REPLICATION OF VIRAL AND CELLULAR GENOMES

DEVELOPMENTS IN MOLECULAR VIROLOGY

Yechiel Becker, Series Editor
Julia Hadar, Managing Editor

Other books in the series:
Herpesvirus DNA, **1981**

REPLICATION OF VIRAL AND CELLULAR GENOMES

Molecular events at the origins of replication and biosynthesis of viral and cellular genomes

Edited by

Yechiel Becker
The Hebrew University of Jerusalem, Israel
Julia Hadar, Managing Editor

Martinus Nijhoff Publishing
a member of the Kluwer Academic Publishers Group
Boston/The Hague/Dordrecht/Lancaster

Distributors for North America:
Kluwer Academic Publishers
190 Old Derby Street
Hingham, MA 02043

Distributors Outside North America:
Kluwer Academic Publishers Group
Distribution Centre
P.O. Box 322
3300 AH Dordrecht
The Netherlands

The figure on the cover is from Tabor, S., et al., "Initiation of Bacteriophage T7 DNA Replication," in The Initiation of DNA Replication *edited by Don Roy and C. F. Fox. N.Y.: Academic Press, 1981. Reprinted with permission. This figure is also reproduced on page 187 of this book.*

Library of Congress Cataloging in Publication Data
Main entry under title:

Replication of viral and cellular genomes.

(Developments in molecular virology)
Includes bibliographical references and index.
1. DNA replication. 2. Viruses—Reproduction. 3. Deoxyribonucleic acid synthesis. 4. Ribonucleic acid synthesis. I. Becker, Yechiel. II. Series.
QH450.R46 1983 574.87'3282 83-11457

ISBN-13: 978-1-4613-3890-1 e-IBSN-13: 978-1-4613-3888-8
DOI: 10.1007/978-1-4613-3888-8

Softcover reprint of the hardcover 1st edition 1983

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording, or otherwise, without written permission of the publisher, Martinus Nijhoff Publishing, 190 Old Derby Street, Hingham, MA 02043.

CONTENTS

Contributing Authors xiii

Preface x

I
Replication of Circular Viral Genomes

1
Replication of ΦX174 DNA Reconstituted
with Purified Enzymes
J. Shlomai, K.-I. Arai, R. Low, J. Kobori, N. Arai,
S. Yasuda, L. Bertsch, and A. Kornberg 1

2
DNA Replication of Papovavarises: In Vivo Studies
M. Seidman and N.P. Salzman 29

3
SV40 Chromatin Replication In Vitro
A. Richter and B. Otto 53

4
Organization and Replication of Chromosome-Associated
Polyoma Virus RNA and Flanking Cellular Sequences in
Polyoma Transformed Rat Cells
H. Manor, A. Tomer, N. Baran, A. Neer, and
E. Mendelsohn 69

II
Biosynthesis of Cellular Genomes

5
Replication of *Escherichia coli* Chromosome
T. Okazaki 93

6
DNA Replication in Mammalian Cells
J.H. Taylor 115

7
Replication of Drosophila Mitochondrial DNA
D.R. Wolstenholme, J.M. Goddard, and C.M.-R Fauron 131

8
Replication of Mammalian Mitochondrial DNA
D.A. Clayton 149

III
Replication of Linear Viral DNA Genomes

9
Replication of Bacteriophage T7 DNA
C.C. Richardson 163

10
Poxvirus DNA Replication
J.A. Holowczak 205

11
Replication of Adenovirus DNA
J.S. Sussenbach and P.C. van der Vliet 237

12
Replication of Parvovirus DNA
L.A. Salzman, R. Mitra, and D. Brown 257

IV
Replication of Retrovirus Genomes: The DNA Stage

13
Organization and Structure of Retrovirus Genomes
I.M. Verma 275

V
Replication of Linear Viral RNA Genomes

14
Host Control of the Replication of Group I RNA Phages
H. Engelberg-Kulka and R. Schoulaker-Schwarz 315

15
The Tobacco Mosaic Virus Genome: Organization
and Gene Function
M. Zaitlin and P. Palukaitis 329

16
Biosynthesis of Influenza Virus RNA
Brian W.J. Mahy and Thomas Barrett 345

Index 367

Contributing Authors

Ken-Ichi Arai
Department of Biochemistry
Stanford University Medical School
Stanford, California 94305, USA

Naoko Arai
Department of Biochemistry
Stanford University Medical School
Stanford, California 94305, USA

Nava Baran
Department of Biology
Technion-Israel Institute of Technology
Haifa, ISRAEL

Thomas Barrett
Division of Virology
Department of Pathology
University of Cambridge,
Laboratories Block
Addenbrooke's Hospital
Hills Road
Cambridge, CB2 2QQ, U.K.

Leroy Bertsch
Department of Biochemistry
Stanford University Medical School
Stanford, California 94305, USA

Dale Brown
Laboratory of the Biology of Viruses
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland 20205, USA

David A. Clayton
Department of Pathology
Stanford University School of Medicine
Stanford, California 94305, USA

Hanna Engelberg-Kulka
Department of Molecular Biology
The Hebrew University-Hadassah Medical School
Jerusalem, ISRAEL

Christiane M.-R. Fauron
Department of Biology
University of Utah
Salt Lake City, Utah, 84112, USA

Judy M. Goddard
Department of Biology
University of Utah
Salt Lake City, Utah 84112, USA

John A. Holowczak
Department of Microbiology
University of Medicine and Dentistry of New Jersey
Rutgers Medical School
Piscataway, New Jersey 08854, USA

Ulrich Hubscher
Department of Biochemistry
Stanford University Medical School
Stanford, California 94305, USA

Joan Kobori
Department of Biochemistry
Stanford University Medical School
Stanford, California 94305, USA

Arthur Kornberg
Department of Biochemistry
Stanford University Medical School
Stanford, California 94035, USA

Robert Low
Department of Biochemistry
Stanford University Medical School
Stanford, California 94305, USA

Brian W.J. Mahy
Division of Virology
Department of Pathology
University of Cambridge,
Laboratories Block
Addenbrooke's Hospital
Hills Road
Cambridge, CB2 2QQ, U.K.

Haim Manor
Department of Biology
Technion-Israel Institute of Technology
Haifa, ISRAEL

Ella Mendelsohn
Department of Biology
Technion-Israel Institute of Technology
Haifa, ISRAEL

Rita Mitra
Laboratory of the Biology of Viruses
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland 20205, USA

Ada Neer
Department of Biology
Technion-Israel Institute of Technology
Haifa, ISRAEL

Tuneko Okazaki
Institute of Molecular Biology
Faculty of Science
Nagoya University
Nagoya 464, JAPAN

Bernd Otto
Biogen S.A.
Route de Troinex 3
1227 Carouge/Geneva, SWITZERLAND

Peter Palukaitis
Department of Plant Pathology
Cornell University
Ithaca, New York 14853, USA

Charles C. Richardson
Department of Biological Chemistry
Harvard Medical School
Boston, Massachusetts 02215, USA

Arndt Richter
Fachbereich Biologie
der Universität Konstanz
D-7750 Konstanz, GERMANY

Lois A. Salzman
Laboratory of the Biology of Viruses
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland 20205, USA

Norman P. Salzman
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland 20205, USA

Rachel Schoulaker-Schwarz
Department of Molecular Biology
The Hebrew University-Hadassah Medical School
Jerusalem, ISRAEL

Michael Seidman
National Institute of Allergy and Infectious Diseases
National Institutes of Health
Bethesda, Maryland 20205, USA

Joseph Shlomai
Department of Protozoology
The Hebrew University-Hadassah Medical School
Jerusalem, ISRAEL

J.S. Sussenbach
Laboratory for Physiological Chemistry
State University of Utrecht
Utrecht, THE NETHERLANDS

J. Herbert Taylor
Institute of Molecular Biophysics
Florida State University
Tallahassee, Florida 32306, USA

Aaron Tomer
Department of Biology
Technion-Israel Institute of Technology
Haifa, ISRAEL

P.C. van der Vliet
Laboratory for Physiological Chemistry
State University of Utrecht
Utrecht, THE NETHERLANDS

Inder M. Verma
Tumor Virology Laboratory
The Salk Institute
Post Office Box 85800
San Diego, California 92138, USA

David R. Wolstenholme
Department of Biology
University of Utah
Salt Lake City, Utah 84112, USA

Seichi Yasuda
Department of Biochemistry
Stanford University Medical School
Stanford, California 94305, USA

Milton Zaitlin
Department of Plant Pathology
Cornell University
Ithaca, New York 14853, USA

PREFACE

Biosynthesis of cellular and viral DNA and RNA has been a major topic in molecular biology and biochemistry. The studies by Arthur Kornberg and his colleagues on the in-vitro synthesis of DNA have opened new avenues to understanding the processes controlling the duplication of the genetic information encoded in the DNA and RNA of bacterial and mammalian cells. Viral nucleic acids are replicated in infected cells (bacterial, plant, and animal) by virus-coded enzymes with or without the involvement of proteins and enzymes coded by the host cells. The ability of the virus to replicate its genome within a relatively short period in the infected cell makes it an excellent biological tool for studying the molecular events in nucleic acid replication. Indeed, the identification of a number of virus-coded proteins that participate in the biosynthesis of ΦX174 and SV40 DNA has led to the construction of in-vitro systems for the study of nucleic acid biosynthesis. Similarly, studies on the replication of other phage, animal and plant viruses have provided an insight into the nucleic acid sequences from which DNA synthesis is initiated, as well as the proteins and enzymes that regulate the catalyse biosynthetic processes. Investigation of the molecular processes involved in the replication of cellular and mitochondrial genomes has gained momentum from the rapid developments in the analyses of viral nucleic acid biosynthesis.

