Edited by Heinz Fraenkel-Conrat and Robert R. Wagner

Comprehensive Virology

8

Regulation and Genetics

Bacterial DNA Viruses

Virology

8

Regulation and Genetics

Bacterial DNA Viruses

PLENUM PRESS • NEW YORK AND LONDON

Library of Congress Cataloging in Publication Data

Fraenkel-Conrat, Heinz, 1910-
Regulation and genetics.

(Their Comprehensive virology; v. 8)
Includes bibliographies and index.
1. Viral genetics. 2. Bacteriophage. 3. Genetic regulation. I. Wagner, Robert R., 1923- II. Title. III. Series.
QR357.F72 vol. 8 [QH434] 576'.64'0ℭ [576'.6482]
ISBN 0-306-35148-X 76-22208

A Division of Plenum Publishing Corporation
227 West 17th Street, New York, N.Y. 10011

Printed in the United States of America

Foreword

The time seems ripe for a critical compendium of that segment of the biological universe we call viruses. Virology, as a science, having passed only recently through its descriptive phase of naming and numbering, has probably reached that stage at which relatively few new—truly new—viruses will be discovered. Triggered by the intellectual probes and techniques of molecular biology, genetics, biochemical cytology, and high-resolution microscopy and spectroscopy, the field has experienced a genuine information explosion.

Few serious attempts have been made to chronicle these events. This comprehensive series, which will comprise some 6000 pages in a total of about 22 volumes, represents a commitment by a large group of active investigators to analyze, digest, and expostulate on the great mass of data relating to viruses, much of which is now amorphous and disjointed, and scattered throughout a wide literature. In this way, we hope to place the entire field in perspective, and to develop an invaluable reference and sourcebook for researchers and students at all levels.

This series is designed as a continuum that can be entered anywhere, but which also provides a logical progression of developing facts and integrated concepts.

Volume 1 contains an alphabetical catalogue of almost all viruses of vertebrates, insects, plants, and protists, describing them in general terms. Volumes 2–4 deal primarily, but not exclusively, with the processes of infection and reproduction of the major groups of viruses in their hosts. Volume 2 deals with the simple RNA viruses of bacteria, plants, and animals; the togaviruses (formerly called arboviruses), which share with these only the feature that the virion's RNA is able to act as messenger RNA in the host cell; and the reoviruses of animals and plants, which all share several structurally singular features, the most important being the double-strandedness of their multiple RNA molecules. This grouping, of course, has only slightly more in its favor than others that could have been, or indeed were, considered.

Volume 3 addresses itself to the reproduction of all DNA-containing viruses of vertebrates, a seemingly simple act of classification, even though the field encompasses the smallest and the largest viruses known. The reproduction of the larger and more complex RNA viruses is the subject matter of Volume 4. These viruses share the property of lipid-rich envelopes with the togaviruses included in Volume 2. They share as a group, and with the reoviruses, the presence of enzymes in their virions and the need for their RNA to become transcribed before it can serve messenger functions.

Volumes 5 and 6 represent the first in a series that focuses primarily on the structure and assembly of virus particles. Volume 5 is devoted to general structural principles involving the relationship and specificity of interaction of viral capsid proteins and their nucleic acids, or host nucleic acids. It deals primarily with helical and the simpler isometric viruses, as well as with the relationship of nucleic acid to protein shell in the T-even phages. Volume 6 is concerned with the structure of the picornaviruses, and with the reconstitution of plant and bacterial RNA viruses.

Volumes 7 and 8 deal with the DNA bacteriophages. Volume 7 concludes the series of volumes on the reproduction of viruses (Volumes 2–4 and Volume 7) and deals particularly with the single- and double-stranded virulent bacteriophages.

In the present volume, the first of the series on regulation and genetics of viruses, the biological properties of the lysogenic and defective phages are covered, the phage–satellite system P2–P4 described, and the regulatory principles governing the development of selected typical lytic phages discussed in depth.

Volume 8 will be followed by others dealing with the regulation of gene expression and integration of animal viruses, the genetics of animal viruses, and regulation of plant virus development, covirus systems, satellitism, and viroids. In addition, it is anticipated that there will be two or three other volumes devoted largely to structural aspects and the assembly of bacteriophages and animal viruses, and to special virus groups.

The complete series will endeavor to encompass all aspects of the molecular biology and the behavior of viruses. We hope to keep this series up to date at all times by prompt and rapid publication of all contributions, and by encouraging the authors to update their chapters by additions or corrections whenever a volume is reprinted.

