "cow" in Latin); Louis Pasteur coined this term in 1881 to honor Jenner's accomplishments.

Initially, the only way to propagate and maintain the cowpox-derived vaccine was by serial infection of human subjects. This method was eventually banned, as it was often associated with transmission of other diseases such as syphilis and hepatitis. By 1860, the vaccine had been passaged in cows; later, horses, sheep, and water buffaloes were also used. The origin of the current vaccine virus, vaccinia virus, is now thought to be horsepox virus (Box 1.3).

The first rabies vaccine was made by Louis Pasteur, although he had no idea at the time that the relevant agent was a virus. In 1885, he inoculated rabbits with material from the brain of a cow suffering from rabies and then used aqueous suspensions of dried spinal cords from these animals to infect other rabbits. After several such passages, the resulting preparations were administered to human subjects, where they produced mild disease but effective immunity against rabies.

Today, viral vaccine strains selected for reduced virulence are called **attenuated**, a term derived from the Latin prefix *ad*, meaning "to," and *tenuis*, meaning "weak." Safer and more efficient methods for the production of larger quantities of these first vaccines awaited the recognition of viruses as distinctive biological entities and parasites of cells in their hosts. Indeed, it took almost 50 years to discover the next antiviral vaccines: a vaccine for yellow fever virus was developed in 1935, and an influenza vaccine was available in 1936. These advances became possible only with radical changes in our knowledge of living organisms and of the causes of disease.

Microorganisms as Pathogenic Agents

The 19th century was a period of revolution in scientific thought, particularly in ideas about the origins of living things. The publication of Charles Darwin's *The Origin of Species* in 1859 crystallized startling (and, to many people, shocking) new ideas about the origin of diversity in plants and animals, until then generally attributed directly to the hand of God. These insights permanently undermined the perception that humans were somehow set apart from all other members of the animal kingdom. From the point of view of the science of virology, the most important changes were in ideas about the causes of disease.

The diversity of macroscopic organisms has been appreciated and cataloged since the dawn of recorded human history. However, a vast new world of organisms too small to be visible to the naked eye was revealed through the microscopes of Antony van Leeuwenhoek (1632–1723). Van Leeuwenhoek's vivid and exciting descriptions of living microorganisms, the "wee animalcules" present in such ordinary materials as rain or seawater, included examples of protozoa, algae, and bacteria. By the early 19th century, the scientific community had accepted the existence of microorganisms and turned to the question of their origin, a topic of fierce debate. Some believed that microorganisms arose spontaneously, for example, in decomposing matter, where they were especially abundant. Others held the view that all were generated by their reproduction, as are macroscopic organisms. The death knell of the spontaneous-generation hypothesis was sounded with the famous experiments of Pasteur. He demonstrated that boiled (i.e., sterilized) medium remained free of microorganisms as long as it was maintained in special flasks with curved, narrow necks designed to prevent entry of airborne microbes (Fig. 1.6). Pasteur also established that distinct microorganisms were associated with specific processes, such as fermentation, an idea that was crucial in the development of modern explanations for the causes of disease.

From the earliest times, poisonous air (miasma) was generally invoked to account for **epidemics** of contagious

BOX 1.3

DISCUSSION

Origin of vaccinia virus

Over the years, many hypotheses have been advanced to explain the curious origin of vaccinia virus. However, recent investigations into this mystery by collaborators in the United States, Germany, and Brazil indicate that horsepox, not cowpox, was the likely precursor of vaccine strains of vaccinia virus.

The proverbial smoking gun was an original wooden and glass container that held capillaries with the smallpox vaccine produced in 1902 by H.K. Mulford in Philadelphia (a company that merged with Sharpe and Dohme in 1929). Sequence analysis of the DNA showed that the core genome of the virus in that vial had the highest degree of similarity (99.7%) to horsepox virus. A review of the historical record shows that during the 19th century, pustular material derived from both cowpox and horsepox lesions was used to immunize against smallpox. The latter technique was called **equination**. Although the disease is now rare in horses and was never reported in the Americas, it was prevalent in Europe, where most vaccine samples were obtained at the time.

Most smallpox vaccines used in the United States, Brazil, and many European countries were produced in the United States from calves inoculated with material collected in 1866 from spontaneous cases of cowpox in France. Genetic analysis of existing samples of these early vaccines indicates that they contained a virus more similar to horsepox and vaccinia viruses than to cowpox virus. While naturally occurring vaccinia viruses are found today only in India (in buffalos) and Brazil (in cows), they can infect horses and people, producing pustular lesions similar to those caused by horsepox and cowpox viruses. One hypothesis is that the ancestor of the current vaccine strain was a naturally occurring vaccinia virus present in the widely distributed French preparation. Alternatively, the vaccine strain may have evolved from horsepox virus during animal passage.

The original wooden (top) and glass (bottom) containers that held capillaries containing the Mulford 1902 smallpox vaccine. Photo kindly provided by Dr. Jose Esparza, Institute of Human Virology, University of Maryland School of Medicine, Baltimore. ©Merck Sharp & Dohme Corp., Merck & Co., Inc.

It is important to consider that development of the smallpox vaccine took place more than a century before modern concepts of virology were established. One can think of other scenarios to explain why the vaccine strain of vaccinia virus is closely related to horsepox and not cowpox, as originally supposed.

- The milkmaid with lesions that were the source of Jenner's original inoculum in 1796 was infected with horsepox, not cowpox. Horsepox can be transmitted to cows, and both animals are common on farms.
- Cows from which pustular material was obtained for vaccination were most often infected with horsepox, transmitted by their handlers or by rodents.

The student is invited to conjure up other plausible explanations.

Damaso CR. 2018. Revisiting Jenner's mysteries, the role of the Beaugency lymph in the evolutionary path of ancient smallpox vaccines. *Lancet Infect Dis* **18:**e55–e63.

Esparza J, Schrick L, Damaso CR, Nitsche A. 2017. Equination (inoculation of horsepox): an early alternative to vaccination (inoculation of cowpox) and the potential role of horsepox virus in the origin of the smallpox vaccine. *Vaccine* **35:**7222–7230.

Schrick L, Tausch SH, Dabrowski PW, Damaso CR, Esparza J, Nitsche A. 2017. An early American smallpox vaccine based on horsepox. *N Engl J Med* **377:**1491–1492.

TWIV 478: A pox on your horse. http://www.microbe.tv/twiv/twiv-478/.

diseases, and there was little recognition of the differences among causative agents. The association of particular microorganisms, initially bacteria, with specific diseases can be attributed to the ideas of the German physician Robert Koch. He developed and applied a set of criteria for identification of the agent responsible for a specific disease (a **pathogen**), articulated in an 1890 presentation in Berlin. These criteria, **Koch's postulates**, can be summarized as follows.

- The organism must be regularly associated with the disease and its characteristic lesions.
- The organism must be isolated from the diseased host and grown in culture.
- The disease must be reproduced when a pure culture of the organism is introduced into a healthy, susceptible host.
- The same organism must be reisolated from the experimentally infected host.

Modern technology has allowed some of Koch's principles to be amended by the application of other types of evidence (Box 1.4). However, by applying his criteria, Koch demonstrated that anthrax, a common disease of cattle, was caused

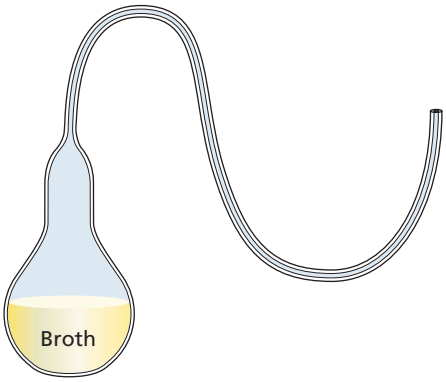


Figure 1.6 Pasteur's famous swan-neck flasks provided passive exclusion of microbes from the sterilized broth. Although the flask was freely open to the air at the end of the long, curved stem, the broth remained sterile, provided that microbe-bearing dust that collected in the neck of the stem did not reach the liquid.

by a specific bacterium (designated *Bacillus anthracis*) and that a second, distinct bacterial species caused tuberculosis in humans. Guided by these postulates and the methods for the sterile culture and isolation of pure preparations of bacteria developed by Pasteur, Joseph Lister, and Koch, many pathogenic bacteria (as well as yeasts and fungi) were identified and classified during the last part of the 19th century (Fig. 1.7). From these beginnings, investigation into the causes of infectious disease was placed on a secure scientific foundation, the first step toward rational treatment and ultimately control. Furthermore, during the last decade of the 19th century, failures of the paradigm that bacterial or fungal agents are responsible for **all** diseases led to the identification of a new class of infectious agents—submicroscopic pathogens that came to be called **viruses**.

Discovery of Viruses

The first report of a pathogenic agent smaller than any known bacterium appeared in 1892. The Russian scientist Dimitrii Ivanovsky observed that the causative agent of tobacco mosaic disease was not retained by the unglazed filters used at that time to remove bacteria from extracts and culture medium (Fig. 1.8A). Six years later in Holland, Martinus Beijerinck independently made the same observation. More importantly, Beijerinck made the conceptual leap that this must be a distinctive agent, because it was so small that it could pass through filters that trapped all known bacteria. However, Beijerinck thought that the agent was an infectious liquid. It was two former students and assistants of Koch, Friedrich Loeffler and Paul Frosch, who in the same year (1898) deduced that such infectious filterable agents comprised small particles: they observed that while the causative agent of foot-and-mouth disease (Box 1.2) passed through filters that held back bacteria, it could be retained by a finer filter.

Not only were the tobacco mosaic and foot-and-mouth disease pathogens much smaller than any previously recognized microorganism, but also they could only reproduce in their host organisms. For example, extracts of an infected tobacco plant diluted into sterile solution produced no additional infectious agents until introduced into leaves of healthy plants, which subsequently developed tobacco mosaic disease. The serial transmission of infection by diluted extracts established that these diseases were not caused by a

BOX 1.4

DISCUSSION

New methods amend Koch's principles

While it is clear that a microbe that fulfills Koch's postulates is almost certainly the cause of the disease in question, we now know that microbes that do not fulfill such criteria may still represent the etiological agents of disease. In the latter part of the 20th century, new methods were developed to associate particular viruses with disease based on immunological evidence of infection, for example, the presence of antibodies in blood. The availability of these methods led to the proposal of modified "molecular Koch's postulates" based on the application of molecular techniques to monitor the role played by virulence genes in bacteria.

The most revolutionary advances in our ability to link particular viruses with disease (or benefit) come from the more recent development of high-throughput nucleic acid sequencing methods and bioinformatics tools that allow detection of viral genetic material directly in environmental or biological samples, an approach called viral metagenomics. Based on these developments, alternative "metagenomic Koch's postulates" have been proposed in which (i) the definitive traits are molecular markers such as genes or full genomes that can uniquely distinguish samples obtained from diseased subjects from those obtained from matched, healthy control subjects and (ii) inoculating a healthy individual with a sample from a diseased subject results in transmission of the disease as well as the molecular markers.

Falkow S. 1988. Molecular Koch's postulates applied to microbial pathogenicity. *Rev Infect Dis* **10**(Suppl 2)**:**S274–S276.

Fredricks DN, Relman DA. 1996. Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* **9:**18–33.

Mokili JL, Rohwer F, Dutilh BE. 2012. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol* **2:**63–77.

Racaniello V. 22 January 2010. Koch's postulates in the 21st century. Virology Blog. http://www.virology.ws/2010/01/22/kochs-postulates-in-the-21st-century/.

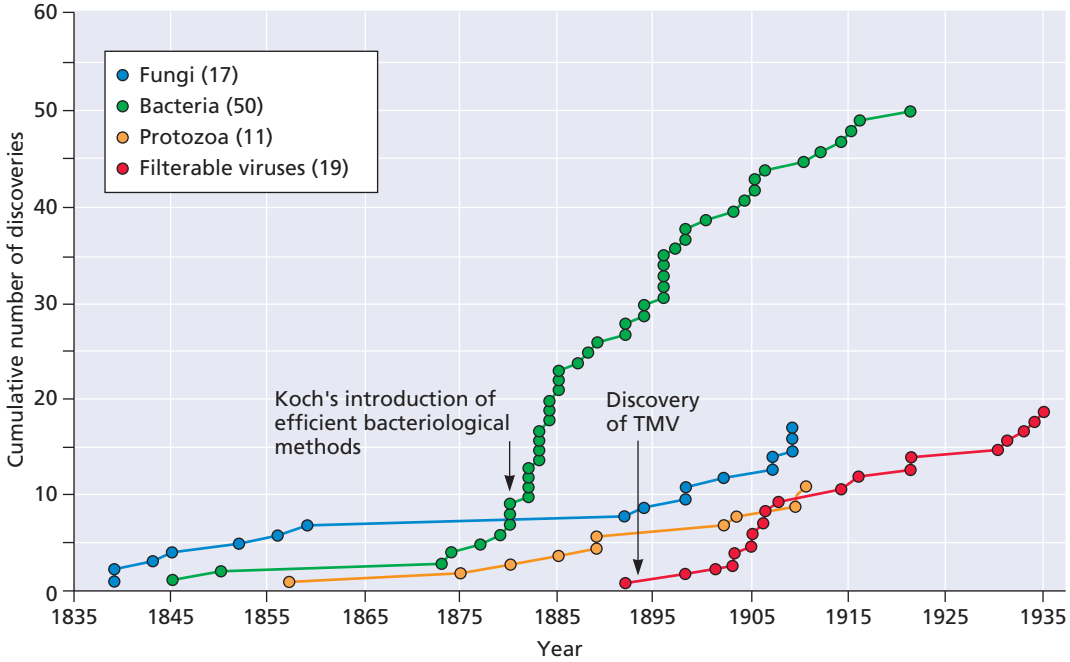


Figure 1.7 The pace of discovery of new infectious agents in the dawn of virology. Koch's introduction of efficient bacteriological techniques spawned an explosion of new discoveries of bacterial agents in the early 1880s. Similarly, the discovery of filterable agents launched the field of virology in the early 1900s. Despite an early surge of virus discovery, only 19 distinct human viruses had been reported by 1935. TMV, tobacco mosaic virus. Data from Burdon KL. 1939. *Medical Microbiology* (Macmillan Co, New York, NY).

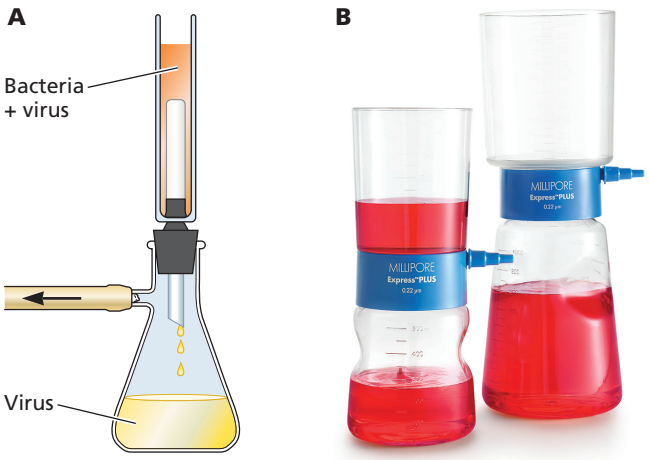


Figure 1.8 Filter systems used to characterize/purify virus particles. (A) The Berkefeld filter, invented in Germany in 1891, was a "candle"-style filter comprising diatomaceous earth (called Kieselguhr), pressed into a hollow candle shape. The white candle in the upper chamber is open at the top to receive the liquid to be filtered. The smallest pore size retained bacteria and let virus particles pass through. Such filters were probably used by Ivanovsky, Loeffler, and Frosch to isolate the first viruses. **(B)** Modern-day filter systems are made of disposable plastic with the upper and lower chambers separated by a biologically inert membrane, available in a variety of pore sizes. Such filtration approaches may have limited our detection of giant viruses. Image courtesy of EMD Millipore Corporation.

bacterial toxin present in the original preparations derived from infected tobacco plants or cattle. The failure of both pathogens to multiply in solutions that readily supported the growth of bacteria, as well as their dependence on host organisms for reproduction, further distinguished these new agents from pathogenic bacteria. Beijerinck termed the submicroscopic agent responsible for tobacco mosaic disease *contagium vivum fluidum* to emphasize its infectious nature and distinctive reproductive and physical properties. Agents passing through filters that retain bacteria came to be called ultrafilterable viruses, appropriating the term *virus* from the Latin for "poison." This term was simplified eventually to "virus."

The discovery of the first virus, tobacco mosaic virus, is often attributed to the work of Ivanovsky in 1892. However, he did not identify the tobacco mosaic disease pathogen as a distinctive agent, nor was he convinced that its passage through bacterial filters was not the result of some technical failure. It may be more appropriate to attribute the founding of the field of virology to the astute insights of Beijerinck, Loeffler, and Frosch, who recognized the distinctive nature of the plant and animal pathogens they were studying more than 120 years ago.

The pioneering work on tobacco mosaic and foot-and-mouth disease viruses was followed by the identification of viruses associated with specific diseases in many other organ-

isms. Important landmarks from this early period include the identification of viruses that cause leukemias or solid tumors in chickens by Vilhelm Ellerman and Olaf Bang in 1908 and Peyton Rous in 1911, respectively. The study of viruses associated with cancers in chickens, particularly Rous sarcoma virus, eventually led to an understanding of the molecular basis of cancer (Volume II, Chapter 6).

The fact that bacteria could also be hosts to viruses was first recognized by Frederick Twort in 1915 and Félix d'Hérelle in 1917. d'Hérelle named such viruses **bacteriophages** because of their ability to cause their bacterial host cells to rupture (a phenomenon called **lysis**; "phage" is derived from the Greek for "eating"). In an interesting twist of serendipity, Twort made his discovery of bacterial viruses while testing the smallpox vaccine virus to see if it would grow on simple media. He found bacterial contaminants, some of which proved to be infected by a bacteriophage. As discussed below, investigation of bacteriophages established not only the foundations for the field of molecular biology but also fundamental insights into how viruses interact with their host cells.

The Defining Properties of Viruses

Throughout the early period of virology when many viruses of plants, animals, and bacteria were cataloged, ideas about the origin and nature of these distinctive infectious agents were quite controversial. Arguments centered on whether viruses originated from parts of a cell or were built from unique components. Little progress was made toward resolving these issues and establishing the definitive properties of viruses until the development of new techniques that allowed their visualization or propagation in cultured cells.

The Structural Simplicity of Virus Particles

Dramatic confirmation of the structural simplicity of virus particles came in 1935, when Wendell Stanley obtained crystals of tobacco mosaic virus. At that time, nothing was known of the structural organization of **any** biologically important macromolecules, such as proteins and DNA. Indeed, the crucial role of nucleic acids as genetic material had not even been recognized. The ability to obtain an infectious agent in crystalline form, a state that was more generally associated with inorganic material, created much wonder and speculation about whether a virus is truly a life form. In retrospect, it is obvious that the relative ease with which this particular virus could be crystallized was a direct result of its structural simplicity.

The 1930s saw the introduction of the instrument that rapidly revolutionized virology: the electron microscope. The great magnifying power of this instrument (eventually more than 100,000-fold) allowed direct visualization of virus particles for the first time. It has always been an exciting experience for investigators to obtain images of viruses, especially as they appear to be remarkably elegant (Fig. 1.9). Images of many different virus particles confirmed that these agents are very small (Fig. 1.10) and that most are far simpler in structure than any cellular organism. Many appeared as regular helical or spherical particles. The description of the morphology of virus particles made possible by electron microscopy also opened the way for the first rational classification of viruses.

The Intracellular Parasitism of Viruses

Organisms as Hosts

A defining characteristic of viruses is their absolute dependence on a living host for reproduction: they are **obligate parasites**. Transmission of plant viruses such as tobacco mosaic virus can be achieved readily, for example, by applying extracts of an infected plant to a scratch made on the leaf of a healthy plant. Furthermore, as a single infectious particle of many plant viruses is sufficient to induce a characteristic lesion (Fig. 1.11), the concentration of the infectious agent could be measured. Plant viruses were therefore the first to be studied in detail. Some viruses of humans and other species could also be propagated in laboratory animals, and methods were developed to quantify them by determining the lethal dose. The transmission of yellow fever virus to mice by Max Theiler in 1930 was an achievement that led to the isolation of an attenuated strain, still considered one of the safest and most effective ever produced for the vaccination of humans.

After specific viruses and appropriate host organisms were identified, it became possible to produce sufficient quantities of virus particles for study of their physical and chemical properties and the consequences of infection for the host. Features such as the incubation period, symptoms of infection, and effects on specific tissues and organs were investigated. Laboratory animals remain an essential tool in investigations of the pathogenesis of viruses that cause disease. However, real progress toward understanding the mechanisms of virus reproduction was made only with the development of cell culture systems. The first and the simplest, but crucial to both virology and molecular biology, were cultures of bacterial cells.

Lessons from Bacteriophages

In the late 1930s and early 1940s, the bacteriophages, or "phages," received increased attention as a result of controversy centering on how they might have arisen. John Northrup, a biochemist at the Rockefeller Institute in Princeton, NJ, championed the theory that a phage was a metabolic product of a bacterium. On the other hand, Max Delbrück, in his work with Emory Ellis and later with Salvador Luria, regarded phages as autonomous, stable, self-replicating entities characterized by heritable traits. According to this paradigm,

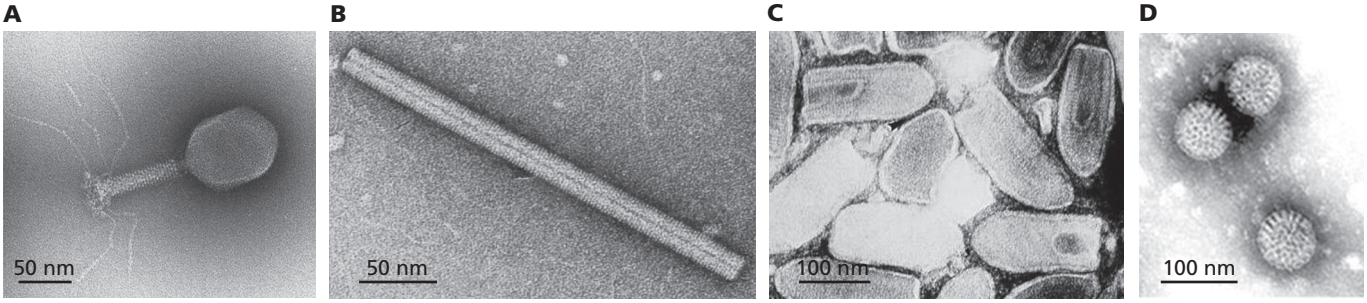


Figure 1.9 Electron micrographs of virus particles following negative staining. **(A)** The complex, nonenveloped virus bacteriophage T4. Note the intricate tail and tail fibers. Reproduced with permission from Dr. Robert L. Duda, University of Pittsburgh, Pittsburgh, PA. **(B)** The helical, nonenveloped particle of tobacco mosaic virus. Courtesy of Plant Resistance Gene Wiki (http://prgdb.crg.eu/wiki/Species:Tobacco_mosaic_virus), licensed under CC BY-SA 3.0. **(C)** Enveloped particles of the rhabdovirus vesicular stomatitis virus. Courtesy of CDC/Dr. Fred. A. Murphy (CDC PHIL ID#5611). **(D)** Nonenveloped, icosahedral human rotavirus particles. Courtesy of F. P. Williams, U.S. Environmental Protection Agency, Washington, DC.

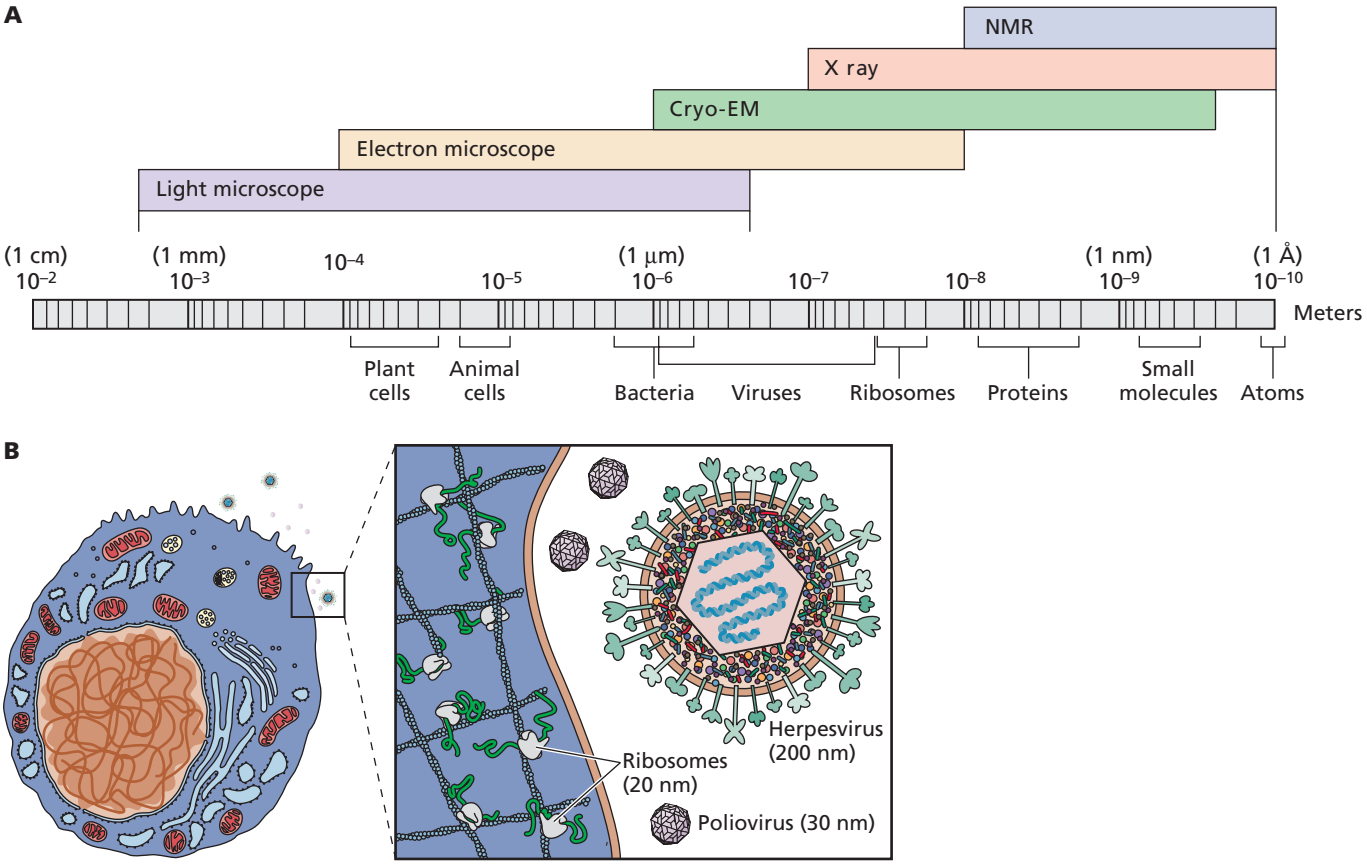


Figure 1.10 Size matters. **(A)** Sizes of animal and plant cells, bacteria, viruses, proteins, molecules, and atoms are indicated. The resolving powers of various techniques used in virology, including light microscopy, electron microscopy, cryo-electron microscopy (Cryo-EM), X-ray crystallography, and nuclear magnetic resonance (NMR) spectroscopy, are indicated. Viruses span a broad range from that equal to some small bacteria to just above ribosome size. The units commonly used in descriptions of virus particles or their components are the nanometer (nm [10^{-9} m]) and the angstrom (Å [10^{-10} m]). **(B)** Illustration of the size differences among two animal viruses and a typical eukaryotic host cell.

Figure 1.11 Lesions induced by tobacco mosaic virus on an infected tobacco leaf. In 1886, Adolph Mayer first described the characteristic patterns of light and dark green areas on the leaves of tobacco plants infected with tobacco mosaic virus. He demonstrated that the mosaic lesions could be transmitted from an infected plant to a healthy plant by aqueous extracts derived from infected plants. Following application of the preparation to healthy plant leaves, the number of characteristic lesions containing dead cells is directly proportional to the number of infectious particles in the test sample. Courtesy of USDA Forest Service, under license CC BY 3.0.

phages were seen as ideal tools with which to investigate the nature of genes and heredity. Probably the most critical early contribution of Delbrück and Ellis was the perfection of the "one-step growth" method for synchronization of the reproduction of phages, an achievement that allowed analysis of a single cycle of phage reproduction in a population of bacteria. This approach introduced highly quantitative methods to virology, as well as an unprecedented rigor of analysis. The first experiments showed that phages indeed multiplied in the bacterial host and were liberated in a "burst" following disruption of the cell.

Delbrück was a zealot for phage research and recruited talented scientists to pursue the fundamental issues of what is now known as the field of molecular biology. This cadre of scientists focused their attention on specific phages of the bacterium *Escherichia coli*. Progress was rapid, primarily because of the simplicity of the phage infectious cycle. By the mid-1950s it was evident that viruses from bacteria, animals, and plants share many fundamental properties. However, the phages provided a far more tractable experimental system. Consequently, their study had a profound impact on the field of virology.

One critical lesson came from a definitive experiment that established that viral nucleic acid carries genetic information. It was known from studies of the "transforming principle" of pneumococcus by Oswald Avery, Colin MacLeod, and Maclyn McCarty (1944) that nucleic acid was both necessary and sufficient for the transfer of genetic traits of bacteria. However, in the early 1950s, protein was still suspected to be an important component of viral heredity. In a brilliantly simple experiment that included the use of a common kitchen food blender, Alfred Hershey and Martha Chase showed that this hypothesis was incorrect; DNA, not protein, carries the information for virus reproduction (Box 1.5).

Bacteriophages were originally thought to be lethal agents, invariably killing their host cells after infection. In the early 1920s, a previously unknown interaction was discovered, in which the host cell not only survived the infection but also stably inherited the genetic information of the virus. It was also observed that certain bacterial strains could lyse spontaneously and produce bacteriophages after a period of growth in culture. Such strains were called **lysogenic**, and the phenomenon, **lysogeny**. Studies of lysogeny revealed many previously unrecognized features of virus-host cell interactions (Box 1.6). Recognition of this phenomenon came from the work of many scientists, but it began with the elegant experiments of André Lwoff and colleagues at the Institut Pasteur in Paris. Lwoff showed that a viral genome exists in lysogenic cells in the form of a silent genetic element called the **prophage**. This element determined the ability of lysogenic bacteria to produce infectious bacteriophages. Subsequent studies of the *E. coli* bacteriophage lambda established a paradigm for one mechanism of lysogeny, the integration of a phage genome into a specific site on the bacterial chromosome.

Bacteriophages became inextricably associated with the new field of molecular biology. Their study established many fundamental principles: for example, control of the decision to enter a lysogenic or a lytic pathway is encoded in the genome of the virus. The first mechanisms discovered for the control of gene expression, exemplified by the elegant operon theory of Nobel laureates François Jacob and Jacques Monod, were deduced in part from studies of lysogeny by phage lambda. The biology of phage lambda provided a fertile ground for work on gene regulation, but study of virulent T phages (T1 to T7, where T stands for "type") of *E. coli* paved the way for many other important advances. As we shall see, these systems also provided an extensive preview of mechanisms of animal virus reproduction (Box 1.7).

Animal Cells as Hosts

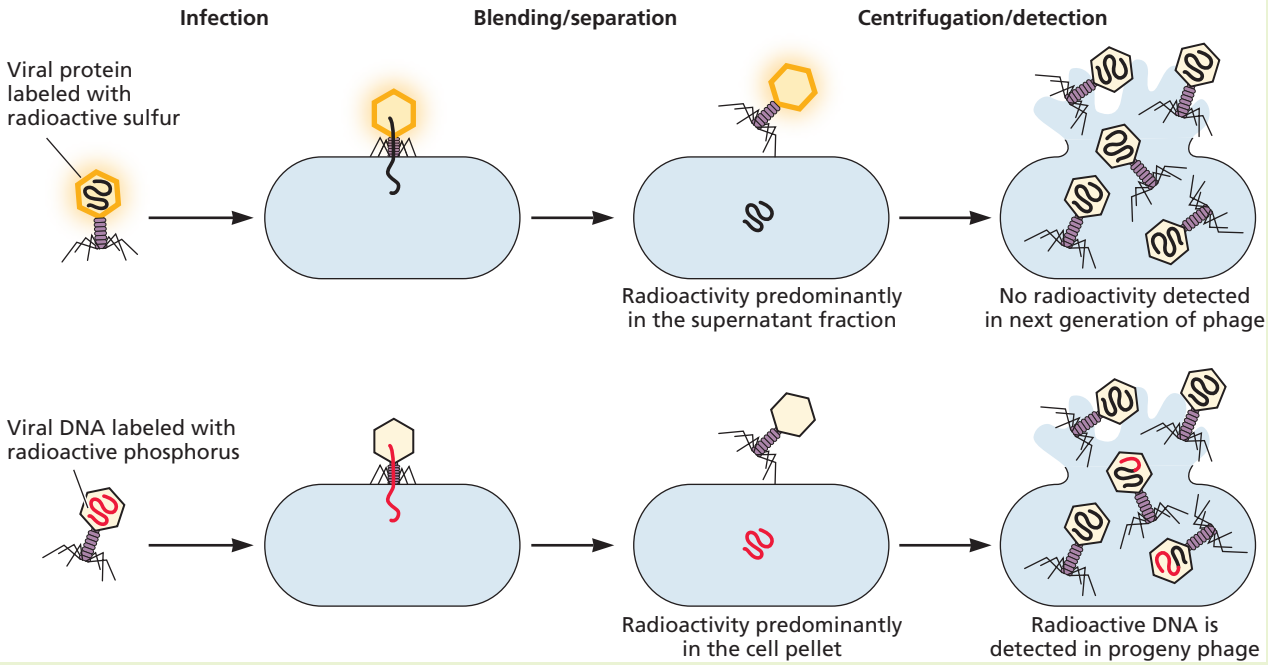
The culture of animal cells in the laboratory was initially more of an art than a science, restricted to cells that grew out of organs or tissues maintained in nutrient solutions under sterile conditions. Cells so obtained from living tissues, called **primary cells**, have a finite life span. Their dependence for growth on natural components in their media such as lymph, plasma, or chicken embryo extracts, and the technical demands

BOX 1.5

EXPERIMENTS

The Hershey-Chase experiment

By differentially labeling the nucleic acid and protein components of virus particles with radioactive phosphorus (^{32}P) and radioactive sulfur (^{35}S), respectively, Alfred Hershey and Martha Chase showed that the protein coat of the infecting virus could be removed soon after infection by agitating the bacteria for a few minutes in a blender. In contrast, ^{32}P-labeled phage DNA entered and remained associated with the bacterial cells under these conditions. Because such blended cells produced a normal burst of new virus particles, it was clear that the DNA contained all of the information necessary to produce progeny phages.



of sterile culture prior to the discovery of antibiotics, made reproducible experimentation very difficult. However, by 1955, the work of many investigators had led to a series of important methodological advances. These included the development of defined media optimal for growth of mammalian cells, incorporation of antibiotics into cell culture media, and development of immortal cell lines such as the mouse L and human HeLa cells that are still in widespread use. These advances allowed growth of animal cells in culture to become a routine, reproducible exercise.

The availability of a variety of well-characterized animal cell cultures had several important consequences for virology. It allowed the discovery and propagation of new human viruses, such as adenovirus, measles virus, and rubella virus, for which animal hosts were not available. In 1949, John Enders and colleagues used cell cultures to propagate poliovirus, a feat that led to the development of polio vaccines a few years later. Cell culture technology revolutionized the ability to investigate the reproduction of viruses. Viral infectious cycles could be studied under precisely controlled conditions by employing the analog of the one-step growth cycle of bacteriophages and simple methods for quantification of infectious particles described in Chapter 2.

Our current understanding of the molecular basis of viral parasitism, the focus of this volume, is based almost entirely on analyses of one-step growth cycles in cultured cells. Such studies established that viruses depend absolutely on the biosynthetic machinery of their host cells for synthesis of the components from which progeny viral particles are built. In contrast to cells, viruses are not reproduced by growth and division. Rather, the infecting genome contains the information necessary to redirect cellular systems to the production of many copies of all the components needed for the *de novo* assembly of new virus particles. It is remarkable, however, that while viruses lack the complex energy-generating and biosynthetic systems necessary

BOX 1.6

BACKGROUND

Properties of lysogeny shared with animal viruses

Lytic versus Lysogenic Response to Infection

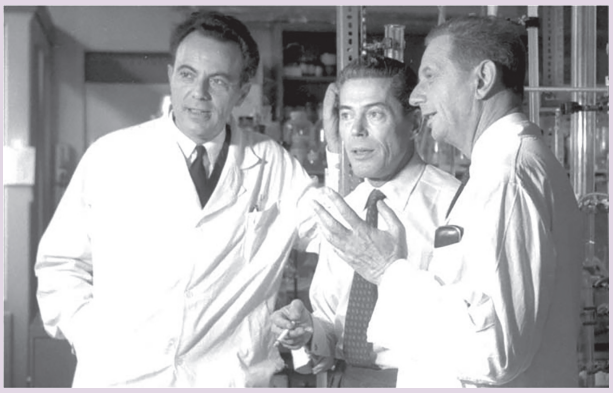
Some bacterial viruses can enter into either destructive (lytic) or relatively benign (lysogenic) relationships with their host cells. Such bacteriophages were called temperate. In a lysogenic bacterial cell, viral genetic information persists but viral gene expression is repressed. Such cells are called lysogens, and the quiescent viral genome, a prophage. By analogy with the prophage, an integrated DNA copy of a retroviral genome in an animal genome is termed a provirus.

Propagation as a Prophage

For some bacteriophages like lambda and Mu (Mu stands for "mutator"), prophage DNA is integrated into the host genome of lysogens and passively replicated by the host. Virally encoded enzymes, known as integrase (lambda) and transposase (Mu), mediate the covalent insertion of viral DNA into the chromosome of the host bacterium, establishing it as a prophage. The prophage DNA of other bacteriophages, such as P1, exists as a plasmid, a self-replicating, autonomous chromosome in a lysogen. Both forms of propagation have been identified in certain animal viruses, for example, retroviruses and a lethal herpesvirus.

Insertional Mutagenesis

Bacteriophage Mu inserts its genome into many random locations on the host chromosome, causing numerous mutations by disrupting host DNA sequences. This process is called insertional mutagenesis and is a phenomenon observed with retroviruses.



Pioneers in the study of lysogeny: Nobel laureates François Jacob, Jacques Monod, and André Lwoff, 1965. Courtesy of the U.S. National Library of Medicine.

Gene Repression and Induction

Prophage gene expression in lysogens is turned off by the action of viral proteins called repressors. Expression can be turned on when repressors are inactivated (a process called induction). The discovery that genes can be regulated by such *trans*-acting proteins, and elucidation of their mechanism, set the stage for later investigation of the control of gene expression with other viruses and their host cells.

Transduction of Host Genes

Bacteriophage genomes can pick up cellular genes and deliver them to new cells (a process known as transduction). For example, occasional mistakes in excision of the lambda prophage from its host chromosome after induction result in production of unusual progeny phages that have lost some of their own DNA but have acquired the bacterial DNA adjacent to the prophage. The acute transforming retroviruses also arise via capture of genes in the vicinity of their integration as proviruses (Volume II, Chapter 6). These cancer-inducing cellular genes are then transduced along with viral genes during subsequent infection.

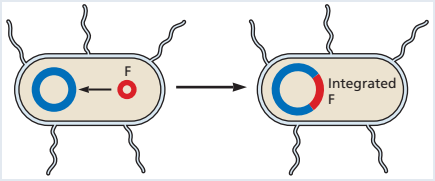
BOX 1.7

TERMINOLOGY

The episome

In 1958, François Jacob and Elie Wollman realized that lambda prophage and the *E. coli* F sex factor had many common properties. This remarkable insight led to the definition of the episome.

An episome is an exogenous genetic element that is not necessary for cell survival. Its defining characteristic is the ability to reproduce in two alternative states: while integrated in the host chromosome or autonomously. However, this term is now most commonly applied to genomes that can be maintained in cells by autonomous replication and never integrate, for example, the DNA genomes of certain animal viruses.

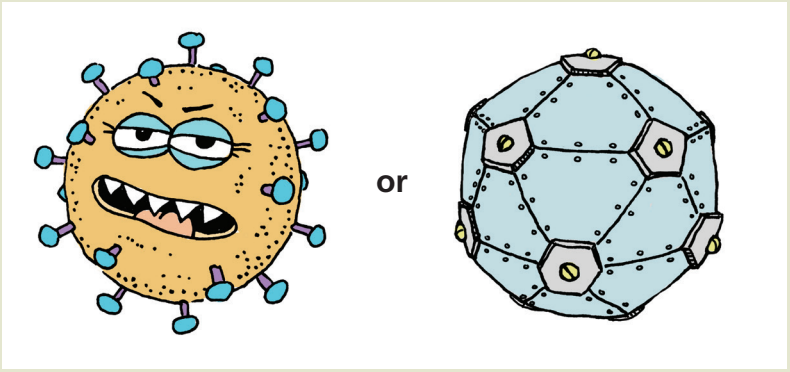


BOX 1.8

DISCUSSION

Are viruses living entities? What can/can't they do?

Viruses can be viewed as microbes that exist in two phases: an inanimate phase, the virion; and a multiplying phase in an infected cell. Some researchers have promoted the idea that viruses are bona fide living entities. According to this notion, inanimate virions may be viewed as "spores" that transform the infected cell into a novel type of organism (termed a virocell), dedicated to the production of new virions. The nature of viruses has long been a topic of intense discussion, stimulated most recently by the discovery of giant viruses such as the mimiviruses and Pandoraviruses, which encode more functions that previously ascribed to viral genomes.



Apart from attributing "life" to viruses, many scientists have succumbed to the temptation of ascribing various **actions** and **motives** when discussing them. While remarkably effective in enlivening a lecture or an article, anthropomorphic characterizations are inaccurate and also quite misleading. Infected cells and hosts respond in many ways after virus infection, but viruses, which are totally at the mercy of their environment, lack the capacity for intentional, goal-directed activity. Therefore, viruses cannot employ, ensure, synthesize, induce, display, destroy, deploy, depend, avoid, retain, evade, exploit, generate, etc.

As virologists can be very passionate about their subject, it is exceedingly difficult to purge such anthropomorphic terms from virology communications. Indeed, hours were spent doing so in the preparation of this textbook, though undoubtedly there remain examples in which actions are attributed to viruses. Should you find them, let us know!

Check out what the contemporary general public feels about this topic at http://www.virology.ws/are-viruses-alive/.

Forterre P. 2016. To be or not to be alive: how recent discoveries challenge the traditional definitions of viruses and life. *Stud Hist Philos Biol Biomed Sci* **59:**100–108.

van Regenmortel MHV. 2016. The metaphor that viruses are living is alive and well, but it is no more than a metaphor. *Stud Hist Philos Biol Biomed Sci* **59:**117–124.

for independent existence (Box 1.8), they are **not** the simplest biologically active agents: **viroids**, which are infectious agents of a variety of economically important plants, comprise a single small molecule of noncoding RNA, whereas agents called **prions**, which cause neurological disease in humans and animals, are thought to be aggregates of single protein molecules (Volume II, Chapter 13).