This book is an authoritative summary of the original studies reported from key laboratories on the biosynthesis of viral DNA and RNA genomes under in-vitro and in-vivo conditions. Molecular events in the nucleic acid biosynthesis of bacteriophages, animal and plant viruses are described, as well as the synthesis of *E. coli* DNA, and cellular and mitochondrial DNA in mammalian cells and Drosophila. Analysis of the molecular events in the synthesis of bacterial and mammalian chromosomal DNA provides an insight into complicated enzymatic processes that are required for the duplication of the genetic information.

I wish to thank all the contributing authors for their fine contributions. I hope that this book will be of use to virologists, molecular biologists, biochemists and geneticists interested in DNA and RNA replication. My thanks to Dr. Julia Hadar for editing and indexing the manuscripts and to Mrs. Esther Herskovics for her excellent secretarial help.

Yechiel Becker
Jerusalem, February, 1983

REPLICATION OF VIRAL AND CELLULAR GENOMES

1

REPLICATION OF ϕX174 DNA RECONSTITUTED WITH PURIFIED ENZYMES

JOSEPH SHLOMAI*, KEN-ICHI ARAI, ROBERT LOW, JOAN KOBORI, NAOKO ARAI, SEICHI YASUDA, ULRICH HUBSCHER, LEROY BERTSCH, and ARTHUR KORNBERG

SUMMARY

Conversion of ϕX174 DNA to its duplex replicative form (SS→RF) is initiated by assembly of a multiprotein unit, the primosome, at a unique site on the chromosome, recognized by prepriming protein n'. At or near this site, the unique origin of ϕX complementary strand replication is located. The primosome and the prepriming intermediate that precedes it contain one molecule each of tightly bound dnaB protein, protein n', protein n, protein i and undetermined stoichiometric quantities of dnaC protein, primase and protein n". The role of dnaB protein as a "replication promotor" is to "engineer" the DNA structure to generate a signal for primase action. Processive translocation of the primosome from its assembly site, at or near the unique origin of replication, is in the 5'→3' direction of the DNA template, and is catalyzed by protein n' utilizing the energy of ATP (dATP) hydrolysis. Priming of synthesis of complementary strands on the nascent viral strand in the multiplication of RF (RF→RF) utilizes the same primosome which is almost completely conserved. The primosome associated with parental viral DNA in the initial SS→RF stage of ϕX174 DNA replication may function repeatedly in the initiation of complementary strands at the subsequent stage of RF→RF multiplication. A nucleoprotein complex functioning in RF replication has been isolated, which combines the priming, helicase, and chain elongation functions in a "repli-

---

*To whom reprints requests should be addressed. Present address: Department of Protozoology, The Hebrew University-Hadassah Medical School, Jerusalem, Israel.

Abbreviations are as follows: ϕX, bacteriophage ϕX174; SS DNA, single-stranded DNA; RF DNA, duplex replicative form of ϕX174 DNA; SSB, single-stranded DNA binding protein.

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

some"-like structure at the replication fork. Its possible role in coupling continuous DNA synthesis of the leading DNA strand and discontinuous synthesis on the lagging strand is discussed.

INTRODUCTION

Replication of bacteriophage øX174 in E. coli proceeds by three stages (1,2,3). In the first stage (SS→RF), the single-stranded viral (+) circle (SS) is converted into a duplex replicative form (RFI), by the action of the DNA replication enzymes of the host cell. In the second stage (RF→RF), multiple copies of RFI are generated, using the newly synthesized RFI as a template. In this and the subsequent stage two additional proteins are employed, the phage-encoded gene A protein and the host rep gene product (4,5). In the last stage (RF→SS), selective synthesis of viral (+) circles is carried out, after further production of RFI has ceased. Using the complementary (-) DNA strand of the RFI as a template, and in the presence of the phage gene products B,D,F,G and H, viral (+) circles are formed and encapsidated into øX174 virions (6,7). Relying almost entirely on the cellular machinery for its replication enzymes, the mechanisms employed by øX174 are pertinent to the process of the host cell chromosome replication. Therefore, the enzymology of E. coli replication proteins employed in øX DNA replication is a guide to their action in replication of the E. coli chromosome. In this contribution we describe the actions of E. coli replication enzymes in the discontinuous and continuous synthetic pathways in øX chromosome replication. The association of these enzymes into the physical and functional prepriming intermediate, priming complex (primosome), and the coordination of leading and lagging strand synthesis in a replisome-like complex is discussed.

Actions of replication proteins dnaB, dnaC, i, n, n' and n" precede the priming event on SSB-coated øX DNA.

Conversion of the single-stranded chromosomes of the three small bacteriophages M13, G4 and øX174 to their duplex replicative forms differ primarily in the mode of initiation of DNA replication. In all three systems elongation of DNA is catalyzed by DNA polymerase III holoenzyme (8,9). However, in each phage system different mechanisms are employed for providing the RNA primer prior to elongation by this enzyme (10-12). Coated with single-stranded DNA binding protein, M13 DNA can be primed directly by the syn-

thesis of a unique RNA transcript by RNA polymerase (13,14); G4 DNA is primed by DNA primase (15,16). However, the SSB-coated ϕX174 chromosome, although primed by the same DNA primase, requires a prepriming stage before the priming event can take place. Six host proteins, n, n', n", i, dnaC and dnaB act upon the SSB-coated ϕX174 DNA and activate it to form a prepriming intermediate complex (17). This activated form of the template promotes synthesis of multiple primers by primase when uncoupled from DNA replication (18,14). Formation of the intermediate complex was recognized by its being the rate limiting step of the SS→RF reaction, and due to its remarkable stability in the presence of ATP, its isolation could be carried out (17). However, it was not until an extensive purification of each of the priming and prepriming proteins was accomplished, that the exact composition of the prepriming intermediate and the priming complex (primosome) derived from it could be accurately determined. Four out of the six prepriming proteins, proteins n, n', n" and i, have not yet been associated with defined genetic loci. However, the other two proteins, the _E. coli_ dnaB and dnaC gene products, could be overproduced by cloning the corresponding _E. coli_ genes into high copy number plasmids (19). Amplification of dnaB (20) and dnaC (21) proteins, and the preparations of proteins i (22), n (23), and n' (24) on a large scale provided suitable amounts of

Table 1. Protein requirements for _in vitro_ conversion of phage single-stranded DNA to duplex RFII.
[Reprinted with permission from: Arai K. et al., Prog. Nuc. Acid Res. Mol. Biol. 26, 9-32 (1981).]

Stage	M13	G4	ϕX174
Prepriming	SSB	SSB	SSB protein i protein n protein n' protein n" dnaB protein dnaC protein
Priming	RNA polymerase	Primase	Primase
Elongation	DNA polymerase III holoenzyme	DNA polymerase III holoenzyme	DNA polymerase III holoenzyme

homogenous prepriming proteins for isotopic labeling, and for tracing their fate during the reaction (see Table 1 and Fig. 1).

The priming system has an intrinsic capacity to synthesize multiple primers on a single-stranded DNA template.

Synthesis of primer RNA by primase occurs on SSB-coated ϕX DNA only in the activated intermediate form, in which one molecule of dnaB protein is bound to the DNA template (17,18). This transcription, when uncoupled from DNA synthesis, is extensive, and multiple primers are made on each DNA circle. The single dnaB protein molecule bound to the DNA template participates in the *de novo* initiation of these oligonucleotide primers (18,14). The presumption was (14) that dnaB protein acts in the nucleoprotein complex as a "mobile promoter" which actively migrates along the DNA template providing signals for primase to produce many primer transcripts.

Upon extensive purification of each of the replication proteins and their use in the ϕX enzyme reconstitution replication reaction, an important feature of the priming reaction became apparent (25). In the absence of SSB, conversion of ϕX174 single-stranded DNA to its duplex form can be initiated simply by the joint action of only dnaB protein and primase. This simple priming system, uncoupled from DNA replication, catalyzes the synthesis of multiple, short primers (10 to 60 residues long) on any uncoated single-stranded phage DNA, and also on poly(dT). This priming activity is inhibited by coating the template with SSB. Thus, on uncoated single-stranded DNA, dnaB protein and primase act as a nonspecific general priming system, while specific initiations on the different single-stranded phage DNAs become evident only when the template is coated with single strand DNA binding protein.

The general priming system provides important clues to the mechanism of initiation of complementary strand replication. It demonstrates that the capacity to synthesize multiple short primers expresses an intrinsic property of the joint action of dnaB protein and primase. The multiple short primers are produced throughout the DNA template length, apparently at a number of preferred initiation points, as judged from the characteristic pattern of priming for each of the different DNA templates. Although the priming system is incapable of directing specific priming at the unique origins of replication when uncoated templates are used, particular regions of DNA are preferentially recognized by dnaB protein and primase (25).

Furthermore, a similar size distribution of short RNA primers (10 to 60 residues) was produced on uncoated ϕX DNA by the specific priming system (26). Inasmuch as the only components common to the general and specific priming systems are dnaB and primase, it may be inferred that these two components are responsible for site selection in priming along the DNA template (26). SSB and the prepriming proteins do not appear to participate significantly in this process. The coupled dnaB protein and primase, acting on single-stranded DNA templates and being inactivated by SSB, initiate primer synthesis primarily on single-stranded regions in the chromosome (27). This is in contrast to the action of RNA polymerase (28) and primase (29,30) which act primarily at base-paired regions which are destabilized by SSB (27).