Contents

Chapter 2

Bacteriophage λ: The Lysogenic Pathway

R. A. Weisberg, S. Gottesman, and M. E. Gottesman

Chapter 3

Defective Bacteriophages and Incomplete Prophages

A. Campbell

Chapter 4

The P 2–P 4 *Trans*activation System

Richard Calendar, Janet Geisselsoder, Melvin G. Sunshine, Erich W. Six, and Björn Lindqvist

CHAPTER 1

Regulation of Gene Action in the Development of Lytic Bacteriophages

Dietmar Rabussay and E. Peter Geiduschek
Department of Biology
University of California at San Diego
La Jolla, California 92093

1. GENERAL INTRODUCTION

In this chapter, we shall analyze the ways in which gene activity is regulated in the development of the larger DNA viruses of bacteria. The kind of analysis that we write about has now been carried out for about 20 years, with increasing sophistication and insight as the understanding of genetic regulation has improved and as the methods of genetics, molecular biology, and enzymology have developed. The immediate goal of this search is to understand, in the molecular terms of protein–nucleic acid interaction, all the strategies of genetic regulation that are encompassed in the development of the phages. That goal is far from realization, although great strides toward it have been taken during the last few years. In fact, we originally hoped to organize our presentation of this subject around a classification of the mechanisms of action of regulatory elements and, in this way, to draw together the common principles of phage development. We eventually decided against the scheme because, with the notable exceptions that we shall emphasize, the mechanisms of gene regulation in phage development are not sufficiently worked out. We have instead chosen several pro-

totypes of lytic phage development in order to analyze the diversity of regulatory mechanisms while avoiding excessive repetition of quasi-equivalent detail. However, one should stress that, although regulatory schemes are wondrously diverse, the general features of the development of the lytic phages fit a common general pattern. We return to a consideration of these common features in the final section of this article, but they are worth noting at the outset.

1. The propagation of the phages that we shall analyze involves time-ordered sequences of gene expression. In general, two or three stages of the development can be distinguished. The earlier stages involve: (a) partial or complete inactivation of host-cell macromolecular metabolism and its replacement by viral genome-directed nucleic acid and protein synthesis; (b) inactivation of host-cell functions that would prevent viral multiplication; (c) preparation for, and initiation of, viral DNA replication and recombination. The late stage of viral gene expression is principally directed toward the production of the components of mature progeny virus particles and toward virion assembly.

2. The time-ordered sequences of viral gene activity are generated primarily by regulation at the transcriptional level.

3. Many different regulatory strategies generate similar temporal sequences of gene action. The common element of these strategies is that at least some viral genes are positively regulated at the transcriptional level by proteins that are coded by the viral genome. These positive regulators are of several kinds. They include genome-specific RNA polymerases and proteins that interact with the RNA polymerase of the host. The latter include some proteins that modify the initiation specificity of RNA polymerase and others that (probably) affect RNA chain termination. In some instances, the regulation of gene activity is partly achieved by DNA processing, i.e., modification of DNA structure.

4. The lytic phages establish diverse regulatory and metabolic interrelations with their hosts. The pattern of total, rapid, and irreversible host genome inactivation and degradation, which was established for the T series of *E. coli* phages in the early years of phage biology, is not universal. Many lytic phages leave host macromolecular synthesis substantially intact until late in infection.

We have chosen our prototypes principally because they embody distinct regulatory patterns of gene expression but also as reflections of our own research activities: phage T 7 has been chosen because it is the best understood of the lytic DNA phages and because it exemplifies regulation of transcription by a phage genome-specific RNA polymerase. We chose the hmU DNA phages (phages which contain

hydroxymethyluridine instead of thymidine in their DNA) of *B. subtilis* because they exemplify regulation of transcription by a phage-specific initiation factor. We chose the T-even phages because of their complexity, because of the great range of regulatory events that occur during their development, and because they permit some discussion of post-transcriptional regulation and replication-transcription coupling. Phage T 5 was included for the sake of its stepwise transfer of DNA. Phages N 4 and PBS 2 were included because of their interesting and only recently studied autarchic regulation and, in the case of N 4, because of the selective shutoff of host genes during viral development. We excluded a detailed analysis of the lytic development of phage λ because it will be treated in Chapter 2 of this volume. We also excluded discussion of phage ϕ 29 despite the detailed analysis to which it is being subjected (e.g., Pène *et al.*, 1973; Schachtele *et al.*, 1973; Carrascosa *et al.*, 1973) simply because it is not yet clear in which of its regulatory principles it differs from our chosen prototypes.