Cataloging Animal Viruses

As new viruses were being discovered and studied by electron microscopy, the virus world was seen to be a veritable zoo of particles with different sizes, shapes, and compositions. With no standard rules for naming isolates, the viral lexicon was, and still is, idiosyncratic (Box 1.9). Constructing a rational scheme by which these agents could be classified became a subject of colorful and quite heated controversy. A traditionalist camp argued that it was impossible to infer, from the known properties of viruses, anything about their evolutionary origin or their relationships to one another—the major goal of classical taxonomy. Others maintained that despite such limitations, there were significant practical advantages in grouping viruses with similar properties. A major sticking point, however, was finding agreement on **which** properties should be considered most important in constructing a scheme for virus classification.

The Classical System

Lwoff, Robert Horne, and Paul Tournier, in 1962, advanced a comprehensive scheme for the classification of all viruses under the classical Linnaean hierarchical system consisting of phylum, class, order, family, genus, and species. Although a subsequently formed international committee on the nomenclature of viruses did not adopt this system *in toto*, its designation of orders, families, genera, and species is used for the classification of animal viruses.

One of the most important principles embodied in the system advanced by Lwoff and his colleagues was that viruses should be grouped according to **their** shared properties rather than those of the cells or organisms they infect. A second principle was a focus on the nature of the nucleic acid

BOX 1.9

TERMINOLOGY

Complexities of viral nomenclature

No consistent system for naming viral isolates has been established by their discoverers. For example, among the vertebrate viruses, some are named for the associated diseases (e.g., poliovirus, rabies virus), for the specific type of disease they cause (e.g., murine leukemia virus), or for the sites in the body that are affected or from which they were first isolated (e.g., rhinovirus and adenovirus). Others are named for the geographic locations from which they were first isolated (e.g., Sendai virus [Sendai, Japan] and Coxsackievirus [Coxsackie, NY]) or for the scientists who first discovered them (e.g., Epstein-Barr virus). In these cases, the virus names are capitalized. Some viruses are even named for the way in which people imagined they were contracted (e.g., influenza, for the "influence" of bad air), how they were first perceived (e.g., the giant mimiviruses [Box 1.10], for the fact that they "mimic" bacteria), or totally by whimsy (e.g., Pandoravirus, after Pandora's jar [later box] of Greek mythology). Finally, combinations of the above designations are also used (e.g., Rous sarcoma virus).

genome as the primary criterion for classification. The importance of the genome had become clear when it was inferred from the Hershey-Chase experiment that viral nucleic acid alone can be infectious (Box 1.5). Four characteristics are used in the taxonomic classification of all viruses:

1. Nature of the nucleic acid in the virus particle (**DNA** or **RNA**)
2. Symmetry of the protein shell (**capsid**)
3. Presence or absence of a lipid membrane (**envelope**)
4. Dimensions of the virion and capsid

The elucidation of evolutionary relationships by analyses of nucleic acid and protein sequence similarities is now **the** standard method for assigning viruses to a particular family and ordering members within a family. For example, hepatitis C virus was classified as a member of the family *Flaviviridae* and MERS was assigned to the *Coronaviridae* based on their genome sequences. However, as our knowledge of molecular properties of viruses and their reproduction has increased, other relationships have become apparent. *Hepadnaviridae*, *Retroviridae*, and some plant viruses are classified as different families on the basis of the nature of their genomes. Nevertheless, they are all related by the fact that reverse transcription is an essential step in their reproductive cycles, and the viral polymerases that perform this task exhibit important similarities in amino acid sequence. Another example is the classification of the giant protozoan *Mimiviridae* as members of a related group called nucleocytoplasmic large DNA viruses (NCLDVs), which includes the *Poxviridae* that infect vertebrates (Box 1.10).

The International Committee on Taxonomy of Viruses (ICTV), founded by André Lwoff, authorizes and organizes the classification and establishes nomenclature for all viruses. Freely available as a periodically updated, online resource (https://ictv.global/taxonomy), the 2018 report lists orders, families, genera, and species for all known viruses. In addition, it describes numerous viruses that are not yet classified and probably representatives of new genera and/or families. The ICTV catalog also includes descriptions of subviral agents (**satellites**, viroids, and prions) and a list of viruses for which information is still insufficient to make assignments. The pace of discovery of new viruses has been accelerated greatly with the application of **metagenomic analyses**, direct sequencing of genomes from environmental samples, suggesting that we have barely begun to chart the viral universe.

The ICTV nomenclature has been applied widely in both the scientific and medical literature, and therefore we adopt it in this text. In this nomenclature, the Latinized virus family names are recognized as starting with capital letters and ending with *-viridae*, as, for example, in the family name *Parvoviridae*. These names are used interchangeably with their common derivatives, as, for example, parvoviruses (see additional examples in the Appendix).

Classification by Genome Type: the Baltimore System

Francis Crick conceptualized the central dogma for flow of information from the DNA genome in all living cells:

$$\text{DNA} \rightarrow \text{mRNA} \rightarrow \text{protein}$$

As intracellular parasites that depend on the host cell's translational machinery for protein production, all viruses must direct the synthesis of mRNAs. But viral genomes comprise both DNA and RNA in a variety of conformations. Appreciation of the essential role of the translational machinery

BOX 1.10

DISCUSSION

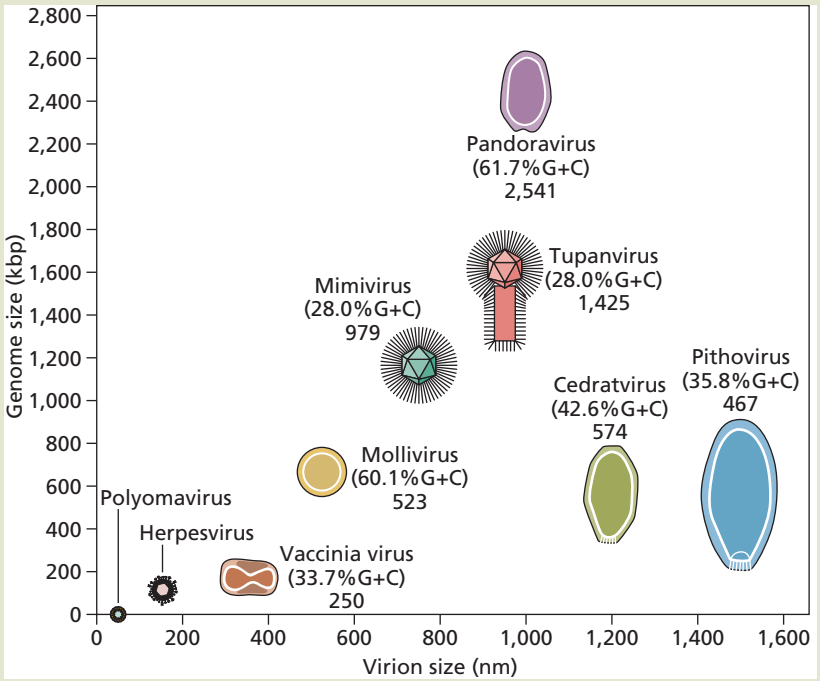
Giant viruses discovered in amoebae

The mimivirus virion, the prototype member of the *Mimiviridae*, was the first giant virus of amoebae to be discovered. Isolated from water in a cooling tower in England in 1992, it is large enough to be visible in a light microscope and was initially thought to be an intracellular bacterium within its host. Not until publication of a brief note in 2003 did it become apparent that this giant was really a virus. The mimivirus genome of 1.2 Mbp was much larger than that of any known virus at the time, exceeding that of some bacteria. This giant encodes more than 900 proteins, many of which are components of the protein translational apparatus, a function for which other viruses rely entirely on the host.

Since reports of the first giant viruses, the use of different strains of amoebae to screen soil and water samples from diverse environments and geographic locations has yielded more than 50 isolates, assigned to nine distinct families. Among the most spectacular is a Pandoravirus isolate, discovered in saltwater off the coast of Chile in 2013. The genome of this giant is twice the size of the mimivirus genome, and contains ~2,500 putative protein-coding sequences, most of them never seen before. Furthermore, while mimivirus has a more or less familiar icosahedral capsid, the Pandoravirus has no regular capsid. Instead, the genomes of these viruses are surrounded by an ovoid envelope, with a pore at the apex that allows delivery of the internal components into the cytoplasm of its host. The following year two additional giant amoeba viruses, a circular mollivirus and ovoid pithovirus, were discovered in a sample of Siberian permafrost more than 30,000 years old.

The unusual properties of the giant viruses of amoebae have prompted the somewhat controversial speculation that they might represent a separate branch in the tree of life, or that they arose by reductive evolution from the nucleus of a primitive cellular life form. However, the discovery in 2017 of another group of these viruses, by metagenomic analysis of samples from a sewer in Klosterneuburg, Austria, has suggested a more pedestrian origin. While the new group, called Klosneuviruses, encode numerous components of translational machinery, comprehensive phylogenetic analyses indicate that these genes were captured from a cellular host by a smaller, precursor virus during evolution of Klosneuviruses. If this is a general phenomenon, the 2018 description of tailed mimivirus relatives, isolated from the extreme environments of an alkaline soda lake in Brazil and from deep in the Atlantic Ocean, must be considered an extraordinary example of such capture. The genomes of these odd-looking isolates, called Tupanviruses, contain nearly all of the necessary translation-associated genes, lacking only ribosomes for protein synthesis. It would seem that there is still much to ponder concerning the evolution of these giant viruses.

For illustrations of giant amoeba virus structures, see http://viralzone.expasy.org/all_by_species/670.html. See also TWiV 261: Giants among viruses. Interview with Drs. Chantal Abergel and Jean-Michel Claverie at http://www.microbe.tv/twiv/twiv-261-giants-among-viruses/.



Properties of some of the largest currently known giants, all of which infect amoebae, with representative vertebrate-infecting DNA viruses, of which poxviruses are the largest. The broad range of nucleic acid composition among the amoeba viral genomes is illustrated by the substantial differences in their G+C content. The number of known or putative coding genes in each viral genome is listed. Examples of small, medium, and large mammalian viruses (poliovirus, herpesvirus, and vaccinia virus, respectively) are included for comparison.

Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E, Colson P, Levasseur A, Kroemer G, Raoult D, La Scola B. 2018. Tailed giant Tupanvirus possesses the most complete translational apparatus of the known virosphere. *Nat Commun* **9:**749–761.

Colson P, La Scola B, Levasseur A, Caetano-Anollés G, Raoult D. 2017. Mimivirus: leading the way in the discovery of giant viruses of amoebae. *Nat Rev Microbiol* **15:**243–254.

Colson P, La Scola B, Raoult D. 2017. Giant viruses of amoebae: a journey through innovative research and paradigm changes. *Annu Rev Virol* **4:**61–85.

Racaniello V. 8 March 2018. Only the ribosome is lacking. Virology Blog. http://www.virology.ws/2018/03/08/only-the-ribosome-is-lacking/.

Schulz F, Yutin N, Ivanova NN, Ortega DR, Lee TK, Vierheilig J, Daims H, Horn M, Wagner M, Jensen GJ, Kyrpides NC, Koonin EV, Woyke T. 2017. Giant viruses with an expanded complement of translation system components. *Science* **356:**82–85.

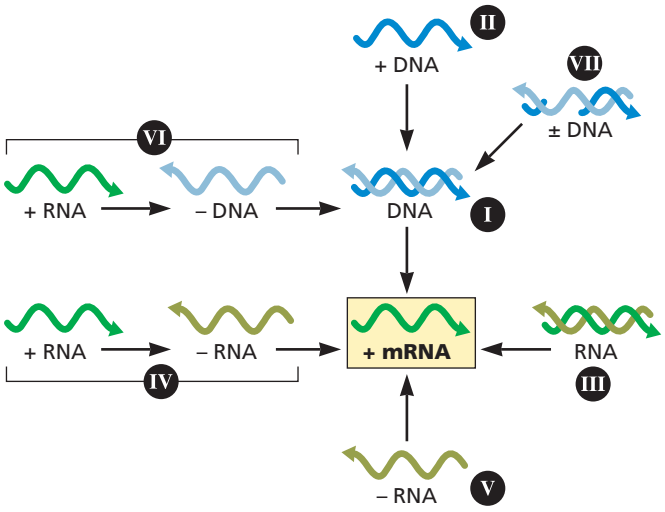


Figure 1.12 The Baltimore classification. The Baltimore classification assigns viruses to seven (I to VII) distinct classes on the basis of the nature and polarity of their genomes. Because all viruses must produce mRNA that can be translated by cellular ribosomes, knowledge of the composition of a viral genome provides insight into the pathways required to produce mRNA, indicated by arrows. See also Baltimore D. 1971. *Bacteriol Rev* 35:235–241.

in virus reproduction inspired David Baltimore, in 1971, to devise a classification scheme for viruses, based on the steps that would be required to produce mRNA from their diverse genomes (Fig. 1.12).

By convention, mRNA is defined as a **positive [(+)] strand** because it contains immediately translatable information. In the Baltimore classification, a strand of DNA that is of equivalent sequence is also designated a (+) strand. The RNA and DNA complements of (+) strands are designated **negative [(–)] strands**.

As originally conceived, the Baltimore scheme included six classes of viral genomes (designated I to VI). When the gapped DNA genome of hepadnaviruses (e.g., hepatitis B virus) was discovered, these viruses were assigned to a seventh class (VII). The DNA and RNA descriptors for the viral classes [single-stranded DNA (ssDNA), double-stranded DNA (dsDNA), (+) RNA, or (–) RNA, etc.], but not the Roman numeral designations, have been adopted universally and are a valuable complement to classical taxonomy. The information embodied in classification by genome type provides virologists with immediate insight into the steps that must take place to initiate the replication and expression of any viral genome.

Because the viral genome carries the entire blueprint for virus propagation, molecular virologists have long considered it the most important characteristic for classification purposes. Although individual virus families are known by their classical designations, they are commonly grouped according to their genome type. In the ICTV compilation, all viral families are assigned to one of the seven classes described in the Baltimore system (Fig. 1.13).

A Common Strategy for Viral Propagation

The basic thesis of this textbook is that **all** viral propagation can be described in the context of three fundamental properties.

- Viral genomes are packaged inside particles that mediate their transmission from host to host.
- The viral genome contains the information for initiating and completing an infectious cycle within a susceptible, permissive cell.
- An infectious cycle includes attachment and entry, decoding of genome information, genome replication, and assembly and release of particles containing the genome.
- Viral propagation is ensured by establishment in a host population.

Perspectives

The study of viruses has increased our understanding of the importance and ubiquitous existence of these diverse agents and, in many cases, yielded new and unexpected insight into the molecular biology of host cells and organisms. Indeed, because viruses are obligate molecular parasites, every tactical solution encountered in their reproduction and propagation must of necessity tell us something about the host as well as the virus. Some of the important landmarks and achievements in the field of virology are summarized in Fig. 1.14. It is apparent that much has been discovered about the biology of viruses and about host defenses against them. Yet the more we learn, the more we realize that much is still unknown.

In the first edition of this textbook (published in 2000), we noted that the most recent (1995) report of the ICTV listed 71 different virus families, which covered most new isolates. We speculated therefore that: "As few new virus families had been identified in recent years, it seems likely that a significant fraction of all existing virus families are now known." In the intervening years, this prediction has been shattered, not only by the discovery of new families of viruses, including giant viruses with genome sizes that surpass those of some bacteria, but also by results from metagenomic analyses. For example, the fact that a high percentage (93%) of protein-coding sequences in the genomes of the giant Pandoraviruses have **no** homologs in the current databases was totally unexpected. The unusual morphological features and atypical reproduction process of these viruses

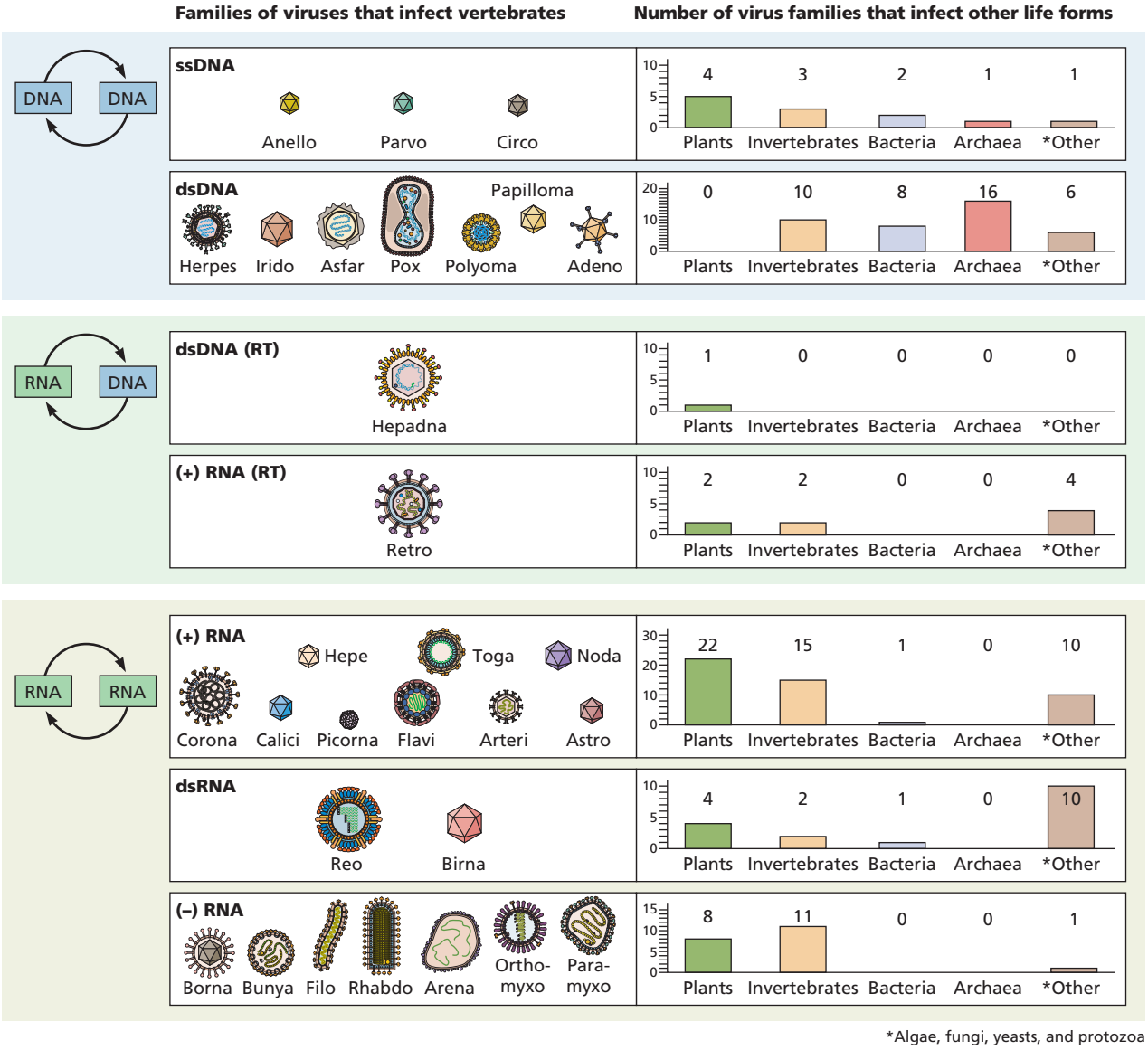


Figure 1.13 Viral families sorted according to the nature of the viral genomes. A wide variety of sizes and shapes are illustrated for the families of viruses that infect vertebrates. Families are identified by Latinized names and organized in seven distinct classes, based on the nature of their genomes. Genome replication cycles are illustrated in the column at the left. Similar diversity exists for the families of viruses that infect other life forms, but the chart lists only the approximate number found to date in each class. As noted in the 9th and 10th ICTV Reports, in some cases there are as yet no examples. Data from King AMQ et al. 2012. *Virus Taxonomy: The Classification and Nomenclature of Viruses* (https://talk.ictvonline.org/ictv-reports/), with assistance from Dr. Elliot J. Lefkowitz, Department of Microbiology, Director of Informatics, UAB Center for Clinical and Translational Science, Birmingham, AL (http://www.uab.edu/bioinformatics/).

were also surprising. In addition, it is mind-boggling to contemplate that of almost 900,000 viral sequences identified in samples of only one type of ecosystem (raw sewage), more than 66% bore **no** relationship to any viral family in the current database. From these analyses, and similar studies of other ecosystems (i.e., oceans and soil), it has been estimated that only a minor percentage of extant viral diversity has been explored to date. Clearly, the viral universe is far more vast and diverse than suspected only a decade ago, and there is much fertile ground for gaining a deeper understanding of the biology of viruses and their host cells and organisms.

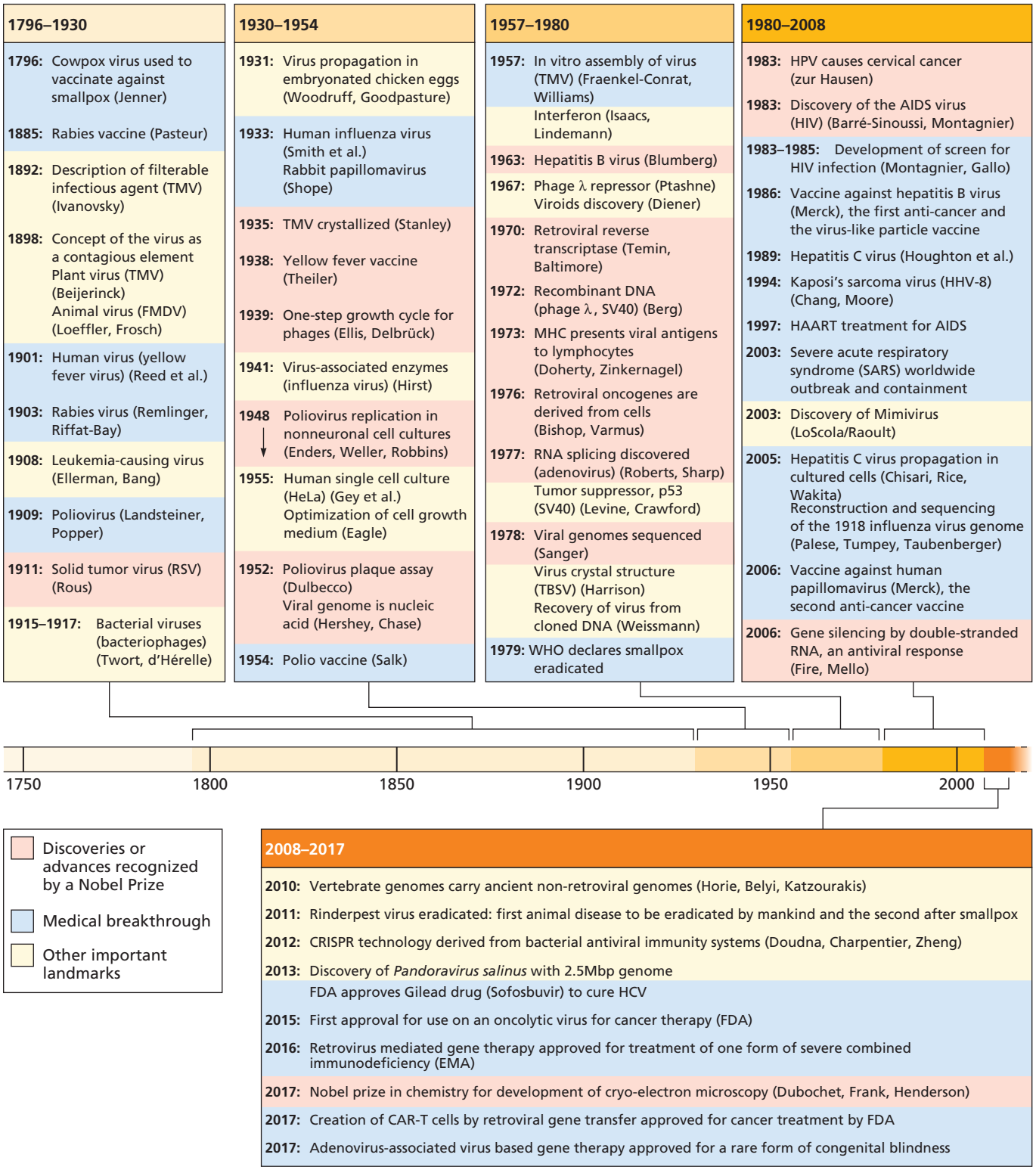


Figure 1.14 Landmarks in the study of viruses. Key discoveries and technical advances are listed for each time interval. The pace of discovery has increased exponentially over time. Abbreviations: AAV, adenovirus-associated virus; EU, European Union; EMA, European Medical Association; FDA, U.S. Food and Drug Administration; FMDV, foot-and-mouth disease virus; HAART, highly active antiretroviral therapy; HCV, hepatitis C virus; HHV-8, human herpesvirus 8; HIV-1, human immunodeficiency virus type 1; HPV, human papillomavirus; MHC, major histocompatibility complex; RSV, Rous sarcoma virus; SV40, simian virus 40; TBSV, tomato bushy stunt virus; TMV, tobacco mosaic virus; WHO, World Health Organization.

REFERENCES

Books

Barry JM. 2005. *The Great Influenza*. Penguin Books, New York, NY.

Brock TD. 1990. *The Emergence of Bacterial Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Brothwell D, Sandison AT (ed). 1967. *Diseases in Antiquity*. Charles C Thomas, Publisher, Springfield, IL.

Cairns J, Stent GS, Watson JD (ed). 1966. *Phage and the Origins of Molecular Biology*. Cold Spring Harbor Laboratory for Quantitative Biology, Cold Spring Harbor, NY.

Creager ANH. 2002. *The Life of a Virus: Tobacco Mosaic Virus as an Experimental Model, 1930–1965*. University of Chicago Press, Chicago, IL.

Denniston K, Enquist L. 1981. *Recombinant DNA. Benchmark Papers in Microbiology*, vol 15. Dowden, Hutchinson and Ross, Inc, Stroudsburg, PA.

Hughes SS. 1977. *The Virus: a History of the Concept*. Heinemann Educational Books, London, United Kingdom.

Karlen A. 1996. *Plague's Progress, a Social History of Man and Disease*. Indigo, Guernsey Press Ltd, Guernsey, Channel Islands.

Knipe DM, Howley PM (ed). 2013. *Fields Virology*, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.

Luria SE. 1953. *General Virology*. John Wiley & Sons, Inc, New York, NY.

Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Rasmussen N. 1997. *Picture Control: the Electron Microscope and the Transformation of Biology in America 1940–1960*. Stanford University Press, Stanford, CA.

Oldstone MBA. 2010. *Viruses, Plagues, & History: Past, Present, and Future*. Oxford University Press, New York, NY.

Papers of Special Interest

Boylston AW. 2018. The myth of the milkmaid. *N Engl J Med* **378:**414–415.

A delightful scientific historian's report on research that debunks the much-cited notion that Edward Jenner was inspired to test the benefits of cowpox by the comments of a milkmaid who claimed to be immune to smallpox because she had had cowpox.

Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, Mead D, Azam F, Rohwer F. 2002. Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* **99:**14250–14255.

Early use of metagenomic analysis to identify viruses in natural marine environments. One of the first to identify these agents using these methods, and to reveal the enormity in number of previously unknown viruses in these environments.

Crick FHC, Watson JD. 1956. Structure of small viruses. *Nature* **177:**473–475.

Authors deduce from X-ray crystal analysis of plant virus particles that virus shells (capsids) are composed of a large number of identical protein molecules, of small or moderate size, packed together in a regular manner.

Murray NE, Gann A. 2007. What has phage lambda ever done for us? *Curr Biol* **17:**R305–R312.

The authors describe how study of the bacteriophage lambda has contributed to an understanding of the molecular basis of numerous fundamental biological processes.

Suttle CA. 2007. Marine viruses—major players in the global ecosystem. *Nat Rev Microbiol* **5:**801–812.

Suttle describes the unappreciated yet enormous contribution that the huge numbers of marine viruses make to the earth's marine and global ecosystems.

Websites

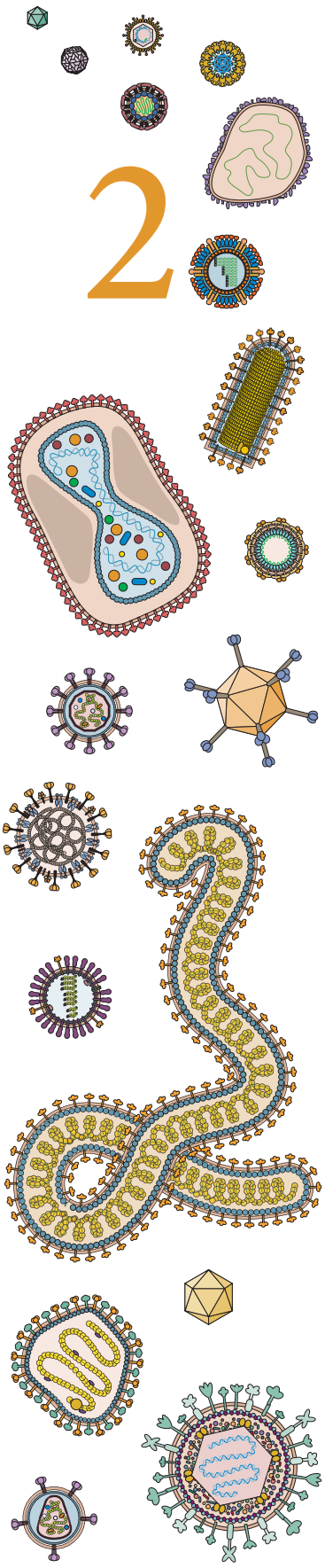
https://talk.ictvonline.org/taxonomy/ Latest update of virus classification from the ICTV.

http://ictvonline.org/ ICTV-approved virus names and other information as well as links to virus databases can be downloaded.

http://microbe.tv/twiv A weekly podcast about viruses featuring informal yet informative interviews with guest virologists who discuss their recent findings and other topics of general interest.

STUDY QUESTIONS

1. What is the definition of a virus?
2. Which is a key property first discovered about viruses that distinguished them from other microorganisms?
   - **a.** They were too large to pass through a 0.2-micron filter
   - **b.** They could reproduce only in broth
   - **c.** They made tobacco plants sick
   - **d.** They were small enough to pass through a 0.2-micron filter
   - **e.** None of the above
3. All of us carry many different viruses throughout our daily lives. Why don't they make us sick?
4. Why do we care that viruses comprise the most biodiversity on the planet?
5. The first viruses were discovered near the end of the 1800s. How was this done?
   - **a.** By transmitting a disease to tobacco plants using a cell-free filtrate of diseased leaves
   - **b.** Pasteur showed that viruses could reproduce in a sterile medium
   - **c.** Leeuwenhoek saw viruses in his microscope
   - **d.** Robert Koch showed that viruses grown in broth could cause disease
   - **e.** All of the above
6. Why were the bacteriophage systems so useful for elucidating principles of viral reproduction? What important features of virus-host interactions were discovered from these studies?
7. How are viruses classified?
8. How does the discovery of new viruses today differ from 100 years ago?
9. Which host cell function is essential for the reproduction of all viruses?
10. What is the basis of the Baltimore classification system? How many genome types are sufficient to describe all viral families in this system?



2 The Infectious Cycle

LINKS FOR CHAPTER 2

- *Video: Interview with Dr. Thomas Hope*
  http://bit.ly/Virology_Hope
- *Cloning HeLa cells with Philip I. Marcus*
  http://bit.ly/Virology_Twiv197
- *Ode to a plaque*
  http://bit.ly/Virology_Twiv68
- *Movie 2.1: Plaque formation by vesicular stomatitis virus*
  http://bit.ly/Virology_VZVGFP
- *Think globally, act locally*
  http://bit.ly/Virology_Twim90

You know my methods, Watson.
SIR ARTHUR CONAN DOYLE

Introduction

Viruses are unique: often made up of nothing more than a nucleic acid molecule wrapped in protein, they parasitize the cellular machinery to produce thousands of progeny. This simplicity is misleading: viruses can infect all known life forms, and they comprise a variety of structures and genomes. Despite such variety, viruses are amenable to study because all viral propagation can be described in the context of three fundamental properties, as noted in Chapter 1: viral genomes are packaged inside particles that mediate their transmission from cell to cell; the viral genome contains the information for initiating and completing an infectious cycle; viruses establish themselves in a host population to ensure virus survival.

How viruses enter individual cells, their genomes are replicated, and new infectious particles are assembled are some of the topics of research in virology. These studies are usually carried out with cell cultures because they are a much simpler and more homogeneous experimental system than animals. Cells can be infected in such a way as to ensure that a single reproduction cycle occurs synchronously in every infected cell, called **one-step growth**. A full understanding of viral infectious cycles also requires knowledge of cell biology. Consequently, to reproduce the diversity of cells and architectures that are typical of tissues and organs, three-dimensional culture systems have been developed. In this chapter we begin with a brief overview of the infectious cycle, followed by a discussion of methods for cultivating and assaying viruses and detecting viral proteins and genomes and a consideration of viral reproduction and one-step growth analysis.

The Infectious Cycle

The production of new infectious particles can take place only within a cell (Fig. 2.1). Virologists divide viral infectious cycles into discrete steps to facilitate their study, although in virus-infected cells no such artificial boundaries occur. The infectious cycle comprises attachment and entry of the particle, production of viral mRNA and its translation by host ribosomes, genome replication, and assembly and release of progeny particles containing the genome. New virus particles produced during the infectious cycle may then infect other cells. The term **virus reproduction** is another name for the sum total of all events that occur during the infectious cycle.

Some events are common to virus replication in animals and in cells in culture, but there are also many important differences. While virus particles readily attach to cells in culture, in nature they must encounter a host, no mean feat for nanoparticles without any means of locomotion. After encountering a host, the virus particle must pass through physical host defenses, such as dead skin, mucous layers, and the extracellular matrix. Such barriers and other host defenses, such as antibodies and immune cells, which exist to combat virus infections, are not found in cell cultures. Virus infection of cells in culture has been a valuable tool for understanding viral infectious cycles, but the dissimilarities with infection of a living animal must always be considered.

The Cell

Viral reproduction requires many different functions of the host cell. Examples include the machinery for translation of viral mRNAs, sources of energy, and enzymes for genome replication. The cellular transport apparatus brings viral genomes to the correct cellular compartment and ensures that viral subunits reach locations where they may be assembled into virus particles. Subsequent chapters include a discussion of

PRINCIPLES ***The infectious cycle***

- Many distinct functions of the host cell are required to complete a viral infectious cycle.
- The synthesis of new virus particles (i.e., a productive infection) requires target cells that are both susceptible (i.e., allow virus entry) and permissive (i.e., support virus reproduction).
- Viral nucleic acids must be shielded from harsh environmental conditions in extracellular particles but be readily accessible for replication once inside the cell.
- Viruses may be studied by propagation in cells within a laboratory animal or in cells in culture.
- The plaque assay is the major way to determine the concentration of infectious virus particles in a sample.
- Methods for quantifying and characterizing virus particles evolve rapidly, based on developments in detection, ease, cost, safety, utility in the field, and amenability to large-scale implementation.
- Relationships among viruses can be deduced from phylogenetic trees generated from protein or nucleic acid sequences.
- Viral reproduction is distinct from cellular or bacterial replication: rather than doubling with each cycle, each single cell cycle of viral reproduction is typically characterized by the release of many (often thousands) of progeny virions.
- The multiplicity of infection (MOI) is the number of infectious units added per cell; the probability that any one target cell will become infected based on the MOI can be calculated from the Poisson distribution.
- Global analysis of viral, cell, and host responses to virus infection can implicate particular cellular pathways in viral reproduction and can reveal signatures of virus-induced lethality or immune protection.

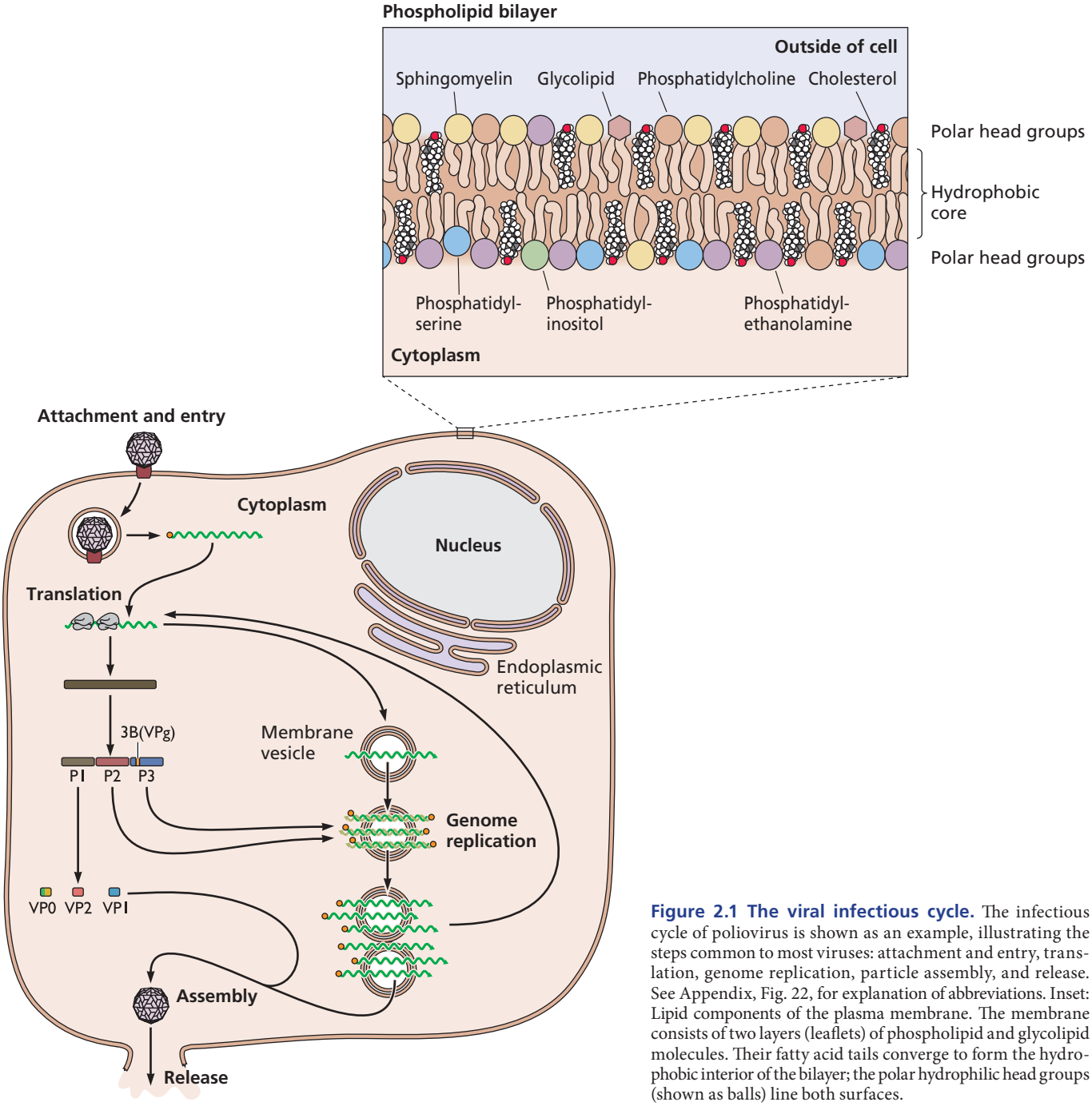


Figure 2.1 The viral infectious cycle. The infectious cycle of poliovirus is shown as an example, illustrating the steps common to most viruses: attachment and entry, translation, genome replication, particle assembly, and release. See Appendix, Fig. 22, for explanation of abbreviations. Inset: Lipid components of the plasma membrane. The membrane consists of two layers (leaflets) of phospholipid and glycolipid molecules. Their fatty acid tails converge to form the hydrophobic interior of the bilayer; the polar hydrophilic head groups (shown as balls) line both surfaces.

cellular functions that are important for individual steps in the viral infectious cycle.

Entering Cells

Viral infection is initiated by a collision between the virus particle and the cell, a process that is governed by chance. A virion may not infect every cell it encounters: it must first come in contact with the tissues that contain cells to which it can bind. Such cells are normally recognized by means of the specific interaction of a virus particle with a cell surface receptor. These cellular molecules do not exist for the benefit of viruses: they all perform functions for the cell. Virus-receptor interactions can be either promiscuous or highly selective, depending on the virus and the distribution of the cell receptor. The presence

of such receptors determines whether the cell will be **susceptible** to the virus. However, whether a cell is **permissive** for the reproduction of a particular virus depends on other, intracellular components found only in certain cell types. Cells must be both susceptible **and** permissive if an infection is to be successful. Virus entry into cells is the topic of Chapter 5.

Viral RNA Synthesis

Although the genomes of viruses come in a number of configurations, they share a common requirement: they must be efficiently copied into mRNAs for the synthesis of viral proteins and progeny genomes for assembly. The synthesis of RNA molecules in cells infected with RNA viruses is a unique process that has no counterpart in the cell (see Chapter 6). With the exception of retroviruses, all RNA viruses encode an RNA-dependent RNA polymerase to catalyze the synthesis of both mRNAs and genomes. For the majority of DNA viruses and retroviruses, synthesis of viral mRNA is accomplished by RNA polymerase II, the enzyme that produces cellular mRNA (see Chapter 7). Much of our current understanding of the mechanisms of cellular transcription comes from study of the transcription of viral templates.

Viral Protein Synthesis

All viruses are parasites of translation: their mRNAs **must** be translated by the host's cytoplasmic protein-synthesizing machinery (see Chapter 11). However, viral infection often results in modification of the host's translational apparatus so that viral mRNAs are translated selectively. The study of such modifications has revealed a great deal about mechanisms of protein synthesis. Analysis of viral translation has also led to the discovery of new mechanisms, such as internal ribosome binding and leaky scanning, that have been subsequently found to occur in uninfected cells.

Viral Genome Replication

Replication of viral genomes requires the cell's synthetic machinery in addition to viral proteins (see Chapters 6, 7, and 9). The cell provides nucleotide substrates, energy, enzymes, and other proteins. Transport systems are required because the cell is compartmentalized: essential components might be found only in the nucleus, the cytoplasm, or within subcellular organelles. Study of the mechanisms of viral genome replication has established fundamental principles of cell biology and nucleic acid synthesis.

Assembly of Progeny Virus Particles

The various components of a virus particle, the nucleic acid genome, capsid protein(s), and in some cases envelope proteins, are often synthesized in different cellular compartments. Their trafficking through and among the cell's compartments and organelles requires that they be equipped with the proper homing signals (see Chapter 12). Components of virus particles must be assembled at some central location, and the information for assembly must be preprogrammed in these molecules (see Chapter 13). The primary sequences of viral structural proteins contain sufficient information to specify assembly; this property is exemplified by the remarkable *in vitro* assembly of tobacco mosaic virus from coat protein and RNA (Box 2.1). Successful virus reproduction depends on redirection of the host cell's metabolic and biosynthetic capabilities, signal transduction pathways, and trafficking systems (see Chapter 14).

Viral Pathogenesis

Viruses command our attention because of their association with animal and plant diseases. **Viral pathogenesis** is the process by which viruses cause disease. The study of viral pathogenesis requires investigating not only the relationships of viruses with the specific cells that they infect but also the

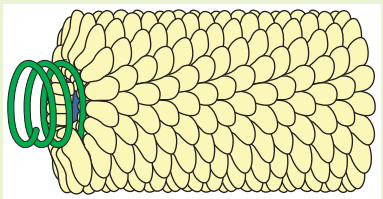
BOX 2.1

EXPERIMENTS

In vitro *assembly of tobacco mosaic virus*

The ability of the primary sequence of viral structural proteins to specify assembly is exemplified by the coat protein of tobacco mosaic virus. Heinz Fraenkel-Conrat and Robley Williams showed in 1955 that purified tobacco mosaic virus RNA and capsid protein assemble into infectious particles when mixed and incubated for 24 h. When examined by electron microscopy, the particles produced *in vitro* were found to be identical to the rod-shaped particles produced from infected tobacco plants (Fig. 1.9B). Neither the purified viral RNA nor the capsid protein alone was infectious. The spontaneous formation of tobacco mosaic particles *in vitro* from protein and RNA components is **the** paradigm for self-assembly in biology.