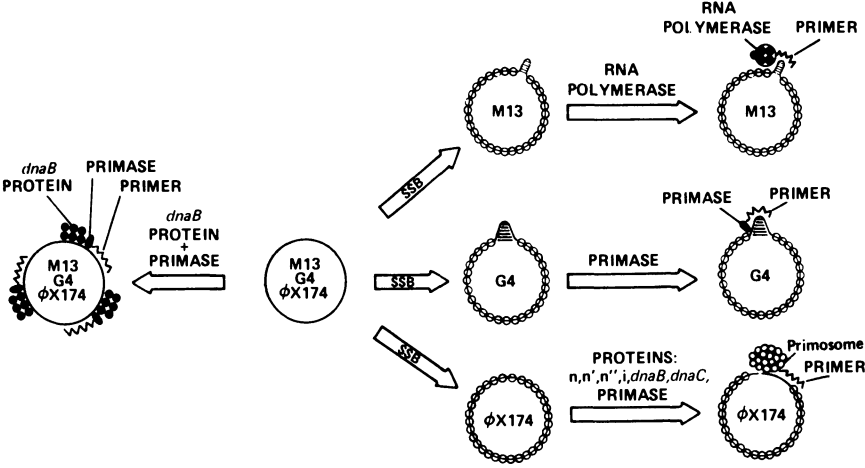


FIGURE 1. Illustration of nonspecific priming by dnaB protein and primase on uncoated DNA and the specific priming systems for M13, G4 and ϕX174 DNAs coated with SSB.
[Reprinted with permission from: Arai and Kornberg (1979) (reference 25)]

Allosteric effect of ATP facilitates the alteration of DNA structure by dnaB protein to generate the replication promoter.

The E. coli dnaB gene product, a hexameric protein of 290,000 daltons, was found to function in the replication of the bacterial chromosome (31-33), ϕX SS→RF (34,35,9,11) and RF→RF (36-38) replication and in plasmid replication in vitro (39-41). The enzyme, a DNA-dependent triphosphatase (42-46,20) was purified to homogeneity (45-47) and also recently crystallized (20). Its role as a replication promoter has been suggested using the specific priming system of ϕX DNA coated with SSB (43,14,18), and using the general priming system of uncoated DNA templates (25). From observations of the general priming system, a model was suggested (47) in which ATP functions as an allosteric effector in a binary complex with dnaB, to alter the structure of the DNA template in priming of replication (Fig. 2). According to this model, binding of one molecule of ATP or another ribonucleoside triphosphate at a single binding site on a protomer of the dnaB

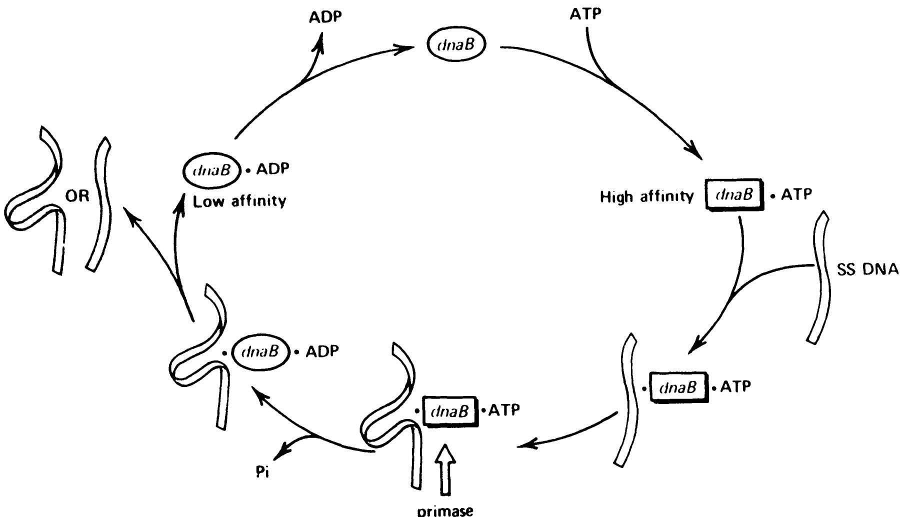


FIGURE 2. Reaction scheme for ATP-induced conformational changes of DNA and dnaB protein as a "replication promoter".
[Reprinted with permission from: Arai and Kornberg (1981) (reference 47)]

protein hexamer (20,44), is the initial event in the formation of a promoter site for priming. A dissociation constant (K_D) for the ATP·dnaB protein binary complex is about $1x10^{-5}$M, and other ribo- and deoxynucleoside triphosphates and ADP compete for the same site with similar affinities (47). The binary dnaB protein·ATP complex has a strong affinity to single-stranded DNA, leading to a ternary complex of DNA·dnaB protein·ATP. This ternary complex is stabilized when ATP is substituted with a nonhydrolyzable analog such as the adenylyl β,γ, imidodiphosphate [App(NH)p]. dnaC stabilizes the ϕX DNA·dnaB protein complex, which can also be stabilized by primase in the presence of ATP. In accord with recent genetic evidence (48), is the formation of a DNA·dnaB protein·primase complex which suggests direct interaction between dnaB protein and primase on the DNA template.

It was suggested (47) that a unique secondary structure is created in the single-stranded DNA by the action of dnaB protein, which is recognized by primase, and therefore dnaB protein can be regarded as a replication promoter for primase (20). What is the nature of the unique structure generated by dnaB protein in the DNA template? The effects of intercalating agents on the general priming system suggest that a secondary structure is created on single-stranded DNA by the dnaB protein·ATP complex. Ethidium bromide strongly inhibited RNA synthesis by dnaB protein and primase on both ϕX and poly(dT) DNA templates. Furthermore, poly(dT), which does not support an increase in fluorescence of ethidium, does so in the presence of dnaB protein and ATP.

Inasmuch as ethidium intercalating effects depend on secondary structure, these results suggest the generation or stabilization of a secondary structure in the single-stranded template. Thus, inhibition of general priming on poly(dT) by ethidium bromide, can be explained by the interaction of the intercalator with unique secondary structures created by the action of dnaB protein upon the single-stranded template. Based on the pattern of protection from nuclease digestion, the DNA template may be wound about the dnaB protein molecule (47) and the intrinsic property of the system to synthesize short primers may result from priming by primase only in the dnaB protein·DNA domain.

Hydrolysis of ATP is not required for the synthesis of multiple primers by the general priming system. Thus, in this system ATP serves as an allosteric effector of dnaB protein, which raises its affinity for the

template and facilitates the interactions described above. ATP hydrolysis results in the conversion of the dnaB protein•ATP complex into a dnaB protein•ADP complex with a low affinity for single—stranded DNA. This facilitates dissociation from DNA, and enables the initiation of another cycle of binding in a distributive model as described in Figure 2.

Stability of dnaB protein in the prepriming intermediate: binding of dnaB protein in a processive model.

Approximately one molecule of ^{3}H-labeled dnaB protein is bound to SSB-coated ϕX DNA when incubated with prepriming proteins n, n', n", i and dnaC (15,17,18,49). ATP was required for the binding of dnaB protein to SSB-coated ϕX DNA, and could not be substituted by ADP. Binding of dnaB protein to DNA in the specific priming system differs from its binding to DNA in the general priming system in its stability (Fig. 3) and template specificity (14,49).

The dnaB protein•DNA complex formed with primase and ATP in the general priming system is unstable and is rapidly exchanged with uncomplexed dnaB protein. (^{3}H)-dnaB protein•ϕX DNA complex formed in the presence of the prepriming proteins but without SSB was also unstable. In contrast, the ^{3}H-labeled dnaB protein•DNA complex formed with App(NH)p is very stable and is not displaced. Similarly, dnaB protein bound in the prepriming intermediate with SSB-coated DNA is remarkably stable kinetically. Only 20 percent of the bound (^{3}H)—dnaB protein was displaced by large excess of unbound dnaB protein in the reaction mixture within 30 min (49). Even the addition of primase and the four rNTPs does not result in the dissociation of dnaB protein. These observations indicate that dnaB protein in the replication intermediate is kinetically stable and does not exchange significantly during the synthesis of multiple primers. This is consistent with the interpretation that dnaB protein in the prepriming replication intermediate catalyzes the synthesis of these multiple primers processively. This processivity is thought to be the result of dnaB protein interaction with the other prepriming proteins in the presence of SSB, and is in contrast with its distributive mode of action in the general priming system.

In addition to its processivity, binding of dnaB protein in the specific priming system differs from its binding in the general system by its remarkable DNA template specificity. Complexing of ^{3}H-labeled dnaB

protein with SSB-coated DNA was specific for ϕX. No complex was formed with M13 DNA or G4 DNA; on SSB-coated ϕX DNA, the binding was absolutely dependent upon the other prepriming proteins n, n', n", i and dnaC. This specificity conferred by the prepriming system upon dnaB protein binding to the DNA template had been puzzling. That is, unlike phages G4 and M13, which were known to possess unique origins for complementary strand starts, both in vivo and in vitro studies had indicated multiple starts on ϕX DNA (18,50,51). This strict specificity of the prepriming system can now be explained by the specific recognition of a unique locus on the ϕX chromosome by prepriming protein n'.

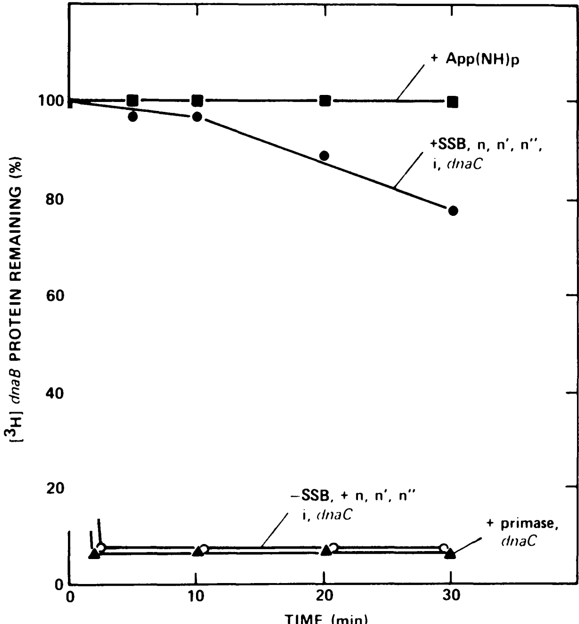


FIGURE 3. Exchangeability of [^{3}H]–dnaB protein associated with ϕX174 DNA.
The reaction mixture contained ϕX174 DNA, ATP, and other components as indicated. After incubation at 30°C for 20 min, 10 µg of unlabeled dnaB protein was added and the incubation continued. At indicated times, reaction mixtures were filtered through Bio-Gel A-5m columns and the ^{3}H-radioactivity associated with the DNA determined.
[Reprinted with permission from: Arai et al. (1981) (reference 49)]

Protein n' recognizes a specific sequence on ϕX DNA.