We have introduced our chapter with a review of closely related aspects of transcription, translation, and their regulation in bacteria. The introduction is deliberately less detailed than the rest of the chapter, though indicating salient features of these subjects. At various stages of our writing, we have been tempted to wait for this or that development which would give us a coherent and simple view of everything that we wanted to write about. At the urging of our editors, we have decided not to wait for the millenium but to write about this subject now, to the best of our abilities. We have been greatly helped by J. Abelson, C. Bordier, J. J. Duffy, S. C. Falco, L. Gold, O. Grau, M. Hayashi, M. H. Malamy, L. Nakada, D. Nierlich, H. Noller, L. B. Rothman-Denes, L. Synder, P. R. Srinivasan, J. A. Steitz, W. Studier, A. Wahba, S. B. Weiss, and C. Yanofsky who provided unpublished material, answered our questions, or provided valuable explanations. We are especially grateful to E. N. Brody, M. Chamberlin, O. Grau, R. Haselkorn, C. K. Mathews, D. J. McCorquodale, D. Nierlich, E. G. Niles, L. B. Rothman-Denes, L. Snyder, J. Steitz, and W. Studier who read and criticized different sections of the text.

1.1. Summary of Certain Aspects of Transcription and Translation Control in Uninfected Bacteria

1.1.1. Regulation of Transcription

Of the three steps in the pathway of genetic information transfer—replication, transcription, and translation—transcription must be the

simplest, and translation the most complex. RNA is synthesized on a DNA template by a single enzyme composed of several subunits. This enzyme, RNA polymerase, can execute the initiation, propagation, and, in some instances, termination of RNA chains in the absence of accessory proteins. The expression of all but the simplest genomes is selective, either intrinsically or as a result of regulation. In *E. coli* this selectivity generates abundances of mRNA that vary by over four orders of magnitude (Kennell, 1968*b*). At least a part of this selectivity must be unregulated* and must result directly from the interactions of RNA polymerase with promoter-region nucleotide sequences in DNA which permit a discrimination among different transcription units based on the strengths of their promoters. The principles relating DNA structure to promoter strength remain to be discovered. Nucleotide sequences are the experimental foundation on which these principles must be based, and promoter sequences are now being amassed. Regulated selectivity of transcription in bacteria involves the action of numerous proteins. Some of these transcription elements, like the *lac* repressor and the *ara*C protein, have specific effects through their action at unique DNA-binding sites. Others may have more general activities either because they interact with DNA at many sites or because they modify RNA polymerase. However, for any single regulated gene or operon, the number of transcription elements is not large.

1.1.1a. RNA Polymerase

All the bacterial RNA polymerases bear strong resemblances to each other. *E. coli* RNA polymerase, which is the best studied, is a zinc metalloenzyme containing three polypeptides, α, β, and β' (molecular weights 40, 155, and 165 $\times$ 10^3) in a tightly associated complex or "core" whose composition is $\alpha_2\beta\beta'$ (Burgess *et al.*, 1969; Zillig *et al.*, 1970). This core complex possesses polymerase activity and is the unit active in elongating RNA chains *in vivo* (Chamberlin, 1974). A fourth subunit, σ (molecular weight 90 $\times$ 10^3), functions as an initiation factor. A σ subunit of different molecular weight (51 $\times$ 10^3) but similar properties is found in *B. subtilis* RNA polymerase (Avila *et al.*, 1970, 1971; Losick *et al.*, 1970). A second σ (σ_{II}) has been reported by one laboratory as a minor species, tightly bound to only a fraction of RNA

* The great majority of *E. coli* genes are transcribed very infrequently *in vivo*. We guess that most of these genes are not specifically regulated; infrequent transcription is likely to be an intrinsic property of these genes and of their promoters.

polymerase (designated as RNA polymerase II; Fukuda *et al.*, 1974; Iwakawa *et al.*, 1974). σ_{II} and RNA polymerase II are reported to have the same spectrum of transcriptional capacity as σ and RNA polymerase I. *E. coli* and *B. subtilis* σ can be readily dissociated from core, but σ_{II} and the initiation factors of a number of bacterial RNA polymerases are less readily dissociable (Herzfeld and Zillig, 1971; Johnson *et al.*, 1971; Bendis and Shapiro, 1973). In *E. coli* and *B. subtilis* only one core RNA polymerase has been found thus far: A single kind of β subunit is responsible for all bacterial messenger and stable RNA synthesis and for the transcription of several bacterial viruses. Another form of RNA polymerase (III) has been reported to participate in primer synthesis for M13 DNA replication (Wickner and Kornberg, 1974). It has been proposed that RNA polymerase III provides one of the two known modes of primer RNA synthesis for DNA replication, having the antibiotic (rifamycin) sensitivity of RNA polymerase I. Another pathway of RNA primer synthesis involves the product of one of the replication genes, *dnaG* (Schekman *et al.*, 1975).