Fraenkel-Conrat H, Williams RC. 1955. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc Natl Acad Sci U S A* **41:**690–698.



consequences of infection for the host organism. The nature of viral disease depends on the effects of viral reproduction on host cells, the responses of the host's defense systems, and the ability of the virus to spread in and among hosts (Volume II, Chapters 1 to 5).

Overcoming Host Defenses

Organisms have many physical barriers to protect themselves from dangers in their environment, such as invading parasites. Vertebrates also possess an immune system to defend against anything recognized as foreign. Studies of the interactions between viruses and the immune system are particularly instructive, because of the many viral countermeasures that can frustrate this system. Elucidation of these measures continues to teach us about the basis of immunity (Volume II, Chapters 2 to 4).

Cultivation of Viruses

Cell Culture

Types of Cell Culture

Although human and other animal cells were first cultured in the early 1900s, contamination with bacteria, mycoplasmas, and fungi initially made routine work with such cultures extremely difficult. For this reason, most viruses were produced in laboratory animals. The use of antibiotics in the 1940s to control microbial infection was crucial to the establishment of the first cell lines, such as mouse L929 cells (1948) and HeLa cells (1951). John Enders, Thomas Weller, and Frederick Robbins discovered in 1949 that poliovirus could multiply in cultured cells. As noted in Chapter 1, this revolutionary finding, for which these three investigators were awarded a Nobel Prize in 1954, led the way to the propagation of many other viruses in cells in culture, the discovery of new viruses, and the development of vaccines such as those against the viruses that cause poliomyelitis, measles, and rubella. The ability to infect cultured cells synchronously permitted studies of the biochemistry and molecular biology of viral reproduction. Large-scale propagation and purification of virus particles allowed studies of the composition of virus particles, leading to the solution of high-resolution, three-dimensional structures (see Chapter 4).

Cells in culture are still the most commonly utilized hosts for the propagation of animal viruses. To prepare a cell culture, tissues are dissociated into a single-cell suspension by mechanical disruption followed by treatment with proteolytic enzymes. The cells are then suspended in culture medium and placed in specialized plastic flasks or covered plates. As the cells divide, they cover the plastic surface. Epithelial and fibroblastic cells attach to the plastic and form a **monolayer**, whereas blood cells such as lymphocytes settle but do not adhere. The cells are grown in a chemically defined and buffered medium optimal for their growth. Commonly used cell lines double in number in 24 to 48 h in such media. Most cells retain viability after being frozen at low temperatures (–70 to –196°C).

There are three main kinds of monolayer cell cultures (Fig. 2.2), each with advantages and disadvantages for virus research. **Primary cell cultures** are prepared from animal tissues as described above. They have a limited life span, usually no more than 5 to 20 cell divisions. Commonly used primary cell cultures are derived from chicken or mouse embryos, monkey kidneys, or human tissues that are otherwise typically disposed of, such as embryonic amnion, kidney, foreskin, and respiratory epithelium. Such cells are used for experimental virology when the state of cell differentiation is important or when appropriate cell lines are not available. They are also used in vaccine production: for example, infectious attenuated poliovirus vaccine strains may be propagated in primary monkey kidney cells. Primary cell cultures are used for the propagation of viruses to be used as human vaccines to avoid contamination of the product with potentially oncogenic DNA from continuous cell lines (see below). Some viral vaccines are now prepared in **diploid cell**

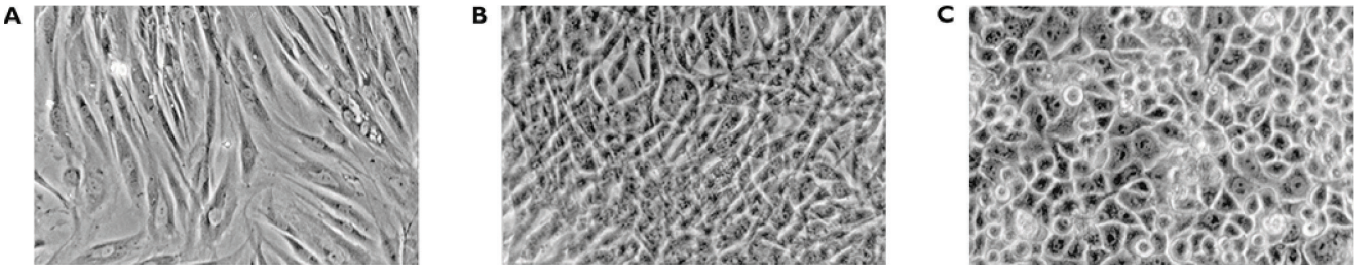


Figure 2.2 Different types of cell culture used in virology. Confluent cell monolayers photographed by low-power light microscopy. **(A)** Primary human foreskin fibroblasts; **(B)** established line of mouse fibroblasts (3T3); **(C)** continuous line of human epithelial cells (HeLa [Box 2.3]). The ability of transformed HeLa cells to overgrow one another is the result of a loss of contact inhibition. Courtesy of R. Gonzalez, Princeton University.

strains, which consist of a homogeneous population of a single cell type and can divide up to 100 times before dying. Despite the numerous divisions, these cells retain the diploid chromosome number. The most widely used diploid cells are those established from human embryos, such as the WI-38 strain derived from human embryonic lung.

Continuous cell lines consist of a single cell type that can be propagated indefinitely in culture. These immortal lines are usually derived from tumor tissue or by treating a primary cell culture or a diploid strain with a mutagenic chemical or an oncogene. Such cell lines often do not resemble the cell of origin; they are less differentiated (having lost the morphology and biochemical features that they possessed in the organ), are often abnormal in chromosome morphology and number (**aneuploid**), and can be tumorigenic (i.e., they produce tumors when inoculated into immunodeficient mice). Examples of commonly used continuous cell lines include those derived from human carcinomas (e.g., HeLa [Henrietta Lacks] cells [Box 2.2]) and from mice (e.g., L and 3T3 cells). Continuous cell lines provide a uniform population of cells that can be infected synchronously for growth curve analysis (see "The One-Step Growth Cycle" below) or biochemical studies of virus replication.

In contrast to cells that grow in monolayers on plastic dishes, others can be maintained in **suspension cultures**, in which a spinning magnet continuously stirs the cells. The advantage of suspension culture is that a large number of cells can be grown in a relatively small volume. This culture method is well suited for applications that require large quantities of virus particles, such as X-ray crystallography or production of vectors.

Despite the wide utility of monolayer and suspension cell cultures in virology, they are not without limitations, including the finite life span of primary cell cultures and the abnormal phenotype of continuous cell lines, such as immortality. These problems can be overcome by the use of **induced pluripotent stem cells (iPSCs)**, which are adult cells that have been reprogrammed genetically to an embryonic stem-cell like state by the introduction of four genes (*Oct4*, *Sox2*, *Kif4*, and *cMyc*). They are most commonly made from human fibroblasts, although other cell types have been used. Such iPSCs can be differentiated into many different cell types, such as

BOX 2.2

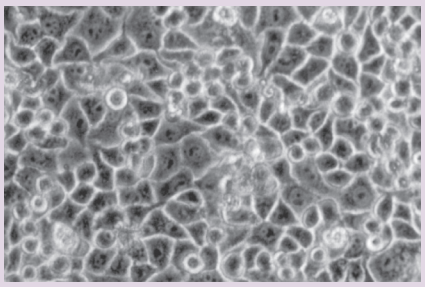
BACKGROUND

The cells of Henrietta Lacks

The most widely used continuous cell line in virology, the HeLa cell line, was derived from Henrietta Lacks. In 1951, the 31-year-old mother of five visited a physician at Johns Hopkins Hospital in Baltimore and was found to have a malignant tumor of the cervix. A sample of the tumor was taken and given to George Gey, head of tissue culture research at Hopkins. Gey had been attempting for years, without success, to produce a line of human cells that would live indefinitely. When placed in culture, Henrietta Lacks' cells propagated as no other cells had before.

On the day in October that Henrietta Lacks died, Gey appeared on national television with a vial of her cells, which he called HeLa cells. He said, "It is possible that, from a fundamental study such as this, we will be able to learn a way by which cancer can be completely wiped out." Soon after, HeLa cells were used to propagate poliovirus, which was causing poliomyelitis throughout the world, and they played an important role in the development of poliovirus vaccines. Henrietta Lacks' HeLa cells started a medical revolution: not only was it possible to propagate many different viruses in these cells, but the work set a precedent for producing continuous cell lines from many human tissues. However, the family of Henrietta Lacks did not learn about HeLa cells, or the revolution they started, until 24 years after her death. Her family members were shocked that cells from Henrietta lived in so many laboratories and that they had not been told that any cells had been taken from her.

The story of HeLa cells is an indictment of the lack of informed consent that pervaded medical research in the 1950s. Since then, biomedical ethics have changed, and there are now strict regulations in clinical research: physicians may not take samples for research from patients without permission. Nevertheless, in early 2013, HeLa cells generated more controversy when a research group published the cells' genome sequence. The Lacks family objected to the publication, claiming that the information could reveal private medical information about surviving family members. As a result, the sequence was withdrawn from public databases. Months later, a second HeLa cell genome sequence was published, but this time the authors were bound by an agreement brokered by the National Institutes of Health,



which required an application process for any individual wishing to view the sequence.

Adey A, Burton JN, Kitzman JO, Hiatt JB, Lewis AP, Martin BK, Qiu R, Lee C, Shendure J. 2013. The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* **500:**207–211.

Callaway E. 2013. Deal done over HeLa cell line. *Nature* **500:**132–133.

Callaway E. 23 March 2013. HeLa publication brews bioethical storm. *Nature*. https://www.nature.com/news/hela-publication-brews-bioethical-storm-1.12689.

Skloot R. April 2000. Henrietta's dance. Johns Hopkins Magazine. http://pages.jh.edu/~jhumag/0400web/01.html.

Skloot R. 2011. *The Immortal Life of Henrietta Lacks*. Broadway Books, New York, NY.

BOX 2.3

EXPERIMENTS

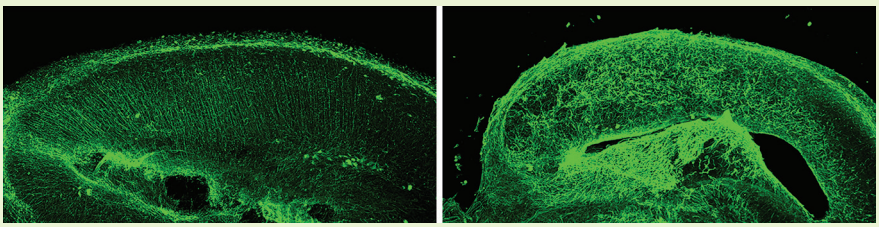
Zika virus blocks the neuronal road

Zika virus infection during pregnancy is a cause of the human birth defect called microcephaly. Babies born with this defect have smaller heads than expected for their age and smaller brains that do not develop normally. Organotypic brain slice cultures from embryonic mice have been used to study the effect of Zika virus on brain development.

To produce organotypic embryonic brain slice cultures, fetal mouse brains were removed, embedded in low-melting-point agarose, and thinly sliced with a vibratome. The slices were placed in cortical culture medium and then infected with Zika virus.

When first- and second-trimester brain slice cultures were infected with different isolates of Zika virus from 1947 to 2016, reproduction was observed as determined by plaque assay. These findings demonstrate that neurotropism of Zika virus is not a recently acquired phenotype.

The small heads observed in microcephalic children reflect a physically smaller brain—specifically, the neocortex is thinner than in a normal brain. The neocortex, the largest part of the cerebral cortex of the brain, is composed of six distinct layers of neurons, which are established during embryonic development. First, glial cells originating from progenitor cells in the ventricular zone extend their processes throughout the cortex and anchor at the pia, the outer surface of the brain. These long fibers provide a scaffold on which neurons, produced from the same progenitor cells, migrate outwards to establish the six layers of the cortex.



Neuronal migration is impaired during Zika virus infection. Brain slice cultures from embryonic day 15 mice were infected with 10^5 PFU of Zika virus and at 4 dpi, were fixed and stained with antibody against vimentin to mark the radial glia progenitor (RGP) basal processes, which are the fibers upon which bipolar neurons migrate. ZIKV infection perturbed the RGP scaffold compared with control slices.

Glial fibers are visible as parallel tracks in the mouse embryonic brain slice cultures stained with an antibody to vimentin, a protein component of the fibers (image, left panel). When embryonic brain slice cultures were infected with Zika virus, the structure of the glial tracks was altered. Instead of parallel tracks, the fibers assumed a twisted morphology that would not allow neurons to travel from the ventricular zone to the developing neocortex (image, right panel). Disruption of glial fibers was observed after infection with Zika viruses isolated from 1947 to 2016.

These results suggest that Zika virus-mediated disruption of glial fibers during embryonic development contributes to microcephaly: if neurons cannot migrate to the pial surface, the neocortex will be thinner.

Rosenfeld AB, Doobin DJ, Warren AL, Racaniello VR, Vallee RB. 2017. Replication of early and recent Zika virus isolates throughout mouse brain development. *Proc Natl Acad Sci U S A* **114:**12273–12278.

cardiomyocytes, neurons, and hepatocytes, by treatment with specific growth factors. Viral reproduction can be studied in specific human cell types using cells derived from iPSCs.

Monolayer and suspension cell cultures do not reproduce the cell type diversity and architecture typical of tissues and organs. One way to overcome this limitation is by the use of **organotypic slice cultures**, which can be produced from a variety of organs, including brain, liver, and kidney. These cultures are prepared by slicing embryonic or postnatal rodent organs into 100- to 400-micrometer slices. They are placed on substrates, such as porous or semiporous membranes, and bathed in cell culture medium. Such cultures remain viable for 1 to 2 weeks. The effect of Zika virus infection on neuronal migration has been examined in organotypic brain slice cultures derived from embryonic mice (Box 2.3).

Another type of three-dimensional cell system is the multicellular, self-organizing **organoid** that approximates the organization, function, and genetics of specific organs. Organoids are derived from either pluripotent stem cells (iPSCs or embryonic stem cells) or adult stem cells from different organs. Organoids that model many organs such as intestine, stomach, esophagus, and brain have been established, and many have been validated for the study of a variety of viral infections (Fig. 2.3). For example, for years propagation of human noroviruses eluded virologists until the development of intestinal organoids.

The differentiation of stem cells into organoids depends on growth conditions and nutrients. For example, one type of brain organoid can be established from human pluripotent stem cells by embedding the cells in a gelatinous protein mixture that resembles the extracellular environment of many tissues. In the absence of further cues, the stem cells differentiate into structures typical of many diverse brain regions, including the cortex. In contrast, the production of intestinal organoids requires agonists of a particular signal transduction pathway. Current attempts to improve organoid cultures

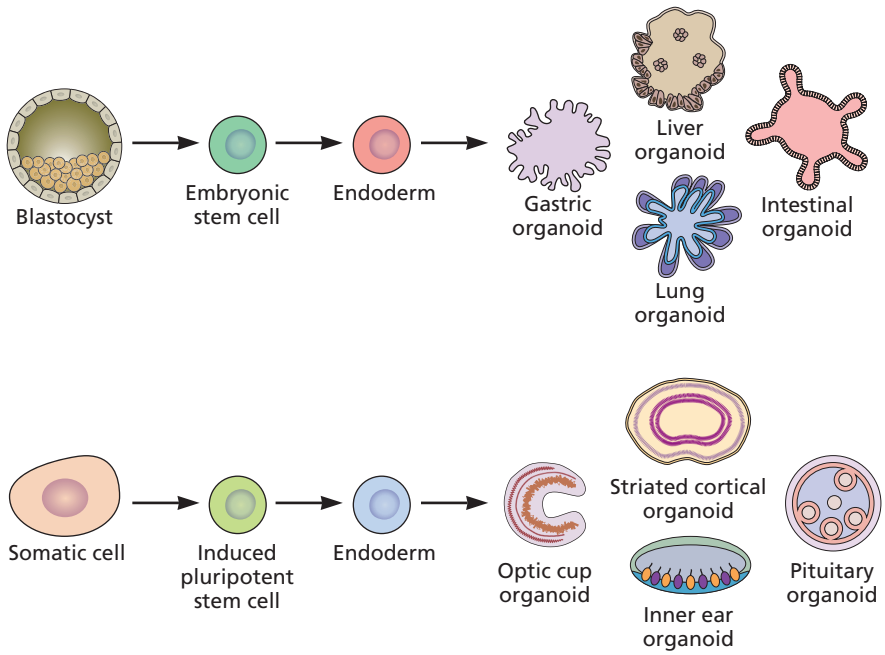


Figure 2.3 Production of organoids from stem cells. The different germ layers shown (endoderm and ectoderm) may be derived from embryonic stem cells (ESCs) or induced pluripotent stem cells (iPSCs) *in vitro* with specific differentiation protocols. After transfer into 3-dimensional systems these cells produce organoids that recapitulate the developmental steps characteristic of various organs.

include the addition of immune cells, vasculature, and commensal microorganisms, to more accurately reflect the details of tissue and organ architectures.

Air-liquid interface cultures are used to model the respiratory tract, a major site of virus entry and infection. This organ presents a challenge because its structure differs from the pharynx to the alveoli. In the trachea and bronchi, the epithelium comprises a single layer of columnar cells which contact the basement membrane. In the alveoli the epithelium is made of a thin, single cell layer to facilitate air exchange. Air-liquid interface cultures may be produced from primary human bronchial cells or respiratory cell lines (Fig. 2.4).

Because viruses are obligatory intracellular parasites, they cannot reproduce outside a living cell. An exception comes from the demonstration in 1991 that infectious poliovirus could be produced in an extract of human cells incubated with viral RNA, a feat that has not been achieved for any other virus. Consequently, most analyses of viral replication have used cultured cells, embryonated eggs, or laboratory animals. For a discussion of whether to call these different systems *in vivo* or *in vitro*, see Box 2.4.

Evidence of Viral Reproduction in Cultured Cells

Before quantitative methods for measuring viruses were developed, evidence of viral propagation was obtained by visual inspection of infected cells. Some viruses kill the cells in which they reproduce, and they may eventually detach from the cell culture plate. As more cells are infected, the changes become visible and are called **cytopathic effects**.

Many types of cytopathic effect can be seen with a simple light or phase-contrast microscope at low power, without fixing or staining the cells. These changes include the rounding up and detachment of cells from the culture dish, cell lysis, swelling of nuclei, and sometimes the formation of a

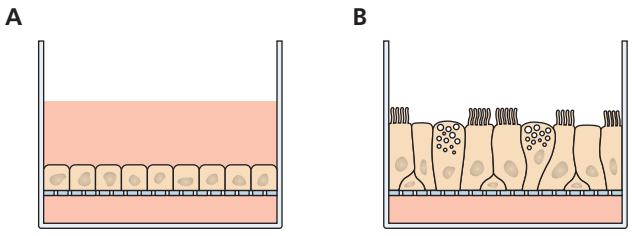


Figure 2.4 Production of airway-liquid interface cultures of bronchial epithelium. (A) Epithelial cells are seeded onto a permeable membrane and cell culture medium is supplied on both apical (top) and basal (bottom) sides. **(B)** When the cells are confluent, medium on the apical side is removed. Contact of the cells with air drives differentiation of cells towards types found in the airways, such as goblet cells, ciliated and nonciliated cells, and basal cells. Cultures may be produced that mimic tracheobronchial cells, with different cell types, or human alveolar cells with only two cell types (not shown).

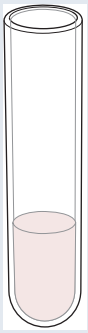
BOX 2.4

TERMINOLOGY

In vitro and *in vivo*

The terms "*in vitro*" and "*in vivo*" are common in the virology literature. *In vitro* means "in glass" and refers to experiments carried out in an artificial environment, such as a glass or plastic test tube. Unfortunately, the phrase "experiments performed *in vitro*" is used to designate not only work done in the cell-free environment of a test tube but also work done within cultured cells. The use of the phrase *in vitro* to describe living cultured cells leads to confusion and is inappropriate. *In vivo* means "in a living organism" but may be used to refer to either cells or animals. Those who work on plants avoid this confusion by using the term "*in planta*."

In this textbook, we use *in vitro* to designate experiments carried out in the absence of cells, e.g., *in vitro* translation. Work done in cells in culture is done *ex vivo*, while research done in animals is carried out *in vivo*.



group of fused cells called a **syncytium** (Fig. 2.5). High-power microscopy is required for the observation of other cytopathic effects, such as the development of intracellular masses of virus particles or unassembled viral components in the nucleus and/or cytoplasm (inclusion bodies), formation of crystalline arrays of viral proteins, membrane blebbing, duplication of membranes, and fragmentation of organelles.

The time required for the development of cytopathology varies considerably among animal viruses. For example, depending on the size of the inoculum, enteroviruses and herpes simplex virus can cause cytopathic effects in 1 to 2 days and destroy the cell monolayer in 3. In contrast, cytomegalovirus, rubella virus, and some adenoviruses may not produce such effects for several weeks.

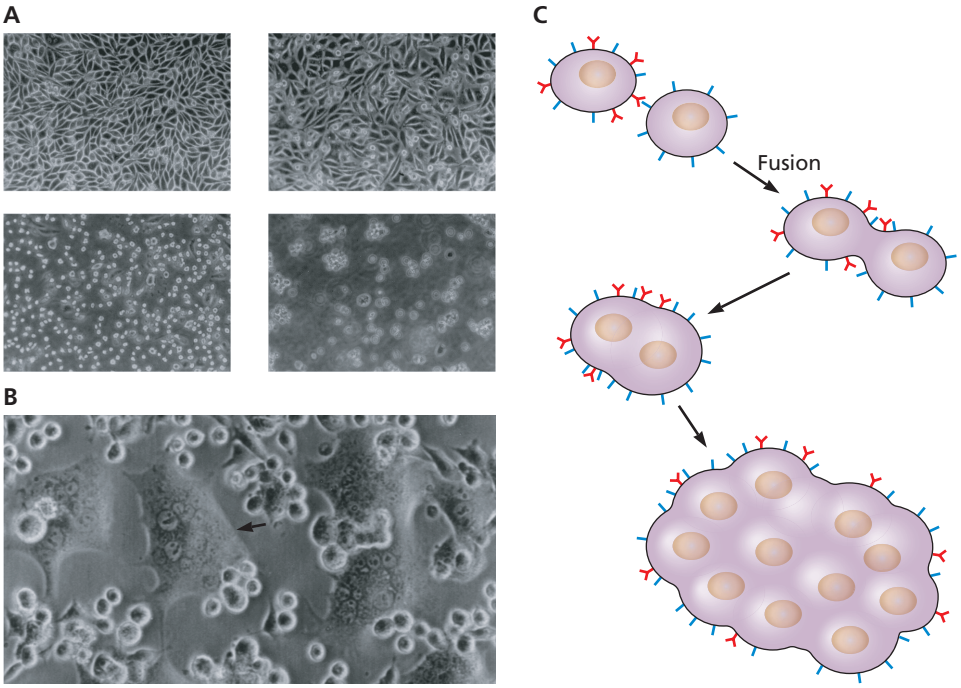


Figure 2.5 Development of cytopathic effect. (A) Cell rounding and lysis during poliovirus infection. Shown are uninfected cells (upper left) and cells 5.5 h after infection (upper right), 8 h after infection (lower left), and 24 h after infection (lower right). **(B)** Syncytium formation induced by murine leukemia virus. The field shows a mixture of individual refractile small cells and flattened syncytia (arrow), which are large, multinucleated cells. Courtesy of R. Compans, Emory University School of Medicine. **(C)** Schematic illustration of syncytium formation. Viral glycoproteins on the surface of an infected cell bind receptors on a neighboring cell, causing fusion.

The development of characteristic cytopathic effects in infected cell cultures is frequently monitored in diagnostic virology after isolation of viruses from specimens obtained from infected patients or animals. In the research laboratory, observation of cytopathic effect can be used to monitor the progress of an infection, and is often one of the phenotypic traits that characterize mutant viruses.

Some viruses multiply in cells without causing obvious cytopathic effects. For example, many members of the families *Arenaviridae*, *Paramyxoviridae*, and *Retroviridae* do not cause obvious damage to cultured cells. Infection by such viruses must therefore be assessed using alternative methods, as described in "Assay of Viruses" below.

Embryonated Eggs

Before the advent of cell culture, many viruses were propagated in embryonated chicken eggs (Fig. 2.6). At 5 to 14 days after fertilization, a hole is drilled in the shell and virus is injected into the site appropriate for its replication. This method of virus propagation is now routine only for influenza virus. The robust yield of this virus from chicken eggs has led to their widespread use in research laboratories and for vaccine production.

Laboratory Animals

In the early 1900s, when viruses were first isolated, freezers and cell cultures were not available, and it was necessary to maintain virus stocks by continuous passage from animal to animal. This practice not only was inconvenient but also, as we shall see, led to the selection of viral mutants (Volume II, Chapter 7). For example, monkey-to-monkey intracerebral passage of poliovirus selected a mutant that could no longer infect chimpanzees by the oral route, the natural means of infection.

Although cell culture has supplanted animals for propagating most viruses, experimental infection of laboratory animals has always been, and will continue to be, obligatory for studying the processes by which viruses cause disease. The study in monkeys of poliomyelitis, the paralytic disease caused by poliovirus, led to an understanding of the basis of this disease and was instrumental in the development of a successful vaccine. Similarly, the development of vaccines against hepatitis B virus would not have been possible without experimental studies with chimpanzees. Understanding how the immune system or any complex organ reacts to a virus cannot be achieved without research on living animals. The development of viral vaccines, antiviral drugs, and diagnostic tests for veterinary medicine has also benefited from research on diseases in laboratory animals. Despite their utility, it must be appreciated that all animal models are surrogates for the events that occur during viral infections of humans.

Assay of Viruses

There are two main types of assay for detecting viruses: biological and physical. Because viruses were first recognized by their infectivity, the earliest assays focused on this most sensitive and informative property. However, biological assays such as the plaque assay and end-point titration methods do not detect noninfectious particles. In contrast, all particles are accounted for with physical assays such as electron microscopy or by immunological methods. Knowledge of the number of noninfectious particles is useful for assessing the quality of a virus preparation.

Measurement of Infectious Units

One of the most important procedures in virology is measuring the **virus titer**, the concentration of infectious virus particles in a sample. This parameter is determined by inoculating serial dilutions of virus into host cell cultures, chicken embryos, or laboratory animals and monitoring for evidence of virus multiplication. The response may be quantitative

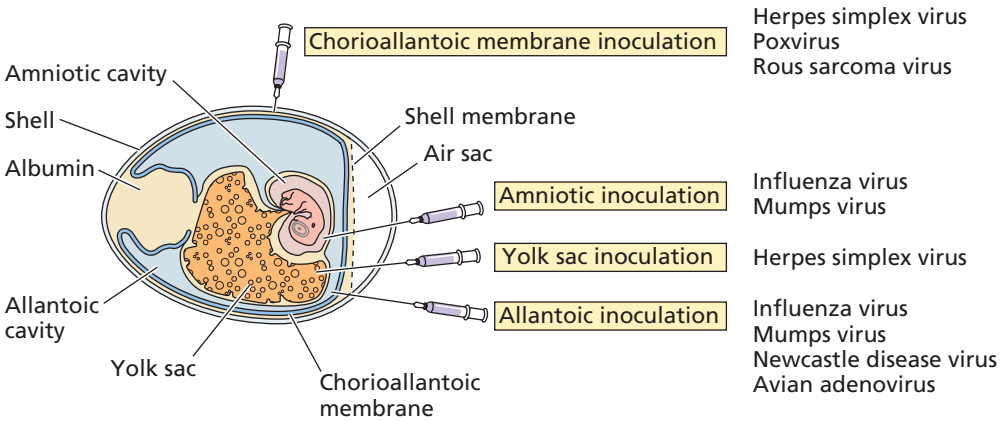


Figure 2.6 Growth of viruses in embryonated eggs. The cutaway view of an embryonated chicken egg shows the different routes by which viruses are inoculated into eggs and the distinct compartments in which particular viruses may propagate.

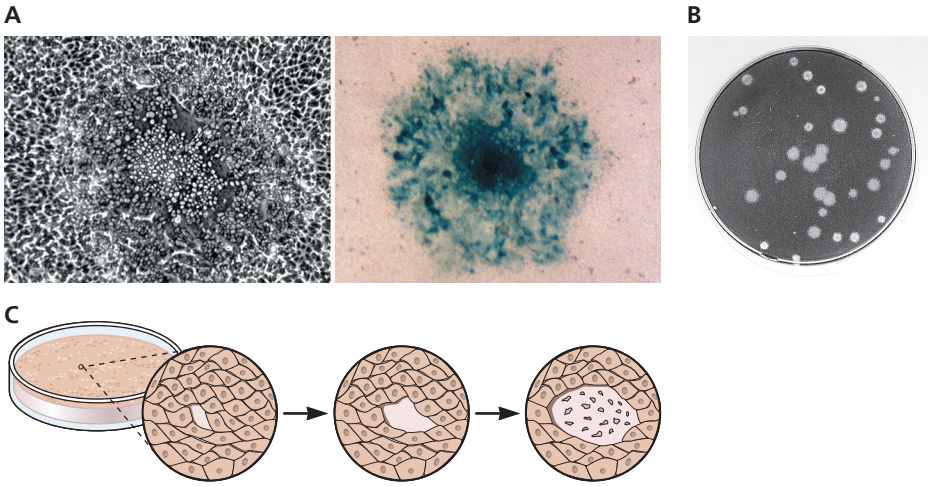


Figure 2.7 Plaques formed by different animal viruses. (A) Photomicrograph of a single plaque formed by pseudorabies virus in bovine kidney cells. Shown are unstained cells (left) and cells stained with the chromogenic substrate X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside), which is converted to a blue compound by the product of the *lacZ* gene carried by the virus (right). Courtesy of B. Banfield, Princeton University. **(B)** Plaques formed by poliovirus on human HeLa cells stained with crystal violet. **(C)** Illustration of the sequential spread of a cytopathic virus from an initial infected cell to neighboring cells, resulting in a plaque.

(as in assays for plaques, fluorescent foci, infectious centers, or abnormal growth and morphology) or all-or-none, in which the presence or absence of infection is measured (as in an end-point dilution assay). Please note that "titer" is not a verb.

Plaque Assay

The measurement of virus titers by plaque assay was first developed for bacteriophages by d'Herelle in 1917 and then modified for animal viruses by Renato Dulbecco in 1952. In this procedure, monolayers of cultured cells are incubated with a preparation of virus to allow adsorption to cells. After removal of the inoculum, the cells are covered with nutrient medium containing a supplement, most commonly agar, which forms a gel. When the original infected cells release new progeny particles, the gel restricts their spread to neighboring uninfected cells. As a result, each infectious particle produces a circular zone of infected cells, a **plaque**. If the infected cells are damaged, the plaque can be distinguished from the surrounding monolayer. In time, the plaque becomes large enough to be seen with the naked eye (Fig. 2.7). Only viruses that cause visible damage of cultured cells can be assayed in this way. A movie that depicts the microscopic development of a plaque can be found at this link: http://bit.ly/Virology_VZVGFP.

For the majority of animal viruses, there is a linear relationship between the number of infectious particles and the plaque count (Fig. 2.8). One infectious particle is therefore sufficient to initiate infection, and the virus is said to infect cells with **one-hit kinetics**. Some examples of **two-hit kinet-**

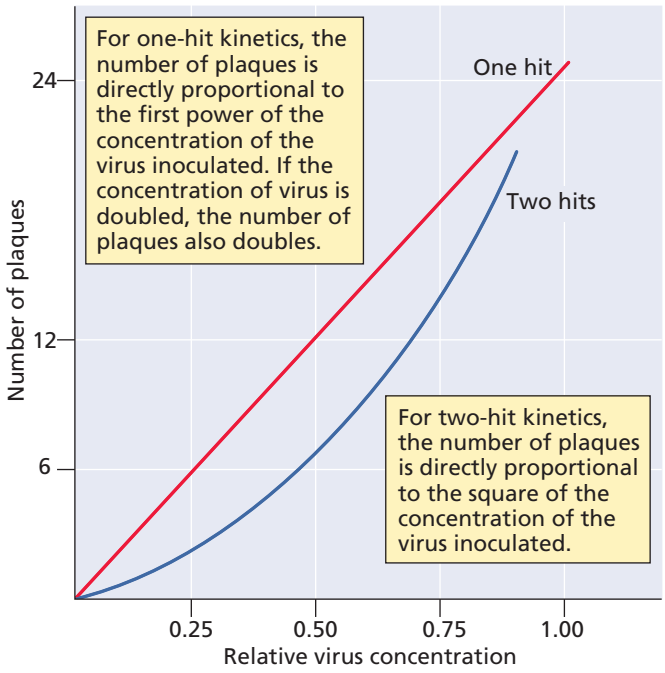


Figure 2.8 The dose-response curve of the plaque assay. The number of plaques produced by a virus with one-hit kinetics (red) or two-hit kinetics (blue) is plotted against the relative concentration of the virus. In two-hit kinetics, there are two classes of uninfected cells, those receiving one particle and those receiving none. The Poisson distribution can be used to determine the proportion of cells in each class: they are e^{-m} and me^{-m} (Box 2.12). Because one particle is not sufficient for infection, $P(0) = e^{-m}(1 + m)$. At a very low multiplicity of infection, this equation becomes $P(\text{i}) = (1/2)m^2$ (where i = infection), which gives a parabolic curve.

ics, in which two different types of virus particle must infect a cell to ensure replication, have been recognized. An example is the genomes of some (+) strand RNA viruses of plants that consist of two RNA molecules which are encapsidated separately. Both RNAs are required for infectivity. The dose-response curve in plaque assays for these viruses is therefore parabolic rather than linear (Fig. 2.8).

The titer of a virus stock can be calculated in **plaque-forming units (PFU) per milliliter** (Box 2.5). The plaque assay may also be used to prepare clonal virus stocks. When one infectious virus particle initiates a plaque, the viral progeny within the plaque are biological clones, and virus stocks prepared from a single plaque are known as **plaque purified**. The tip of a small pipette is plunged into the overlay above the plaque, and the plug of agar containing the virus is recovered. The virus within the agar plug is eluted into buffer and used to prepare virus stocks. To ensure purity, this process is usually repeated at least one more time.

Fluorescent-Focus Assay

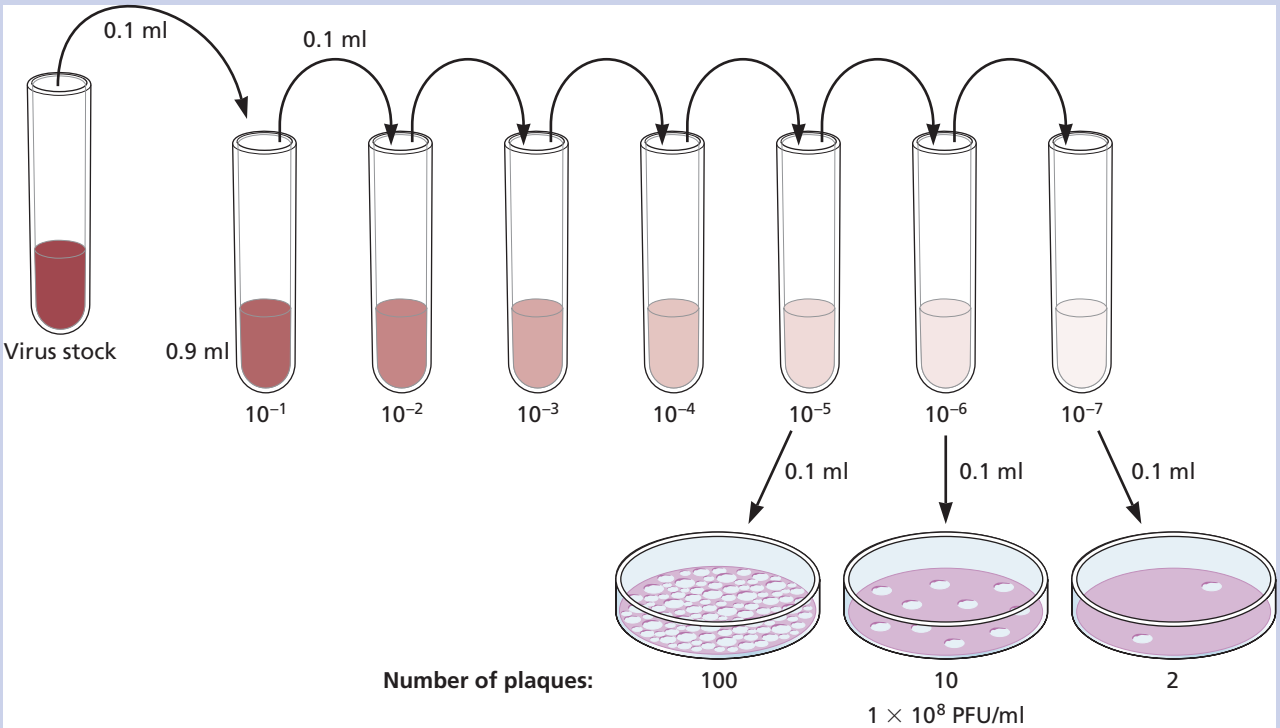
The fluorescent-focus assay, a modification of the plaque assay, can be done more rapidly and is useful in determining the titers of viruses that do not form plaques. The initial procedure is the same as in the plaque assay. However, after a period sufficient for adsorption and gene expression, cells are made permeable and incubated with an antibody raised against a viral protein. A second antibody, which recognizes the first, is then added. This second antibody is usually conjugated to a fluorescent molecule.

BOX 2.5

METHODS

Calculating virus titer from the plaque assay

To calculate the titer of a virus in plaque-forming units (PFU) per milliliter, 10-fold serial dilutions of a virus stock are prepared in a buffer, and suitable aliquots are inoculated onto susceptible cell monolayers which are covered with an agar overlay (see figure). After a suitable incubation period, the monolayers are stained and the plaques are counted. To minimize error in calculating the virus titer, only plates containing between 10 and 100 plaques are counted, depending on the area of the cell culture vessel. Plates with >100 plaques are generally not counted because the plaques may overlap, causing inaccuracies. According to statistical principles, when 100 plaques are counted, the sample titer varies by ±10%. For accuracy, each dilution is plated in duplicate or triplicate (not shown in the figure). In the example shown, 10 plaques are observed on the plate produced from the 10^{-6} dilution. Therefore, the 10^{-6} dilution tube contains 10 PFU per 0.1 ml, or 100 PFU per ml, and the titer of the virus stock is 100×10^6 or 1×10^8 PFU/ml.



The cells are then examined under a microscope at an appropriate wavelength. The titer of the virus stock is expressed in fluorescent-focus-forming units per milliliter. When the gene encoding a fluorescent protein is incorporated into the viral genome, foci may be detected without the use of antiviral antibodies.

Infectious-Centers Assay

Another modification of the plaque assay, the infectious-centers assay, is used to determine the fraction of cells in a culture that are infected with a virus. Monolayers of infected cells are suspended before progeny viruses are produced. Dilutions of a known number of infected cells are then plated on monolayers of susceptible cells, which are covered with an agar overlay. The number of plaques that form on the indicator cells is a measure of the number of cells infected in the original population. The fraction of infected cells can therefore be determined. A typical use of the infectious-centers assay is to measure the proportion of virus-producing cells in persistently infected cultures.

Transformation Assay

The transformation assay provides a method for determining the titers of some retroviruses that do not form plaques. For example, when Rous sarcoma virus transforms chicken embryo cells, the cells lose their contact inhibition (the property that governs whether cells in culture grow as a single monolayer [see Volume II, Chapter 6]) and become heaped up on one another. The transformed cells form small piles, or **foci**, that can be distinguished easily from the rest of the monolayer (Fig. 2.9). Infectivity is expressed in focus-forming units per milliliter.

End-Point Dilution Assay

The end-point dilution assay provided a means to determine virus titer before the development of the plaque assay. It is still used for measuring the titers of certain viruses that do not form plaques or for determining the virulence of a virus in animals. Serial dilutions of a virus stock are inoculated into replicate test units (typically 8 to 10), which can be cell cultures, eggs, or animals. The number of test units that have become infected is then determined for each virus dilution. In cell culture, infection may be determined by the development of cytopathic effect; in eggs or animals, infection may be gauged by virus titer, death, or disease. An example of an end-point dilution assay using cell cultures is shown in Box 2.6, with results expressed as 50% infectious dose (ID_{50}) per milliliter. This type of assay is also suitable for high-throughput applications.

When the end-point dilution assay is used to assess the virulence of a virus or its capacity to cause disease (Volume II, Chapter 1), the result can be expressed in terms of 50% lethal dose (LD_{50}) per milliliter or 50% paralytic dose (PD_{50}) per milliliter, end points of death and paralysis, respectively. The 50% end point determined in an animal host can be related to virus titer, determined separately by plaque assay or other means. In this way, the effects of the route of inoculation or specific mutations on viral virulence can be quantified.

Efficiency of Plating

Efficiency of plating is defined as the infectious virus titer (in PFU/ml) divided by the total number of virus particles in the sample. The **particle–to–plaque-forming-unit**

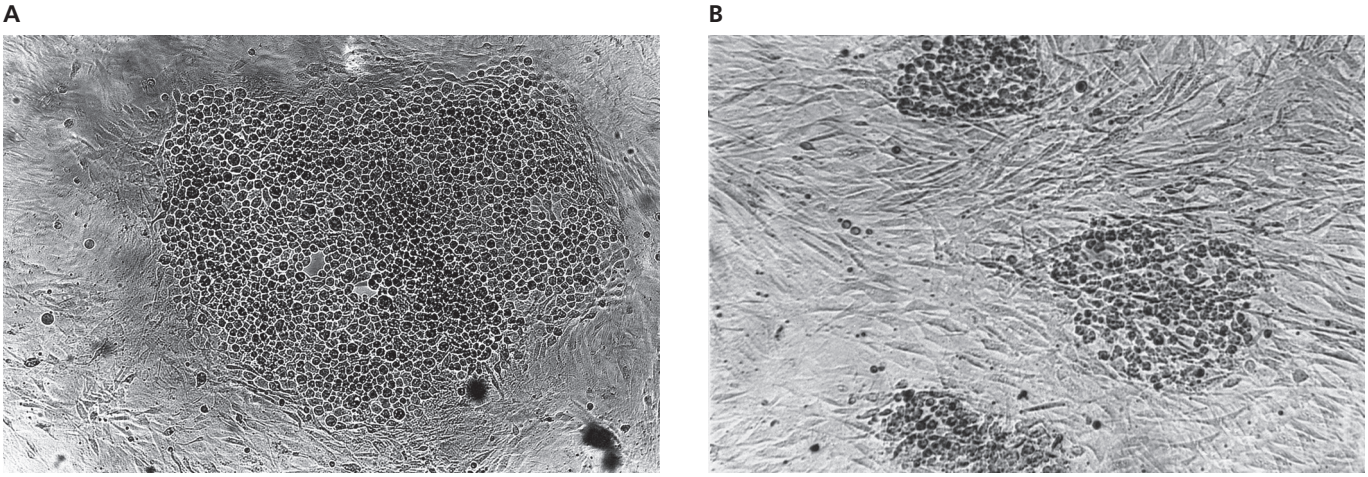
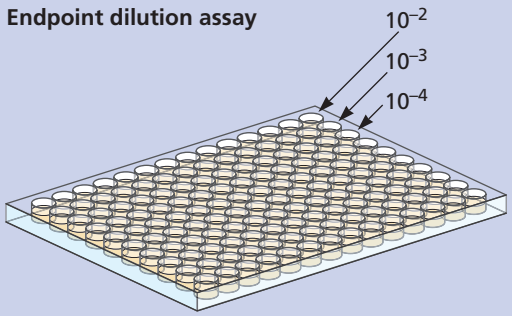


Figure 2.9 Transformation assay. Chicken cells transformed by two different strains of Rous sarcoma virus are shown. Loss of contact inhibition causes cells to pile up rather than grow as a monolayer. One focus is seen in panel **A** and three foci are seen in panel **B** at the same magnification. Courtesy of H. Hanafusa, Osaka Bioscience Institute.