Purification of the proteins required for the in vitro conversion of SS→RF has resolved one of the fractions, called protein n (N-ethylmaleimide-sensitive) (52), into three distinct components designated proteins n, n', and n". All three proteins are required in prepriming of ϕX DNA replication (24). Protein n', a single polypeptide of 76,000 daltons has been recognized in three functional ways (53): (i) participation in the formation

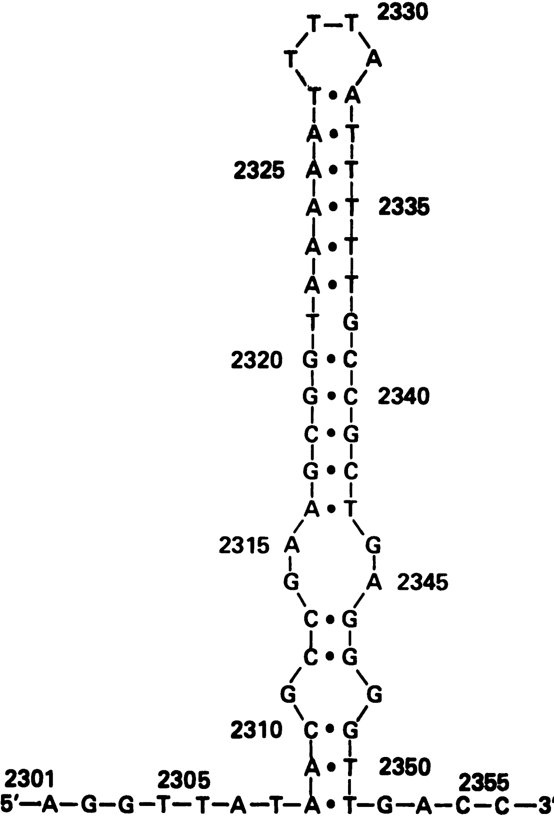


FIGURE 4. Possible secondary structure in recognition locus of protein n'. The proposed structure of base-paired region separated by two internal loops and a hairpin loop is based on a calculation of free energy contributions of base-paired regions and loops of the sequence within the 55-nucleotide fragment possessing the protein n' recognition site.
[Reprinted with permission from: Shlomai and Kornberg (1980) (reference 54)]

of a øX174 viral strand prepriming complex of replication; (ii) a sequence-specific, DNA-dependent ATPase activity; and (iii) an ATP-stimulated destabilization of SSB bound to øX DNA (53).

When limiting amounts of ^{3}H-labeled n' protein were used and prepriming intermediate formation was monitored by gel filtration, one molecule of n' was found associated with each øX DNA circle replicated. As will be discussed, this single n' protein molecule plays an important role in "fueling" the mobility of the priming complex and displacing the SSB during its translocation. Its strict requirement at the prepriming stage however, stems from its specific recognition of a unique locus in the øX chromosome. Unlike most other DNA-dependent ATPases, protein n' is strictly specific in its requirements for a DNA effector (24,53,54). Among many single-stranded and duplex DNAs tested, only øX DNA was active, and its effector activity was not abolished by coating with SSB (53). The specific DNA sequence recognized by n' was determined after fragmentation of øX DNA by restriction endonuclease HaeIII, followed by digestion with exonuclease VII (54). A 55-nucleotide fragment isolated from a øX viral DNA strand specifically supported ATPase activity of protein n'. Within this fragment there is a 44-nucleotide sequence with a potential for forming a stable hairpin structure (Fig. 4). This sequence is contained within an untranslated region of the øX chromosome between structural genes F and G. As described in the following section, assembly of the priming complex is carried out at or near this n' recognition site.

Priming of øX DNA replication starts at a unique site.

Unlike phages M13 (56,28) and G4 (29,55,50), which possess unique origins for priming of complementary strand replication, priming on øX DNA occurs at multiple sites, as indicated by both in vivo (50) and in vitro studies (14,18). However, the strict specificity of the prepriming system for øX DNA (14,25) and the specific recognition by protein n' of a 55-nucleotide sequence in øX DNA, located at an analogous intergenic region with the G4 origin (54), suggested a unique origin for øX complementary strand replication. If SSB-coated øX DNA has a unique origin, then the dnaB protein, serving as the replication promoter for primase should migrate from this origin along the DNA template. Since primers arise at many sites during the continuous and rapid movement of the replication promoter, the true origin of replication may be obscured. Thus, to detect and locate

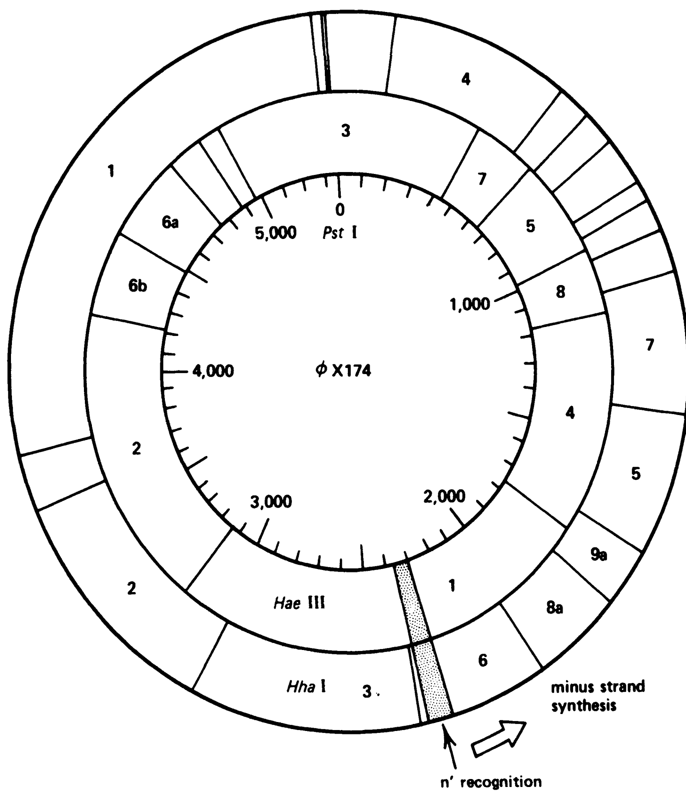


FIGURE 5. Physical map of ϕX174 DNA. The n' recognition site is located at nucleotides 2301-2354 (54). The open arrow indicates the DNA elongation direction.
[Reprinted with permission from: Arai and Kornberg (1981) (reference 58)

such a unique start the mobility of the priming system had to be restricted. ϕX DNA was digested by restriction endonucleases HaeIII or HhaI (which cleave single-stranded ϕX DNA at specific sites, identical to the cleavage sites on the duplex RF form (57)), and the resulting fragments were used as templates for complementary DNA strand synthesis in the presence of SSB (58). That a unique primed start of DNA replication can be substituted by a fragment of a phage circle, was first determined with G4 DNA. It was found that a restriction fragment (HhaI, 6) which contains the unique origin of G4 complementary DNA replication (59), specifically directed DNA synthesis dependent on primase (58). With SSB-coated ϕX DNA only the 1353-nucleotide HaeIII Z1 fragment and the 300-nucleotide HhaI fragment 6 (Fig-

ure 5) could specifically direct DNA synthesis (58). These two ØX DNA fragments were activated as templates for DNA replication by the prepriming proteins and primase. Each of these latter two fragments contains the 55-nucleotide locus recognized by protein n'. Thus, when movement of the priming system is restrained, only fragments containing the protein n' recognition site were effective in sustaining prepriming, priming and elongation reactions. Analysis of the DNA products synthesized on these fragments, suggests that primers are made at or near the hairpin within the 55-nucleotide protein n' recognition locus, where presumably assembly of the priming complex takes place (58). Thus it seems that the strict specificity of the prepriming system for ØX DNA coated with SSB, is derived from the capacity of protein n' to recognize and bind a specific site in ØX DNA. The series of reactions that culminates in the priming of ØX complementary strand synthesis must therefore be initiated by protein n' at that specific locus, which serves as the origin of DNA replication. The remarkable feature of the unique origin for ØX complementary strand synthesis is that this site is not necessarily defined as the site in which the first primer is transcribed. This origin may be defined as the site for assembly of the priming system (primosome) from which migration along the DNA template is started (58).

Mobility of priming system is in a direction opposite chain synthesis.

Since SSB-coated ØX DNA has a single origin for complementary strand replication, at or near the protein n' recognition site, the priming system must migrate processively along the DNA template to accomplish primer synthesis at multiple sites. Two models for the polarity of this migration could be considered: migration in a direction opposite to the direction of primer synthesis and DNA elongation (antielongation), or migration of the priming system in the same direction as that of primer transcription and DNA chain elongation. The following observations have revealed (58) that polarity of the priming system migration is uniquely in the antielongation direction (5'→3' direction of the template):

(A) If mobility of the priming system were in the 3'→5' polarity of the template then DNA products shorter than 600 nucleotides must be expected when HaeIII fragment Z1 was used or shorter than 250 nucleotides when HhaI fragment 6 was used as a template in an SS→RF reaction (58) (Figure 5). Instead, the most abundant size of products were 1200-1300 nucleotides

when Z1 fragment was used, and longer than 250 nucleotides with HhaI fragment 6, suggesting chain initiations at the 3' side of the protein n' recognition site (Figure 5).

(B) DNA products synthesized on the HaeIII fragment Z1 were found, upon their digestion with HhaI, to contain not only the fragments 6, 8a, and 9a, but also fragment 3 which is located at the 3' side of the protein n' recognition site (namely, in an antielongation direction relative to this site).

(C) In accord with the latter finding is the observation that the DNA products synthesized on the Z1 fragment hybridized not only to the HhaI fragment 6, but also to fragment 3 at the 3' side of the n' protein recognition site.