1.1.1b. Initiation, Growth, and Termination of RNA Chains

In bacteria, different segments of the genome are expressed at enormously different rates, and most bacterial genes are transcribed very rarely under any growth conditions. Electron micrographs of lysed bacteria, in which coupled transcription and translation can be examined *in situ*, provide the most dramatic illustration of this relationship (Miller *et al.*, 1970). Variations in the frequency of initiation rather than chain growth rate are responsible for this range of genetic activity. The initiation process has been analyzed in some detail *in vitro*, particularly through the use of phage DNA templates which yield only a relatively small number of different transcripts. One can summarize* the current understanding of this subject, which has recently been reviewed in a particularly lucid way (Chamberlin, 1974) in terms of the following reaction sequence: One can conceive of the binding of RNA polymerase to promoter regions of DNA as occurring in at least two steps: DNA recognition, which provides RNA polymerase access to the double helix, followed by tight binding. Little is known about the tight-binding step and one surmises, mainly on the basis of genetic evidence, that it is distinct from recognition-entry. However, no information

* This summary refers only to RNA polymerase I, which has been thoroughly studied. Throughout our text we shall refer to this enzyme simply as "RNA polymerase."

about the chemical nature of promoter recognition by RNA polymerase is as yet available.

1. *Binding.* *E. coli* RNA polymerase holoenzyme binds very tightly to limited numbers of sites on double-helical DNA. This binding is strongly dependent on temperature and ionic strength. The interaction between T 7 DNA and *E. coli* RNA polymerase is characterized by the following properties: At 37°C and low ionic strength, the RNA polymerase–DNA complex is extraordinarily stable (dissociation constant of approximately 10^{-14} *M*). RNA polymerase also binds much more weakly (dissociation constant approximately 10^{-6} to 10^{-8} *M*) to DNA at very many sites, perhaps at all nucleotide sequences. The σ subunit evidently determines this dichotomy of binding, since RNA polymerase core binds generally to DNA with a single intermediate affinity (dissociation constant approximately 10^{-11} *M*). At least some of the tight-binding sites on DNA for RNA polymerase are located at initiation sequences for RNA chains. It seems very likely, therefore, that a tight-binding site for RNA polymerase is a part of any strong promoter.

2. *Promoter Activation.* In order for tight binding to occur, RNA polymerase and DNA must undergo an interaction which leads to some unwinding of the DNA double helix (Saucier and Wang, 1972). This process of forming an "open" promoter or "rapid start" complex has the high temperature-dependence characteristic of a cooperative transition (Walter *et al.*, 1967; Zillig *et al.*, 1970; Mangel and Chamberlin, 1974*c*). Different transcription units and RNA polymerases have different transition temperatures (Remold-O'Donnell and Zillig, 1969; Pène and Barrow-Carraway, 1972; Travers *et al.*, 1973). This must reflect differences in nucleotide sequence and RNA polymerase–DNA interaction. The ease of opening probably helps determine promoter strength.

It is generally assumed that the initiation factor σ is required for formation of open, rapidly initiating complexes between RNA polymerase and promoters. Chamberlin (1974) has pointed out that, plausible as this notion seems, it is not yet supported by experimental evidence. Specifically, it is not yet known whether an open complex would form spontaneously in the absence of σ. The outcome of this experiment would decide whether σ enhances initiation directly by acting at the promoter, or only indirectly by preventing the sequestering of RNA polymerase at nonpromoter sites. (This ambiguity about the role of σ recapitulates past arguments about the role of the initiation factor IF3 in protein synthesis.) Experiments with very small restriction frag-

ments of double-stranded DNA containing intact promotor regions could help to resolve remaining doubts about the mechanism of action of σ.