BOX 2.6

METHODS

End-point dilution assays



Virus dilution	Cytopathic effect									
10^{-2}	+	+	+	+	+	+	+	+	+	+
10^{-3}	+	+	+	+	+	+	+	+	+	+
10^{-4}	+	+	−	+	+	+	+	+	+	+
10^{-5}	−	+	+	−	+	−	−	+	−	+
10^{-6}	−	−	−	−	−	−	+	−	−	−
10^{-7}	−	−	−	−	−	−	−	−	−	−

End-point dilution assays are usually carried out in multiwell plastic plates (see the figure above). In the example shown in the adjacent table above, 10 monolayer cell cultures were infected with each virus dilution. After the incubation period, plates that displayed cytopathic effect were scored +. At high dilutions, none of the cell cultures are infected because no infectious particles are delivered to the cells; at low dilutions, every culture is infected. The **end point** is the dilution of virus that affects 50% of the test units. This number can be calculated from the data and expressed as 50% infectious dose (ID_{50}) per milliliter. Fifty percent of the cell cultures displayed cytopathic effect at the 10^{-5} dilution, and therefore, the virus stock contains 10^5 $TCID_{50}$ (tissue culture infectious dose) units.

In most cases, the 50% end point does not fall on a dilution tested as shown in the example; for this reason, various statistical procedures have been developed to calculate the end point of the titration. In one popular method, the dilution containing the ID_{50} is identified by interpolation between the dilutions on either side of this value. The assumption is made that the location of the 50% end point varies linearly with the log of the dilution. Because the number of test units used at each dilution is usually small, the accuracy of this method is relatively low. For example, if six test units are used at each 10-fold dilution, differences in virus titer of only 50-fold or more can be detected reliably. The method is illustrated in the second example below, in which the lethality of poliovirus in mice is the end point. Eight mice were inoculated per dilution. In the method of Reed and Muench, the results are pooled, as shown in the table below, which equalizes chance variations (another way to achieve the same result would be to utilize greater numbers of animals at each dilution). The interpolated value of the 50% end point, which in this case falls between the 5th and 6th dilutions, is calculated to be $10^{-6.5}$. The virus sample therefore contains $10^{6.5}$ LD_{50} (50% lethal dose). The LD_{50} may also be calculated as the concentration of the stock virus in PFU per milliliter (1×10^9) times the 50% end-point titer. In the example shown, the LD_{50} is 3×10^2 PFU.

Reed LJ, Muench H. 1938. A simple method of estimating fifty percent endpoints. *Am J Hyg* **27**:493–497.

Dilution	Alive	Dead	Total alive	Total dead	Mortality ratio	Mortality (%)
10^{-2}	0	8	0	40	0/40	100
10^{-3}	0	8	0	32	0/32	100
10^{-4}	1	7	1	24	1/25	96
10^{-5}	0	8	1	17	1/18	94
10^{-6}	2	6	3	9	3/12	75
10^{-7}	5	3	8	3	8/11	27

(PFU) ratio, a term more commonly used today, is the inverse value (Table 2.1). For many bacteriophages, the particle-to-PFU ratio approaches 1, the lowest value that can be obtained. However, for animal viruses, this value can be much higher, ranging from 1 to 10,000. These high values have complicated the study of animal viruses. For example, when the particle-to-PFU ratio is high, it may not be clear that properties measured biochemically are in fact those of the infectious particle or those of the noninfectious component.

Although the linear nature of the dose-response curve indicates that a single particle is capable of initiating an infection (one-hit kinetics) (Fig. 2.8), the high particle-to-PFU ratio of many viruses demonstrates that not all virus particles are successful. High values are sometimes caused by the presence of noninfectious particles with genomes that harbor lethal mutations or that have been damaged during growth or purification (**defective particles**). An alternative explanation is that although all viruses in a preparation are in fact capable of initiating infection, not all of them succeed because of the complexity of the infectious cycle. Failure at any one step in the cycle prevents completion. In this case, a high particle-to-PFU ratio indicates not that most particles

Table 2.1 Particle-to-PFU ratios of some animal viruses

Virus	Particle/PFU ratio
Papillomaviridae	
Papillomavirus	10,000
Picornaviridae	
Poliovirus	30–1,000
Herpesviridae	
Herpes simplex virus	50–200
Polyomaviridae	
Polyomavirus	38–50
Simian virus 40	100–200
Adenoviridae	20–100
Poxviridae	1–100
Orthomyxoviridae	
Influenza virus	20–50
Reoviridae	
Reovirus	10
Alphaviridae	
Semliki Forest virus	1–2

are defective but, rather, that they failed to complete the infection.

Measurement of Virus Particles

Although the numbers of virus particles and infectious units are often not equal, assays for particle number are frequently used to approximate the number of infectious particles present in a sample. For example, assuming that the ratio of infectious units to physical particles is constant, the concentration of viral DNA or protein can be used to estimate the number of infectious particles. Biochemical or physical assays are usually more rapid and easier to carry out than those for infectivity, which may be slow, cumbersome, or impossible. Assays for subviral components also provide information on particle number if the amount of these components in each virus particle is known.

Electron Microscopy

With few exceptions, virus particles are too small to be observed directly by light microscopy. However, they can be seen readily in the electron microscope. If a sample contains only one type of virus, the particle count can be determined. A virus preparation is mixed with a known concentration of latex beads, and the numbers of virus particles and beads are then counted, allowing the concentration of the virus particles in the sample to be determined by comparison.

Hemagglutination

Members of the *Adenoviridae*, *Orthomyxoviridae*, and *Paramyxoviridae*, among others, contain proteins that bind to erythrocytes (red blood cells); these viruses can link multiple cells, resulting in formation of a lattice. This property is called **hemagglutination**. For example, influenza viruses contain an envelope glycoprotein called hemagglutinin (HA), which binds to *N*-acetylneuraminic acid-containing glycoproteins on erythrocytes. In practice, 2-fold serial dilutions of the virus stock are prepared, mixed with a known quantity of red blood cells, and added to small wells in a plastic tray (Fig. 2.10). Unlinked red blood cells tumble to the bottom of the well and form a sharp dot or button. In contrast, agglutinated red blood cells form a diffuse lattice that coats the well. Because the assay is rapid (30 min), it is often used as a quick indicator of the relative quantities of virus particles. However, it is not sufficiently sensitive to detect small numbers of particles.

Centrifugation

The use of centrifugal force to separate particles from solution according to size, shape, or density has been a staple of virology. The instrument used for such separations is called a **centrifuge**, which can range from small tabletop devices that accommodate small tubes to large floor models with greater capacity and to ultracentrifuges that can achieve revolutions per minute in excess of 70,000. The ultracentrifuge was invented by Theodor Svedberg in 1925, and it is the first initial of his last name that is used to describe the sedimentation coefficient of a particle as measured by centrifugation, e.g., the 16S ribosomal subunit.

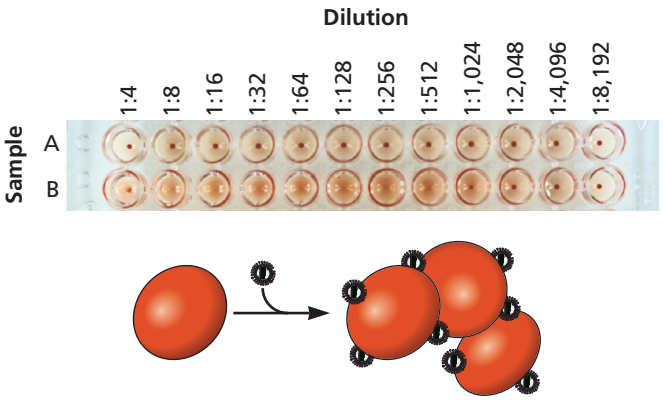


Figure 2.10 Hemagglutination assay. (**Top**) Samples of different influenza viruses were diluted, and a portion of each dilution was mixed with a suspension of chicken red blood cells and added to the wells. After 30 min at 4°C, the wells were photographed. Sample A does not contain virus. Sample B causes hemagglutination until a dilution of 1:512 and therefore has a hemagglutination titer of 512. Elution of the virus from red blood cells at the 1:4 dilution is caused by neuraminidase in the virus particle. This enzyme cleaves *N*-acetylneuraminic acid from glycoprotein receptors and elutes bound viruses from red blood cells. (**Bottom**) Schematic illustration of hemagglutination of red blood cells by influenza virus. **Top**, Courtesy of C. Basler and P. Palese, Mount Sinai School of Medicine of the City University of New York.

It would not be wrong to state that every virology laboratory is in possession of at least one centrifuge and probably has access to more. The uses of the centrifuge in virology are manifold: from low-speed separation of virus particles from infected cell debris in cell culture medium to fractionation of infected cells to isolate nuclei, cytoplasm, or ribosomes, and to purification of virus particles.

Differential centrifugation is used to separate viruses, organelles, or subcellular structures from cells. Preformed gradients of sucrose are often used because particles that move with various velocities can be separated differentially in the increasing viscosity of the solution. One application of sucrose gradients is the purification of virus particles. Another is **polysome profiling**, an analysis of the mRNAs associated with ribosomes (Fig. 2.11). Because mRNAs undergoing translation can be associated with different numbers of ribosomes, they can be separated on a sucrose gradient. A more modern use of the polysome profile is to extract the RNA from each fraction and determine which mRNAs are being actively translated.

Another method for purifying viruses is by **isopycnic centrifugation**, which separates particles solely on the basis of their density. A virus preparation is mixed with a compound (e.g., cesium chloride) that forms a density gradient during centrifugation. Virus particles move down the tube until they reach the point at which their density is the same as the gradient medium. Structural studies of virus particles often require highly purified preparations which can be made by differential or isopycnic centrifugation.

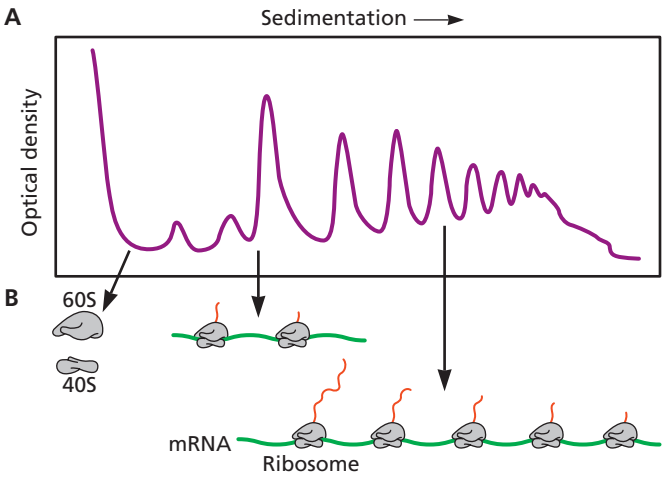


Figure 2.11 Polysome analysis. To study the association of mRNAs with ribosomes, cell lysates are prepared and separated by centrifugation through sucrose gradients. Fractions are collected and their optical density measured to locate mRNAs bound to one or more ribosomes. The graph shows the optical density of fractions from the top (left) to the bottom (right) of the gradient. The slower-moving materials at the top of the gradient are ribosomal subunits, while mRNAs associated with one or more ribosomes move faster in the sucrose gradient.

Measurement of Viral Enzyme Activity

Some animal virus particles contain nucleic acid polymerases, which can be detected by mixing permeabilized particles with precursors and measuring their incorporation into nucleic acid. This type of assay is used most frequently for retroviruses, many of which neither transform cells nor form plaques. The reverse transcriptase incorporated into the virus particle is assayed by mixing cell culture supernatants with a mild detergent (to permeabilize the viral envelope), an RNA template and primer, and a radioactive nucleoside triphosphate. If reverse transcriptase is present, a radioactive product will be produced by priming on the template. This product can be detected by precipitation or bound to a filter and quantified. Because enzymatic activity is proportional to particle number, this assay allows rapid tracking of virus production in the course of an infection. Many of these assays have been modified to permit the use of safer, nonradioactive substrates. For example, when nucleoside triphosphates conjugated to biotin are used, the product can be detected with streptavidin (which binds biotin) conjugated to a fluorochrome. Alternatively, the reaction products may be quantified by quantitative real-time PCR (see "Detection of Viral Nucleic Acids" below).

Serological Methods

The specificity of the antibody-antigen reaction has been used to design a variety of assays for viral proteins and antiviral antibodies. These techniques, such as immunostaining, immunoprecipitation, immunoblotting, and the enzyme-linked immunosorbent assay, are by no means limited to virology: all these approaches have been used extensively to study the structures and functions of cellular proteins.

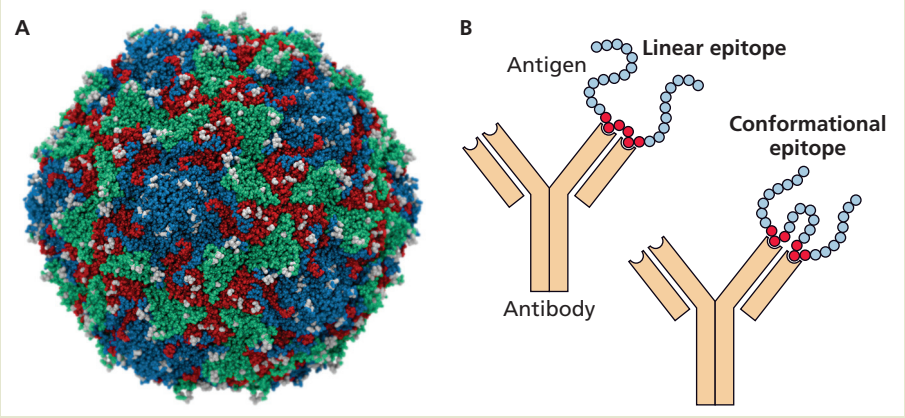
Virus neutralization. When a virus preparation is inoculated into an animal, an array of antibodies is produced. These antibodies can bind to virus particles, but not all of them can block infectivity (**neutralize**), as discussed in Volume II, Chapter 4. Virus neutralization assays are usually conducted by mixing dilutions of antibodies with virus, incubating them, and assaying for remaining infectivity in cultured cells, eggs, or animals. The end point is defined as the highest dilution of antibody that inhibits the development of cytopathic effect in cells or virus reproduction in eggs or animals.

Some neutralizing antibodies define **type-specific antigens** on the virus particle. For example, the three **serotypes** of poliovirus are distinguished on the basis of neutralization tests: type 1 poliovirus is neutralized by antibodies to type 1 virus but not by antibodies to type 2 or type 3 poliovirus. The results of neutralization tests were once used for virus classification, a process now accomplished largely by comparing viral genome sequences. Nevertheless, the detection of antiviral

BOX 2.7

DISCUSSION

Neutralization antigenic sites



Antigenic sites defined by antibodies. (A) Locations of neutralization antigenic sites on the capsid of poliovirus type 1. Amino acids that change in viral mutants selected for resistance to neutralization by monoclonal antibodies are shown in white on a model of the viral capsid. These amino acids are in VP1 (blue), VP2 (green), and VP3 (red) on the surface of the virus particle. Figure courtesy of Jason Roberts, Victorian Infectious Diseases Reference Laboratory, Doherty Institute, Melbourne, Australia. **(B)** Conformational and linear epitopes bound to antibody molecules. Linear epitopes are made of consecutive amino acids, while conformational epitopes are made of amino acids from different parts of the protein.

Knowledge of the antigenic structure of a virus is useful in understanding the immune response to these agents and in designing new vaccination strategies. The use of **monoclonal antibodies** (antibodies of a single specificity made by a clone of antibody-producing cells) in neutralization assays permits mapping of antigenic sites on a virus particle or of the amino acid sequences that are recognized by neutralizing antibodies.

Each monoclonal antibody binds specifically to 8 to 12 residues that fit into the antibody-combining site. These amino acids are either next to one another either in primary sequence (**linear epitope**) or in the folded structure of the native protein (**nonlinear** or **conformational epitope**). In contrast, **polyclonal antibodies** comprise the repertoire produced in an animal against the many epitopes of an antigen. Antigenic sites may be identified by cross-linking a monoclonal antibody to the virus and determining which protein is the target of that antibody. Epitope mapping may also be performed by assessing the abilities of monoclonal antibodies to bind synthetic peptides representing viral protein sequences. When the monoclonal antibody recognizes a linear epitope, it may react with the protein in immunoblot analysis, facilitating direct identification of the viral protein harboring the antigenic site.

An elegant understanding of antigenic structures has come from the isolation and study of variant viruses that are resistant to neutralization with specific monoclonal antibodies (called **monoclonal antibody-resistant variants**). By identifying the amino acid change(s) responsible for this phenotype, the antibody-binding site can be located and, together with three-dimensional structural data, can provide detailed information on the nature of antigenic sites that are recognized by neutralizing antibodies (see the figure).

antibodies in animal sera is still extremely important for identifying infected hosts. These antibodies may also be used to map the three-dimensional structure of neutralization antigenic sites on the virus particle (Box 2.7).

Hemagglutination inhibition. Antibodies against viral proteins with hemagglutination activity can block the ability of virus to bind red blood cells. In this assay, dilutions of antibodies are incubated with virus, and erythrocytes are added as outlined above. After incubation, the titer is read as the highest dilution of antibody that inhibits hemagglutination. This test is sensitive, simple, inexpensive, and rapid, and can be used to detect antibodies to viral hemagglutinin in animal and human sera. For example, hemagglutination inhibition assays were used to identify individuals who had been infected with the newly discovered avian influenza A (H7N9) virus in China during the 2013 outbreak.

Visualization of proteins. Antibodies can be used to visualize viral or cellular proteins in infected cells or tissues. In direct immunostaining, an antibody that recognizes a viral protein is coupled directly to an indicator such as a fluorescent dye or an enzyme (Fig. 2.12). A more sensitive approach is indirect immunostaining, in which a second antibody is coupled to the indicator. The second antibody recognizes a common region on the virus-specific antibody.

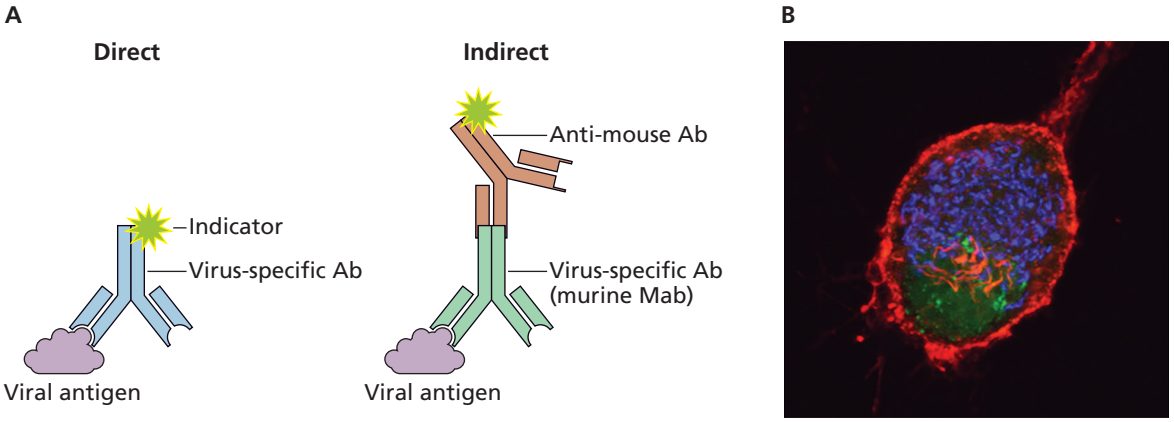


Figure 2.12 Direct and indirect methods for antigen detection. (A) The sample (tissue section, smear, or bound to a solid phase) is incubated with a virus-specific antibody (Ab). In direct immunostaining, the antibody is linked to an indicator such as fluorescein. In indirect immunostaining, a polyclonal antibody, which recognizes several epitopes on the virus-specific antibody, is coupled to the indicator. Mab, monoclonal antibody. **(B)** Use of immunofluorescence to visualize pseudorabies virus replication in neurons. Superior cervical ganglion neurons were grown in culture and infected with a recombinant virus that produces green fluorescent protein (GFP) fused to the VP26 capsid protein. Neurons were stained with AF568-phalloidin, which stains actin red, and anti-GM130 to stain the Golgi blue. GFP-VP26 is visualized by direct fluorescence. Courtesy of L. Enquist, Princeton University.

Multiple second-antibody molecules bind to the first antibody, resulting in an increased signal from the indicator compared with that obtained with direct immunostaining. Furthermore, a single indicator-coupled second antibody can be used in many assays, avoiding the need to purify and couple an indicator to multiple first antibodies.

In practice, virus-infected cells (unfixed or fixed with acetone, methanol, or paraformaldehyde) are incubated with polyclonal or monoclonal antibodies (Box 2.7) directed against viral antigen. Excess antibody is washed away, and in direct immunostaining, cells are examined by microscopy. For indirect immunostaining, the second antibody is added before examination of the cells by microscopy. Commonly used indicators fluoresce on exposure to UV light. Filters are placed between the specimen and the eyepiece to remove blue and UV light so that the field is dark, except for cells to which the antibody has bound, which emit light of distinct colors (Fig. 2.12). Today's optics are much better at keeping the wavelengths separated, permitting the use of different colors to detect various components in the same specimen. Antibodies can also be coupled to molecules other than fluorescent indicators, including enzymes such as alkaline phosphatase, horseradish peroxidase, and β-galactosidase, a bacterial enzyme that in a test system converts the chromogenic substrate X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) to a blue product. In these instances, excess antibody is washed away, a suitable chromogenic substrate is added, and the presence of the indicator antibody is revealed by the development of a color that can be visualized.

Immunostaining has been applied widely in the research laboratory for determining the subcellular localization of cellular and viral proteins (Fig. 2.12), monitoring the synthesis of viral proteins, determining the effects of mutation on protein production, localizing the sites of viral genome replication in animal hosts, and determining the effect of infection on structure of the tissue. It is the basis of the fluorescent-focus assay.

Immunostaining of viral antigens in smears of clinical specimens may be used to diagnose viral infections. For example, direct and indirect immunofluorescence assays with nasal swabs or washes can detect a variety of viruses, including influenza virus and measles virus. Viral proteins or nucleic acids may also be detected in infected animals by immunohistochemistry. In this procedure, tissues are embedded in a solid medium such as paraffin, and thin slices are produced using a microtome. Viral antigens can be detected within the cells in the sections by direct and indirect immunofluorescence assays.

Enzyme immunoassay. Detection of viral antigens or antiviral antibodies can be accomplished by solid-phase methods, in which an antiviral antibody or protein is adsorbed to a plastic surface (Fig. 2.13A). To detect antibodies to viruses, viral protein is first linked to the plastic support, and then the specimen is added (Fig. 2.13B). Like other detection methods, enzyme immunoassays are used in both experimental and diagnostic virology. In the clinical laboratory, enzyme immunoassays are used to detect a variety of viruses, including rotavirus, herpes simplex virus, and human immunodeficiency viruses. A modification of the enzyme immunoassay is the lateral flow immunochromatographic assay, which has

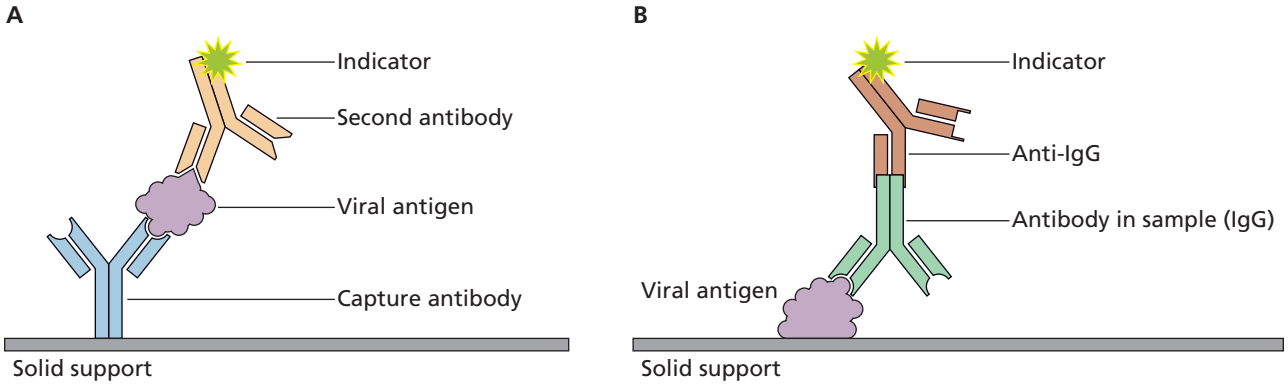


Figure 2.13 Detection of viral antigen or antibodies against viruses by enzyme-linked immunosorbent assay (ELISA). **(A)** To detect viral proteins in serum or clinical samples, antibodies specific for the virus are immobilized on a solid support such as a plastic well. The sample is placed in the well, and viral proteins are "captured" by the immobilized antibody. After washing to remove unbound proteins, a second antibody against the virus is added, which is linked to an indicator. The second antibody will bind if viral antigen has been captured by the first antibody. Unbound second antibody is removed by another washing, and when the indicator is an enzyme, a chromogenic molecule that is converted by the enzyme to an easily detectable product is then added. The enzyme amplifies the signal because a single catalytic enzyme molecule can generate many product molecules. Another wash is done to remove unbound second antibody. If viral antigen has been captured by the first antibody, the second antibody will bind and the complex will be detected by the indicator. **(B)** To detect antibodies to a virus in a sample, viral antigen is immobilized on a solid support such as a plastic well. The test sample is placed in the well, and antiviral IgG antibodies present in the sample will bind the immobilized antigen. After washing to remove unbound components in the sample, a second antibody, directed against a general epitope on the first antibody, is added. Unbound second antibody is removed by another wash. If antibodies against the virus are present in the specimen, the second antibody will bind to them and the complex will be detected via the indicator attached to the second antibody, as described in (A).

been used in rapid antigen detection test kits (Fig. 2.14). The lateral flow immunochromatographic assay does not require instrumentation and can be read in 5 to 20 min in a physician's office or in the field. Commercial rapid antigen detection assays are currently available for influenza virus, respiratory syncytial virus, and rotavirus.

Fluorescent Proteins

The discovery of green fluorescent protein revolutionized the study of the cell biology of virus infection. This protein, isolated from the jellyfish *Aequorea victoria*, is a convenient reporter for monitoring gene expression, because it is directly visible in living cells without the need for fixation, substrates, or coenzymes. Mutagenesis of the gene encoding this protein has led to the development of new fluorescent probes ranging in color from blue to yellow (Fig. 2.15A). Additional fluorescent proteins emitting in the red, deep red, cyan, green, yellow, and orange spectral regions have been isolated from other marine species. Codon optimization for maximum translation in specific cell types and improved stability and brightness are other modifications that have broadened the utility of these proteins.

Fluorescence Microscopy

Fluorescence microscopy allows virologists to study all steps of virus reproduction, including cell surface attachment, cell entry, trafficking, replication, assembly, and egress. Single virus particle tracking can be achieved by inserting the coding sequence for a fluorescent protein into the viral genome, often fused to the coding region of a viral protein. The fusion protein is incorporated into the viral particle, which is visible in cells by fluorescence microscopy (Fig. 2.15B). An alternative approach is to attach small-molecule fluorophores to viral capsid proteins. Light microscopy has a resolution in the range of 200 to 500 nm, whereas most viruses are between 20 and 400 nm in size and are therefore below the diffraction limit. However, when the virus particle emits a high fluorescent signal in a low background, it is possible to use a computational point tracking algorithm to locate the particle with greater precision than the diffraction limit of the light microscope. This technique allows single particle tracking with accuracy in the range of low tens of nanometers.

Recent improvements in microscopy technology and computational image manipulation have led to unprecedented levels of resolution and contrast and an ability to reconstruct three-dimensional structures from captured images. The first advance was **confocal microscopy**, which utilizes a scanning point of light instead of full-sample illumination. In a conventional light microscope, light can penetrate the specimen only to a fixed depth. In a confocal microscope, a small beam of light is focused to multiple narrow depths. By capturing multiple two-dimensional images

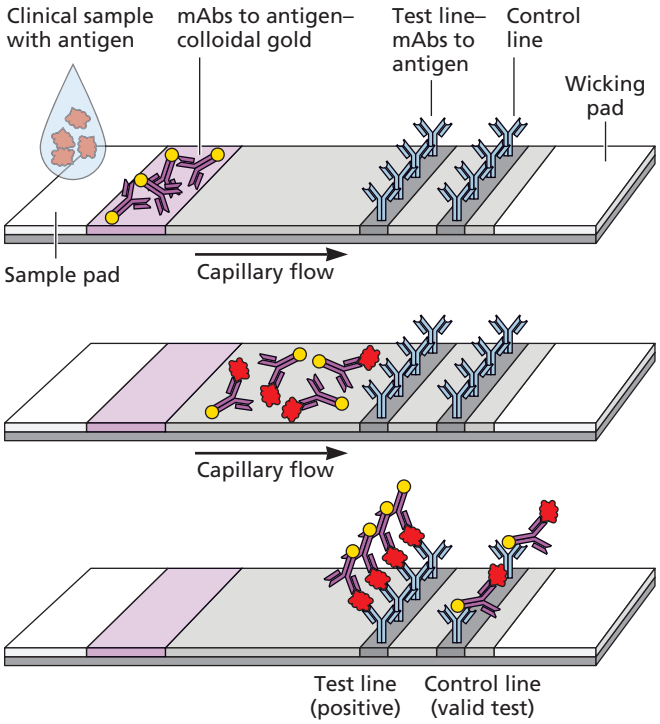


Figure 2.14 Lateral flow immunochromatographic assay. A slide or "dipstick" covered with a membrane is used to assay for the presence of viral antigens. The clinical specimen is placed on an absorbent pad at one end and is drawn across the slide by capillary action. Antigens in the sample react with a virus-specific antibody, which is linked to an indicator, in this example, colloidal gold. The antigen-antibody complexes move across the membrane until they are captured by a second virus-specific antibody in a test line. If viral antigen is present in the sample, an indicator line becomes visible in the test line. Accumulation of the indicator-containing antibody at the control line provides validation that the assay is functioning.

at different depths, it is possible to reconstruct high-resolution three-dimensional structures, a process known as optical sectioning.

Superresolution microscopy combines the advantages of fluorescent imaging (multicolor labeling and live-cell imaging) while breaking the resolution limit of light microscopy. Different formats include **single molecule localization microscopy**, in which only a subset of fluorophores are turned on during each imaging cycle, thus allowing position determination with nanometer accuracy. Fluorophore positions from a series of images are then used to reconstruct the final image. **Structured illumination microscopy** utilizes standing waves formed by interference in laser illumination to create an excitation field that allows optical sectioning at very high resolution. These approaches can achieve resolution below 1 nm, well below the limit of light microscopy. This resolution is achieved by combining sequential acquisition of images with random switching of fluorophores on and off. From several hundred to thousands of images are collected and processed to generate a superresolution data set that can resolve cellular ultrastructure.

These superresolution microscopy methods are well suited for providing high-resolution images of static sections. Because these methods acquire images slowly, are phototoxic, and require computationally intensive image processing, their use for time-lapse imaging of live cells is impractical.

Fluorescence resonance energy transfer (FRET) microscopy can be used to examine protein-protein and protein-DNA or RNA interactions and conformational changes in these molecules. FRET solves the problem encountered in conventional fluorescence microscopy, which is of insufficient resolution to determine if molecules interact. The method is based on the principle that fluorescent emissions of one wavelength can excite a second distinct fluorophore at a distance of approximately 10 nm. For example, if two proteins are thought to interact under certain conditions, one can be labeled with a donor fluorophore that will emit light of a certain wavelength. If the two proteins are farther apart than 10 nm, only the donor color will be observed. However, if the two proteins are in close contact, then fluorescence of the second protein, which is linked to an acceptor fluorophore, will take place.

Another commonly used fluorescent microscopy technique in virology is **fluorescence recovery after photobleaching (FRAP)**, a method for determining the kinetics of diffusion in cells. A viral or cellular protein is labeled with a fluorescent molecule, a portion of the cell is photobleached to eliminate fluorescence, and then recovery of fluorescence is observed over time. Fluorescence in the bleached area recovers as bleached fluorophore-linked proteins are replaced with unbleached molecules from a different part of the cell.

Detection of Viral Nucleic Acids

The detection of viruses in cell cultures is being increasingly supplanted by molecular methods such as the polymerase chain reaction and high-throughput sequencing, especially for discovery of new viruses associated with human diseases. These methods can be used to identify viruses that cannot be propagated in cell culture, offering new ways to fulfill Koch's postulates (Box 1.4).

Polymerase chain reaction. In this technique, specific oligonucleotides are used to amplify viral DNA sequences from infected cells or clinical specimens. Amplification is done in cycles, using a thermostable DNA polymerase (Fig. 2.16). Each cycle consists of thermal denaturation, primer annealing, and extension, carried out by automated cycler machines. The result is exponential amplification (a $2n$-fold increase after n cycles of amplification) of the target sequence that is located between the two DNA primers.

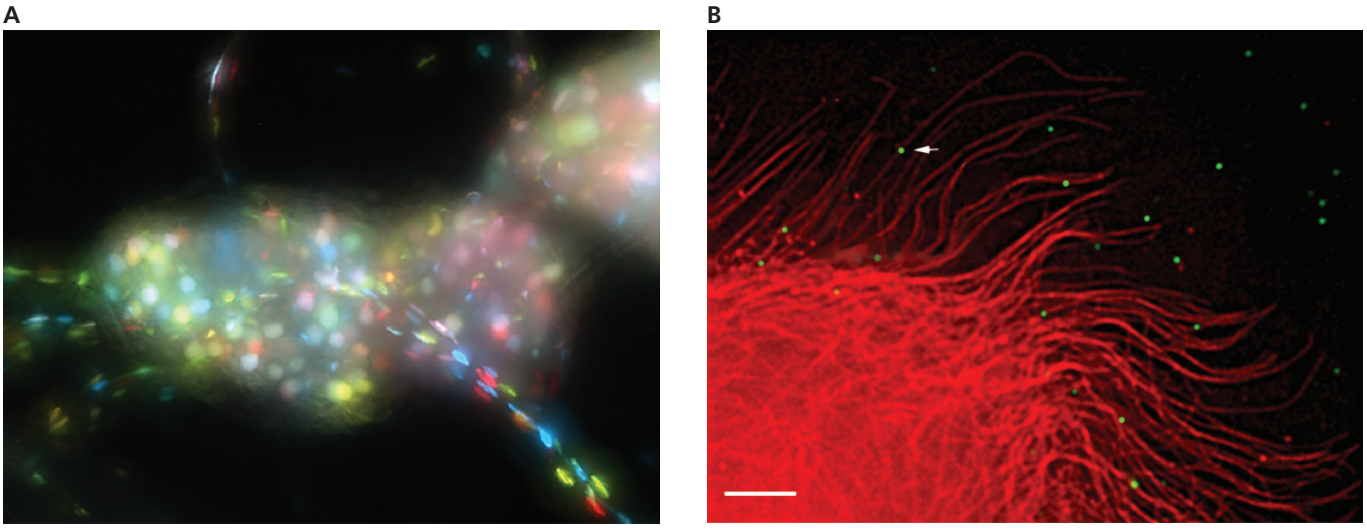


Figure 2.15 Using fluorescent proteins to study virus particles and virus-infected cells. (A) Submandibular ganglia after infection of the salivary gland with three recombinant pseudorabies viruses, each expressing a different color fluorescent protein. Courtesy of Lynn Enquist, Princeton University. **(B)** Single-virus-particle imaging with green fluorescent protein illustrates microtubule-dependent movement of human immunodeficiency virus type 1 particles in cells. The cells were infected with virus particles that contain a fusion of green fluorescent protein with a viral protein. Rhodamine-tubulin was injected into cells to label microtubules (red). Virus particles can be seen as green dots (white arrow). Bar, 5 µm. Courtesy of David McDonald, University of Illinois.

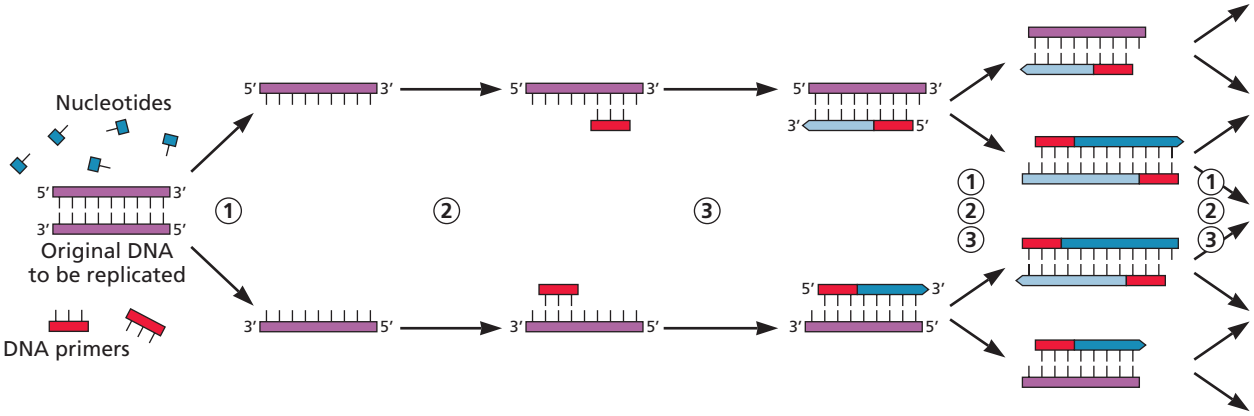


Figure 2.16 Polymerase chain reaction. The DNA to be amplified is mixed with nucleotides, thermostable DNA polymerase, and a large excess of DNA primers. DNA polymerase initiates synthesis at the primers bound to both strands of denatured DNA, which are then copied. The product DNA strands are then separated by heating. Primer annealing, DNA synthesis steps, and DNA duplex denaturation steps are repeated multiple times, leading to geometric amplification of a specific DNA.

Clinical laboratories employ PCR assays to detect evidence for infection by a single type of virus (singleplex PCR), while screening for the presence of hundreds of different viruses can be accomplished with multiplex PCR. In contrast to conventional PCR, real-time PCR can be used to quantitate the amount of DNA or RNA in a sample. In this procedure, also called quantitative PCR, the amplified DNA is detected as the reaction progresses, not after it is completed as in conventional PCR. The product is detected either by incorporation of a ds-DNA specific dye or by release of a fluorescence resonance energy transfer probe via the 5′-to-3′ exonuclease activity of DNA polymerase. The number of cycles needed to detect fluorescence above background can then be compared between standard and experimental samples. Quantitative PCR is widely used in research and clinical applications for genotyping, gene expression analysis, copy number variation assays, and pathogen detection. While PCR is often used to detect viral genomes in clinical specimens or during experimental research, it is important to recognize that the nucleic acid detected does not necessarily correspond to infectious virus (Box 2.8).

BOX 2.8

EXPERIMENTS

Viral RNA is not infectious virus

A study of sexual transmission of Zika virus among mice demonstrates beautifully that viral nucleic acid detected by polymerase chain reaction (PCR) is not the same as infectious virus particles.

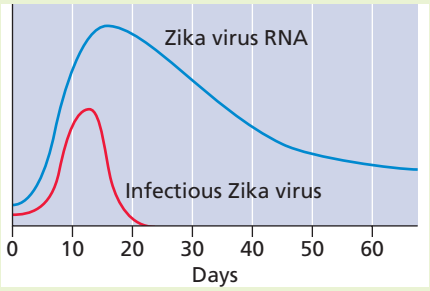
Male mice were infected with Zika virus and then mated with female mice. Efficient sexual transmission of the virus from males to females was observed. To understand the dynamics of sexual transmission, the authors measured Zika virus shedding in seminal fluid, by PCR to detect viral RNA and by plaque assay to detect infectious virus particles. The results (see figure) show that Zika virus RNA persisted in semen for up to 60 days, far longer than did infectious virus, which could not be detected after about three weeks.

There is a lower limit of detection of virus via the plaque assay of approximately 10 plaque forming units/ml. Whether this low concentration of infectious particles would be sufficient to transmit the virus is not known. However, it seems unlikely that these mice are able to transmit virus after a few weeks, despite the presence of Zika virus RNA in seminal fluid for at least 60 days after infection.

Recently many papers have been published demonstrating that Zika virus and Ebolavirus can persist in a variety of human fluids for extended periods of time. These results have been interpreted with alarm by both by scientists and science writers. However, in most cases detection was by PCR, not by plaque assay, and therefore, we do not know if infectious virus particles were present. Viral RNA would not constitute a threat to transmission, while infectious virus would.

Many laboratories choose to assay the presence of viral genomes by PCR. This is an acceptable technique as long as the limitations are understood—it detects nucleic acids, not infectious virus.

The lesson from this study is very clear: in novel experimental or epidemiological studies it is important to prove that any viral nucleic acid detected by PCR represents infectious virus. Failing to do so clouds the conclusions of the study.



Detection of Zika virus RNA and infectious virus in seminal fluid. Male mice were infected with Zika virus. At different times after infection, viral RNA and infectious virus particles were measured in seminal fluid by PCR (blue line) and by plaque assay (red line).

Duggal NK, Ritter JM, Pestorius SE, Zaki SR, Davis BS, Chang GJ, Bowen RA, Brault AC. 2017. Frequent Zika virus sexual transmission and prolonged viral RNA shedding in an immunodeficient mouse model. *Cell Rep* **18:**1751–1760.

High-throughput sequencing. The development of DNA sequencing methods in the 1970s revolutionized biology by allowing the decoding of viral genes and entire viral genomes. While powerful, these methods were laborious: in 1980 it took one year for a single person to determine the nucleotide sequence of the 7,440-nucleotide genome of poliovirus. Today the same result could be achieved in less than one hour.

The difference is a consequence of the development of second- and third-generation sequencing methods, spurred by the desire to sequence larger and larger virus and cell genomes. These methods were originally called next-generation sequencing, because they followed the very first sequencing methods. The first of these new methods to be developed, 454 sequencing, was released in 2005 and could produce 200,000 reads of 110 base pairs. Other technologies that generated larger numbers of sequence reads soon followed (Solexa/Illumina, SOLiD, and Ion Torrent) which generated larger numbers of reads, but the number of bases in each read was much shorter. These technologies relied on amplification of the target DNA and optical detection of incorporated fluorescent nucleotides. Third-generation sequencing methods can not only detect single molecules (e.g., amplification is not required) but also carry out sequencing in real time. PacBio instruments can achieve maximum read lengths of 20 kb, and those from Illumina can generate 1.8 terabytes of sequence per run. The latter reduces the cost of sequencing a human genome to below $1,000, a 10,000-fold reduction in price since 2004, when the first human genome was deciphered.

These technologies have not only made sequencing of DNA cheaper and faster but also helped create innovative experimental approaches to study genome organization, function, and evolution. Their use has led to the discovery of new viruses and has given birth to the field of **metagenomics**, the analysis of sequences directly from clinical or environmental samples. These sequencing technologies can be used to study the **virome**, the genomes of all viruses in a specific environment, such as sewage, the human body, or the intestinal tract. While these virus detection technologies are extremely powerful, the results obtained must be interpreted with caution. It is very easy to detect traces of a viral contaminant when searching for new agents of human disease (Box 2.9).