(D) Substitution of App(NH)p for ATP during prepriming intermediate formation minimizes movement of the replication promoter thereby synchronizing the priming process (58). Abundancy of primers on the intact ϕX circle could be monitored after the addition of ATP and the components for priming and elongation. When analyzed by restriction endonuclease digestion it was found that the abundancy of primers on each fragment decreased in the order: Z1>Z2>Z3>Z4, consistent with the prediction of migration of the priming system in the direction Z1→Z2→Z3→Z4, (5'→3' direction of the template) (58).

These data, indicating polarity of the priming system migration uniquely in the antielongation direction, are consistent with the role presumed for a mobile replication promoter, entailing synchronous and linked movement at the replication fork.

Association of the prepriming proteins in the primosome of DNA replication.

Like the prepriming replication intermediate which precedes its assembly, the primosome contains one molecule each of tightly bound dnaB protein and n'. It is suggested that assembly of the prepriming proteins and primase into a primosome take place at or near protein n' recognition site (54). At or near this locus transcription of the first primer is also carried out (58). As illustrated schematically in Figure 6, initial reaction in the assembly of the primosome is the recognition of the unique origin by protein n'. It seems that protein n and maybe also n" are involved in this initial stage of the assembly of the primosome. Protein n (23) a dimer of 12,000 dalton subunits, is essential for the formation of the

prepriming intermediate, and is bound to ϕX DNA only when coated with SSB. Its binding at a level of one monomer per circle, required for primosome formation, is stimulated by the presence of bound n' and is retained and conserved in the primosome complex (23). Protein n', bound to its recognition site, may guide catalytic levels of protein n to the n' recognition site. Protein n is thought to promote DNA-protein or protein-protein interactions subsequent to protein n' binding, and thus would function in the placement of proteins i, n", dnaC and dnaB at the assembly site. In this initial stage, designated the recognition stage, at which proteins n', n and n" act, ATP is not required, and a complex of ϕX DNA·protein intermediate could be isolated in its absence (22,26). At the subsequent stage, termed the assembly, ATP but not its hydrolysis is required; binding of dnaB protein at this stage (which is dependent on proteins i and dnaC), can be supported by the nonhydrolyzable analog App(NH)p. dnaC protein, a single polypeptide of 23,000 daltons (21), forms a tight complex with the dnaB protein in the presence of ATP (60). A functional dnaB-dnaC protein complex was recently isolated by DEAE-cellulose chromatography, and was found to contain six dnaC protein monomers per dnaB protein hexamer (21). It was suggested that dnaB protein is delivered in this complex form to the SSB-coated ϕX DNA. dnaC protein is not firmly held by the prepriming intermediate (preprimosome), but upon addition of primase and DNA polymerase III holoenzyme, in the subsequent stages, it becomes more stably bound in the primosome (21). Either protein i or dnaC protein or both, recognize the structure of the nucleoprotein complex formed in the first stage, interact with it, and trigger the sequence of reactions which result in the binding of dnaB protein in the complex (22). Protein i, a trimeric protein of 61,000 daltons is bound at ~1.7 molecules per replication intermediate. However it is not yet clear whether it constitutes an integral part of the primosome since it has not yet been found in the isolated primosome after DNA replication has been terminated. It is presumed that at this stage of the assembly at the origin, all the prepriming proteins are included in the complex, which is ready in this form and in the presence of ATP for the subsequent migration stage. In this stage the primosome migrates processively in a direction opposite to DNA elongation. Since n' is the only primosomal enzyme capable of catalyzing the hydrolysis of ribo- as well as deoxy ATP (53), and since dATP hydrolysis can support the mobility of the primosome (26) it is deduced that protein n' catalyzes the translocation of

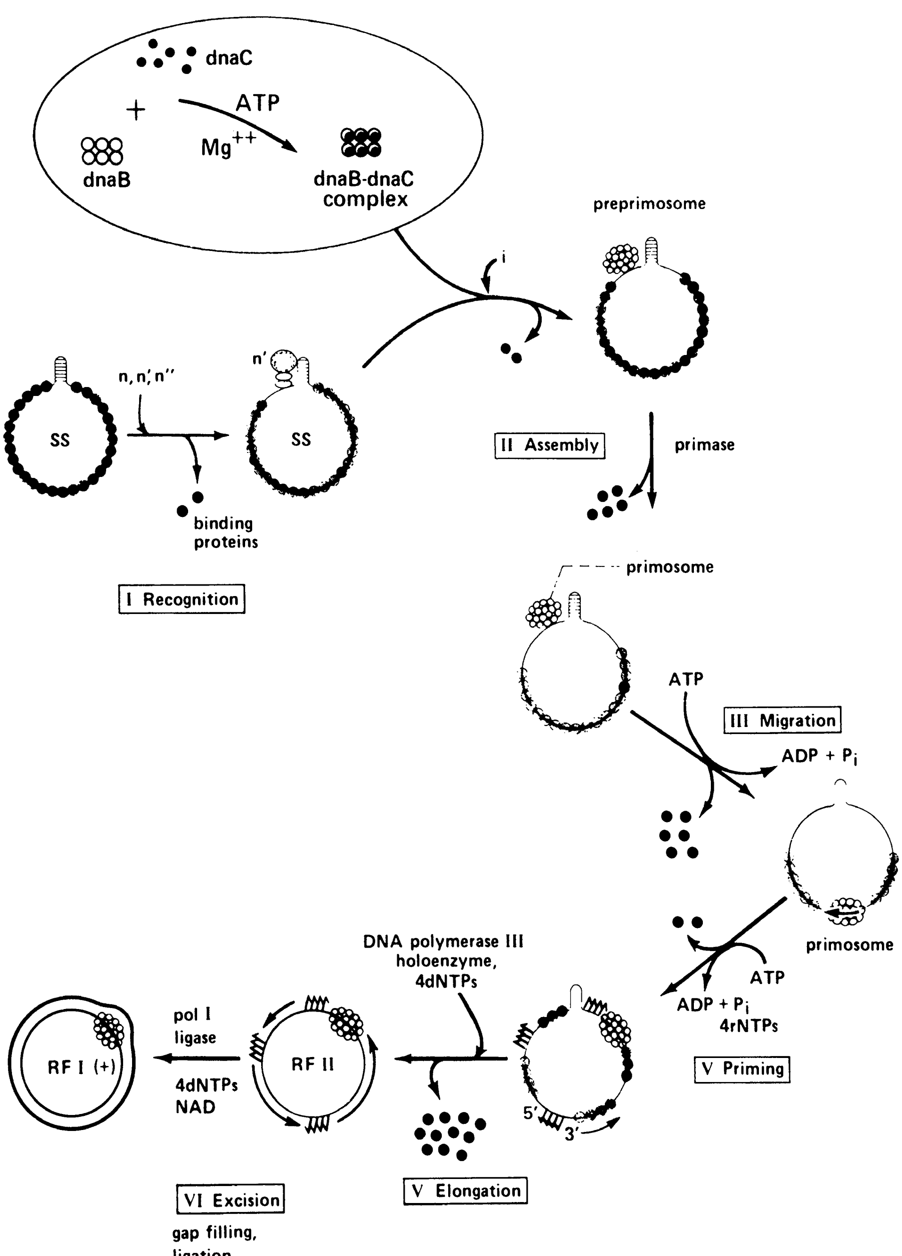

dnaC
+
ATP
Mg++
dnaB
dnaB-dnaC
complex
preprimosome
i
n, n′, n″
n′
SS
SS
binding
proteins
I Recognition
II Assembly
primase
primosome
ATP
III Migration
ADP + Pi
primosome
ATP
ADP + Pi
4rNTPs
V Priming
DNA polymerase III
holoenzyme,
4dNTPs
pol I
ligase
4dNTPs
NAD
RF I (+)
RF II
5′
3′
VI Excision
gap filling,
ligation
V Elongation

FIGURE 6. Scheme for assembly and migration of the primosome and the stepwise displacement of SSB in the ϕX174 SS→RF replication. [Reprinted with permission from: Kobori and Kornberg (reference 21)]

the primosome along the template. It is not yet clear how n' ATPase activity is carried out during primosome migration away from n' recognition site. During the assembly and migration stages, a quantitative destabilization of bound SSB takes place. Protein n' is the only primosomal protein capable of destabilizing SSB bound to ϕX DNA. The reaction is ATP (dATP)-stimulated and about one third of the SSB molecules bound to a ϕX circle were released by a level of n' needed for the formation of the prepriming intermediate (53). With the assembly of the complex, about half of the SSB molecules bound to DNA are released, whether coupled to RNA priming or not (58,54). It is presumed that it is the n' action of displacing SSB that removes the inhibitory effect of SSB on the replication promoter activity of dnaB protein (26,49). At this priming stage, dnaB protein in the primosome functions as the "replication promoter" to provide recognition signals for primase to synthesize multiple short primers (Fig. 7). The observation

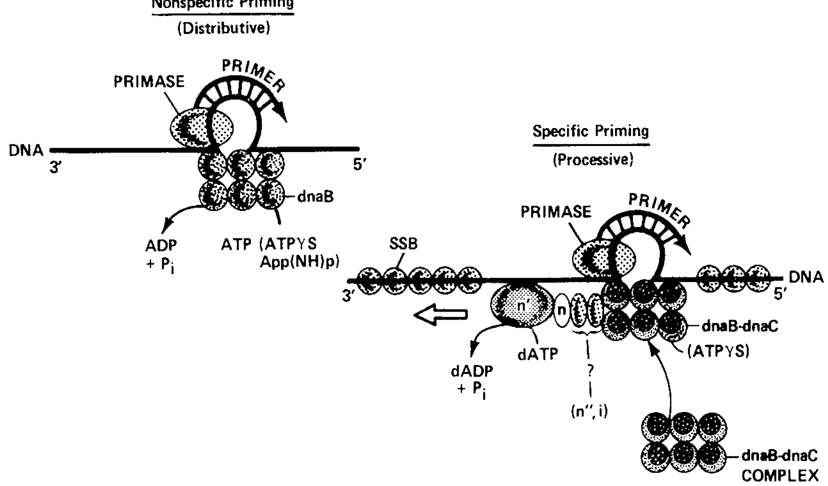


FIGURE 7. Nonspecific priming by dnaB protein and primase and specific priming by the primosome. [Reprinted with permission from: Arai et al. (1981) (reference 26)]

that ^{3}H-labeled dnaB protein in the primosome was not released during this stage is in support of the processive model for primosome mobility as was discussed in the preceding sections. The 3'-OH terminus of the primer formed is available for elongation by DNA polymerase III holoenzyme, provided that ATP is present. If ATP is substituted with a nonhydrolyzable analog and dATP is present, priming by the primosome is carried out (52). However, under these conditions the primer termini formed will not be elongated. Thus the role of ATP hydrolysis by dnaB protein in the primosome is to displace dnaB protein and primase from the newly synthesized primer terminus in order to make it accessible to the active site of DNA polymerase III haloenzyme (47,49). Hydrolysis of ATP by dnaB protein in the primosome, unlike in the general priming system, does not result in its release from the DNA template. Action of dnaB protein in the primosome remains processive due to its linkage to the other primosomal proteins. DNA polymerase III holyenzyme extends but does not excise the RNA primers (13,18,49). Excision and gap-filling activities are carried out by DNA polymerase I to produce RFII duplexes, followed by ligation by DNA ligase, to produce covalently sealed RFI.