The notion that promotor activation involves changes of DNA conformation is consistent with the effects of negatively superhelically twisting (supercoiling) DNA on transcription *in vitro*. Supercoiling increases the template activity of circular double-stranded DNA (Hayashi and Hayashi, 1971; Botchan *et al.*, 1973; Wang, 1974; Zimmer and Millette, 1975). It appears to increase the rate of formation and the stability of "open" promoter complexes between *E. coli* RNA polymerase and phage PM2 DNA, allows these complexes to form at lower temperatures, and increases the number of enzyme molecules that can form open complexes on each DNA molecule. Supercoiling evidently does not increase the rate of RNA chain initiation at open complexes (Richardson, 1975). In a general sense, DNA supercoiling produces effects that are analogous to increasing promoter strength. Indeed it has been proposed that the introduction of single-strand breaks into circular double-stranded phage S 13 DNA deactivates transcription as it initiates replication in the infected cell (Puga and Tessman, 1973*a*).

3. *Initiation.* RNA polymerase that is bound to DNA in an "open" complex can initiate RNA chains extremely rapidly, within a fraction of a second at sufficiently high nucleotide concentrations (Mangel and Chamberlin, 1974*a,b,c*). The initiation involves the joining of the 5′ terminal nucleotide pppX to its 3′ first neighbor. X is always a purine, *in vitro* and *in vivo* (Maitra and Hurwitz, 1965; Bremer *et al.*, 1965; Maitra *et al.*, 1967; Jorgenson *et al.*, 1969; Konrad *et al.*, 1975). It is the great rate of RNA initiation from these "open" or "rapid start" complexes which evidently makes them so resistant to inhibition by the antibiotic rifamycin, an inhibitor of RNA chain initiation by procaryotic RNA polymerase (Sippel and Hartmann, 1970). The term "initiation" is used loosely here. In fact, rifamycin has recently been shown to block the first translocation step of RNA synthesis: the addition of the third nucleotide to the initial dinucleotide (Johnston and McClure, 1976). *In vitro* and, probably, *in vivo*, RNA synthesis is kinetically limited by the process of RNA polymerase binding to RNA at sites where a rapid start complex can be formed. For this reason, it has been proposed that binding sites for RNA polymerase might be clustered in a "storage region" near those transcription units that initiate RNA chains most rapidly, about once every one or two seconds (Schäfer *et al.*, 1973*a*). However, there is as yet no compelling experi-

mental evidence in favor of this attractive notion and, in the case of the T 7 early promoter there is evidence against it (Bordier and Dubochet, 1974; Chamberlin, 1974). At the tight-binding sites, RNA polymerase protects approximately 40-nucleotide-long segments of double-stranded DNA from digestion by endonucleases (Okamoto *et al.*, 1972; Heyden *et al.*, 1972). The protected segments include the template for approximately fifteen 5′ terminal nucleotides of the transcript. The transcribed sections of different polymerase-protected segments share no common sequences; presumably each messenger-generating sequence is merely occluded by an RNA polymerase molecule which is specifically bound at the adjacent tight-binding site. This tight-binding site may be specified by a staggered sequence (e.g., xAxxGCx-xCTAAxTx in one strand) resembling the irregular but specific ridges on a key rather than by a contiguous sequence of nucleotides. Among the RNA polymerase tight-binding sites that have been analyzed thus far, an AT-rich sequence of seven nucleotides is relatively "conserved" (Pribnow, 1975*a*,*b*).

The varying strengths of different promoters probably arise in one of the following ways: (a) by variations in the frequency of promoter recognition or of the rate with which the transition from the (hypothetical) promoter recognition state to the tight-binding state of RNA polymerase–DNA complexes occurs; (b) by variations in the rate of RNA chain initiation by tightly bound RNA polymerase–promoter complexes. The location of the known promoter strength mutations in the lactose promoter, both within and adjacent to the tight-binding region (Dickson *et al.*, 1975; Gralla, 1976) is consistent with all the above alternatives. However, it may be significant that the two known mutations in the tight-binding site increase promoter strength, while the two adjacent mutations decrease promoter strength.

4. *Elongation.* Bacterial RNA chain elongation, *in vivo*, proceeds at varying average rates for different transcription units: Messenger RNA chains grow 35–50 nucleotides per second (at 37°C), while ribosomal RNA chains can grow up to twice as fast, depending on the physiological state of the cell (for recent data on this subject and a review of prior work see Dennis and Bremer, 1974; Dougan and Glaser, 1974). *In vitro* experiments (Darlix and Fromageot, 1972; Dahlberg and Blattner, 1973; Maizels, 1973) suggest that RNA chain elongation is pulsatile and that there are unique hold-up points. Nevertheless, the nucleotide sequences and composition at these hold-up points near the beginning of the *lac* operon are diverse (Maizels, 1973) and offer no insight into the mechanism of saltatory chain elon-