It should be noted that metagenomics is not limited to DNA viruses. Nucleic acids extracted from clinical or environmental samples may be treated with DNase, and the remaining RNAs converted to DNA with reverse transcriptase for sequencing and identification.

BOX 2.9

EXPERIMENTS

Pathogen de-discovery

High-throughput sequencing of nucleic acids has accelerated the pace of virus discovery, but at a cost: contaminants are much easier to detect.

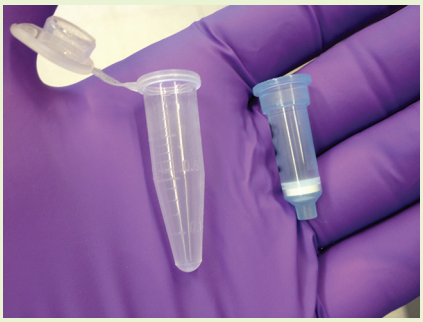
During a search for the causative agent of seronegative hepatitis (disease not caused by hepatitis A, B, C, D, or E virus) in Chinese patients, a new virus with a single-stranded DNA genome was discovered in sera by high-throughput sequencing. Seventy percent of 90 patient serum samples were positive for viral DNA by PCR, and sera from 45 healthy controls were negative. Furthermore, 84% of patients were positive for antibodies against the virus. Among healthy controls, 78% were antibody positive. The authors concluded that this virus was highly prevalent in some patients with seronegative hepatitis. A second independent laboratory identified the same virus in sera from patients in the United States with non-A-to-E hepatitis, while a third group identified the virus in diarrheal stool samples from Nigeria.

The first clue that something was amiss was the observation that the new virus identified in all three laboratories shared 99% nucleotide and amino acid identity: this similarity would not be expected in virus samples from such geographically, temporally, and clinically diverse samples. Another problem was that in the U.S. non-A-to-E hepatitis study, all pools of patient sera were positive for viral sequences. These observations suggested the possibility of viral contamination.

When nucleic acids were repurified from the U.S. non-A-to-E hepatitis samples using a different method, **none** were positive for the new virus. The presence of the virus was traced to the use of column-based purification kits manufactured by Qiagen, Inc. (pictured). Nearly the entire viral genome could be detected by deep sequencing of sterile water that was passed through these columns. The nucleic acid purification columns contaminated with the new virus were used to purify nucleic acid from patient samples. These columns, produced by a number of manufacturers, are typically an inch in length and contain a silica gel membrane that binds nucleic acids. The clinical samples are added to the column, which is then centrifuged briefly to remove liquids (hence the name "spin" columns). The nucleic acid adheres to the silica gel membrane. Contaminants are washed away, and the nucleic acids are then released from the silica by the addition of a buffer.

Why were the Qiagen spin columns contaminated with viral DNA? A search of the publicly available environmental metagenomic data sets revealed the presence of sequences highly related to this virus (87 to 99% nucleotide identity). The data sets containing these sequences were obtained from seawater collected off the Pacific coast of North America and coastal regions of Oregon and Chile. The source of contamination could be explained if the silica in the Qiagen spin columns was produced from ocean-dwelling diatoms that were infected with the virus.

In retrospect, it was easy to be fooled into believing that the novel virus might be a human pathogen because it was detected only in sick and not healthy patients. Why antibodies to the virus were detected in samples from both



sick and healthy patients remains to be explained. However, the virus is not likely to be associated with any human illness: when non-Qiagen spin columns were used, the viral sequences were not found in any patient sample.

The lesson to be learned from this story is clear: high-throughput sequencing is a very powerful and sensitive method but must be applied with great care. Every step of the virus discovery process must be carefully controlled, from the water used to the plastic reagents. Most importantly, laboratories carrying out pathogen discovery must share their sequence data, something that took place during this study.

Naccache SN, Greninger AL, Lee D, Coffey LL, Phan T, Rein-Weston A, Aronsohn A, Hackett J, Jr, Delwart EL, Chiu CY. 2013. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J Virol* **87:**11966–11977.

Xu B, Zhi N, Hu G, Wan Z, Zheng X, Liu X, Wong S, Kajigaya S, Zhao K, Mao Q, Young NS. 2013. Hybrid DNA virus in Chinese patients with seronegative hepatitis discovered by deep sequencing. *Proc Natl Acad Sci U S A* **110:**10264–10269.

Computational biology. The generation of nucleotide sequences at an unprecedented rate has spawned a new branch of bioinformatics to develop algorithms for assembling sequence reads into continuous strings and to determine whether they are from a new or previously discovered virus. Storing, analyzing, and sharing massive quantities of data constitute an immense challenge: the number of bases in GenBank, an open-access, annotated collection of all publicly available nucleotide sequences produced and maintained by the National Center for Biotechnology Information, has doubled every 18 months since 1982. As of June 2019 GenBank held 329,835,282,370 bases.

Computational problems must be solved at multiple steps during the process of genome sequencing. The initial problem is that sequence reads are typically short, and there are many of them (e.g., high throughput). These short sequences must be overlapped and, if possible, mapped to a genome. Many computer programs have been developed to address this problem. Some carry out alignment of sequence reads to a reference genome, while others perform this process *de novo*, i.e., in the absence of a reference genome.

When clinical or environmental samples are subjected to high-throughput sequencing for pathogen discovery, it is essential to identify viral sequences in what is typically a mix of host,

bacterial, and fungal sequences. This task relies on alignment of sequences to reference viral databases. However, such databases are limited because most of the sequences retrieved in metagenomic studies are unknown (so-called "dark matter") and therefore cannot be annotated. Consequently, computational pipelines have been designed to analyze high-throughput sequencing data to search for those likely to be of viral origin.

Some computational pipelines are designed to define the abundance and types of viruses in a sample, such as Viral Informatics Resource for Metagenome Exploration (VIROME), the Viral MetaGenome Annotation Project (VMGAP), and Basic Local Alignment Search Tool (BLAST). Other virus discovery programs (MePIC, READSCAN, CaPSID, VirusFinder, and SRSA) rely on nucleotide sequence alignment and will work only for the detection of viruses with high sequence similarity to known viruses. PathSeq, SURPI, VirFind, and VirusHunter identify viruses by amino acid searches, a computationally demanding exercise that is critical for new virus identification. VirusSeeker-Virome (VS-Virome) is a computational pipeline designed for defining both the type and abundance of known and novel viral sequences in metagenomic data sets (Fig. 2.17).

Genome sequences can provide considerable insight into the evolutionary relationships among viruses. Such information can be used to understand the origin of viruses and how selection pressures change viral genomes and to assist in epidemiological investigations of viral outbreaks. When few viral genome sequences were available, pairwise homologies were often displayed in simple tables. As sequence databases increased in size, tables of multiple alignments were created, but these were still based only on pairwise comparisons. Today, phylogenetic trees are used to illustrate the relationships among numerous viruses or viral proteins (Box 2.10). Not only are such trees important tools for understanding evolutionary relationships, but they may allow conclusions to be drawn about biological functions: examination of a phylogenetic tree may allow determination of how closely or distantly a sequence relates to one of known function. Software programs such as AdaPatch, AntiPatch, and AntigenicTree have been developed to produce phylogenetic trees. However, these approaches do not account for horizontal gene transfer, recombination, or the evolutionary relationships between viruses and their hosts, which will require unconventional computational methods to resolve.

Algorithms have also been written to apply high-throughput sequencing methods to a variety of genome-wide analyses, including detection of single-nucleotide polymorphisms (SNP), RNA-seq, ChiP-seq, CLIP, and more (see below).

Viral Reproduction: the Burst Concept

A fundamental and important principle is that viruses are reproduced via the assembly of preformed components into particles: the parts are first made in cells and then assembled into the final product. This simple build-and-assemble strategy is unique to all viruses, but the details of how this process transpires are astonishingly diverse among members of different virus families. There are many ways to build a virus particle, and each one tells us something new about virus structure and assembly.

Modern investigations of viral reproduction strategies have their origins in the work of Max Delbrück and colleagues, who studied the T-even bacteriophages starting in 1937. Delbrück believed that these bacteriophages were perfect models for understanding the basis of heredity. He focused his attention on the fact that one bacterial cell usually makes hundreds of progeny virus particles. The yield from one cell is one viral generation; it was called the **burst** because the viruses that he studied literally burst from the infected cell. Under carefully controlled laboratory conditions, most cells make, on average, about the same number of bacteriophages per cell. For example, in one of Delbrück's experiments, the average number of bacteriophage T4 particles produced from individual single-cell bursts from *Escherichia coli* cells was 150 particles per cell.

Another important implication of the burst is that a cell has a finite capacity to produce virus. Multiple parameters limit the number of particles produced per cell. These include metabolic resources, the number of sites for genome replication in the cell, the regulation of release of virus particles, and host defenses. In general, larger cells (e.g., eukaryotic cells) produce more virus particles per cell: yields of 1,000 to 10,000 virions per eukaryotic cell are not uncommon.

A burst occurs for viruses that kill the cell after infection, namely, cytopathic viruses. However, some viruses do **not** kill their host cells, and virus particles are produced as long as the cell is alive. Examples include filamentous bacteriophages, most retroviruses, and hepatitis viruses.

The One-Step Growth Cycle

The idea that one-step growth analysis can be used to study the single-cell reproductive cycle of viruses originated from the work on bacteriophages by Emory Ellis and Delbrück. In their classic experiment, they added virus particles to a culture of rapidly growing *E. coli*. These particles adsorbed quickly to the cells. The infected culture was then diluted, preventing further adsorption of unbound particles. This simple dilution step is the key to the experiment: it reduces further binding of virus to cells and effectively synchronizes the infection. Samples of the diluted culture were then taken every few minutes and analyzed for the number of infectious bacteriophages.

When the results of this experiment were plotted, several key observations emerged. The graphs were surprising in that they did not resemble the growth curves of bacteria or cultured cells. After a short lag, bacterial cell growth becomes exponential (i.e., each progeny cell is capable of dividing into two cells) and follows a straight line (Fig. 2.18A). Exponential growth continues until the nutrients in the medium are exhausted. In contrast, numbers of new viruses do not increase

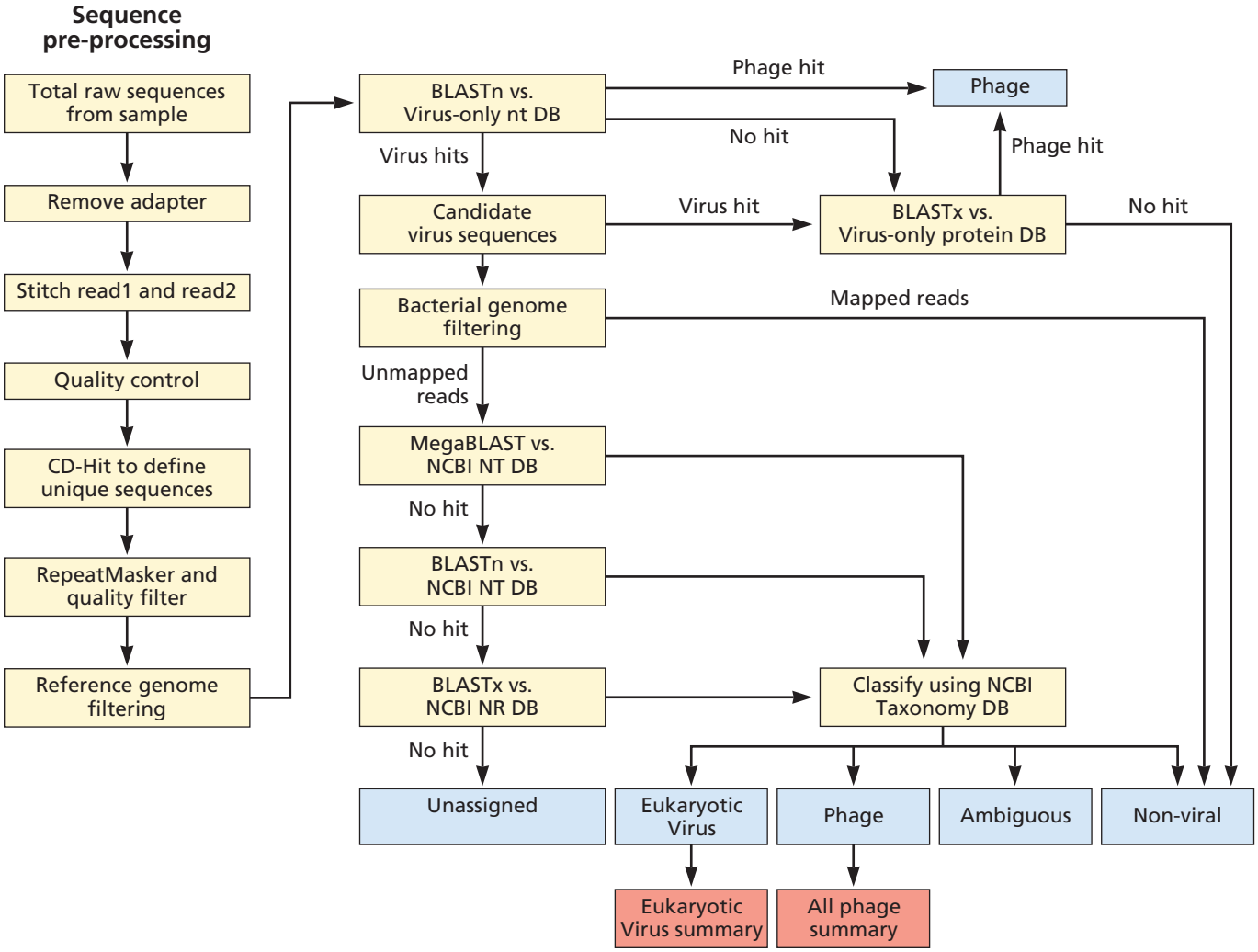


Figure 2.17 Workflow for VS-Virome. Shown is the computational pipeline designed for defining the type and abundance of known and novel viral sequences in metagenomic data sets. VS-Virome first pre-processes the sequences (left) to remove adapter sequences (these are added to every DNA in the sample, and contain barcoding sequences, primer binding sites, and sequences for immobilizing the DNA), joins paired end reads if they overlap, performs quality control on sequences, and identifies low-complexity sequences and host sequences before subjecting all the sequences to BLAST (right) to detect viral sequences. Because integrated prophage are found in bacterial genomes, alignment to comprehensive databases could lead to removal of bona fide bacteriophage sequences. Bacteriophage hits are therefore placed into a separate output file. Candidate eukaryotic viral sequences are filtered to remove sequences that have high identity to bacterial genomes. Remaining reads are then aligned to the more comprehensive GenBank NT and NR databases to identify reads or contigs that have greater similarity to nonviral sequences than to viral sequences (i.e., increased likelihood of being a false positive). To have a high degree of confidence in viral classification, sequences that have significant hits to both viral and any nonviral reference sequence are placed in an "ambiguous" bin. Sequences in the viral bin only have significant alignment to viral sequences.

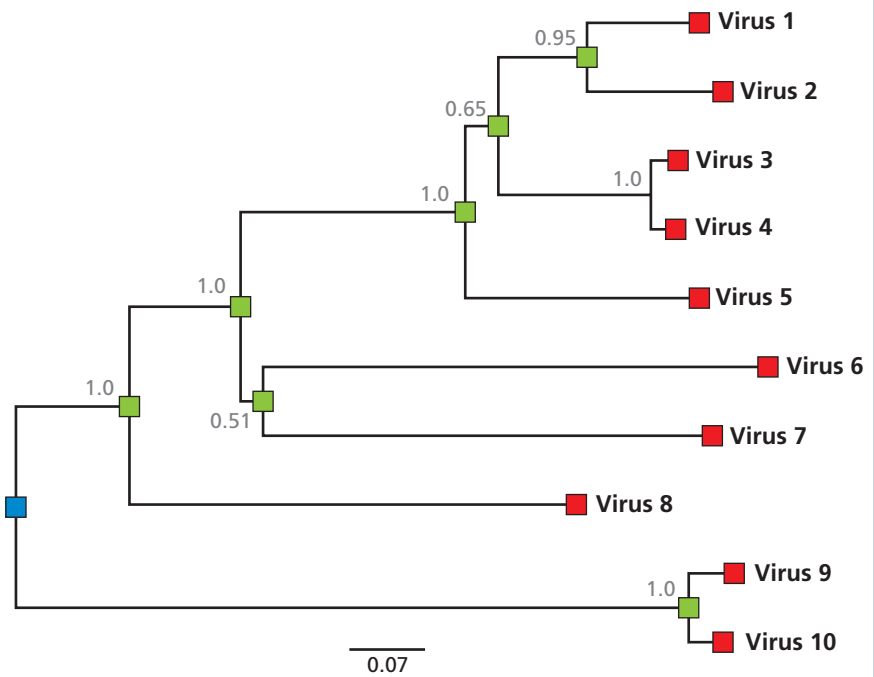
in a linear fashion from the start of the infection (Fig. 2.18B, left). There is an initial lag period in which no infectious viruses can be detected. This lag period is followed by a rapid increase in the number of infectious particles, which then plateaus. The single cycle of virus reproduction produces this "burst" of virus progeny. If the experiment is repeated, such that only a few cells are initially infected, the graph looks different (Fig. 2.18B, right). Instead of a single cycle, there is a stepwise increase in numbers of new viruses with time. Each step represents one cycle of virus infection.

Once the nature of the viral propagation cycle was explored using the one-step growth curve, questions emerged about what was happening in the cell before the burst. What was the fate of the incoming virus? Did it disappear? How

BOX 2.10

METHODS

How to read a phylogenetic tree



Rooted phylogenetic tree of 10 viral genome sequences.

Phylogenetic dendrograms, or trees, provide information about the inferred evolutionary relationships between viruses. The example shown in the figure is a phylogenetic tree for sequenced viral isolates from 10 different individuals. The horizontal dimension of the tree represents the degree of genetic change, and the scale (0.07) is the number of changes divided by the length of the sequence (in some trees this may be expressed as % change). The blue circles, called nodes, represent putative ancestors of the sampled viruses. Therefore, the branches represent chains of infections that have led to sampled viruses. The vertical distances have no significance.

The tree in the figure is *rooted*, which means that the root of the tree represents the common ancestor of all the sampled viruses. As we move from the root to the tips, we are moving forward in time, although the unit of time might not be known. The numbers next to each node represent the measure of support; these are computed by a variety of statistical approaches including "bootstrapping" and "Bayesian posterior probabilities." A value close to 1 indicates strong evidence that sequences to the right of the node cluster together better than any other sequences. Often there is no known isolate corresponding to the root of the tree; in this case, an arbitrary root may be estimated, or the tree will be unrooted. In these cases, it can no longer be assumed that the order of ancestors proceeds from left to right.

Phylogenetic trees can also be constructed by grouping sampled viruses by host of isolation. Such an arrangement sometimes makes it possible to identify the animal source of a human virus. Circular forms, such as a radial format tree, are often displayed when the root is unknown.

Trees relating nucleic acid sequences depict the relationships as if sampled and intermediary sequences were on a trajectory to the present. This deduction is an oversimplification, because any intermediate that was lost during evolution will not be represented in the tree. In addition, any recombination or gene exchange by coinfection with similar viral genomes will scramble ordered lineages.

A fair question is whether we can predict the future trajectory or branches of the tree. We can never answer this question for two reasons: any given sample may not represent the diversity of any given virus population in an ecosystem, and we cannot predict the selective pressures that will be imposed.

Hall BG. 2011. *Phylogenetic Trees Made Easy: A How-to Manual*, 4th ed. Sinauer Associates, Sunderland, MA.

ViralZone. Phylogenetics of animal pathogens: basic principles and applications (a tutorial). http://viralzone.expasy.org/e_learning/phylogenetics/content.html

were more virus particles produced? These questions were answered by looking inside the infected cell. Instead of sampling the diluted culture for virus after various periods of infection, researchers prematurely lysed the infected cells as the infection proceeded and then assayed for infectious virus. The results were extremely informative. Immediately after dilution, there was a complete loss, or eclipse, of infectious virus for 10 to 15 min (Fig. 2.18B). In other words, input virions disappeared, and no new phage particles were produced during this period. It was shown later that the loss of infectivity is a consequence of the

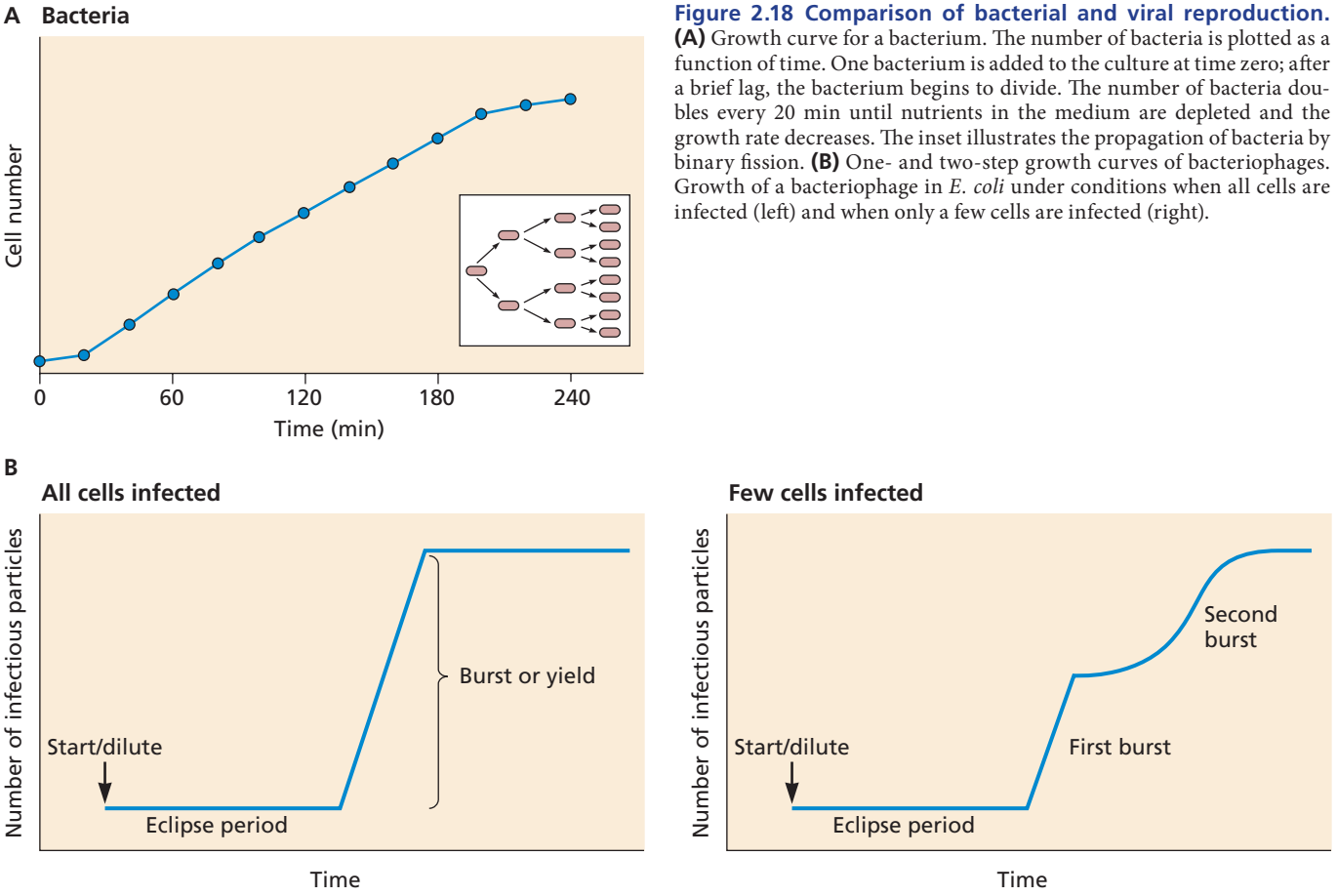


Figure 2.18 Comparison of bacterial and viral reproduction. **(A)** Growth curve for a bacterium. The number of bacteria is plotted as a function of time. One bacterium is added to the culture at time zero; after a brief lag, the bacterium begins to divide. The number of bacteria doubles every 20 min until nutrients in the medium are depleted and the growth rate decreases. The inset illustrates the propagation of bacteria by binary fission. **(B)** One- and two-step growth curves of bacteriophages. Growth of a bacteriophage in *E. coli* under conditions when all cells are infected (left) and when only a few cells are infected (right).

release of the genome from the virion, to allow for subsequent transcription of viral genes. Particle infectivity is lost during this phase because the released genome is not infectious under the conditions of the plaque assay. Later, newly assembled infectious particles could be detected inside the cell that had not yet been released by cell lysis.

The results of these experiments defined two new terms in virology: the **eclipse period**, the phase in which infectivity is lost when virions are disassembled after penetrating cells, and the **latent period**, the time it takes to replicate and assemble new virus particles before lysis, ~20 to 25 min for some *E. coli* bacteriophages.

Synchronous infection, the key to the one-step growth cycle, is usually accomplished by infecting cells with a sufficient number of virus particles to ensure that most of the cells are infected rapidly. Exactly how many virus particles must be added is described by the **multiplicity of infection** (Box 2.11).

One-Step Growth Analysis: a Valuable Tool for Studying Animal Viruses

One-step growth analysis soon became adapted for studying the reproduction of animal viruses. The experiment begins with removal of the medium from the cell monolayer and addition of virus in a small volume to promote rapid adsorption. After ~1 h, unadsorbed inoculum containing virus particles is removed, the cells are washed, and fresh medium is added. At different times after infection, samples of the cell culture supernatant are collected and the virus titer is determined. The kinetics of intracellular virus production can be monitored by removing the medium containing extracellular particles, scraping the cells into fresh medium, and lysing them. A cell extract is prepared after removal of cellular debris by centrifugation, and the virus titer in the extract is measured.

The results of a one-step growth experiment establish a number of important features about viral reproduction. In the example shown in Fig. 2.19A, the first 11 h after infection constitute the eclipse period, during which the viral nucleic acid is uncoated from its protective shell and no infectious virus can be detected inside cells. The small number of infectious particles detected during this period probably represents adsorbed virus that was not uncoated. Beginning at 12 h after adsorption, the quantity of intracellular infectious virus particles begins to increase, marking the onset of the synthetic phase, during which assembly begins. During the latent period, no extracellular virus

BOX 2.11

DISCUSSION

Multiplicity of infection (MOI)

Infection depends on the random collision of cells and virus particles. When susceptible cells are mixed with a suspension of virus particles, some cells are uninfected and other cells receive one, two, three, etc., particles. The distribution of virus particles per cell is best described by the Poisson distribution:

$$P(k) = e^{-m}m^k/k!$$

In this equation, $P(k)$ is the fraction of cells infected by k virus particles. The multiplicity of infection, m, is calculated from the proportion of uninfected cells, $P(0)$, which can be determined experimentally. If k is made 0 in the above equation, then

$$P(0) = e^{-m} \text{ and } m = -1n\ P(0)$$

The fraction of cells receiving 0, 1, and >1 virus particle in a culture of 10^6 cells infected with an MOI of 10 can be determined as follows.

The fraction of cells that receive 0 particles is

$$P(0) = e^{-10} = 4.5 \times 10^{-5}$$

and in a culture of 10^6 cells, this equals 45 uninfected cells.

The fraction of cells that receive 1 particle is

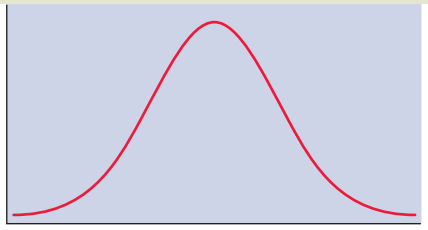
$$P(1) = 10 \times 4.5 \times 10^{-5} = 4.5 \times 10^{-4}$$

and in a culture of 10^6 cells, 450 cells receive 1 particle.

The fraction of cells that receive >1 particle is

$$P(>1) = 1 - e^{-m}(m + 1) = 0.9995$$

and in a culture of 10^6 cells, 999,500 cells receive >1 particle. [The value in this equation is obtained by subtracting from 1 (the sum of all probabilities for any value of k) the probabilities $P(0)$ and $P(1)$.]



The fraction of cells receiving 0, 1, and >1 virus particle in a culture of 10^6 cells infected with an MOI of 0.001 is

$P(0) = 99.99\%$
$P(1) = 0.0999\%$ (for 10^6 cells, 10^4 are infected)
$P(>1) = 10^{-6}$

The MOI required to infect 99% of the cells in a cell culture dish is
$P(0) = 1\% = 0.01$
$m = -1n\ (0.01) = 4.6$ PFU per cell.

can be detected. At 18 h after adsorption, virions are released from cells into the extracellular medium. Ultimately, virus numbers plateau as the cells become metabolically and structurally incapable of supporting additional reproduction.

The yield of infectious virus per cell can be calculated from the data collected during a one-step growth experiment. This value varies widely among different viruses and with different virus-host cell combinations. For many viruses, increasing the multiplicity of infection above a certain point does not increase the yield: cells have a finite capacity to produce new virus particles. In fact, infecting at a very high multiplicity of infection can cause premature cell lysis and decrease virus yields.

The kinetics of the one-step infectious cycle can vary dramatically among viruses. For example, enveloped viruses that mature by budding from the plasma membrane, as discussed in Chapter 13, generally become infectious only as they leave the cell, and therefore, little intracellular infectious virus can be detected (Fig. 2.19B). The curve shown in Fig. 2.19A illustrates the pattern observed for a DNA virus with the long latent and synthetic phases typical of many DNA viruses, some retroviruses, and reovirus. For small RNA viruses, the entire growth curve is complete within 6 to 8 h, and the latent and synthetic phases are correspondingly shorter.

One-step growth curve analysis can provide quantitative information about different virus-host systems. It is frequently employed to study mutant viruses to determine what parts of the infectious cycle are affected by a particular genetic lesion. It is also valuable for studying the multiplication of a new virus or viral reproduction in a new virus-host cell combination.

When cells are infected at a low multiplicity of infection, several cycles of viral reproduction may occur (Fig. 2.18B). Growth curves established under these conditions can also provide useful information. When infection is carried out at a high multiplicity, a mutation may fail to have an obvious effect on viral reproduction. The defect may only become evident following a low-multiplicity infection. Because the effect of a mutation in each cycle is multiplied, a small effect can be amplified after several cycles. Defects in the ability of viruses to spread from cell to cell may also be revealed when multiple cycles of reproduction occur.

Global Analysis

The study of replication cycles of many viruses with one-step growth analysis has allowed a reductionist approach to understanding and defining the steps of virus attachment, entry, replication, and assembly. In contrast, new experimental and computational tools permit global analysis of viral, cellular, and host responses to infection. Global analyses apply a dizzying array of different high-throughput technologies to measure system-wide changes in DNA, RNA, proteins, and metabolites during virus infection of cells, tissues, or entire organisms. Data obtained from high-throughput measurements are integrated and analyzed using mathematical algorithms

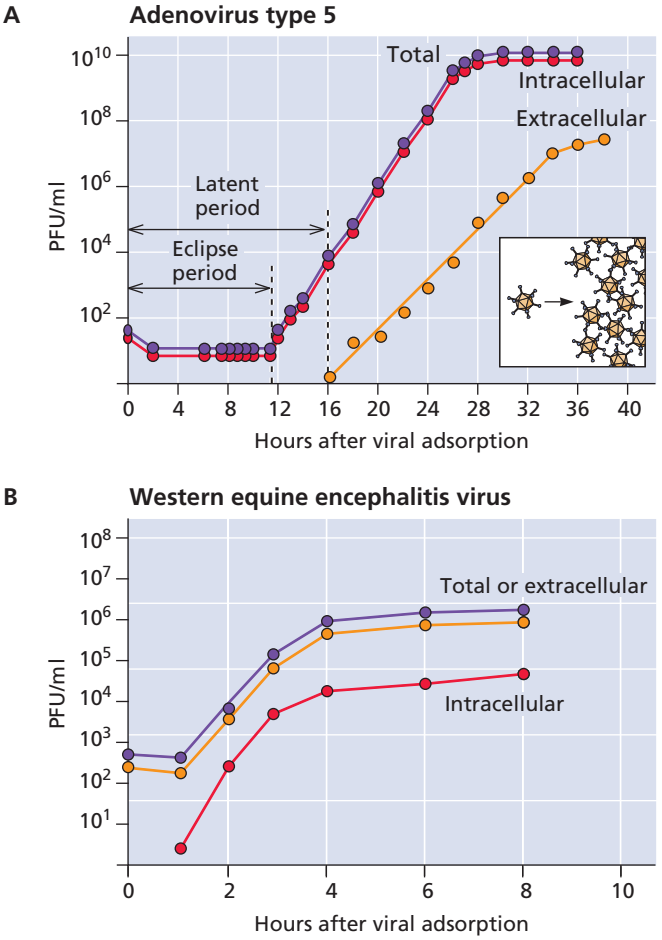


Figure 2.19 One-step growth curves of animal viruses. (A) Growth of a nonenveloped virus, adenovirus type 5. The inset illustrates the concept that viruses multiply by assembly of preformed components into particles. **(B)** Growth of an enveloped virus, Western equine encephalitis virus, a member of the *Togaviridae*. This virus acquires infectivity after maturation at the plasma membrane, and therefore, little intracellular virus can be detected. The small quantities observed at each time point probably represent released virus contaminating the cell extract.

to generate models that are predictive of the system. For example, virus infections of different animals are characterized by the induction of distinct sets of cytokine genes, a property that can be correlated with different pathogenic outcomes. When a model has been developed, it can be further refined by the use of viral mutants or targeted inhibition of host genes or pathways. Global analysis is therefore a holistic, host-directed approach that complements traditional methods for studying viruses.

Examples of global analyses include genome-wide transcriptional profiling to study the host response to infection. Introduction of the 1918 strain of influenza virus into mice leads to a rapidly fatal disease characterized by sustained induction of proinflammatory cytokine and chemokine genes. Understanding the gene expression signature that correlates with lethality is one goal of these studies. Global analysis can also predict signatures of vaccine efficacy. In one study, transcriptional profiling of peripheral blood mononuclear cells from vaccinated subjects revealed that the yellow fever virus vaccine induces the expression of genes encoding members of the complement system and stress response proteins. This pattern accurately predicts $CD8^+$ T cell and antibody responses that are thought to mediate protection from infection with yellow fever virus. A separate signature that accurately predicts neutralizing antibody synthesis during infection was also identified.

Some of the methods used in global analysis are described below.

DNA Microarrays

An early staple of global analyses, this method enables the study of the gene expression profile of a cell in response to virus infection (Chapter 14) and can also be used to discover new viruses. In this method, millions of unique viral DNA sequences fixed to glass or silicon wafers are incubated with sequences complementary to DNAs or RNAs, which have been amplified from clinical and environmental samples by PCR. Binding is usually detected by using fluorescent molecules incorporated into amplified nucleic acids. Microarrays have been largely supplanted by high-throughput sequencing, which allows identification of transcripts and their quantification in an unbiased manner, e.g., without prior assumption of what genes are involved.

In **RNAseq**, RNAs extracted from cells or tissues are converted by reverse transcription to complementary DNAs, which are then subjected to high-throughput DNA sequencing. The results provide insight into sequences and quantity of RNAs in a cell at a given time under specific conditions. It allows detection and quantification of transcripts that are not represented on microarrays. Information on transcriptional activity is provided by **native elongating transcript sequencing** (NET-seq), in which immunoprecipitation of RNA polymerase is followed by high-throughput sequencing of the 3′ ends of the associated RNAs. A method to study the association of RNAs with ribosomes is **ribo-seq**, in which polysomes are treated with RNases and the 20- to 30-nucleotide ribosome-protected fragments are sequenced. The information provides insight into translational control of gene expression and the mechanism of protein synthesis and allows annotation of translated sequences.

A number of methods yield global views of protein-nucleic acid interactions at unprecedented levels of resolution. **Chromatin-immunoprecipitation sequencing** (**ChiP-seq**) can localize protein-DNA interactions with single-nucleotide precision (Fig. 2.20). In this method, protein-DNA complexes are immunoprecipitated with antibodies to DNA binding proteins, such as transcription proteins, histones, or even specific methyl groups on histones. The DNAs are then sub-

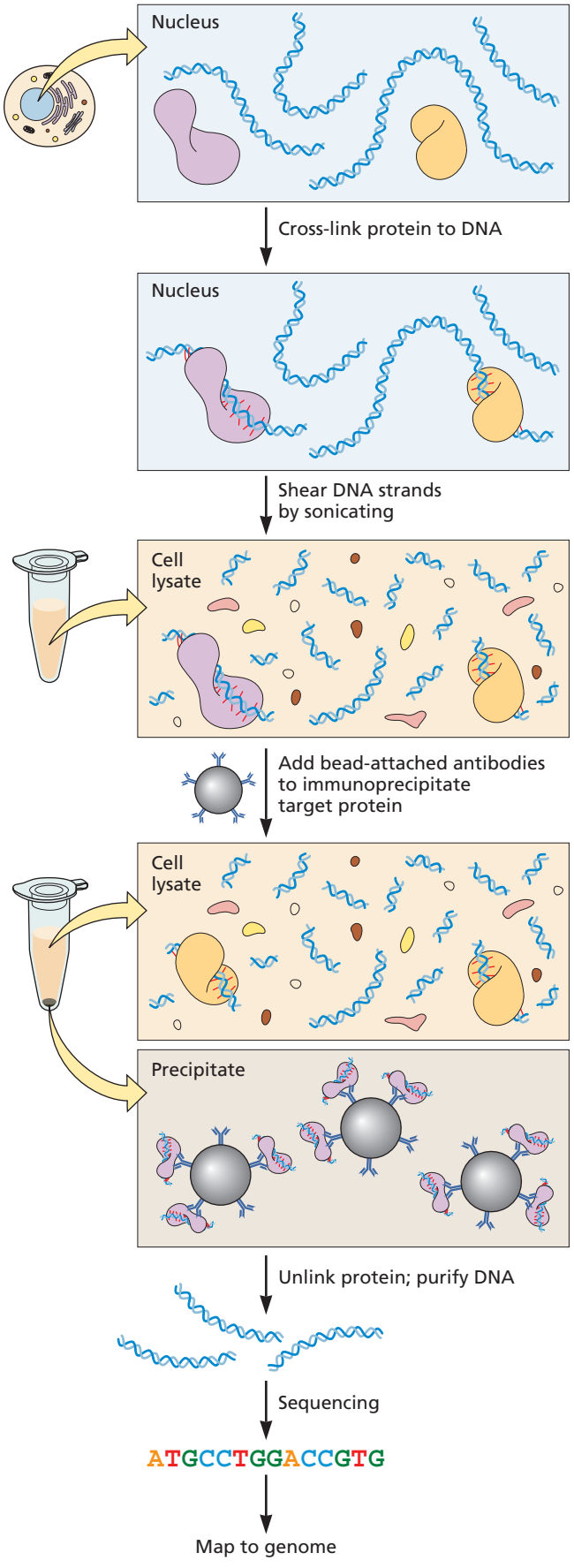


Figure 2.20 Chromatin immunoprecipitation and DNA sequencing, ChiP-seq. This technique is used to identify the precise binding sites of proteins on DNA. DNA is cross-linked to proteins by treating cells with formaldehyde, followed by sonication to shear DNA to 200 to 1,000 bp. Beads coated with antibody to the DNA binding protein of interest are added and precipitated. The protein is removed and DNA purified and subjected to high-throughput sequencing to identify protein binding sites on the DNA.

jected to high-throughput sequencing to identify the sites on DNA to which these proteins bind. An early variant called **ChiP on chip** employed microarrays to identify protein binding sites on DNA.

Many protocols have been devised for genome-wide analysis of RNA-protein interactions that are based on **cross-linking immunoprecipitation** (CLIP). In CLIP-seq, RNA-protein complexes are cross-linked in cells in culture with UV light. Cells are lysed and proteins of interest are immunoprecipitated. Proteins are removed by digestion with protease, DNA is synthesized from the previously bound RNA with reverse transcriptase, and the product is subjected to high-throughput sequence analysis. Interaction sites are identified by mapping the nucleic acid sequence reads to the transcriptome. A modification of this technique is called **photoactivatable ribonucleoside-enhanced cross-linking and immunoprecipitation**, PAR-CLIP. In this method, photoreactive ribonucleoside analogs such as 4-thiouridine are incorporated into RNA transcripts in living cells. Irradiation with UV light induces efficient cross-linking of RNAs containing these analogs to interacting proteins. Immunoprecipitation and sequencing are then carried out as in other CLIP methods.

Other genome-wide mapping analyses that can be performed include identifying the binding sites for long noncoding RNAs (lncRNA) on chromatin using **capture hybridization analysis of RNA targets (CHART)**. In this method, biotin-linked oligonucleotides that are complementary to the target RNA are designed. These are added to reversibly cross-linked chromatin extracts, and the target RNA is purified with streptavidin beads, which bind with high affinity to biotin. The sequences of the RNA targets identify the genomic binding sites of endogenous RNAs. A related method is **chromatin isolation by RNA purification (ChIRP)**, in which tiled oligonucleotides labeled with biotin are used to retrieve specific lncRNA bound to protein and DNAs.

How DNA is organized in virus particles and in the cell nucleus is being studied using **chromosome conformation capture** technology, abbreviated as 3C, 4C, 5C, and Hi-C, which differ in scope. For example, 3C identifies interactions between a single pair of genomic loci. Chromosome conformation capture on chip (4C) studies the interaction of one genomic locus and all other genomic loci, while chromosome conformation capture carbon copy (5C) detects interactions between all restriction fragments in a given region. In HiC, high-throughput sequencing is used to identify the restriction fragments studied. These methods begin with cross-linking of

cell genomes with formaldehyde and digestion with restriction endonucleases, followed by random ligation under conditions where joining of cross-linked fragments is favored over those that are not. PCR is then used to amplify ligated junctions and identify interacting loci. The open or closed state of chromatin can be measured by **DNaseI-seq** (DNaseI hypersensitive sites sequencing) and **FAIRE-seq** (formaldehyde-assisted isolation of regulatory elements). These protocols are based on the use of formaldehyde to cross-link DNA: this reaction is more efficient in nucleosome-rich regions than in nucleosome-poor areas. The non-cross-linked DNA, typically from open chromatin, is then purified and its sequence is determined. The two protocols differ in that FAIRE-seq does not require permeabilization of cells or the isolation of nuclei. The methylation state of DNA can be assessed using **bisulfite sequencing**. Treatment of DNA with bisulfite converts C to U but does not affect 5-methylated cytosines. A variety of sequencing methods that can use this change to provide single-nucleotide resolution information about DNA methylation have been developed. As might be expected, interpreting the growing sets of data on chromatin structure has required the development of new statistical and computational approaches.

Mass Spectrometry

Mass spectrometry (MS) is a technique that can identify the chemical constituents of complex and simple mixtures. It has emerged as a powerful tool for detecting and quantifying thousands of proteins in biological samples, including viruses and virus-infected cells.

A mass spectrometer ionizes the chemical constituents of a mixture and then sorts the ions based on their mass-to-charge ratio. Identification of the components is done by comparison with the patterns generated by known materials.

The total protein content of a cell or a virus particle is called the **proteome**. Human cells have been estimated to contain from 500,000 to 3,000,000 proteins per cubic micrometer, encoded by ~20,000 open reading frames, and their products are further diversified by transcriptional, posttranscriptional, translational, and posttranslational regulation. The cell proteome may be further altered during virus infection. The proteome of virus particles is far less complex, but the very largest viruses can still contain hundreds of proteins. Mass spectrometry can be used to identify proteins and their concentrations in cells and in virus particles and also to reveal protein localization, protein-protein interactions, and posttranslational modifications in infected and uninfected cells.