Multiplication of ϕX duplex replicative form is carried out by coupling of continuous and discontinuous synthetic pathways.

Crude, cell-free preparations for replication of RFI (61,62) have been fractionated into two separate systems (63,64): the RF→SS, and the RF→RF. The RF→SS requires the participation of ϕX-encoded gene A protein, host rep protein, SSB, and DNA polymerase III holoenzyme to produce multiple copies of covalently closed, single-stranded circles from RFI. Replication is initiated by gene A protein, which creates a nick in the viral strand at the origin of RF replication (65-67). Gene A protein binds covalently to the 5'phosphoryl end of the nick (69). Rep protein recognizes gene A protein at the origin, binds to it, and catalyzes an ATP-dependent separation of the DNA strands to generate a fork for replication. When coupled to replication, a "looped" intermediate is formed (64,67,69). It was suggested (65,81) that ATP binding induces a conformational change of rep protein in the complex prior to the melting of a nucleotide base pair by the tightly bound gene A protein•rep protein•ADP complex. ATP hydrolysis by rep protein (of two ATP molecules for every base pair melted) (72, 73,70), promotes the displacement of the gene A protein•rep protein•ADP

complex from the single-stranded region. As suggested (70,71), the gene A protein•rep protein•ATP complex is regenerated and is then translocated to a base-paired region. Repetition of this sequence enables a single rep protein molecule to migrate processively and to separate the strands in advance of the replication fork. Binding of SSB to the separated strands prevents their reannealing and enables use of the exposed circular template by DNA polymerase III holoenzyme, while the 3'-OH terminus at the origin is covalently extended. When the origin sequence is regenerated gene A protein nicks it again and then ligates to produce a unit-length viral circle.

The synthesis of viral strand circles from RF in the cell-free system reconstituted with the gene A protein-rep protein system (63), and the capacity of the synthetic circles isolated from this reaction to serve as templates for RF synthesis by purified proteins (74), suggested that coupling of the RF→SS system with the SS→RF system could account for multiplication of RF *in vivo*. However, although replication of RF had been observed with crude enzyme fractions (61,62), net quantities of RF had not been synthesized with such impure preparation, nor could much be learned about the molecular details of the reaction. It was not until an extensive purification of the enzyme preparations, that a system synthesizing net RF could be reconstituted (38). Multiplication of RF has been reconstituted by coupling the RF→SS system (gene A protein, rep protein, SSB, and holoenzyme) with SS→RF system (protein n', n, n", i, dnaC, dnaB, primase, SSB and holoenzyme). The principal product of the reaction are RFII molecules (at about 10 duplex circles per input duplex, in 40 min). Omission of any one of the proteins required for priming the SS→RF reaction reduces DNA synthesis in RF multiplication to a level of 13-35 percent, and results in the synthesis of only single-stranded viral circles. RNA polymerase could not substitute for the prepriming and priming proteins, and the omission of gene A protein, rep protein, or SSB completely abolished DNA synthesis in the coupled system. When DNA polymerase I and DNA ligase were included together with the other replication proteins, gaps were filled, RNA primer excised and the strands were sealed to form closed duplex RFI DNA (71). In addition to the RF, more complex forms, including ϕX-unit length duplex circles with multigenome-length tails accumulated during RF multiplication (38,71). The synthesis of these multigenome forms is especially pronounced at late stages of the reaction and at low levels of SSB.

Analysis of the location of RNA primers in the various products and

intermediates showed that RNA primers were associated with RFII and multi-genome RF when DNA polymerase I and DNA ligase were absent. Location of ^{32}P-labeled primers as determined by annealing to viral DNA on paper, annealing to poly(UG) in isopycnic banding, and by competitive annealing to viral DNA in solution, was found to be exclusively on the complementary strands (71). The lack of primers associated with viral strands made it clear that their synthesis is continuous in this rolling circle system. The wide distribution of primers around the genome and average frequency of primed starts of complementary strands of one per genome are in accord with in vivo observation (75,76) and are similar to priming of the complementary strand in the SS→RF reaction. Thus, the evidence from the purified in vitro system suggests, in contrast to earlier proposals (82-84), that only the complementary strand is synthesized discontinuously, and that synthesis of the viral strand is largely or completely continuous (71).

A functional primosome is conserved at the replication fork of duplex ϕX DNA.

A nearly intact primosome is retained with synthetic RF produced in the in vitro reconstituted SS→RF reaction, when isolated by procedures that preserve protein components bound to DNA (80). As judged by physical and functional studies, the primosome associated with synthetic RF, after its covalent closure by DNA ligase, contains the prepriming proteins except for proteins i (22) and n (23). Protein n is not required for continuous RF multiplication after the primosome has been assembled. Protein i however, is needed for continuous function of the primosome and has to be supplied (80). Thus synthetic RFI, unlike the phenol-extracted supercoiled RFI from infected cells that requires the primosomal proteins, undergoes repeated rounds of RF duplication with the addition of only protein i. It was also found, that unlike protein-free RFI which can be used as a template for RF→SS replication only after supercoiling, the synthetic RFI can be used without further supercoiling. In fact, it can promote a rate of RF→SS production 28-fold faster than that obtained using cell-extracted, protein-free RFI. Retention of the primosome may lead to a DNA topology that promotes efficient cleavage by gene A protein and thereby abolishes the need for gyrase (80). The findings (81) that during ϕX infection it is the parental RF, rather than the progeny RF, that undergoes duplication can also be explained by conservation of the primosome on the infecting viral

DNA strand of the parental RF.

Replication intermediates in RF replication have provided important clues to the mechanism of RF replication and to the employment of the conserved primosome in this process (82). The multigenome forms accumulated during RF replication have revealed that synthesis of complementary strands can be initiated on a viral strand template before it is completed. This was also supported by the remarkable efficiency of RF multiplication (38). The primosome, beginning at the protein n' site may migrate processively along the viral strand, coupled to the movement of the replication fork (26) generated by the action of gene A and rep protein (68,71). Coupling of the priming system movement with unwinding of the duplex appears to take place in other replication systems in E. coli. The T7-encoded gene 4 protein was shown to be a helicase (83) as well as primase (84), and the T4-encoded gene 41 and 61 proteins appear to be analogous to dnaB and primase (85). In ϕX RF multiplication, a gene A•RF•primosome complex could be used catalytically as an intermediate, avoiding the rate-limiting prepriming stage for primosome assembly. The isolation of such an RF replication intermediate strongly supports such a model (82). The functional DNA•protein complex isolated by gel filtration during RF replication was found to contain the tightly bound proteins n', dnaB, primase, rep protein, DNA polymerase III holoenzyme, SSB and most probably n and n". This complex is capable of continuing DNA synthesis upon addition of proteins i and dnaC. These studies also showed the existence of an intermediate complex that contains only n', rep, SSB and DNA polymerase III holoenzyme, which sustains the formation of SS DNA and which can convert SS DNA to RFII upon addition of proteins n, n", i, dnaC, dnaB and primase (82).

These studies have led to the interpretation of the pathways of RF replication as described in the scheme in Fig. 8. The rolling circle intermediate containing proteins gene A and rep, DNA polymerase III holoenzyme and SSB, identical to the replication intermediate for the RF→SS system (71) is first formed (complex I). This complex associates with protein n' to form the intermediate of RF multiplication (complex II). A viral circle containing protein n' is produced by complex II and is converted to RFII by the association of the other primosomal proteins, to generate complex III. This latter complex containing an assembled primosome is predominant in the late stages of the RF multiplication reaction, and can catalyze the initiation of complementary strand synthesis before the viral