Mass spectrometry may be combined with biochemical and genomic techniques to provide global views of viral reproduction cycles. For example, changes in proteins secreted by host cells upon virus infection can be readily characterized by performing mass spectrometry on supernatants from infected cells. Another application is to identify protein-protein interactions in virus-infected cells: a promiscuous biotinylating enzyme can be directed to a subcellular compartment, where it biotinylates adjacent molecules. These can be purified by attachment to streptavidin-containing beads and identified by mass spectrometry. Integration of mass spectrometry with some of the methods described above for genome analysis can be used to identify proteins that participate in the regulation of gene expression.

At one time the mass spectrometer was a very expensive instrument restricted to chemistry laboratories. Recent advances in the instrumentation, including cost reduction, as well as sample preparation and computational biology have propelled this technology into the virology research laboratory.

Protein-Protein Interactions

A major goal of virology research is to understand how protein-protein interactions modulate reproduction cycles and pathogenesis. Consequently, multiple experimental approaches have been devised to identify the entire set of interactions among viral proteins and between viral and cell proteins. The yeast two-hybrid screen, a complementation assay which was designed to discover protein-protein interactions, has been adapted to high-throughput applications. In this assay, a transcriptional regulatory protein is split into two fragments, the DNA-binding domain and the activating domain. The coding sequences of two different proteins are fused with the two domains. If the two proteins interact, when the fusion proteins are produced in cells, transcriptional activation (leading to the transcription of a reporter gene) will take place. For high-throughput applications, libraries of protein-coding DNAs are screened against a single viral protein or all viral proteins. This method was used to describe the virus-host interactome of two herpesviruses.

Other approaches to defining interactomes include coimmunoprecipitation, affinity purification of tagged proteins (Fig. 2.21), and labeling of cell proteins with chemical cross-linkers (used to identify plant proteins that interact with plant virus proteins), followed by mass spectrometry.

While these methods allow definition of virus-cell interactomes, they are not unambiguous. For at least one virus, interactomes determined in different laboratories are very diverse. Most importantly, the observation of a protein-protein interaction does not confirm biological relevance: the roles of such interactions in viral reproduction must be determined by other means (Box 2.12).

Single-Cell Virology

Much of virology research is carried out by using populations of cells in culture or in animals. However, as discovered by virologists in the 1950s, individual cells of the same type can behave very differently with respect to susceptibility and permissiveness to infection and the kinetics of virus production.

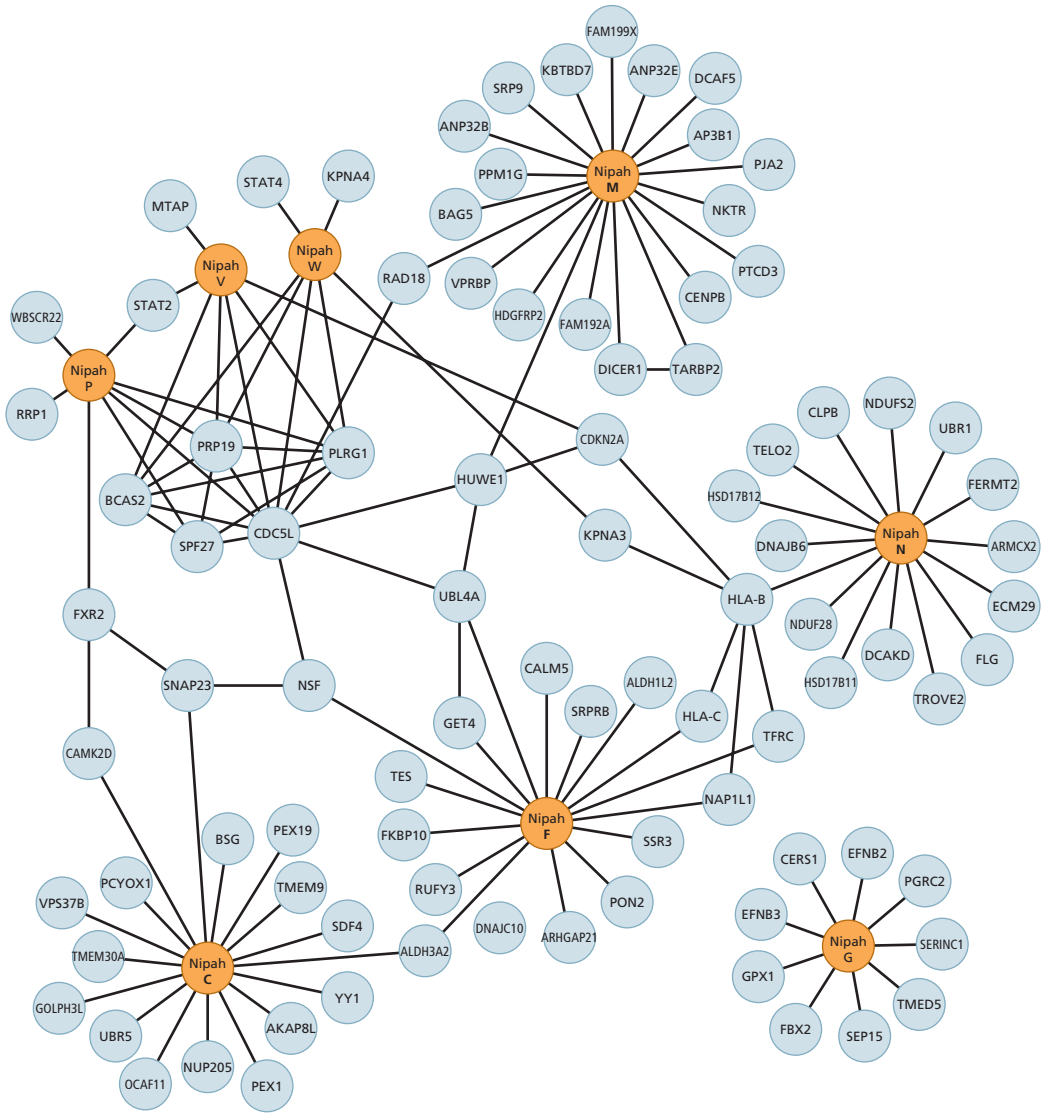


Figure 2.21. Interactions between human proteins and Nipah virus proteins. Network representation of interactions of Nipah virus and human proteins determined by affinity purification and mass spectrometry. Nipah virus proteins are shown in orange. Cellular proteins are shown in gray. Protein names (from UniProt) are shown. Adapted from Martinez-Gil L, Vera-Velasco NM, Mingarro I. 2017. *J Virol* 91:e01461-17, with permission.

As early efforts to study virus infections in single cells were hampered by technical difficulties, the field failed to progress. This situation has changed with the development of flow cytometry and microfluidics and the adaptation of high-throughput methods, such as genome sequencing and mass spectrometry, to single cells.

Initially, micropipettes were used to aspirate a single cell at a time from a population, using a microscope. This labor-intensive method was supplanted by fluorescence-activated cell sorting to allow isolation of up to millions of cells in a few hours, according to size, morphology, or synthesis of specific proteins. More recently, automated microfluidic devices have been developed to allow automated capture of single cells using integrated fluidic circuits. Infection, cell lysis, reverse transcription, and amplification are all performed in these systems before high-throughput sequencing.

The study of virus infections in single cells is expected to provide information that explains why some cells are not infected, why the kinetics of viral reproduction may be so different, and how genomes change in a single cell. An example is the study of poliovirus infection of single cells, using a microfluidics platform installed on a fluorescent microscope (Fig. 2.22). This approach revealed observations otherwise masked in population-based studies, including the unique and independent

BOX 2.12

WARNING

Determining a role for cellular proteins in viral reproduction can be quite difficult

Understanding the roles of both viral and cellular proteins at various stages of viral reproduction is essential for elucidating molecular mechanisms and for developing strategies for blocking pathogenic infections. As viral genomes have a limited set of genes, the viral proteins or genetic elements that are essential at each step can be deduced by introducing mutations and observing phenotypes. Identifying critical cellular genes begins with the identification of cellular proteins that are included in virus particles and/or bind to viral proteins (*in vitro* or in cells).

Once candidates are identified, the contribution of the cellular protein to viral reproduction may be evaluated by observing the effects of

- specific small-molecule inhibitors of the protein's function (inhibitory drugs)
- synthesis of an altered protein, known to have a dominant negative effect on its normal function
- treatment with small RNAs that induce mRNA degradation (see Chapter 10) and reduce the concentration of the cellular protein
- reproduction in cells in which the candidate gene has been mutated or deleted

Even after applying the multiple approaches and methods described above, identifying relevant cellular proteins and evaluating their roles in viral reproduction is seldom easy. The problems encountered include the following.

- More than one protein may provide the required function (redundancy).
- The function of the protein might be essential to the cell, and mutation of the gene that encodes it (or inhibition of protein production) could be lethal.
- Only small quantities of the protein might be required, and reducing its activity with an inhibitor, or its concentration may be insufficient to induce a defect in viral reproduction.
- The cellular protein might provide a slight enhancement to viral reproduction that could be difficult to detect but may be physiologically significant.
- Synthesis of an altered cellular gene or overexpression of a normal cellular gene may produce changes that affect virus reproduction for reasons that are irrelevant to the natural infection (artifacts).

Given these difficulties, it is not surprising that the literature in this area is sometimes contradictory and the results can be controversial.

contribution of viral and cell parameters to reproduction kinetics, the wide variation in reproduction start times, and the finding that reproduction begins later and with greater speed in single cells than in populations. A study of influenza virus infection of single cells revealed a wide variation in the yield, from 1 to 970 PFU per cell. Furthermore, the amounts of viral RNAs within individual cells varied by three orders of magnitude.

Infection of single cells with vesicular stomatitis virus identified 496 mutations that arose in 24 hours during genome replication within 90 cells. The rates of mutation varied among individual cells, and this high value represents an average for all of the cells. In addition, preexisting viral genetic diversity was used to track infection in single cells. These investigations revealed that even though viruses were added at a low multiplicity of infection, most cells had acquired more than one virus particle. The results suggested that virus particles have a tendency to stick to one another, raising further challenges to determining multiplicities of infection.

Single-cell studies have demonstrated that measurements of virus reproduction in populations of cells do not represent the diversity that exists among individual cells. Consequently, they will likely become a complementary tool to the one-step growth experiment for studying virus infection.

Perspectives

One-step growth analysis, while simple, remains a powerful tool for studying virus reproduction. When cells are infected at a high multiplicity of infection, sufficient viral nucleic acid or protein can be isolated to allow a study of their production during the infectious cycle. The ability to synchronize infection is the key to this approach. Many of the experimental results discussed in subsequent chapters of this book were obtained using such one-step growth analysis. The power of this approach is such that it reports on all stages of the reproduction cycle in a simple and quantitative fashion. With modest expenditure of time and reagents, virologists can deduce a great deal about viral translation, genome replication, and assembly. It has long been assumed from such one-step growth analyses that the same steps of the viral reproduction cycle occur at the same time in every infected cell. However, results from analyses of single infected cells demonstrate that the same steps can take place

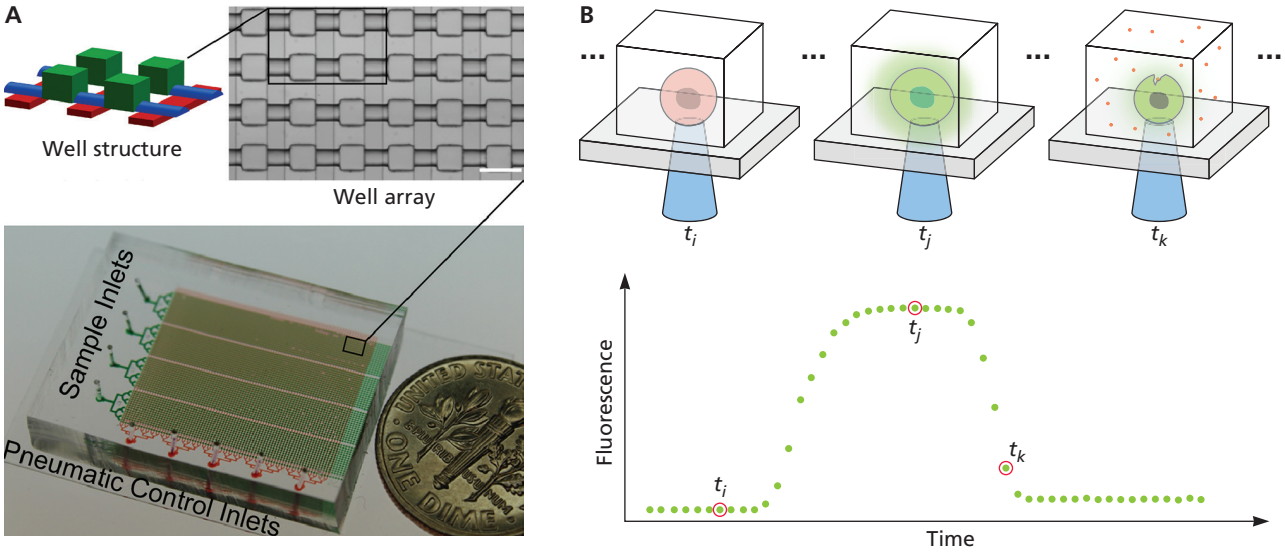


Figure 2.22. Single-cell virology. (A) A microfluidic device with 6,400 wells is fitted with four separate sample inlets (green) and pneumatic control lines (red) that permit each well to be sealed and isolated. A small part of the device is magnified at the top, showing an array of 24 wells, and four wells are further magnified to the left. **(B)** The device can be used to measure real-time fluorescence in cells infected with a virus encoding a fluorescent reporter. The production of fluorescence is shown in the graph and illustrated in the views of single cells in individual cells above. There is a lag in the detection of fluorescence in an infected cell (t_i), followed by virus reproduction (t_j) and a decline in fluorescence caused by cell lysis (t_k). Reprinted from Guo F et al. 2017. *Cell Rep* 21:1692–1704, with permission.

at vastly different times in individual cells in the population. We now understand that results from population-based studies of viral reproduction comprise an average of events occurring in individual cells. One-step growth analyses with single cells have the potential of unraveling the viral and cellular basis for such individual heterogeneity.

From the humble beginnings of the one-step growth curve, many new methods have propelled our understanding of viruses and infected cells to greater depths and at unprecedented speed. An astounding array of technologies, including high-throughput sequencing, proteomics, and single-cell approaches, have been developed. These methods have already led to significant discoveries about viral evolution, reproduction, and pathogenesis. We are truly in a remarkable era, when few experimental questions are beyond the reach of the techniques that are currently available.

REFERENCES

Books

Green MR, Sambrook J. 2012. *Molecular Cloning: A Laboratory Manual*, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Maramorosch K, Koprowski H (ed). 1968. *Methods in Virology.* Academic Press, New York, NY.

A series of volumes begun in the 1960s that contain review articles on classic virological methods.

Review Articles

Dulbecco R, Vogt M. 1953. Some problems of animal virology as studied by the plaque technique. *Cold Spring Harb Symp Quant Biol* **18:**273–279.

Kuadkitkan A, Wikan N, Smith DR. 2016. Induced pluripotent stem cells: a new addition to the virologists armamentarium. *J Virol Methods* **235:**191–195.

Ramani S, Crawford SE, Blutt SE, Estes MK. 2018. Human organoid cultures: transformative new tools for human virus studies. *Curr Opin Virol* **29:**79–86.

van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. 2018. The third revolution in sequencing technology. *Trends Genet* **34:**666–681.

Wang I-H, Burckhardt CJ, Yakimovich A, Greber UF. 2018. Imaging, tracking and computational analyses of virus entry and egress with the cytoskeleton. *Viruses* **10:**1–29.

Papers of Special Interest

Ellis EL, Delbrück M. 1939. The growth of bacteriophage. *J Gen Physiol* **22:**365–384.

A landmark paper describing the development of the plaque assay to measure infectivity of bacteriophage and its use in studying viral reproduction by one-step analysis.

Reed LJ, Muench H. 1938. A simple method for estimating fifty percent endpoints. *Am J Hyg* **27:**493–497.

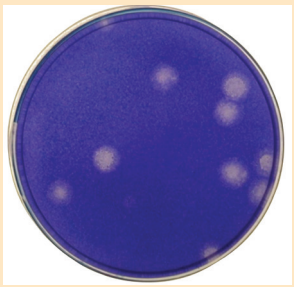
An important paper that shows how to estimate 50% endpoints in titrations of viruses in animals.

Guo F, Li S, Caglar MU, Mao Z, Liu W, Woodman A, Arnold JJ, Wilke CO, Huang TJ, Cameron CE. 2017. Single-cell virology: on-chip investigation of viral infection dynamics. *Cell Rep* **21:**1692–1704.

A seminal paper that presents a pneumatic device for measuring infection of single cells with poliovirus and its use to measure the heterogeneity of infection.

STUDY QUESTIONS

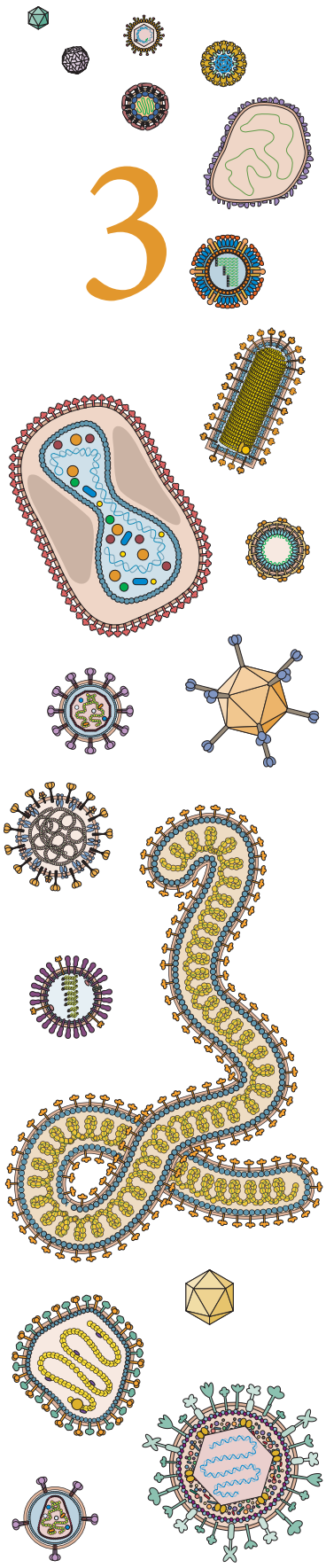
1. Which of the following statements is not correct?
   - **a.** The infectious cycle and virus reproduction are synonymous
   - **b.** A susceptible cell has a functional receptor for a given virus
   - **c.** A cell that is resistant to infection has the viral receptor but the virus cannot be reproduced in it
   - **d.** Virus can be reproduced in a permissive cell
   - **e.** A susceptible and permissive cell is the only cell in which a virus can enter and be reproduced
2. Which of the following statements about cytopathic effects (CPE) is correct?
   - **a.** Some viruses may not produce cytopathic effects in certain cells
   - **b.** Syncytium formation is a type of CPE
   - **c.** Rounding up and detachment of cultured cells is a type of CPE
   - **d.** Formation of a viral plaque depends upon CPE
   - **e.** All of the above
3. Which of the following assays determines the number of infectious viral particles?
   - **a.** ELISA
   - **b.** Hemagglutination assay
   - **c.** Plaque assay
   - **d.** High-throughput sequencing
   - **e.** Polymerase chain reaction
4. If one million infectious virus particles are added to a culture dish of one million cells, which of the following will happen:
   - **a.** Each cell will receive one virus particle because the MOI is 1
   - **b.** None of the cells will receive any virus particles
   - **c.** 37% of the cells are uninfected
   - **d.** Only one infectious cycle will take place
   - **e.** None of the above
5. When doing a plaque assay, what is the purpose of adding a semisolid agar overlay on the monolayer of infected cells?
   - **a.** To stabilize progeny virions
   - **b.** To ensure that cells remain susceptible and permissive
   - **c.** To act as a pH indicator
   - **d.** To keep cells adherent to the plate during incubation
   - **e.** To restrict viral diffusion after lysis of infected cells
6. In the particle-to-PFU ratio, "particle" can best be described as:
   - **a.** One of the proteins which makes up the virion
   - **b.** A virus which may or may not be infectious
   - **c.** A virus which is infectious
   - **d.** A virus which is not infectious
   - **e.** Elementary or composite
7. The plaque assay plate below was made from a dilution of 10^{-6} and 0.1 ml of the dilution was plated on the cell monolayer. What is the titer in PFU/ml?



8. Explain why no infectious viruses are observed in the cell culture medium during the latent phase of a one-step growth curve.
9. You infect a plate of one million cells at an MOI of 100. The particle-to-PFU ratio for this virus is 1,000. How many total virus particles did you add to the cells?

PART II

Molecular Biology



3 Genomes and Genetics

LINKS FOR CHAPTER 3

⏭ *Virocentricity with Eugene Koonin*
http://bit.ly/Virology_Twiv275

⏭ *CRISPR-Cas immune systems*
http://microbe.tv/twim/twim184

... everywhere an interplay between nucleic acids and proteins; a spinning wheel in which the thread makes the spindle and the spindle the thread.

ERWIN CHARGAFF, 1955

Introduction

Earth abounds with uncountable numbers of viruses of great diversity. However, because taxonomists have devised methods of classifying viruses, the number of identifiable groups is manageable (Chapter 1). One of the contributions of molecular biology has been a detailed analysis of the genetic material of representatives of major virus families. From these studies emerged the principle that the **viral genome** is the nucleic acid-based repository of the information needed to build, reproduce, and transmit a virus (Box 3.1). These analyses also revealed that the thousands of distinct viruses defined by classical taxonomic methods can be organized into seven groups, based on the structures of their genomes.

Genome Principles and the Baltimore System

A universal function of viral genomes is to specify proteins. However, none of these genomes encode the complete machinery needed to carry out protein synthesis. Consequently, one important principle is that all viral genomes must be copied to produce messenger RNAs (mRNAs) that can be read by host ribosomes. Literally, all viruses are parasites of their host cells' translation system.

A second principle is that there is unity in diversity: evolution has led to the formation of only seven major types of viral genome. The Baltimore classification system integrates these two principles to construct an elegant molecular algorithm for virologists (Fig. 3.1). When the bewildering array of viruses is classified by this system, we find seven pathways to mRNA. The value of the Baltimore system is that by knowing only the nature of the viral genome, one can deduce the basic steps that must take place to produce mRNA. Perhaps more pragmatically, the system simplifies comprehension of the extraordinary reproduction cycles of viruses.

The Baltimore system omits the second universal function of viral genomes, to serve as a template for synthesis of progeny genomes. Nevertheless, there is also a finite number of nucleic acid-copying strategies, each with unique primer, template, and termination requirements. We shall combine this principle with that embodied in the Baltimore system to define seven strategies based on mRNA synthesis **and** genome replication. The Baltimore system has stood the test of time: despite the discovery of multitudes of viral genome sequences, they all fall into one of the seven classes.

Replication and mRNA synthesis present no obvious challenges for most viruses with DNA genomes, as all cells use DNA-based mechanisms. In contrast, animal cells possess no known systems to copy viral RNA templates and to produce mRNA from them. For RNA viruses to propagate, their RNA genomes must, by definition, encode a nucleic acid polymerase.

Structure and Complexity of Viral Genomes

Despite the simplicity of expression strategies, the composition and structures of viral genomes are far more varied than those seen in the entire archaeal, bacterial, or eukaryotic domains. Nearly every possible method for encoding information in nucleic acid can be found in viruses. Viral genomes can be

- DNA or RNA
- DNA with short segments of RNA
- DNA or RNA with covalently attached protein
- single-stranded (+) strand, (–) strand, or ambisense (Box 3.2)
- double stranded
- linear
- circular
- segmented
- gapped

PRINCIPLES *Genomes and Genetics*

- The genomes of viruses range from the extraordinarily small (<2 kb) to the extraordinarily large (>2,500 kbp); the diversity in size likely provides advantages in the niches in which particular viruses exist.
- Viral genomes specify some, but never all, of the proteins needed to complete the viral reproductive cycle.
- That only seven viral genome replication strategies exist for all known viruses implies unity in viral diversity.
- Some genomes can enter the reproduction cycle upon entry into a target cell, whereas others require prior repair or synthesis of viral gene products before replication can proceed.
- Although the details of replication differ, all viruses with RNA genomes must encode either an RNA-dependent RNA polymerase to synthesize RNA from an RNA template or a reverse transcriptase to convert viral RNA to DNA.
- The information encoded in viral genomes is optimized by a variety of mechanisms; the smaller the genome, the greater the compression of genetic information.
- The genome sequence of a virus is at best a biological "parts list" and tells us little about how the virus interacts with its host.
- Technical advances allowing the introduction of mutations into any viral gene or genome sequence are responsible for much of what we know about viruses.

BOX 3.1

BACKGROUND

What information is encoded in a viral genome?

Gene products and regulatory signals required for

- replication of the genome
- efficient expression of the genome
- assembly and packaging of the genome
- regulation and timing of the reproduction cycle
- modulation of host defenses
- spread to other cells and hosts

Information **not** contained in viral genomes:

- genes encoding a complete protein synthesis machine (e.g., no ribosomal RNA and no ribosomal or translation proteins)
- genes encoding proteins of membrane biosynthesis
- telomeres (to maintain genomes) or centromeres (to ensure segregation of genomes)
- this list becomes shorter with each new edition of this textbook!

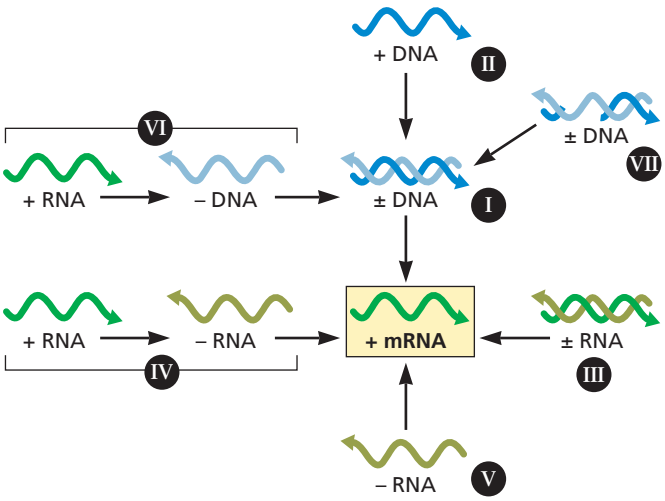


Figure 3.1 The Baltimore classification. All viruses must produce mRNA that can be translated by cellular ribosomes. This classification system traces the pathways from viral genomes to mRNA for the seven classes of viral genomes.

The seven strategies for expression and replication of viral genomes are illustrated in Fig. 3.2 to 3.8. In some cases, genomes can enter the replication cycle directly, but in others, genomes must first be repaired, and viral gene products that participate in the replication cycle must first be synthesized. Examples of specific viruses in each class are provided.

BOX 3.2

TERMINOLOGY

Important conventions: plus (+) and minus (–) strands

mRNA is defined as the positive (+) strand, because it can be translated. A strand of DNA of the equivalent polarity is also designated as a (+) strand; i.e., if it were mRNA, it would be translated into protein.

The RNA or DNA complement of the (+) strand is called the (–) strand. The (–) strand cannot be translated; it must first be copied to make the (+) strand. Ambisense RNA contains both (+) and (–) sequences.

A color key for nucleic acids, proteins, membranes, cells, and more is located in the front of this book.

DNA Genomes

The strategy of having DNA as a viral genome appears at first glance to be the ultimate in genetic efficiency: the host genetic system is based on DNA, so viral genome replication and expression could simply emulate the host system. While the replication of viral and cellular DNA genomes is fundamentally similar, the mechanistic details are varied because viral genomes are structurally diverse.

Double-Stranded DNA (dsDNA) (Fig. 3.2)

There are 38 families of viruses with dsDNA genomes. Those that include vertebrate viruses are the *Adenoviridae*, *Alloherpesviridae*, *Asfarviridae*, *Herpesviridae*, *Papillomaviridae*, *Polyomaviridae*, *Iridoviridae*, and *Poxviridae*. These genomes may be linear or circular. Genome replication and mRNA synthesis are accomplished by host or viral DNA-dependent DNA and RNA polymerases.

Gapped DNA (Fig. 3.3)

Members of two virus families, *Caulimoviridae* and *Hepadnaviridae*, have a gapped DNA genome. The *Hepadnaviridae* include viruses that infect vertebrates. As the gapped DNA genome is partially double stranded, the gaps must be filled to produce perfect duplexes. This repair process must precede mRNA synthesis because the host RNA polymerase can transcribe only fully dsDNA. The unusual gapped DNA genome is produced from an RNA template by a virus-encoded enzyme, reverse transcriptase.

Single-Stranded DNA (ssDNA) (Fig. 3.4)

Thirteen families of viruses containing ssDNA genomes have been recognized; the families *Anelloviridae*, *Circoviri-*

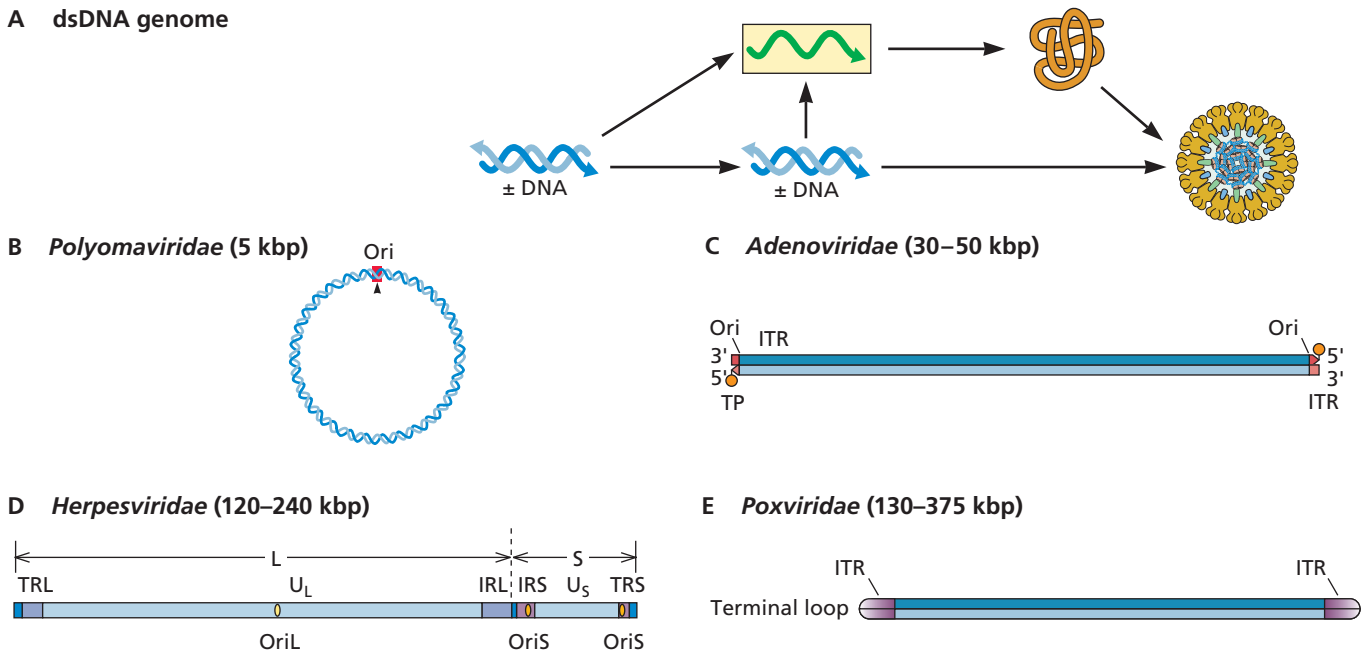


Figure 3.2 Structure and expression of viral double-stranded DNA genomes. (A) Synthesis of genomes, mRNA (shown as green line in yellow box), and protein (shown as brown line). The icon represents a polyomavirus particle. **(B to E)** Genome configurations. Ori, origin of replication; ITR, inverted terminal repeat; TP, terminal protein; L, long region; S, short region; U_L and U_S, long and short unique regions; IRL, internal repeat sequence, long region; IRS, internal repeat sequence, short region; TRL, terminal repeat sequence, long region; TRS, terminal repeat sequence, short region; OriL, origin of replication of the long region; OriS, origin of replication of the short region.

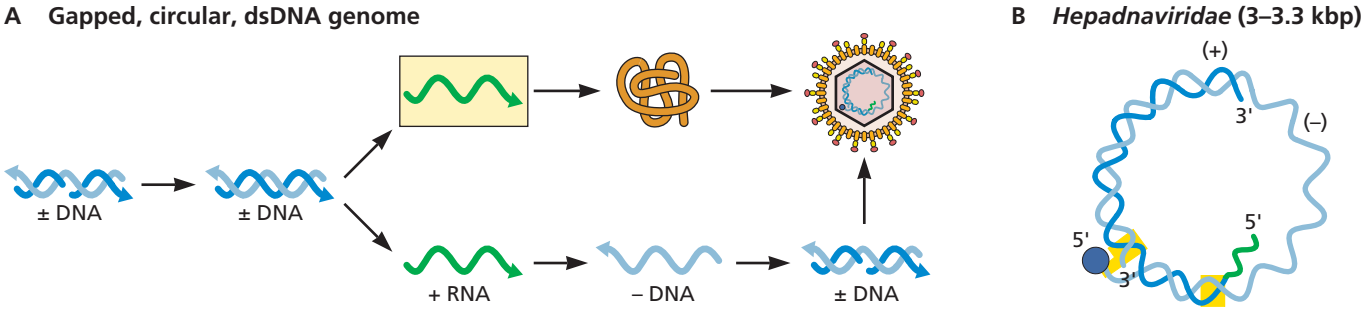


Figure 3.3 Structure and expression of viral gapped, circular, double-stranded DNA genomes. (A) Synthesis of genome, mRNA, and protein. **(B)** Configuration of the hepadnavirus genome.

dae, *Genomoviridae*, and *Parvoviridae* include viruses that infect vertebrates. ssDNA must be copied into mRNA before proteins can be produced. However, RNA can be made only from a dsDNA template, whatever the sense of the ssDNA. Consequently, DNA synthesis **must** precede mRNA production in the replication cycles of these viruses. All synthesis of viral DNA is catalyzed by cellular DNA polymerases.

RNA Genomes

Cells have no RNA-dependent RNA polymerases that can replicate the genomes of RNA viruses or make mRNA from RNA templates (Box 3.3). One solution to this problem is that RNA virus genomes encode RNA-dependent RNA polymerases that produce RNA from RNA templates. The other solution, exemplified by retrovirus genomes, is reverse transcription of the genome to dsDNA, which can be transcribed by host RNA polymerase.

dsRNA (Fig. 3.5)

There are twelve families of viruses with linear dsRNA genomes. The number of dsRNA segments in the virus particle may be 1 (*Totiviridae, Hypoviridae,* and *Endornaviridae*,

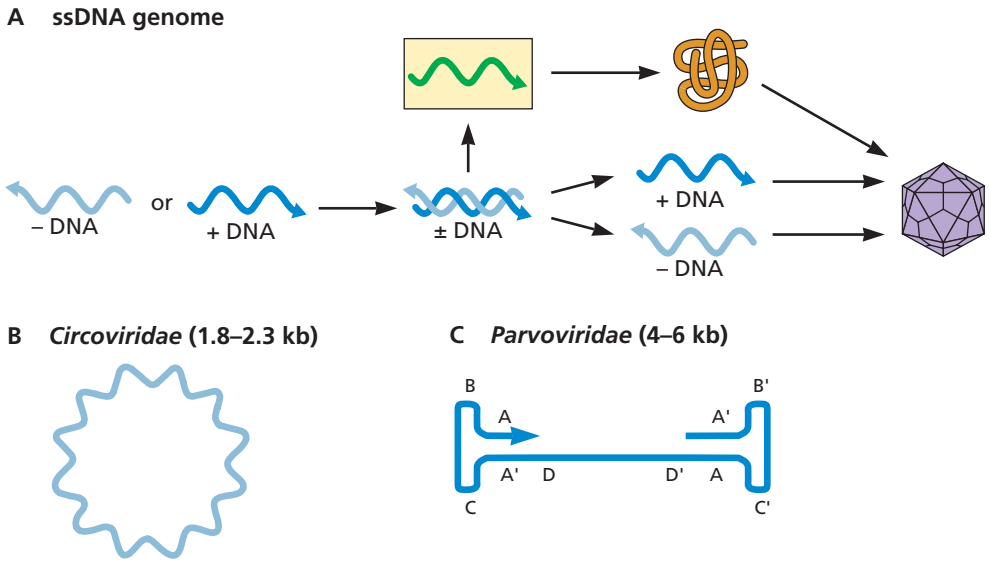


Figure 3.4 Structure and expression of viral single-stranded DNA genomes. (A) Synthesis of genomes, mRNA, and protein. **(B and C)** Genome configurations.

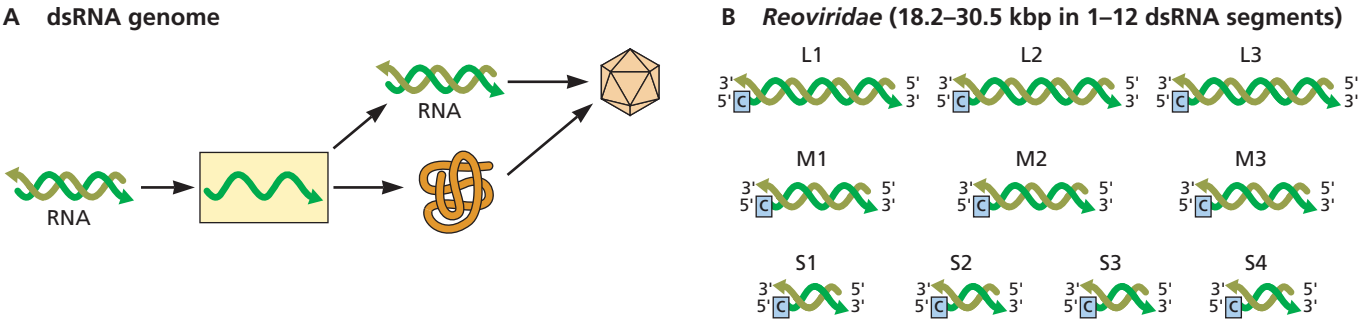


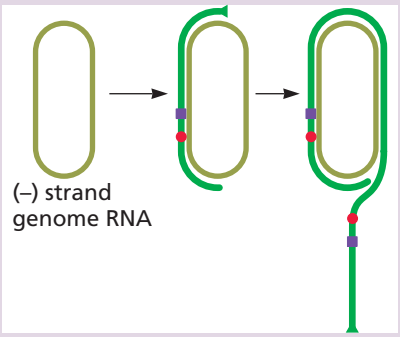
Figure 3.5 Structure and expression of viral double-stranded RNA genomes. (A) Synthesis of genomes, mRNA, and protein. **(B)** Genome configuration.

BOX 3.3

BACKGROUND

RNA synthesis in cells

There are no known host cell enzymes that can copy the genomes of RNA viruses. However, at least one enzyme, RNA polymerase II, can copy an RNA template. The 1.7-kb circular, ssRNA genome of hepatitis delta satellite virus is copied by RNA polymerase II to form multimeric RNAs (see the figure). How RNA polymerase II, an enzyme that produces pre-mRNAs from DNA templates, is reprogrammed to copy a circular RNA template is not known.



Hepatitis delta satellite (–) strand genome RNA is copied by RNA polymerase II at the indicated position. The polymerase passes the poly(A) signal (purple box) and the self-cleavage domain (red circle). For more information, see Fig. 6.25. Redrawn from Taylor JM. 1999. *Curr Top Microbiol Immunol* 239:107–122, with permission.

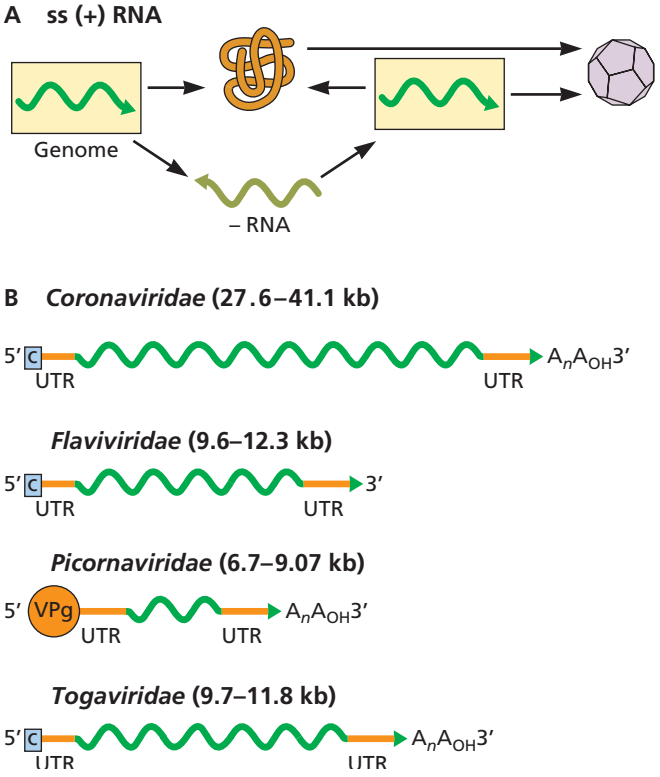


Figure 3.6 Structure and expression of viral single-stranded (+) RNA genomes. (A) Synthesis of genomes, mRNA, and protein. **(B)** Genome configurations. UTR, untranslated region; VPg, virion protein, genome linked.

viruses of fungi, protozoa, and plants); 2 (*Partitiviridae*, *Birnaviridae*, and *Megabirnaviridae*, viruses of fungi, plants, insects, fish, and chickens); 3 (*Cystoviridae*, viruses of *Pseudomonas* bacteria); 4 (*Chrysoviridae*, viruses of fungi); or 10 to 12 (*Reoviridae*, viruses of protozoa, fungi, invertebrates, plants, and vertebrates). While dsRNA contains a (+) strand, it cannot be translated to synthesize viral proteins as part of a duplex. The (–) strand of the genomic dsRNA is first copied into mRNAs by a viral RNA-dependent RNA polymerase. Newly synthesized mRNAs are encapsidated and then copied to produce dsRNAs.

(+) Strand RNA (Fig. 3.6)

There are more different types of (+) strand RNA viruses than any other, and 38 families have been recognized [not counting (+) strand RNA viruses with DNA intermediates]. These genomes are linear and may be single molecules (nonsegmented) or segmented, depending on the family. The families *Arteriviridae*, *Astroviridae*, *Caliciviridae*, *Coronaviridae*, *Flaviviridae*, *Hepeviridae*, *Nodaviridae*, *Picornaviridae*, and *Togaviridae* include viruses that infect vertebrates. (+) strand RNA genomes usually can be translated directly into protein by host ribosomes. The genome is replicated in two steps. The (+) strand genome is first copied into a full-length (–) strand, and the (–) strand is then copied into full-length (+) strand genomes. In some cases, a subgenomic mRNA is produced.

(+) Strand RNA with a DNA Intermediate (Fig. 3.7)

Members of four virus families are (+) strand RNA viruses with a DNA intermediate; those viruses within *Retroviridae* infect vertebrates. In contrast to other (+) strand RNA viruses, the (+) strand RNA genome of retroviruses is converted to a dsDNA intermediate by viral RNA-dependent DNA polymerase (reverse transcriptase). Following integration into host DNA, the viral DNA then serves as the template for viral mRNA and genome RNA synthesis by cellular enzymes.