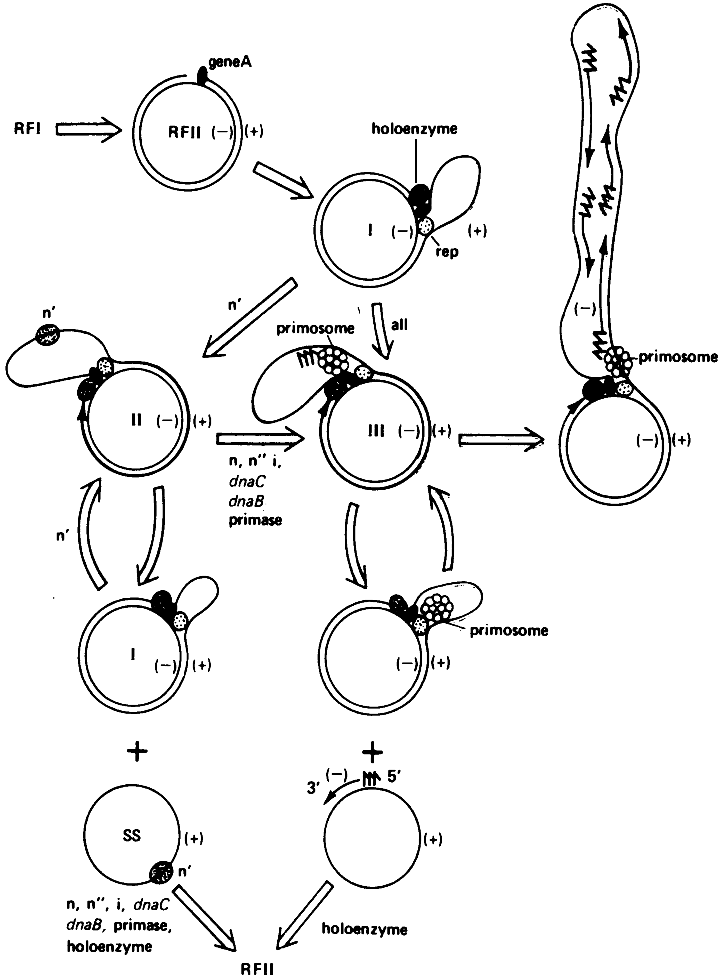


FIGURE 8. Scheme for RF multiplication in vitro. Complex I is a rolling circle intermediate containing the proteins: gene A, rep, DNA polymerase III holoenzyme, and SSB. Complex II is complex I associated with proteins n, n", dnaB and primase. [Reprinted with permission from: Arai et al. (1981) (reference 82)]

circle has been completed. In such a complex, the numerous components needed for both continuous and discontinuous DNA synthesis may be organized into a major "replisome"-like complex (86).

Coupling of continuous and discontinuous synthetic pathways in ϕX DNA replication as a model for semidiscontinuous replication of the E. coli chromosome.

A possible structure for the replication fork of the E. coli chromosome emerges from these studies of ϕX DNA replication (Fig. 9). The leading strand is synthesized by a relatively continuous mechanism and the lagging strand discontinuously. The primosome, assembled at or near the origin of replication, migrates processively, coupled to the movement of the replication fork. Multiple short primers, synthesized by primase at

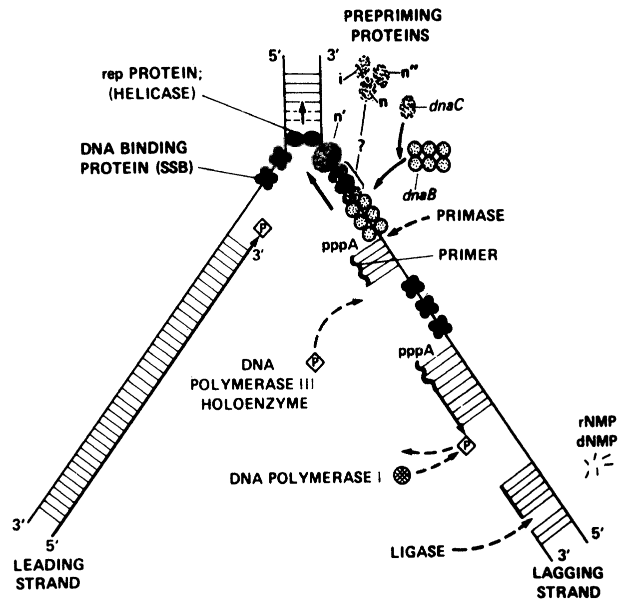


FIGURE 9. Role of the primosome at the replication fork. The primosome, assembled at or near the DNA replication origin, migrates processively on the lagging strand coupled to movement of the replication fork.
[Reprinted with permission from: Arai et al. (1981) (reference 26)]

numerous preferred regions, are elongated into Okazaki fragments by DNA polymerase III holoenzyme. Excision of primers and gap filling by DNA polymerase I and ligation of the nascent DNA fragments result in the formation of a continuous lagging DNA strand. It seems plausible that the primosome facilitates the unwinding of the duplex at the replication fork for efficient coupling of the lagging and leading strands replication. The association of a helicase (rep protein), the primosome, and DNA polymerase III holoenzyme in a functional and physical complex, observed in ϕX RF multiplication, suggests that these components interact with each other to form a replisome-like complex (86) at the replication fork. Such a complex functioning at the E. coli replication fork may link the processive migration of the primosome on the lagging strand to the movement of the replication fork and to the elongation by DNA polymerase III holoenzyme on the leading strand in the semidiscontinuous mode of replication of the E. coli chromosome.

REFERENCES

1. Sinsheimer, R.L. Prog. Nucleic Acid Res. Mol. Biol. 8, 115-169, 1968.
2. Denhardt, D.T. Comp. Virol. 7, 1-104, 1977.
3. Denhardt, D.T. CRC Crit. Rev. Microbiol. 4, 161-223, 1975.
4. Tessman, E.S. J. Mol. Biol. 17, 218-236, 1966.
5. Denhardt, D.T., Iwaya, M., and Larison, L.L. Virology 49, 486-496, 1972.
6. Fujisawa, H., and Hayashi, M. J. Virol. 19, 416-424, 1976.
7. Tessman, E.S., and Peterson, P.K. J. Virol. 20, 400-412, 1976.
8. McHenry, C., and Kornberg, A. J. Biol. Chem. 252, 6478-6484, 1977.
9. Meyer, R.R., Shlomai, J.; Kobori, J., Bates, D.L., Rowen, L., McMacken, R., Ueda, K., and Kornberg, A. Cold Spring Harbor Symp. Quant. Biol. 43, 289-293, 1978.
10. Kornberg, A. Cold Spring Harbor Symp. Quant. Biol. 43, 1-9, 1978.
11. Wickner, S. Ann. Rev. Biochem. 47, 1163-1191, 1978.
12. Kornberg, A. in RNA Polymerase, eds. Losick, R., and Chamberlin, M. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.) pp. 331-352, 1976.
13. Geider, K. and Kornberg, A. J. Biol. Chem. 249, 3999-4005, 1974.
14. McMacken R., and Kornberg, A. J. Biol. Chem. 253, 3313-3319, 1978.
15. Bouche, J.-P., Zechel, K., and Kornberg, A. J. Biol. Chem. 250, 5995-6001, 1975.
16. Zechel, K., Bouche, J.-P., and Kornberg, A. J. Biol. Chem. 250, 4684-4689, 1975.
17. Weiner, J.H., McMacken, R., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 73, 752-756, 1976.
18. McMacken, R., Ueda, K., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 74, 4190-4194, 1977.

19. Uhlin, B.E., Molin, S., Gustafsson, P., and Nordstrom, K. Gene 6, 91-106, 1979.
20. Arai, K., Yasuda, S., and Kornberg, A. J. Biol. Chem. 256, 5247-5252, 1981.
21. Kobori, J., and Kornberg, A. J. Biol. Chem., in press, 1982.
22. Arai, K., McMacken, R., Yasuda, S., and Kornberg, A. J. Biol. Chem. 256, 5247-5252, 1981.
23. Low, R., Shlomai, J., and Kornberg, A. J. Biol. Chem. 257, 6242-6250, 1982.
24. Shlomai, J., and Kornberg, A. J. Biol. Chem. 255, 6789-6793, 1980.
25. Arai, K., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 76, 4308-4312, 1979.
26. Arai, K., Low, R., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 78, 707-711, 1981.
27. Arai, K., and Kornberg, A. J. Biol. Chem. 256, 5267-5272, 1980.
28. Geider, K., Beck, E., and Schaller, H. Proc. Natl. Acad. Sci. U.S.A. 76, 645-649, 1978.
29. Bouche, J.-P., Rowen, L., and Kornberg, A. J. Biol. Chem. 253, 765-769, 1978.
30. Hourcade, D., and Dressler, D. Proc. Natl. Acad. Sci. U.S.A. 75, 1652-1656, 1978.
31. Gross, J.D. Curr. Top. Microbiol. Immunol. 57, 39-74, 1972.
32. Zyskind, J.W., and Smith, D.W. J. Bacteriol. 129, 1476-1486, 1977.
33. Wechsler, J.A., and Gross, J.D. Mol. Gen. Genet. 113, 273-284, 1971.
34. Schekman, R., Wickner, W., Westergaard, O., Brutlag, D., Geider, K., Bertsch, L.L., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 61, 2691-2695, 1972.
35. Wickner, R.B., Wright, M., Wickner, S., and Hurwitz, J. Proc. Natl. Acad. Sci. U.S.A. 69, 3233-3237, 1972.
36. Eisenberg, S., Scott, J., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 73, 3151-3155, 1976.
37. Sumida-Yasumoto, C., Ikeda, J.-E., Benz, E., Marians, K.J., Vicuna, R., Surgue, S., Ziprusky S.L., and Hurwitz, J. Cold Spring Harbor Symp. Quant. Biol. 43, 311-329, 1978.
38. Arai, K., Arai, N., Shlomai, J., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 77, 3322-3326, 1980.
39. Tomizawa, J., and Selzer, G. Annu. Rev. Biochem. 48, 999-1034, 1979.
40. Staudenbauer, W.L. Mol. Gen. Genet. 162, 243-249, 1978.
41. Conrad, S.E., and Campbell, J.L. Nucleic Acid Res. 6, 3289-3304, 1979.
42. Wright, M., Wickner, S., and Hurwitz, J. Proc. Natl. Acad. Sci. U.S.A. 70, 3120-3124, 1973.
43. Ueda, K., McMacken, R., and Kornberg, A. J. Biol. Chem. 253, 261-269, 1978.
44. Reha-Krantz, L.J., and Hurwitz, J. J. Biol. Chem. 253, 4043-4050, 1978.
45. Reha-Krantz, L.J., and Hurwitz, J. J. Biol. Chem. 253, 4051-4057, 1978.
46. Lanka, E., Edelbluth, C., Schlicht, M., and Schuster, H. J. Biol. Chem. 253, 5847-5851, 1978.
47. Arai, K., and Kornberg, A. J. Biol. Chem. 256, 5260-5266, 1981.
48. Masukata, H., and Ogawa, T. Mol. Gen. Genet. 180, 531-538, 1980.
49. Arai, K., Low, R., Kobori, J., Shlomai, J., and Kornberg, A. J. Biol. Chem. 256, 5273-5280, 1981.
50. Sims, J., and Dressler, D. Proc. Natl. Acad. Sci. U.S.A. 75, 3094-3098, 1972.
51. Martin, D.M., and Godson, G.N. J. Mol. Biol. 117, 321-325, 1977.