(–) Strand RNA (Fig. 3.8)

Viruses with (–) strand RNA genomes are found in 19 families. These genomes are linear and may be single molecules

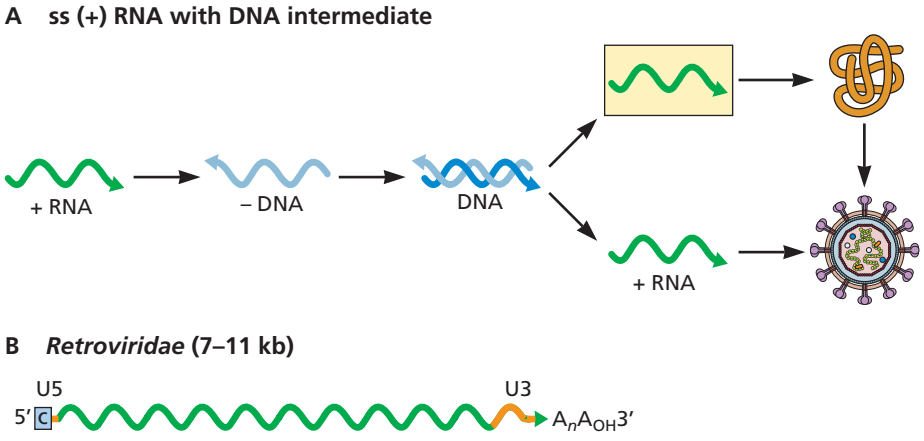


Figure 3.7 Structure and expression of viral single-stranded (+) RNA genomes with a DNA intermediate. (A) Synthesis of genomes, mRNA, and protein. **(B)** Genome configuration.

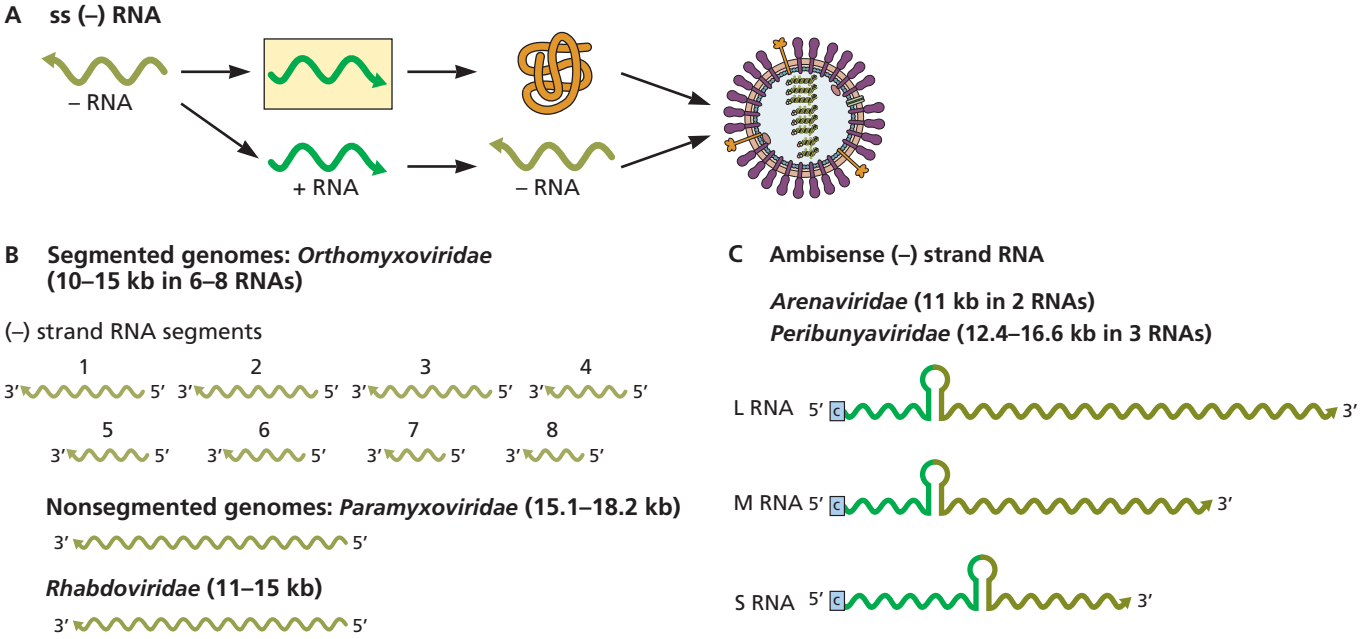


Figure 3.8 Structure and expression of viral single-stranded (–) RNA genomes. (A) Synthesis of genomes, mRNA, and protein. The icon represents an orthomyxovirus particle. **(B and C)** Genome configurations.

(nonsegmented; some viruses with this configuration have been classified in the order *Mononegavirales*) or segmented. Viruses of this type that can infect vertebrates include members of the *Arenaviridae*, *Bornaviridae*, *Filoviridae*, *Hantaviridae*, *Orthomyxoviridae*, *Paramyxoviridae*, *Pneumoviridae*, and *Rhabdoviridae* families. Unlike (+) strand RNA, (–) strand RNA genomes cannot be translated directly into protein but must be first copied to make (+) strand mRNA. There are no enzymes in the cell that can make mRNAs from the RNA genomes of (–) strand RNA viruses. These virus particles therefore contain virus-encoded RNA-dependent RNA polymerases. The genome is also the template for the synthesis of full-length (+) strands, which, in turn, are copied to produce (–) strand genomes.

The genomes of certain (–) strand RNA viruses (e.g., members of the *Arenaviridae* and *Bunyaviridae*) are ambisense: they contain both (+) and (–) strand information on a single strand of RNA (Fig. 3.8C). The (+) sense information in the genome is translated upon entry of the viral RNA into cells. Replication of the RNA genome yields additional (+) sense sequences, which are then translated.

What Do Viral Genomes Look Like?

Some small RNA and DNA genomes enter cells from virus particles as naked molecules of nucleic acid, whereas others are always associated with specialized nucleic acid-binding proteins or enzymes. A fundamental difference between the genomes of viruses and those of their hosts is that although viral genomes are often covered with proteins, they are usually not bound by histones in the virus particle (polyomaviral and papillomaviral genomes are exceptions). However, it is likely that all viral DNAs become coated with histones shortly after they enter the nucleus.

While viral genomes are all nucleic acids, they should **not** be thought of as one-dimensional structures. Virology textbooks (this one included) often draw genomes as straight, one-dimensional lines, but this notation is for illustrative purposes only; physical reality is certain to be dramatically different. Genomes have the potential to adopt amazing secondary and tertiary structures in which nucleotides may engage in long-distance interactions (Fig. 3.9).

The sequences and structures near the ends of viral genomes are often indispensable for replication. For example, the DNA sequences at the ends of parvovirus genomes form T-shaped structures that are required for priming during DNA synthesis. Proteins covalently attached to 5′ ends, inverted and tandem repeats, and bound tRNAs may also participate in the replication of RNA and DNA genomes. Secondary RNA structures may facilitate translation (the internal ribosome entry site [IRES] of picornavirus genomes) and genome packaging (the structured packaging signal of retroviral genomes, [Fig. 3.9]).

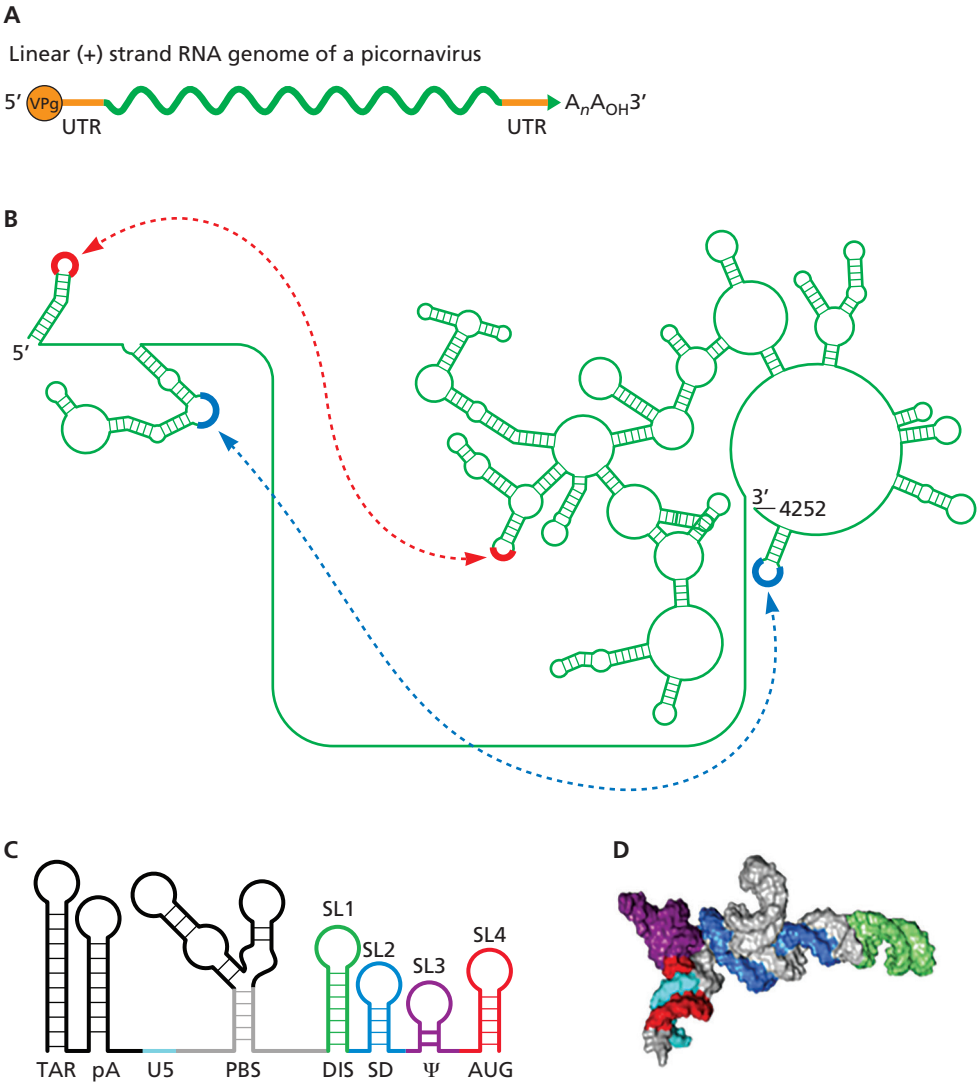


Figure 3.9 Genome structures in cartoons and in real life. **(A)** Linear representation of a picornavirus RNA genome. UTR, untranslated region. **(B)** Long-distance RNA-RNA interactions in a tombusvirus RNA genome. The 4,252-nucleotide viral genome is shown with secondary RNA structures at the 5′ and 3′ ends. Sequences that base-pair are shown in blue (required for RNA frameshifting) and red (required to bring ribosomes from the 3′ end to the 5′ end). Courtesy of Anne Simon, University of Maryland. **(C)** Schematic representation of RNA secondary-structure elements in the human immunodeficiency virus type 1 5′ leader, including the core packaging signal. **(D)** NMR structure of the RNA shown in C, without elements colored black. Courtesy of Paul Bieniasz, Rockefeller University.

Coding Strategies

The compact genome of most viruses renders the "one gene, one mRNA" dogma inaccurate. Extraordinary tactics for information retrieval, such as the production of multiple subgenomic mRNAs, alternative mRNA splicing, RNA editing, and nested transcription units (Fig. 3.10), allow the production of multiple proteins from a single viral genome. Further expansion of the coding capacity of the viral genome is achieved by posttranscriptional mechanisms, such as polyprotein synthesis, leaky scanning, suppression of termination, and ribosomal frameshifting. In general, the smaller the genome, the greater the compression of genetic information.

What Can Viral Sequences Tell Us?

Knowledge about the physical nature of genomes and coding strategies was first obtained by the study of the nucleic acids of viruses. Indeed, DNA sequencing technology was perfected on viral genomes. The first genome of any kind to be

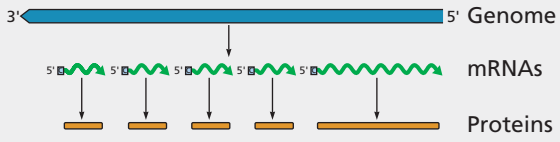
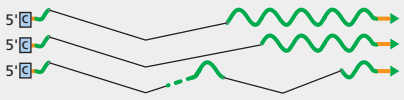
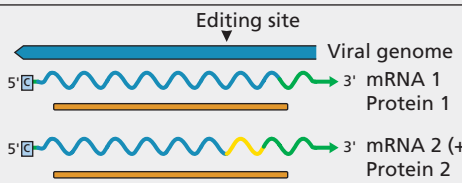
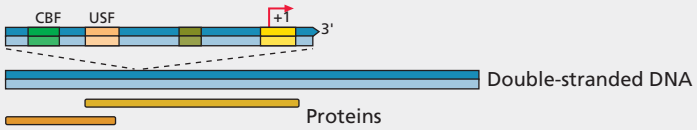
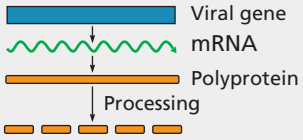
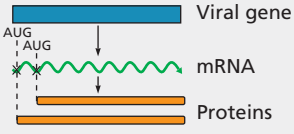
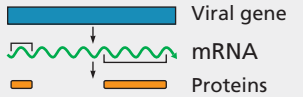
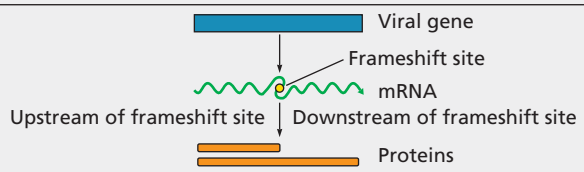
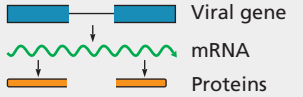
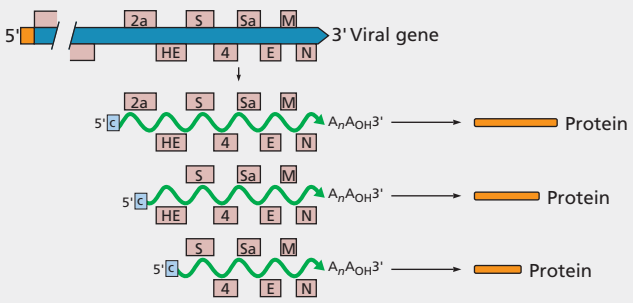
Mechanism	Diagram	Virus	Chapter(s)	Figures in appendix
Multiple subgenomic mRNAs		*Adenoviridae* *Hepadnaviridae* *Herpesviridae* *Paramyxoviridae* *Poxviridae* *Rhabdoviridae*	7, 8 7, 10 7 6 7 6	1, 2 11, 12 17, 18 25, 26 31, 32
Alternative mRNA splicing		*Adenoviridae* *Orthomyxoviridae* *Papillomaviridae* *Polyomaviridae* *Retroviridae*	7, 8 8 7, 8 7, 8 8, 10	1, 2 15, 16 23, 24 29, 30
RNA editing		*Paramyxoviridae* *Filoviridae* Hepatitis delta virus	6, 8 8 8	
Information on both strands		*Adenoviridae* *Polyomaviridae* *Retroviridae*	7–9 7–9 10	1, 2 23, 24 29, 30
Polyprotein synthesis		Alphaviruses *Flaviviridae* *Picornaviridae* *Retroviridae*	6, 11 6, 11 6, 11 6, 11	33, 34 9, 10 21, 22 29, 30
Leaky scanning		*Orthomyxoviridae* *Paramyxoviridae* *Polyomaviridae* *Retroviridae*	11 11 11 11	15, 16 29, 30
Reinitiation		*Orthomyxoviridae* *Herpesviridae*	11 11	15, 16
Suppression of termination		Alphaviruses *Retroviridae*	11 11	33, 34 29, 30
Ribosomal frameshifting		*Astroviridae* *Coronaviridae* *Retroviridae*	11 11 11	 5, 6 29, 30
IRES		*Flaviviridae* *Picornaviridae*	11 11	 21, 22
Nested mRNAs		*Coronaviridae* *Arteriviridae*	6 6	5, 6 5, 6

Figure 3.10 Information retrieval from viral genomes. Different strategies for decoding the information in viral genomes are depicted. CBF, CCAAT-binding factor; USF, upstream stimulatory factor; IRES, internal ribosome entry site.

sequenced was that of the *Escherichia coli* bacteriophage MS2, a linear ssRNA of 3,569 nucleotides. dsDNA genomes of larger viruses, such as herpesviruses and poxviruses (vaccinia virus), were sequenced completely by the 1990s. Since then, high-throughput sequencing has revolutionized the biological sciences, allowing rapid determination of genome sequences from clinical and environmental samples. Organ- and tissue-specific viromes of many organisms have been determined. In one study, over 186 host species representing the phylogenetic diversity of vertebrates, including lancelets (chordates, but considered invertebrates), jawless fish, cartilaginous fish, ray-finned fish, amphibians, and reptiles, all ancestral to birds and mammals, were sampled. RNA was extracted from multiple organs and subjected to high-throughput sequencing. Among 806 billion bases that were read, 214 new viral genomes were identified. The results show that in vertebrates other than birds and mammals, RNA viruses are more numerous and diverse than suspected. Every viral family or genus of bird and mammal viruses is also represented in viruses of amphibians, reptiles, or fish. Arenaviruses, filoviruses, and hantaviruses were found for the first time in aquatic vertebrates. The genomes of some fish viruses have now expanded so that their phylogenetic diversity is larger than in mammalian viruses. New relatives of influenza viruses were found in hagfish, amphibians, and ray-finned fish. As of this writing, the complete sequences of >8,000 different viral genomes have been determined. Published viral genome sequences can be found at http://www.ncbi.nlm.nih.gov/genome/viruses/.

The utility of viral genome sequences extends well beyond building a catalog of viruses. These sequences are the primary basis for classification and also provide information on the origin and evolution of viruses. In outbreaks or epidemics of viral disease, even partial genome sequences can provide information about the identity of the infecting virus and its spread in different populations. New viral nucleic acid sequences can be associated with disease and characterized even in the absence of standard virological techniques (Volume II, Chapter 10). For example, human herpesvirus 8 was identified by comparing sequences present in diseased and nondiseased tissues, and a novel member of the parvovirus family was identified as the cause of unexpected deaths of laboratory mice in Australia and the United States.

Despite their utility, genome sequences cannot provide a complete understanding of how viruses reproduce. The genome sequence of a virus is at best a biological "parts list": it provides some information about the intrinsic properties of a virus (for example, predicted sequences of viral proteins and particle composition), but says little or nothing about how the virus interacts with cells, hosts, and populations. This limitation is best illustrated by the results of environmental metagenomic analyses, which reveal that the number of viruses around us (especially in the sea) is astronomical. Most are uncharacterized and, because their hosts are also unknown, cannot be investigated. A reductionist study of individual components in isolation provides few answers. Although the reductionist approach is often the simplest experimentally, it is also important to understand how the genome behaves among others (population biology) and how the genome changes with time (evolution). Nevertheless, reductionism has provided much-needed detailed information for tractable virus-host systems. These systems allow genetic and biochemical analyses and provide models of infection *in vivo* and in cells in culture. Unfortunately, viruses and hosts that are difficult or impossible to manipulate in the laboratory remain understudied or ignored.

The "Big and Small" of Viral Genomes: Does Size Matter?

The question "does genome size matter" is difficult to answer considering the three orders of magnitude in genome length that separate the largest and the smallest viral genomes. The two largest viral genomes known are those of Pandoravirus salinus (2.4 million bases of dsDNA) and Pandoravirus dulcis (1.9 million bases of dsDNA), encoding 2,541 and 1,487 open reading frames, respectively. The largest RNA virus genomes are far behind (Box 3.4). At the other end are anelloviruses, with a 1,759-base ssDNA genome encoding two proteins (Fig. 3.3B), and viroids, circular, single-stranded RNA molecules of 246 to 401 nucleotides that encode no protein (Volume II, Chapter 13). Anelloviruses include agriculturally important pathogens of chickens and pigs and torque teno (TT) virus, which infects >90% of humans with no known consequence. Viroids cause economically important diseases of crop plants.

All viruses with genome sizes spanning the range from the biggest to the smallest are successful as they continue to reproduce and spread within their hosts. Despite detailed analyses, there is no evidence that one size is more advantageous than another. All viral genomes have evolved under relentless selection, so extremes of size must provide particular advantages. One feature distinguishing large genomes from smaller ones is the presence of many genes that encode proteins for viral genome replication, nucleic acid metabolism, and countering host defense systems. When mimiviruses were first discovered, the surprise was that their genomes encoded components of the protein synthesis system, such as tRNAs and aminoacyl-tRNA synthetases. Tupanviruses, isolated from soda lakes in Brazil and deep ocean sediments, encode all 20 aminoacyl-tRNA synthetases, 70 tRNAs, multiple translation proteins, and more. Only the ribosome is lacking. Why would large viral genomes carry these genes when they are available in their cellular hosts? Perhaps by producing a large part of the translational machinery, viral mRNAs can be more efficiently translated. This explanation is consistent with the finding that

BOX 3.4

EXPERIMENTS

Planaria and mollusks yield the biggest RNA genomes

In the past 20 years the development of high-throughput nucleic acid sequencing methods has rapidly increased the pace of virus discovery. Yet in that time, while the largest DNA genomes have increased nearly ten times, the largest known RNA viral genome has only increased in size by ten percent. This situation has now changed with the discovery of new RNA viruses of planarians and mollusks.

Until very recently, the biggest RNA virus genome known was 33.5 kb (ball python nidovirus), which is much larger than the average sized RNA virus genome of 10 kb. The reason for the difference is that RNA polymerases make errors, and most do not have proofreading capabilities. Nidovirus genomes encode a proofreading exoribonuclease which improves replication fidelity and presumably allows for larger genomes. Even with a proofreading enzyme, the biggest RNA virus genome is much smaller than the minimal cellular DNA genome, which is 200 kb. The results of two new studies show that we can find larger virus RNAs, suggesting that we have not yet reached the size limit of RNA genomes.

A close study of the transcriptome of a planarian revealed a new nidovirus, planarian secretory cell nidovirus, with an RNA genome of 41,103 nucleotides. This viral genome is unusual because it encodes a single, long open reading frame of 13,556 amino acids—the longest viral open reading frame (ORF) discovered so far. All the other known nidoviruses encode multiple open reading frames. Phylogenetic analysis of known nidoviruses suggests that the planarian virus arose from viruses with multiple ORFs, after which their single ORF expanded in size.

The other nidovirus with a large RNA genome was discovered by searching all the available RNA sequences of the mollusk *Aplysia californica*. With a simple nervous system of 20,000 neurons, this mollusk has been studied as a model system in many laboratories. Aplysia californica nido-like virus has an RNA genome of 35,906 nucleotides with ORFs that encode two polyproteins.

From the perspective of genome size, the discovery of these nidovirus genomes suggests that viruses with even larger RNAs remain to be discovered. In both cases the viruses were identified from sequences that had been deposited in public databases, although in both cases, infectious viruses were not reported. Nevertheless, many organisms have not yet had their genomes sequenced and it is likely that many RNA viruses remain to be discovered. Declaring an upper limit on RNA genome size does not seem reasonable if we have not sampled every species.

Saberi A, Gulyaeva AA, Brubacher JL, Newmark PA, Gorbalenya AE. 2018. A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog* **14:**e1007314.

Debat HJ. 2018. Expanding the size limit of RNA viruses: evidence of a novel divergent nidovirus in California sea hare, with a ~39.5 kb virus genome. *bioRxiv* 307678.

the codon and amino acid usage of tupanvirus is different from that of the amoeba that it infects.

Another intriguing set of genes belongs to tetraselmis virus 1, which infects green algae. These hosts, found in nutrient-rich marine and fresh waters, are photosynthetic. The viral genome encodes pyruvate formate-lyase and pyruvate formate-lyase-activating enzyme, which are key members of cellular anaerobic respiration pathways and allow energy production when no oxygen is available. Green algae may use this system in waters depleted of oxygen by exuberant algal growth. If this process occurs in cells, why does the viral genome carry some of the genes involved? The answer is not known, but it is possible that the extra metabolic demands placed on cells during virus replication—especially at night—require additional fermentation enzymes for energy production. The presence of these genes suggests that tetraselmis virus 1 can change host metabolism, perhaps facilitating its reproduction.

These large viruses therefore have sufficient coding capacity to escape some restrictions imposed by host cell biochemistry. The smallest genome of a free-living cell is predicted to comprise <300 genes (based on bacterial genome sequences). Remarkably, this number is smaller than the genetic content of large viral DNA genomes. Nevertheless, the big viruses are **not** cells: their reproduction absolutely requires the cellular translation machinery, as well as host cell systems to make membranes and generate energy.

The parameters that limit the size of viral genomes are largely unknown. There are cellular DNA and RNA molecules that are much longer than those found in virus particles. Consequently, the rate of nucleic acid synthesis is not likely to be limiting. Nor does the capsid volume appear to limit genome size: the icosahedral shell of Mimivirus, which houses a 1.2 million-base-pair DNA genome, is constructed mainly of a single major capsid protein. For larger genomes, the solution is helical symmetry, which can in principle accommodate very large genomes. The Pandoraviruses, with the largest known DNA viral genomes (2,500 kbp), are housed in decidedly nonisometric ovoid particles 1 μm in length and 0.5 μm in diameter.

There is no reason to believe that the upper limit in viral particle and genome size has been discovered. The core compartment of a mimivirus particle is larger than needed to accommodate the 1,200-kbp DNA genome. A particle of this size could, in principle, house a genome of 6 million bp if the DNA were packed at the same density as in polyomaviruses. Indeed, if the genome were packed into the particle at the density reached in some bacteriophages, it could be >12 million bp, the size of that of the smallest free-living unicellular eukaryote.

In cells, DNAs are much longer than RNA molecules. RNA is less stable than DNA, but in the cell, much of the RNA is used

for the synthesis of proteins and therefore need not exceed the size needed to specify the largest polypeptide. However, this constraint does not apply to viral genomes. Yet the largest viral single-molecule RNA genomes, the 41-kb (+) strand RNAs of the nidoviruses (Box 3.4), are dwarfed by the largest (2,500-kbp) DNA virus genomes. Susceptibility of RNA to chemical and nuclease attack might limit the size of viral RNA genomes. However, the most likely explanation is that there are few known enzymes that can correct errors introduced during RNA synthesis. An exonuclease encoded in the coronavirus genome is one exception: its presence could explain the large size of these RNAs. DNA polymerases can eliminate errors during polymerization, a process known as proofreading, and remaining errors can also be corrected after DNA synthesis is complete. The average error frequencies for RNA genomes are about 1 misincorporation in 10^4 or 10^5 nucleotides polymerized. In an RNA viral genome of 10 kb, a mutation frequency of 1 in 10^4 would produce about 1 mutation in every replicated genome. Hence, very long viral RNA genomes, perhaps longer than 40 kb, would sustain too many mutations that would be lethal. Even the 7.5-kb genome of poliovirus exists at the edge of infectivity: treatment of the virus with the RNA mutagen ribavirin causes a >99% loss in a single round of replication.

When new viral genomes are discovered, often many of the putative genes are previously unknown. For example, >93% of the >2,500 genes of Pandoravirus salinus resemble nothing known, and 453 of the 663 predicted open reading frames of tetraselmis virus 1 show no sequence similarity to known proteins. The implication of these findings is clear: our exploration of global genome sequences is far from complete, and viruses with larger genomes might yet be discovered.

The Origin of Viral Genomes

The absence of *bona fide* viral fossils, i.e., ancient material from which viral nucleic acids can be recovered, might appear to make the origin of viral genomes an impenetrable mystery. The oldest viruses recovered from environmental samples, the 30,000-year-old Pithovirus sibericum and Mollivirus sibericum, isolated from Late Pleistocene Siberian permafrost, are simply too rare and too young to provide much information on viral evolution. However, the discovery of fragments of viral nucleic acids integrated into host genomes, coupled with the advances in determining genome sequences of viruses and their hosts, has provided an improved understanding of the evolutionary history of viruses, a topic discussed in depth in Volume II, Chapter 10.

How viruses with DNA or RNA genomes arose is a compelling question. A predominant hypothesis is that RNA viruses are relics of the "RNA world," a period populated only by RNA molecules that catalyzed their own replication in the absence of proteins. During this time, billions of years ago, cellular life could have evolved from RNA, and the earliest cellular organisms might have had RNA genomes. Viruses with RNA genomes might have evolved during this time. Later, DNA replaced RNA as cellular genomes, perhaps through the action of reverse transcriptases. With the emergence of DNA genomes probably came the evolution of DNA viruses. However, those with RNA genomes were and remain evolutionarily competitive, and hence they continue to survive to this day.

Analysis of sequences of more than 4,000 RNA-dependent RNA polymerases is consistent with the hypothesis that the first RNA viruses to emerge after the evolution of translation were those with (+) strand RNA genomes. The last common ancestor of these viruses encoded only an RNA-dependent RNA polymerase and a single capsid protein. Double-stranded RNA viruses evolved from (+) strand RNA viruses on at least two different occasions, and (–) strand RNA viruses evolved from dsRNA viruses. The emergence of viruses with the latter genome types was likely facilitated by the capture of genes such as those encoding RNA helicases, to allow for the production of larger genomes.

Single-stranded DNA viruses of eukaryotes appear to have evolved from genes contributed from both bacterial plasmids and (+) strand RNA viruses. Different dsDNA viruses originated from bacteriophages at least twice. The larger eukaryotic DNA viruses form a monophyletic group based on analysis of 40 genes that derive from a last common ancestor. These viruses appear to have emerged from smaller DNA viruses by the capture of multiple eukaryotic and bacterial genes, such as those encoding translation system components.

There is no evidence that viruses are monophyletic, i.e., descended from a common ancestor: there is no single gene shared by all viruses. Nevertheless, viruses with different genomes and replication strategies do share a small set of viral hallmark genes that encode icosahedral capsid proteins, nucleic acid polymerases, helicases, integrases, and other enzymes. For example, as discussed above, the RNA-dependent RNA polymerase is the only viral hallmark protein conserved in RNA viruses. Examination of the sequences of viral capsid proteins reveals at least 20 distinct varieties that were derived from unrelated genes in ancestral cells on multiple occasions. The emerging evidence therefore suggests that viral replication enzymes arose from precellular self-replicating genetic elements, while capsid protein genes were captured from unrelated genes in cellular hosts.

The compositions of the eukaryotic and bacterial viromes differ substantially (Chapter 1, Fig. 1.13). In bacteria, most known viruses possess dsDNA genomes; fewer viruses have ssDNA genomes, and there is a very limited number of viruses with RNA genomes. In eukaryotes, most of the virome diversity is accounted for by RNA viruses, but ssDNA and dsDNA viruses are common (Chapter 1, Fig. 1.13). The reasons for this difference are unclear, but one possibility is that the formation of the eukaryotic nucleus erected a

barrier for DNA virus reproduction. On the other hand, the eukaryotic cytoplasm with its extensive membranous system might have been a hospitable location for RNA virus replication.

Viral genomes display a greater diversity of genome composition, structure, and reproduction than any organism. Understanding the function of such diversity is an intriguing goal. As viral genomes are survivors of constant selective pressure, all configurations must provide benefits. One possibility is that different genome configurations allow unique mechanisms for control of gene expression. These mechanisms include synthesis of a polyprotein from (+) strand RNA genomes or production of subgenomic mRNAs from (–) strand RNA genomes (see Chapter 6). There is some evidence that segmented RNA genomes might have arisen from monopartite genomes, perhaps to allow regulation of the production of individual proteins (Box 3.5). Segmentation probably did not emerge to increase genome size, as the largest RNA genomes are monopartite.

Genetic Analysis of Viruses

The application of genetic methods to study the structure and function of animal viral genes and proteins began with development of the plaque assay by Renato Dulbecco in 1952. This assay permitted the preparation of clonal stocks of virus, the

BOX 3.5

EXPERIMENTS

Origin of segmented RNA virus genomes

Segmented genomes are plentiful in the RNA virus world. They are found in virus particles from different families and can be double stranded (*Reoviridae*) or single stranded, with (+) (*Closteroviridae*) or (–) (*Orthomyxoviridae*) polarity. Some experimental findings suggest that monopartite viral genomes emerged first and then later fragmented to form segmented genomes.

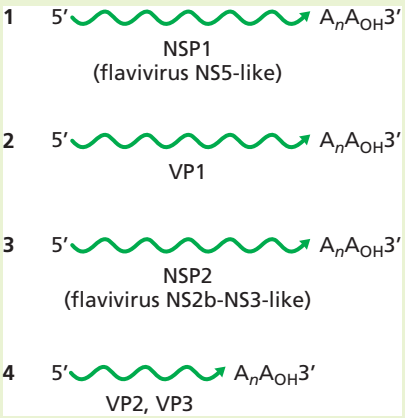
Insight into how such segmented genomes may have been formed comes from studies with the picornavirus foot-and-mouth disease virus. The genome of this virus is a single molecule of (+) strand RNA. Serial passage of the virus in baby hamster kidney cells led to the emergence of genomes with two different large deletions (417 and 999 nucleotides) in the coding region. Neither mutant genome is infectious, but when they are introduced together into cells, an infectious virus population is produced. This population comprises a mixture of each of the two mutant genomes packaged separately into virus particles. Infection is successful because of complementation: when a host cell is infected with both particles, each genome provides the proteins missing in the other.

Further study of the deleted viral genomes revealed the presence of point mutations in other regions of the genome. These mutations had accumulated before the deletions appeared and increased the fitness of the deleted genome compared with the wild-type genome.

These results show how monopartite viral RNAs may be divided, possibly a pathway to a segmented genome. It is interesting that the point mutations that gave the RNAs a fitness advantage over the standard RNA arose before fragmentation occurred, implying that the changes needed to occur in a specific sequence. The authors of the study conclude: "Thus, exploration of sequence space by a viral genome (in this case an unsegmented RNA) can reach a point of the space in which a totally different genome structure (in this case, a segmented RNA) is favored over the form that performed the exploration." While the fragmentation of the foot-and-mouth disease virus genome may represent a step on the path to segmentation, its relevance to what occurs in nature is unclear, because the results were obtained in cells in culture.

A compelling picture of the genesis of a segmented RNA genome comes from the discovery of a new tick-borne virus in China, Jingmen tick virus. The genome of this virus comprises four segments of (+) strand RNA. Two of the RNA segments have no known sequence homologs, while the other two are related to sequences of flaviviruses. The RNA genome of flaviviruses is not segmented: it is a single strand of (+) sense RNA. The proteins encoded by RNA segments 1 and 3 are nonstructural proteins that are clearly related to the flavivirus NS5 and NS3 proteins.

The genome structure of this virus suggests that at some point in the past a flavivirus genome fragmented to produce the RNA segments encoding the NS3- and NS5-like proteins. This fragmentation might have initially taken place as shown for foot-and-mouth disease virus in cells in culture, by fixing of deletion mutations that complemented one another. Next, coinfection of this segmented flavivirus with another unidentified virus could have produced the precursor of Jingmen tick virus.



RNA genome of JMTV virus. The viral genome comprises four segments of single-stranded, (+) sense RNA. Proteins encoded by each RNA are indicated. RNA segments 1 and 3 encode flavivirus-like proteins.

The results provide new clues about the origins of segmented RNA viruses.

Moreno E, Ojosnegros S, García-Arriaza J, Escarmís C, Domingo E, Perales C. 2014. Exploration of sequence space as the basis of viral RNA genome segmentation. *Proc Natl Acad Sci U S A* **111:**6678–6683.

Qin XC, Shi M, Tian JH, Lin XD, Gao DY, He JR, Wang JB, Li CX, Kang YJ, Yu B, Zhou DJ, Xu J, Plyusnin A, Holmes EC, Zhang YZ. 2014. A tick-borne segmented RNA virus contains genome segments derived from unsegmented viral ancestors. *Proc Natl Acad Sci U S A* **111:**6744–6749.

measurement of virus titers, and a convenient system for studying viruses with conditional lethal mutations. Although a limited repertoire of classical genetic methods was available, the mutants that were isolated (Box 3.6) were invaluable in elucidating many aspects of infectious cycles and cell transformation. Contemporary methods of genetic analysis based on recombinant DNA technology confer an essentially unlimited scope for genetic manipulation; in principle, any viral gene of interest can be mutated, and the precise nature of the mutation can be predetermined by the investigator. Much of the large body of information about viruses and their reproduction that we now possess can be attributed to the power of these methods.

Classical Genetic Methods

Mapping Mutations

Before the advent of recombinant DNA technology, it was extremely difficult for investigators to determine the locations of mutations in viral genomes. The **marker rescue** technique (described in "Introducing Mutations into the Viral Genome" below) was a solution to this problem, but before it was developed, other, less satisfactory approaches were exploited.

Recombination mapping can be applied to both DNA and RNA viruses. Recombination results in genetic exchange between genomes within the infected cell. The frequency of recombination between two mutations in a linear genome increases with the physical distance separating them. In practice, cells are coinfected with two mutants, and the frequency of recombination is calculated by dividing the titer of phenotypically wild-type virus (Box 3.7) obtained under restrictive conditions (e.g., high temperature) by the titer measured under permissive conditions (e.g., low temperature). The recombination frequency between pairs of mutants is determined, allowing the mutations to be placed on a contiguous map. Although a location can be assigned for each mutation relative to others, this approach does not result in a physical map of the actual location of the base change in the genome.

In the case of RNA viruses with segmented genomes, the technique of **reassortment** allows the assignment of mutations to specific genome segments. When cells are coinfected with both mutant and wild-type viruses, the progeny includes **reassortants** that inherit RNA segments from either parent. The origins of the RNA segments can be deduced from their migration patterns during gel electrophoresis (Fig. 3.11) or by nucleic acid hybridization. By analyzing a panel of such reassortants, the segment responsible for the phenotype can be identified.

Functional Analysis

Complementation describes the ability of gene products from two different mutant viruses to interact functionally in the same cell, permitting viral reproduction. It can be distinguished from recombination or reassortment by examining the progeny produced by coinfected cells. True complementation yields only the two parental mutants,

BOX 3.6

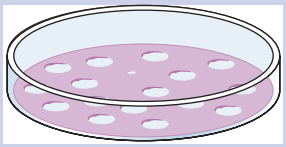
METHODS

Spontaneous and induced mutations

In the early days of experimental virology, mutant viruses could be isolated only by screening stocks for interesting phenotypes, for none of the tools that we now take for granted, such as restriction endonucleases, efficient DNA sequencing methods, and molecular cloning procedures, were developed until the mid to late 1970s. RNA virus stocks usually contain a high proportion of mutants, and it is only a matter of devising the appropriate selection conditions (e.g., high or low temperature or exposure to drugs that inhibit viral reproduction) to select mutants with the desired phenotype from the total population. For example, the live attenuated poliovirus vaccine strains developed by Albert Sabin are mutants that were selected from a virulent virus stock (Volume II, Fig. 7.11).

The low spontaneous mutation rate of DNA viruses necessitated random mutagenesis by exposure to a chemical mutagen. Mutagens such as nitrous acid, hydroxylamine, and alkylating agents chemically modify the nucleic acid in preparations of virus particles, resulting in changes in base-pairing during subsequent genome replication. Base analogs, intercalating agents, or UV light are applied to the infected cell to cause changes in the viral genome during replication. Such agents introduce mutations more or less at random. Some mutations are lethal under all conditions, while others have no effect and are said to be silent.

To facilitate identification of mutants, the population must be screened for a phenotype that can be identified easily in a plaque assay. One such phenotype is temperature-sensitive



viability of the virus. Virus mutants with this phenotype reproduce well at low temperatures, but poorly or not at all at high temperatures. The permissive and nonpermissive temperatures are typically 33 and 39°C, respectively, for viruses that replicate in mammalian cells. Other commonly sought phenotypes are changes in plaque size or morphology, drug resistance, antibody resistance, and host range (that is, loss of the ability to reproduce in certain hosts or host cells).

BOX 3.7

TERMINOLOGY

What is wild type?

Terminology can be confusing. Virologists often use terms such as "strains," "variants," and "mutants" to designate a virus that differs in some heritable way from a parental or wild-type virus. In conventional usage, the **wild type** is defined as the original (often laboratory-adapted) virus from which mutants are selected and which is used as the basis for comparison. A wild-type virus may **not** be identical to a virus isolated from nature. In fact, the genome of a wild-type virus may include numerous mutations accumulated during propagation in the laboratory. For example, the genome of the first isolate of poliovirus obtained in 1909 undoubtedly is very different from that of the virus we call wild type today. We distinguish carefully between laboratory wild types and new virus isolates from the natural host. The latter are called **field isolates** or **clinical isolates**.

The field of viral taxonomy has its own naming conventions which can cause some confusion. Viruses are classified into orders, families, subfamilies, genera, and species. These names are always italicized and start with a capital letter (e.g., *Picornaviridae*). To ensure clarity, the names of viruses (like poliovirus) should be written differently from the names of species (which are constructs that assist in the cataloging of viruses). A species name is written in italics with the first word beginning with a capital letter (other words should be capitalized if they are proper nouns). For example, the causative agents of poliomyelitis, poliovirus types 1, 2, and 3, are members of the species *Enterovirus C*. A virus name should never be italicized, even when it includes the name of a host species or genus, and should be written in lowercase: for example, Sida ciliaris golden mosaic virus. A good exercise would be to see how often we have accidentally violated these rules in this textbook.

while wild-type genomes result from recombination or reassortment. If the mutations being tested are in separate genes, each virus is able to supply a functional gene product, allowing both viruses to be reproduced. If the two viruses carry mutations in the same gene, no reproduction will occur. In this way, the members of collections of mutants obtained by chemical mutagenesis were initially organized into complementation groups defining separate viral functions. In principle, there can be as many complementation groups as genes.

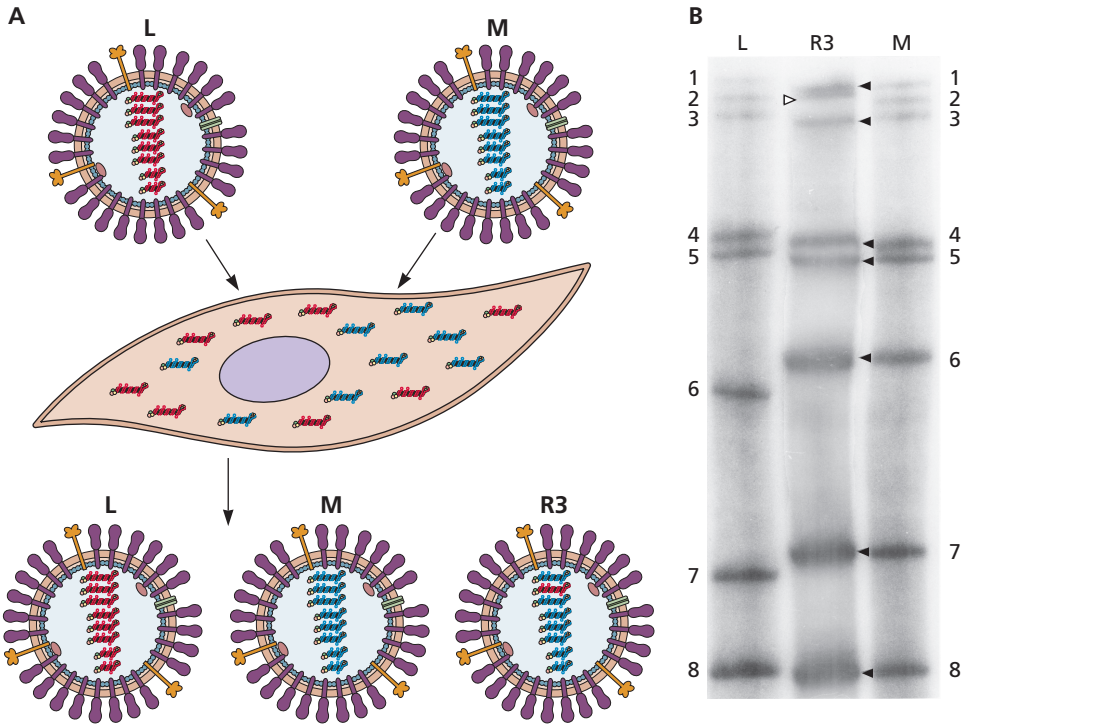


Figure 3.11 Reassortment of influenza virus RNA segments. (A) Progeny viruses of cells that are coinfected with two influenza virus strains, L and M, include both parents and viruses that derive RNA segments from them. Recombinant R3 has inherited segment 2 from the L strain and the remaining seven segments from the M strain. **(B)** ^{32}P-labeled influenza virus RNAs were fractionated in a polyacrylamide gel and detected by autoradiography. Migration differences of parental viral RNAs (M and L) permitted identification of the origin of RNA segments in the progeny virus R3. Panel B reprinted from Racaniello VR, Palese P. 1979. *J Virol* 29:361–373.

Engineering Mutations into Viral Genomes

Infectious DNA Clones

Recombinant DNA techniques have made it possible to introduce any kind of mutation anywhere in the genome of most animal viruses, whether that genome comprises DNA or RNA. The quintessential tool in virology today is the **infectious DNA clone**, a dsDNA copy of the viral genome that is carried on a bacterial vector such as a plasmid. Infectious DNA clones, or *in vitro* transcripts derived from them, can be introduced into cultured cells by **transfection** (Box 3.8) to recover infectious virus. This approach is a modern validation of the Hershey-Chase experiment described in Chapter 1. The availability of site-specific bacterial restriction endonucleases, DNA ligases, and an array of methods for mutagenesis has made it possible to manipulate these infectious clones at will. Infectious DNA clones also provide a stable repository of the viral genome, a particularly important advantage for vaccine strains. As oligonucleotide synthesis has become more efficient and less costly, the assembly of viral DNA genomes up to 212 kbp has become possible (Box 3.9).

DNA viruses. Current genetic methods for the study of most viruses with DNA genomes are based on the infectivity of viral DNA. When deproteinized viral DNA molecules are introduced into permissive cells by transfection, they generally initiate a complete infectious cycle, although the infectivity (number of plaques per microgram of DNA) may be low. For example, the infectivity of deproteinized human adenoviral DNA is between 10 and 100 PFU per µg. When the genome is isolated by procedures that do not degrade the covalently attached terminal protein, infectivity is increased by 2 orders of magnitude, probably because this protein facilitates the assembly of initiation complexes on the viral origins of replication.

The complete genomes of polyomaviruses, papillomaviruses, and adenoviruses can be cloned in plasmid vectors, and such DNA is infectious under appropriate conditions. The DNA genomes of herpesviruses and poxviruses are too large to insert into conventional bacterial plasmid vectors, but they can be cloned into vectors that accept larger insertions (e.g., cosmids and bacterial artificial chromosomes). The plasmids containing such cloned herpesvirus genomes are infectious. In contrast, poxvirus DNA is not infectious, because the viral promoters cannot be recognized by cellular DNA-dependent RNA polymerase. Poxvirus DNA is infectious when early functions (viral DNA-dependent RNA polymerase and transcription proteins) are provided by complementation with a helper virus.

RNA viruses. ***(i) (+) strand RNA viruses.*** The genomic RNA of retroviruses is copied into dsDNA by reverse transcriptase early during infection, a process described in Chapter 10. Such DNA is infectious when introduced into cells, as are molecularly cloned forms inserted into bacterial plasmids.

Infectious DNA clones have been constructed for many (+) strand RNA viruses. An example is the introduction of a plasmid containing cloned poliovirus DNA into cultured mammalian cells, which leads to the production of progeny virus (Fig. 3.12A). The mechanism by which cloned poliovirus DNA initiates infection is not known, but it has been suggested that the DNA enters the nucleus, where it is transcribed by cellular DNA-dependent RNA polymerase from cryptic, promoter-like sequences on the plasmid. The resulting (+) strand RNA transcripts initiate an infectious cycle. During genome replication, the extra terminal nucleotide sequences transcribed from the vector must be removed or ignored, because the virus particles that are produced contain RNA with the authentic 5′ and 3′ termini.

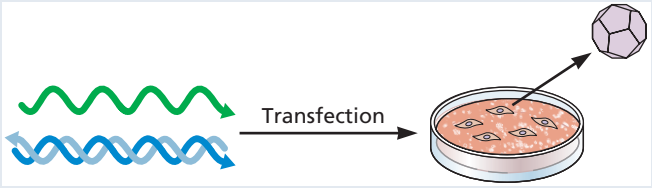
By incorporating promoters for bacteriophage T7 DNA-dependent RNA polymerase in plasmids containing poliovirus

BOX 3.8

TERMINOLOGY

DNA-mediated transformation and transfection

The introduction of foreign DNA into cells is called DNA-mediated transformation to distinguish it from the oncogenic transformation of cells caused by tumor viruses and other insults. The term "transfection" (transformation-infection) was coined to describe the production of infectious virus after transformation of cells by viral DNA, first demonstrated with bacteriophage lambda. Unfortunately, the term "transfection" is now routinely used to describe the introduction of any DNA or RNA into cells.



In this textbook, we use the correct nomenclature: the term "transfection" is restricted to the introduction of viral DNA or RNA into cells with the goal of obtaining virus reproduction.

BOX 3.9

METHODS

Synthesis of infectious horsepox virus from chemically synthesized DNA

Although smallpox has been eradicated, vaccination against the disease is still carried out in certain populations, e.g., the military. The modern smallpox vaccine, which has some undesirable side effects, shares common ancestry with horsepox virus. However, horsepox virus is extinct, so the necessary experiments to determine if it has a better safety profile could not be done. Might the 212,000-bp horsepox dsDNA genome sequence, available in public databases since 1993, be of use?

To rescue horsepox virus from DNA, ten large DNA fragments from 10 to 30 kb were synthesized by a commercial facility (at a cost of $150,000). The DNAs were transfected into cells that were also infected with a related poxvirus, Shope fibroma virus. The latter is needed to provide proteins necessary for transcription of the viral DNA, which contains promoters that are not recognized by the cellular machinery. The medium from the transfected cells was subjected to plaque assay, and single plaques were shown to contain horsepox virus, as determined by viral genome sequencing. The rescued horsepox virus protected immunized mice against a lethal challenge with vaccinia virus.

This work is the first complete synthesis of a poxvirus using synthetic biology methodology. Some argued that the work enabled the rescue of smallpox virus. However, these concerns are spurious, as no new methods were developed by this work. The infectivity of DNA copies of viral genomes had been known for many years when this work was undertaken.

Noyce RS, Lederman S, Evans DH. 2018. Construction of an infectious horsepox virus vaccine from chemically synthesized DNA fragments. *PLoS One* **13:**e0188453.

DNA, full-length (+) strand RNA transcripts can be synthesized *in vitro*. The specific infectivity of such RNA transcripts resembles that of genomic RNA (10^6 PFU per μg), which is higher than that of cloned DNA (10^3 PFU per μg).

(ii) (–) strand RNA viruses. Genomic RNA of (–) strand RNA viruses is not infectious, because it can be neither translated nor copied into (+) strand RNA by host cell RNA polymerases (Chapter 6). Two different experimental approaches have been used to develop infectious DNA clones of these viral genomes (Fig. 3.12B and C).

The recovery of influenza virus from cloned DNA is achieved using an expression system in which cloned DNA copies of the eight RNA segments of the viral genome are inserted between two cellular promoters, so that complementary RNA strands can be synthesized (Fig. 3.12B). When all eight plasmids carrying DNA for each viral RNA segment are introduced into cells, infectious influenza virus is produced.

When the full-length (–) strand RNA of viruses with a nonsegmented genome, such as vesicular stomatitis virus (a rhabdovirus), is introduced into cells containing plasmids that produce viral proteins required for production of mRNA, no infectious virus is recovered. Lack of infectivity is thought to be a consequence of the hybridization of full-length (–) strand RNA with (+) strand mRNAs produced from plasmids encoding viral proteins. Such hybridization might interfere with association of the (–) strand RNA with the N protein, which is required for copying by the viral RNA-dependent RNA polymerase. In contrast, when a full-length (+) strand RNA is transfected into cells that have been engineered to synthesize the vesicular stomatitis virus nucleocapsid protein, phosphoprotein, and polymerase, the (+) strand RNA is copied into (–) strand RNAs. These RNAs initiate an infectious cycle, leading to the production of new virus particles.

dsRNA viruses. Genomic RNA of dsRNA viruses is not infectious because ribosomes cannot access the (+) strand in the duplex. The recovery of reovirus from cloned DNA is achieved by an expression system in which cloned DNA copies of the 10 RNA segments of the viral genome are inserted under the control of a promoter for bacteriophage T7 RNA polymerase (Fig. 3.12D). When all 10 plasmids carrying DNA for each viral dsRNA segment are introduced into cells, infectious reovirus is produced.

Types of Mutation

Recombinant DNA techniques allow the introduction of many kinds of mutation at any desired site in cloned DNA (Box 3.10). **Deletion mutations** can be used to remove an entire gene to assess its role in reproduction, to produce truncated gene products, or to assess the functions of specific segments of a coding sequence. Noncoding regions can be deleted to identify and characterize regulatory sequences such as promoters. **Insertion mutations** can be made by the addition of any desired sequences and may be used to produce **fusion proteins**. **Substitution mutations**, which can correspond to one or more nucleotides, are often made in coding or noncoding regions. Included in the former class are **nonsense mutations**, in which a termination codon is introduced, and **missense mutations**, in which a single nucleotide or a codon is changed, resulting in the synthesis of a protein with a single amino acid substitution. The introduction of a termination codon is frequently exploited to truncate a membrane protein so that it is secreted or to eliminate the synthesis of a protein without changing the

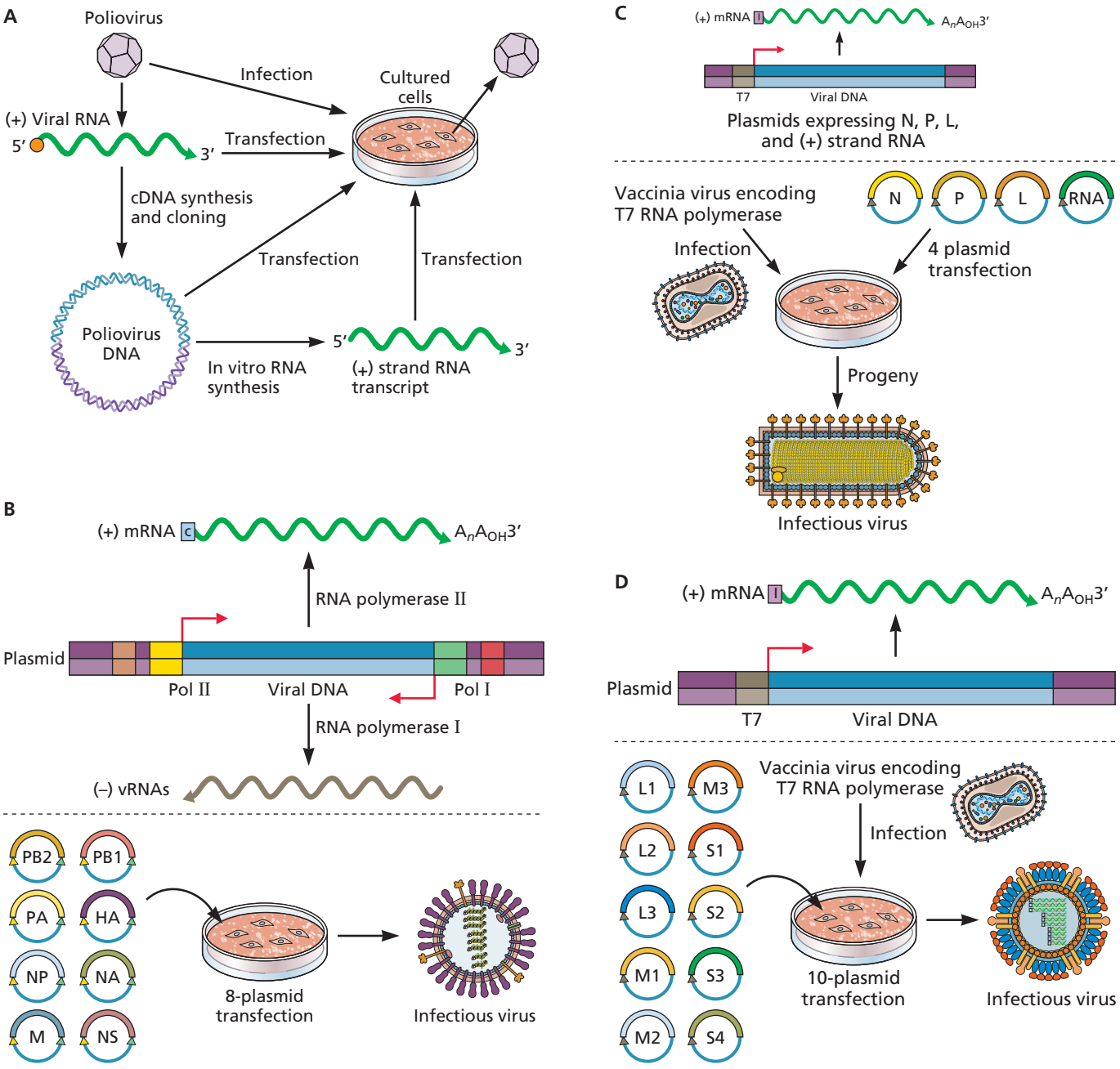


Figure 3.12 Recovery of infectivity from cloned DNA of RNA viruses. (A) The infectivity of cloned DNA of the (+) strand poliovirus RNA genome, which is infectious when introduced into cultured cells by transfection. A complete DNA clone of the viral RNA (blue strands), carried in a plasmid, is also infectious, as are RNAs derived by *in vitro* transcription of the full-length DNA. **(B)** Recovery of influenza viruses by transfection of cells with eight plasmids. Cloned DNA of each of the eight influenza virus RNA segments is inserted between an RNA polymerase I promoter (Pol I [green]) and terminator (brown), and an RNA polymerase II promoter (Pol II [yellow]) and a polyadenylation signal (red). When the eight plasmids are introduced into mammalian cells, (–) strand viral RNA (vRNA) molecules are synthesized from the RNA polymerase I promoter, and mRNAs are produced by transcription from the RNA polymerase II promoter. The mRNAs are translated into viral proteins, and infectious virus is produced from the transfected cells. For clarity, only one cloned viral RNA segment is shown. **(C)** Recovery of infectious virus from cloned DNA of viruses with a (–) strand RNA genome. Cells are infected with a vaccinia virus recombinant that synthesizes T7 RNA polymerase and transformed with plasmids that encode a full-length (+) strand copy of the viral genome RNA and proteins required for viral RNA synthesis (N, P, and L proteins). Production of RNA from these plasmids is under the control of the bacteriophage T7 RNA polymerase promoter (brown). Because bacteriophage T7 RNA transcripts are uncapped, an internal ribosome entry site (I) is included so the mRNAs will be translated. After the plasmids are transfected into cells, the (+) strand RNA is copied into (–) strands, which serve as templates for mRNA synthesis and genome replication. The example shown is for viruses with a single (–) strand RNA genome (e.g., rhabdoviruses and paramyxoviruses). A similar approach has been demonstrated for bunyamwera virus, with a genome comprising three (–) strand RNAs. **(D)** Recovery of infectious virus from cloned DNA of dsRNA viruses. Cloned DNA of each of the 10 reovirus dsRNA segments is inserted under the control of a bacteriophage T7 RNA polymerase promoter (brown). Because bacteriophage T7 RNA transcripts are uncapped, an internal ribosome entry site (I) is included so the mRNAs will be translated. Cells are infected with a vaccinia virus recombinant that synthesizes T7 RNA polymerase and transformed with all 10 plasmids. For clarity, only one cloned viral RNA segment is shown.

size of the viral genome or mRNA. Substitutions are used to assess the roles of specific nucleotides in regulatory sequences or of amino acids in protein function, such as polymerase activity or binding of a viral protein to a cell receptor.

Introducing Mutations into the Viral Genome

Mutations can be introduced into a viral genome when it is cloned in its entirety. Mutagenesis is usually carried out on cloned subfragments, which are then substituted into full-length cloned DNA. This step can now be bypassed by using CRISPR/Cas9 to introduce mutations into complete DNA copies of viral genomes. Viruses are then recovered by introduction of the mutagenized DNA into cultured cells by transfection. This approach has been applied to cloned DNA copies of RNA and DNA viral genomes.

Introduction of mutagenized viral nucleic acid into cultured cells by transfection may have a variety of outcomes, ranging from no effect to a complete block of viral reproduction. Whether the introduced mutation is responsible for an observed phenotype deserves careful scrutiny (Box 3.10).

Reversion Analysis

The phenotypes caused by mutation can **revert** in one of two ways: by change of the mutation to the wild-type sequence or by acquisition of a mutation at a second site, either in the same gene or a different gene. Phenotypic reversion caused by second-site mutation is known as **suppression**, or **pseudoreversion**, to distinguish it from reversion at the original site of mutation. Reversion has been studied since the beginnings of classical genetic analysis. In the modern era of genetics, cloning and sequencing techniques can be used to demonstrate suppression and to identify the nature of the suppressor mutation (see below). The identification of suppressor mutations is a powerful tool for studying protein-protein and protein-nucleic acid interactions. Some mutations complement changes made at several sites, whereas **allele-specific** suppressor mutations complement only a specific change. The allele specificity of second-site mutations provides evidence for physical interactions among proteins and nucleic acids.

Phenotypic revertants can be isolated either by propagating the mutant virus under restrictive conditions or, in the

BOX 3.10

TERMINOLOGY

Operations on nucleic acids and proteins

A mutation is a change in DNA or RNA comprising base changes and nucleotide additions, deletions, and rearrangements. When mutations occur in open reading frames, they can be manifested as changes in the synthesized proteins. For example, one or more base changes in a specific codon may produce a single amino acid substitution, a truncated protein, or no protein. The terms "mutation" and "deletion" are often used incorrectly or ambiguously to describe alterations in proteins. In this textbook, these terms are used to describe genetic changes and the terms "amino acid substitution" and "truncation" to describe protein alterations.

BOX 3.11

DISCUSSION

Is the observed phenotype due to the mutation?

In genetic analysis of viruses, mutations are made *in vitro* by a variety of techniques, all of which can introduce unintended changes. Errors can be introduced during cloning, PCR, or sequencing and when the viral DNA or plasmid DNA is introduced into the cell.

With these potential problems in mind, how can it be concluded that a phenotype arises from the planned mutation? Here are some possible solutions.

- Test several independent DNA clones for the phenotype.
- Repeat the plasmid construction. It is unlikely that an unlinked mutation with the same phenotype would occur twice.
- Look for marker rescue. Replace the mutation and all adjacent DNA with parental DNA. If the mutation indeed causes the phenotype, the wild-type phenotype should be restored in the rescued virus.
- Allow synthesis of the wild-type protein in the mutant background. If the wild-type phenotype is restored (complemented), then the probability is high that the phenotype arises from the mutation. The merit of this method over marker rescue is that the latter shows only that unlinked mutations are probably not the cause of the phenotype.

Each of these approaches has limitations, and it is therefore prudent to use more than one.

case of mutants exhibiting phenotypes (e.g., small plaques), by searching for wild-type properties. Chemical mutagenesis may be required to produce revertants of DNA viruses but is not necessary for RNA viruses, which spawn mutants at a higher frequency. Nucleotide sequence analysis is then used to determine if the original mutation is still present in the genome of the revertant. The presence of the original mutation indicates that reversion has occurred by second-site mutation. The suppressor mutation is identified by nucleotide sequence analysis. The final step is introduction of the suspected suppressor mutation into the genome of the original mutant virus to confirm its effect. Several specific examples of suppressor analysis are provided below.

Some mutations within the origin of replication (Ori) of simian virus 40 reduce viral DNA replication and induce the formation of small plaques (see Chapter 9 for more information on the Ori). Pseudorevertants of Ori mutants were isolated by random mutagenesis of mutant viral DNA followed by introduction into cultured cells and screening for viruses that form large plaques. The second-site mutations that suppressed the replication defects were localized to a specific region within the gene for large T antigen. These results indicated that a specific domain of large T antigen interacts with the Ori sequence during viral genome replication.

The 5′ untranslated region of the poliovirus genome contains elaborate RNA secondary-structural features, which are important for RNA replication and translation, as discussed in Chapters 6 and 11, respectively. Disruption of such features by substitution of a short nucleotide sequence produces a virus that replicates poorly and readily gives rise to pseudorevertants that reproduce more efficiently. Nucleotide sequence analysis of the genomes of two pseudorevertants revealed base changes that restore the disrupted secondary structure. These results confirm that the RNA secondary structure is important for the biological activity of this untranslated region.

RNA Interference (RNAi)

RNA interference (Chapter 8) has become a powerful and widely used tool because it enables targeted loss of gene function. In such analyses, duplexes of 21-nucleotide RNA molecules, called **small interfering RNAs (siRNAs)**, which are complementary to small regions of the mRNA, are synthesized chemically or by transcription reactions. siRNAs or plasmids or viral vectors that encode them are then introduced into cultured cells by transformation or infection. The small molecules then block the production of specific proteins by inducing sequence-specific mRNA degradation or inhibition of translation. Duplex siRNAs are unwound from one 5′ end, and one strand becomes tightly associated with a member of the argonaute (Ago) family of proteins in the RNA-induced silencing complex, RISC. The small RNA acts as a "guide," identifying the target mRNA by base-pairing to specific sequences within it prior to cleavage of the mRNA or inhibition of its translation.

To determine the role of a viral gene in the reproduction cycle, siRNA targeting the mRNA is introduced into cells. Reduced protein levels are verified (e.g., by immunoblot analysis) and the effect on virus reproduction is determined. The same approach is used to evaluate the role of cell proteins such as receptors or antiviral proteins.

In another application of this technology, libraries of thousands of siRNAs directed at all cellular mRNAs or a specific subset can be introduced into cells to identify genes that stimulate or block viral reproduction. The siRNAs are produced from lentiviral vectors as short hairpin RNAs (shRNAs) that are processed into dsRNAs that are then targeted to mRNAs by RISC. In one approach, cells are infected with pools of shRNA-containing lentivirus vectors (Fig. 3.13). The cells are placed under selection and infected with virus to identify changes in reproduction caused by the integrated vector. If necessary, pools of vectors that have an effect on virus reproduction can be further subdivided and rescreened. Enriched shRNAs are detected by high-throughput sequencing and bioinformatic programs that quantitate the number of reads per shRNA compared with the starting population. The likelihood that knockdown of a specific mRNA is a valid result increases as the number of enriched orthologous shRNAs for the targeted gene increases. In other words, a gene targeted by three different shRNAs established by sequencing data is more likely to be a true positive than a gene targeted by only one. Another approach, arrayed RNAi screening, uses transfection of siRNAs into cells grown in a multiwell format (Fig. 3.13). As a record is kept of which siRNAs are added to each well, targeted genes can be readily identified after their effect on virus infection has been ascertained.

No matter which method is used to identify genes that affect viral reproduction, the most convincing confirmation of the result is restoration of the phenotype by expression of a gene containing a mutation that makes the mRNA resistant to silencing.

Targeted Gene Editing with CRISPR-Cas9

Bacteria and archaea possess an endogenous system of defense in which short single-stranded guide RNAs (sgRNAs) are used to target and destroy invading DNA (Volume II, Chapter 3, Box 3.9). One embodiment of this defense, the CRISPR-Cas9 (clustered regularly interspersed short palindromic repeat [CRISPR]-associated nuclease 9) system, has been adapted for effective and efficient targeting gene disruptions and mutations in any genome. The specificity depends on the ability of the sgRNAs to hybridize to the correct DNA sequence within the chromosome. Once annealed, the endonuclease Cas9 catalyzes formation of a double-strand break,

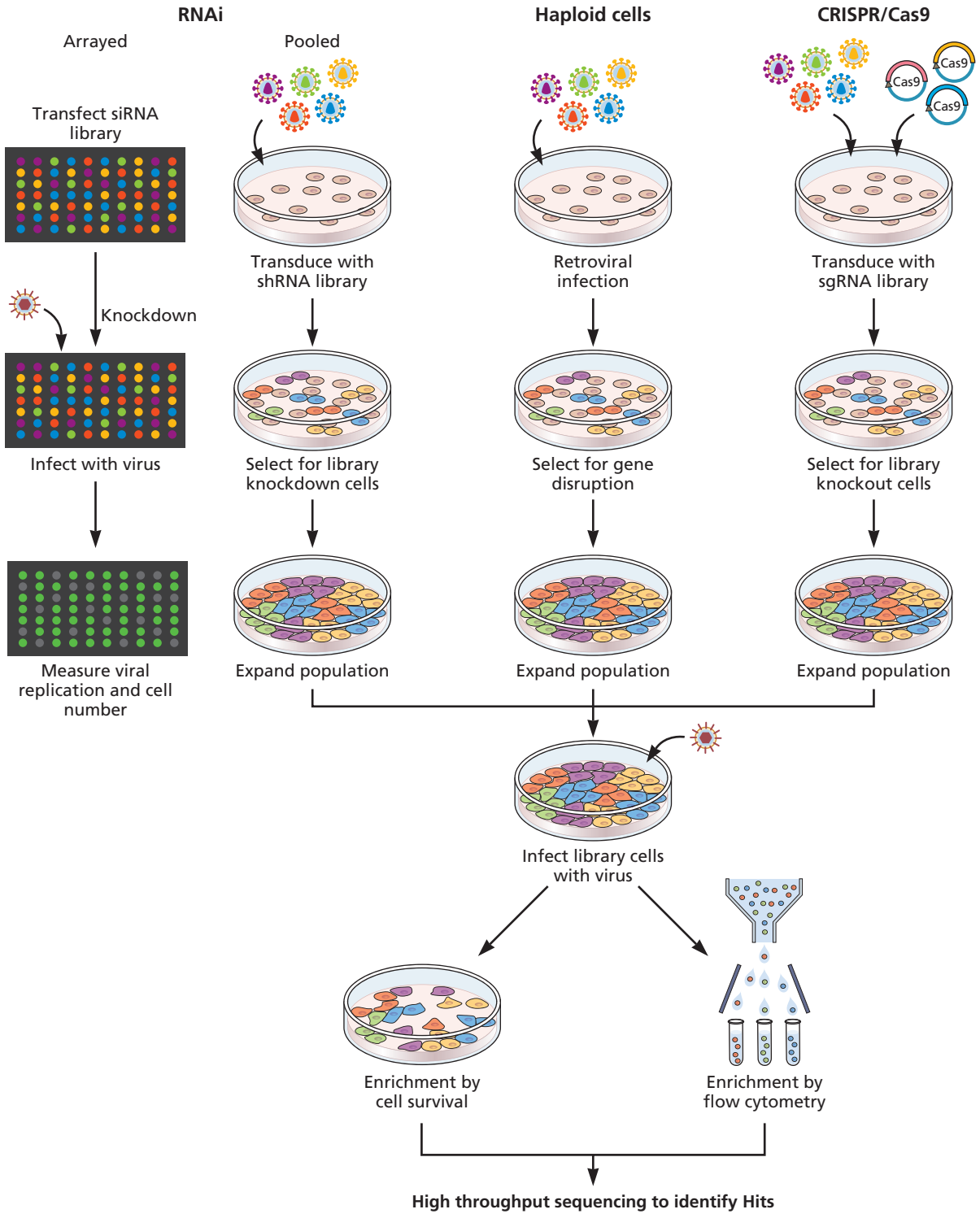


Figure 3.13 Use of RNAi, haploid cells, and CRISPR-Cas9 to study virus-host interactions. In arrayed screens, siRNAs are introduced into cells growing in wells that are subsequently infected with virus. Production of infectious virus or a viral protein is quantified by plaque assay or measurement of a fluorescent protein. Individual siRNA with the desired effect can be identified based on their location in the multiwell plate. In pooled RNAi screens, collections of shRNA producing lentiviral vectors are used to infect cells. After selection for cells with integrated vectors, the cells are infected with the test virus and the production of a viral protein or infectious virus is monitored. In pooled haploid cell screens, cells are infected with lentiviruses at a low multiplicity of infection so that on average one viral genome integration per cell takes place. In pooled CRISPR-Cas9 screens, libraries of sgRNAs are introduced, via lentivirus vector, into cells that produce Cas9. After selection for lentiviral integration, cells are infected with virus. Cell survival and production of infectious virus or a viral protein may be measured depending on what types of genes are sought (e.g., those that are essential for reproduction). In each screen, the cell gene that is disrupted is identified by nucleotide sequencing.

which is then repaired, creating frameshifting insertion/deletion mutations within the gene. One advantage of using CRISPR-Cas9 methodology to modify cell genomes is that the method can be applied to any cell type. Like siRNAs, CRISPR-Cas9 can be used to affect individual mRNAs or to carry out genome-wide screens to identify cell genes that stimulate or block viral reproduction (Fig. 3.13). As with RNAi screens, the most convincing confirmation of the result is restoration of the phenotype by expression of a gene containing a mutation that makes it resistant to Cas9, via changes in the sgRNA target sequence.

While the experimental use of RNAi can lead to reduced protein production, genomic manipulation by CRISPR-Cas9 has advantages of complete depletion of the protein through the production of a homozygous null genotype and fewer off-target effects. With CRISPR-Cas9, the expression of a gene can be permanently extinguished. In contrast, the shRNA-expressing provirus must continually silence the product of ongoing transcription.

Haploid Cell Screening

Haploid cell lines have been used to identify genes required for viral reproduction. These cells, which have only one copy of each chromosome, are infected with retroviruses under conditions where one integration event occurs per cell. The disruption of individual genes that are essential for viral replication can be identified by the isolation of cells resistant to infection (Fig. 3.13). Surviving cells are expanded and the site of proviral integration is determined by PCR and high-throughput sequencing. This approach has been used to identify receptors for viruses, including ebolavirus, Lassa virus, and hantavirus, and genes required for receptor modification and endosomal trafficking.

While powerful, a drawback of this approach is that only a few haploid cell lines are available, and not all viruses can infect these cells.

Engineering Viral Genomes: Viral Vectors

Naked DNA can be introduced into cultured animal cells as complexes with calcium phosphate or lipid-based reagents or directly by electroporation. Such DNA can direct synthesis of its gene products transiently or stably from integrated or episomal copies. Introduction of DNA into cells is a routine method in virological research and is also employed for certain clinical applications, such as the production of a therapeutic protein or a vaccine or the engineering of primary cells, progenitor cells, and stem cells for subsequent introduction into patients. However, this approach is not suitable for all applications. In some cases, gene delivery by viral vector is preferred. Viral vectors have also found widespread use in the research laboratory, including applications in which the delivery of a gene to specific cells, or at high efficiency, is desired. The use of viral vectors for gene therapy, the delivery of a gene to patients who either lack the gene or carry defective versions of it, or to destroy tumors typically employs viral vectors, not naked DNA (see Volume II, Chapter 9). In one application, DNA including the gene is introduced and expressed in cells obtained from the patient. After infusion into patients, the cells can become permanently established. If the primary cells to be used are limiting in a culture (e.g., stem cells), it is not practical to select and amplify the rare cells that receive naked DNA. Recombinant viruses carrying foreign genes can infect a greater percentage of cells and thus facilitate generation of the desired population. A complete understanding of the structure and function of viral vectors requires knowledge of viral genome replication, a topic discussed in subsequent chapters for selected viruses and summarized in the Appendix.

Design requirements for viral vectors include the use of an appropriate promoter, maintenance of genome size within the packaging limit of the particle, and elimination of viral virulence, the capacity of the virus to cause disease. Expression of foreign genes from viral vectors may be controlled by homologous or heterologous promoters and enhancers chosen to support efficient or cell-type-specific transcription, depending on the goals of the experiment. Such genes can be built directly into the viral genome or introduced by recombination in cells, as described above (see "Engineering Mutations into Viral Genomes"). The viral vector genome generally carries deletions and sometimes additional mutations. Deletion of some viral sequences is often required to overcome the limitations on the size of viral genomes that can be packaged in virus particles.

When viral vectors are designed for therapeutic purposes, it is essential to prevent their reproduction as well as destruction of target host cells. The deletions necessary to accommodate a foreign gene may contribute to such disabling of the vector. For example, the E1A protein-coding sequences that are always deleted from adenovirus vectors are necessary for efficient transcription of viral early genes; in their absence, viral yields from cells in culture are reduced by about 3 to 6 orders of magnitude (depending on the cell type). Removal of E1A-coding sequences from adenovirus vectors is therefore doubly beneficial, although it is not sufficient to ensure that the vector cannot reproduce or induce damage in a host animal. Adenovirus-associated virus vectors are not lytic, obviating the need for such manipulations. As discussed in detail in Volume II, Chapter 9, production of virus vectors that do not cause disease can be more difficult to achieve.

A summary of viral vectors is presented in Table 3.1, and examples are discussed below.

DNA Virus Vectors

One goal of gene therapy is to introduce genes into terminally differentiated cells. Such cells normally do not divide,

Table 3.1 Some viral vectors

Virus	Insert size	Integration	Duration of expression	Advantages	Potential disadvantages
Adeno-associated virus	~5 kb	No	Long	Nonpathogenic, episomal, infects nondividing and dividing cells, broad tropism, low immunogenicity	Small transgene capacity, helper virus needed for vector production
Adenovirus	~8–38 kb	No	Short	Broad tropism, efficient gene delivery, infects nondividing and dividing cells, large cargo capacity	Transient, immunogenic, high levels of preexisting immunity
Baculovirus	No known upper limit	No	Short	High levels of protein synthesis, recombinant viruses easily made, more than one protein can be made in same cells	Insect cells typically used, no replication in mammalian cells, human type protein glycosylation not 100% efficient, paucimannose structures present
Gammaretrovirus (murine leukemia virus)	8 kb	Yes	Short	Stable integration, broad tropism possible via pseudotyping, low immunogenicity, low preexisting immunity	Risk of insertional mutagenesis, poor infection of nondividing cells, faulty reverse transcription
Herpes simplex virus	~50 kb	No	Long in central nervous system, short elsewhere	Infects nondividing cells, large capacity, broad tropism, latency	Virulence, persistence in neurons, high levels of preexisting immunity, may recombine with genomes in latently infected cells
Lentivirus	9 kb	Yes	Long	Stable integration, transduces nondividing and dividing cells	Potential insertional mutagenesis; none detected in clinical trials
Rhabdovirus	~4.5 kb	No	Short	High-level expression, rapid cell killing, broad tropism, lack of preexisting immunity	Virulence, highly cytopathic, neurotropism, immunogenic
Vaccinia virus	~30 kb	No	Short	Wide host range, ease of isolation, large capacity, high-level expression, low preexisting immunity	Transient, immunogenic

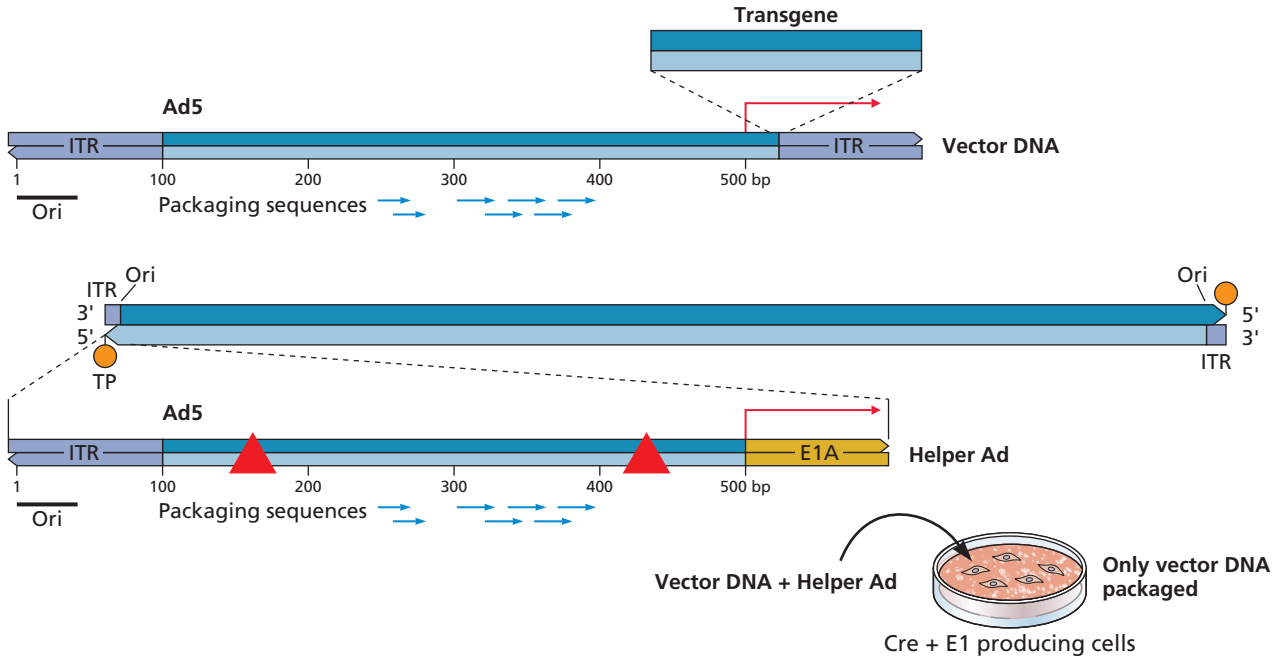


Figure 3.14 Adenovirus vectors. High-capacity adenovirus "gutless" vectors contain only the origin-of-replication-containing inverted terminal repeats (ITR), the packaging signal (blue arrows), the viral E4 transcription unit (red arrow), and the transgene with its promoter. Additional DNA flanking the foreign gene must be inserted to allow packaging of the viral genome (not shown). A helper virus (bottom) is required to package the recombinant vector genome. Two *loxP* sites for cleavage by the Cre recombinase have been introduced into the adenoviral helper genome (red arrowheads). Infection of cells that produce Cre leads to excision of sequences flanked by the *loxP* sites so that the helper genome is not packaged.

and they cannot be propagated in culture. Moreover, the organs they comprise cannot be populated with cells infected by viruses *ex vivo*. DNA virus vectors have been developed to overcome some of these problems.

Adenovirus vectors were originally developed for the treatment of cystic fibrosis because of the tropism of the virus for the respiratory epithelium. Adenovirus can infect terminally differentiated cells, but only transient gene expression is achieved, as infected cells are lysed. Yields of particles are high and these viruses can infect many replicating and nonreplicating cell types. In the earliest vectors that were designed, foreign genes were inserted into the E1 and/or E3 regions. As these vectors had limited capacity, genomes with minimal adenovirus sequences have been designed (Fig. 3.14). This strategy allows up to 38 kb of foreign sequence to be introduced into the vector. In addition, elimination of most viral genes reduces cytotoxicity and the host immune response to viral proteins, simplifying multiple immunizations. Considerable efforts have been made to modify the adenovirus capsid to target the vectors to different cell types. For example, the fiber protein, which mediates adenovirus binding to cells, has been altered by insertion of ligands that bind particular cell surface receptors. Such alterations could increase the cell specificity of adenovirus attachment and the efficiency of gene transfer, thereby decreasing the dose of virus that need be administered.

Adenovirus-associated virus has attracted much attention as a vector for gene therapy. This virus requires a helper virus for replication; in its absence the genome remains episomal and persists, in some cases with high levels of expression, in many different tissues. There has been increasing interest in these vectors to target therapeutic genes to smooth muscle and other differentiated tissues, which are highly susceptible and support sustained high-level expression of foreign genes. Although the first-generation adenovirus-associated virus vectors were limited in the size of inserts that could be transferred, other systems have been developed to overcome the limited genetic capacity (Fig. 3.15). The cell specificity of adenovirus-associated virus vectors has been altered by inserting receptor-specific ligands into the capsid. In addition, many new viral serotypes that vary in their tropism and ability to trigger immune responses have been identified or generated.

Vaccinia virus and other animal poxvirus vectors offer the advantages of a wide host range, a genome that accepts very large fragments, high expression of foreign genes, and relative ease of preparation. Foreign DNA is usually inserted into the viral genome by homologous recombination, using an approach similar to that described for marker transfer. Because of the relatively low pathogenicity of the virus, poxvirus recombinants have been considered candidates for human and animal vaccines.

Baculoviruses, which infect arthropods, have large circular dsDNA genomes. These viruses have been modified to become versatile and powerful vectors for the production of proteins for research and clinical use. The general approach is to replace the viral polyhedron gene with the gene of interest. Recombinant viruses are produced in *E. coli* using a bacmid vector that harbors the baculovirus genome. The gene to be introduced is inserted into the baculovirus genome by recombination. Strong viral promoters are used to obtain high levels of protein production. Recombinant baculoviruses are obtained after transfection of bacmids into insect cells and have been used for protein production for research purposes and for large-scale synthesis for commercial uses. Examples include the influenza virus vaccine FluBlok, which consists of the viral HA proteins produced in insect cells via a baculovirus vector, and porcine circovirus 2 vaccine for the prevention of fatal disease in swine.

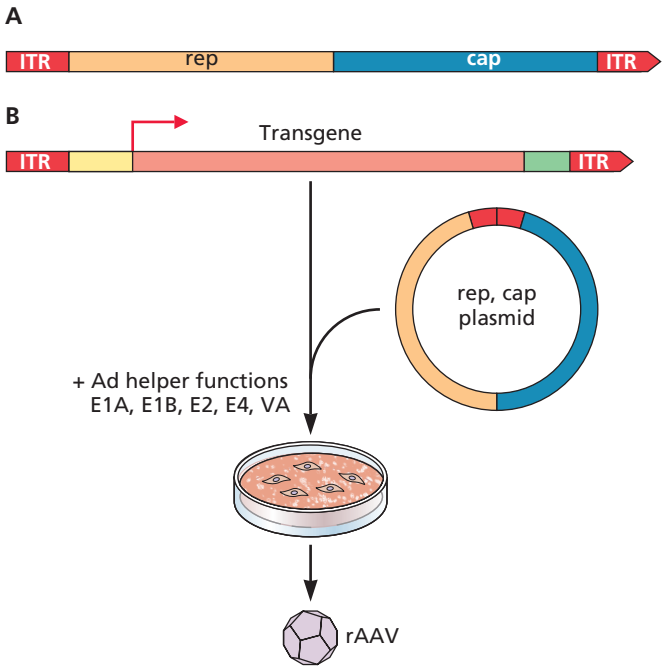


Figure 3.15 Adeno-associated virus vectors. (A) Map of the genome of wild-type adeno-associated virus. The viral DNA is single stranded and flanked by two inverted terminal repeats (ITR); it encodes capsid (blue) and nonstructural (orange) proteins. **(B)** In one type of vector, the viral genes are replaced with the transgene (pink) and its promoter (yellow) and a poly(A) addition signal (green). These DNAs are introduced into cells that have been engineered to produce capsid proteins, and the vector genome is encapsidated into virus particles. A limitation of this vector structure is that only 4.1 to 4.9 kb of foreign DNA can be packaged efficiently. Ad, adenovirus; rAAV, recombinant adenovirus-associated virus.