52. Schekman, R., Weiner, J.H., Weiner, A., and Kornberg, A. J. Biol. Chem. 250, 5859-5865, 1975.
53. Shlomai, J., and Kornberg, A. J. Biol. Chem. 255, 6794-6798, 1980.
54. Shlomai, J., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 77, 799-803, 1980.
55. Fiddes, J.C., Barrell, B.G., and Godson, G.N. Proc. Natl. Acad. Sci. U.S.A. 75, 1081-1085.
56. Tabak, H.F., Griffith, J., Geider, K., Schaller, H., and Kornberg, A. J. Biol. Chem. 249, 3049-3054, 1974.
57. Blakesley, R.W., Dogson, J.B., Nes, I.F., and Wells, R.D. J. Biol. Chem. 252, 7300-7306, 1977.
58. Arai, K., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 78, 69-73, 1981.
59. Godson, G.N., Barell, B.G., Staden, R., and Fiddes, J.C. Nature 276, 236-247, 1978.
60. Widkner, S. and Hurwitz, J. Proc. Natl. Acad. Sci. U.S.A. 72, 921-925, 1975.
61. Eisenberg, S., Scott, J.F., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 73, 1594-1597, 1976.
62. Sumida-Yasumoto, C., Yudelevich, A., and Hurwitz, J. Proc. Natl. Acad. Sci. U.S.A. 73, 1887-1891, 1976.
63. Eisenberg, S., Scott, J.F., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 73, 3151-3155, 1976.
64. Eisenberg, S., Griffith, J., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 74, 3198-3202, 1977.
65. Henry, T.J., and Knippers, R. Proc. Natl. Acad. Sci. U.S.A. 71, 1549-1553, 1974.
66. Ikeda, J., Yudelevich, A., and Hurwitz, J. Proc. Natl. Acad. Sci. U.S.A. 73, 2669-2673, 1976.
67. Longeveld, S.A., von Mansfeld, A.D.M., Baas, P.D., Jansz, H.S., von Arkel, G.A., and Weisbeek, P.J. Nature (London) 271, 417-420, 1978.
68. Eisenberg, S., Scott, J.F., and Kornberg, A. Cold Spring Harbor Sym. Quant. Biol. 43, 295-302, 1978.
69. Eisenberg, S., and Kornberg, A. J. Biol. Chem. 254, 5328-5332, 1979.
70. Arai, N., Arai, K., and Kornberg, A. J. Biol. Chem. 256, 5287-5293, 1981.
71. Arai, N., and Kornberg, A. J. Biol. Chem. 256, 5294-5298, 1981.
72. Scott, J.F., Eisenberg, S., Bertsch, L.L., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 74, 193-197, 1977.
73. Kornberg, A., Scott, J.F., and Bertsch, L.L. J. Biol. Chem. 253, 3298-3304, 1978.
74. Shlomai, J., Polder, L., Arai, K., and Kornberg, A. J. Biol. Chem. 256, 5233-5238, 1981.
75. Eisenberg, S., and Denhardt, D.T. Proc. Natl. Acad. Sci. U.S.A. 71, 984-988, 1974.
76. Keegstra, W., Baas, P.D., and Jansz, H.S. J. Mol. Biol. 135, 69-89, 1979.
77. Mathes, M., and Denhardt, D.T. J. Mol. Biol. 136, 45-63, 1980.
78. Machida, Y., Okazaki, T., and Okazaki, R. Proc. Tatl. Acad. Sci. U.S.A. 74, 2776-2779, 1977.
79. Baas, P.D., Teetstra, W.R., and Jansz, H.S. J. Mol. Biol. 125, 167-185, 1978.
80. Low, R.L., Arai, K., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 78, 1436-1440, 1981.
81. Yarsus, M., and Sinsheimer, R. J. Virol. 1, 135-144, 1967.

82. Arai, N., Polder, L., Arai, K., and Kornberg, A. J. Biol. Chem. 256, 5239-5246, 1981.
83. Kolodner, R., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 74, 1525-1529, 1977.
84. Scherzinger, E., Lanka, E., Morelli, G., Seiffert, D., and Yuki, A. Eur. J. Biochem. 72, 543-558, 1977.
85. Liu, C.C., Burke, R.L., Hibner, U., Barry, J., and Alberts, B. Cold Spring Harbor Symp. Quant. Biol. 43, 469-487, 1978.
86. Kornberg, A. DNA Replication, W.H. Freeman and Co., San Francisco, 1980.

DNA REPLICATION OF PAPOVAVIRUSES: IN VIVO STUDIES

Michael Seidman, Ph.D. and Norman P. Salzman, Ph.D.

INTRODUCTION

The replication of papovaviruses has been the subject of recent review articles (1, 2, 3). In this discussion, we will consider initiation, chain elongation, termination and replication in permissive cells and nonpermissive cells, with an emphasis on studies concerned with events in vivo. The reviews by DePamphilis and his colleagues (3, 4, 5) are recommended for a more detailed report of data from the in vitro replication systems, an area which receives little attention here. We have not discussed the replication of papovaviruses as nucleoproteins, as that will be considered elsewhere in this volume.

INITIATION

The mechanisms for initiation of replication of SV40 and polyoma DNA have been areas of active investigation since molecular studies on viral replication began. In early work, it was found that replication began at a fixed origin and was bi-directional (6, 7, 8, 9, 10). It has also been known for some time that T antigen, the product of the A gene, is required for the initiation of replication (11, 12, 13). This protein has been shown to bind to DNA sequences at the origin of replication (14, 15, 16, 17). Recent work has been directed towards the delineation of the precise nucleotides in the origin region required for replication, the nucleotides necessary for T antigen binding, and the relationship between T antigen binding and replication.

The experiments of Danna and Nathans (7) and Fareed et al., (18) demonstrated that the replication of SV40 DNA starts at about 0.67 units. Similar experiments established the origin of polyoma DNA to be at map position 71 (10, 8). Although these regions of the respective genomes clearly function as origins in the majority of replicating molecules, about one percent of the replicating molecules in SV40-infected cells initiate bidirectional synthesis at an origin about 2,400 bp from the normal origin (19). Because the initiation process regulates the viral replication cycle, special attention

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

T I　　　　T II

15　　17　　15　　27

5' GCTTTTTGCAAAAGCCTAGGCCTCCAAAAAAGCCTCCTCACTACTTCTGGAATAGCTCAGAGGCCGAGGCGGC
5180　5190　5200　5210　5220　5230　5240　1

Early

ori

T III

17 AT　　"21"　　21

CTCGGCCTCTGCATAAATAAAAAAAATTAGTCAGCCATGGGGCGGAGAATGGGCGGAACTGGGCGGAGTTAGGGG
10　20　30　40　50　60　70

21　　72

CGGGATGGGCGGAGTTAGGGGCGGGACTATGGTTGCTGACTAATTGAGATGCATGCTTTGCATACTTCTGCCTGCT
80　90　100　110　120　130　140　150

72

GGGGAGCCTGGGGACTTTCCACACCTGGTTGCTGACTAATTGAGATGCATGCTTTGCATACTTCTGCCTGCTGGGG
160　170　180　190　200　210　220

AGCCTGGGGACTTTCCACACC 3'
230　240　250

Late

FIGURE 1. Sequence of SV40 origin of replication. The DNA sequence shown is the E (minus) strand. The sequence also corresponds in both composition and direction to late mRNA transcripts. The following elements in the sequences are marked: a) 15 bp symmetrical sequence; b) 17 bp true palindrome; c) 15 bp palindrome; d) 27 bp perfect palindrome containing the BglI site (.......); e) 17 bp AT stretch; f) 21 bp repeats; g) 72 bp repeats; h) minimum sequence for origin function (ori). The T antigen binding sites described by Tjian (47) are also indicated. It should be noted that studies of T antigen binding to sequence at the origin are in progress in many labs and there is some disagreement in the published literature. In particular, Tegtmeyer *et al*. (17) do not find evidence for a third binding site.

was given to the description and interpretation of the origin sequences of these viruses (20, 21, 22, 23, 24) and subsequently, the total sequence of SV40, polyoma, and BKV virus was determined (25, 26, 27, 28). In SV40, the origin region sequence contains two convenient landmark restriction enzyme sites, HindIII (bp 5171) and BglI. In the numbering scheme employed here, the central nucleotide in the BglI recognition site is assigned nucleotide

number 1 (2). There are a number of interesting features in the sequence of the region (see figure). Starting near the HindIII site on the early gene side of the origin and moving from the early to the late side, there is a 15 bp sequence with a two-fold rotational axis of symmetry (#5173-5187) followed by a 17 bp true palindrome with only one "incorrect" base (#5193-5209). Then there is a 15 bp palindrome (#5213-5227) and a 27 nucleotide sequence which contains the BglI site and is a perfect palindrome (#5230-13). A 17 bp stretch of AT appears (15-30) flanked on either side by GC-rich sequences. A series of repeat units are then found beginning with three 21 bp repeats (starting with #39-58), the first of which contains a two base "insert." Adjacent to these repeats are the two tandem repeats of 72 bp (107-250). Another feature of interest within the region are six copies of the sequence GGGCGGPuPu (29, 28). The presence of the palindromes allows the origin region to be drawn in the form of hairpin loop structures and this has prompted discussion about the interaction of regulatory proteins with structures such as these, as much as with specific sequences (20). The origin sequences of polyoma and BK virus display similar configurational possibilities (21, 22).

Although the appearance of symmetrical sequences with attendant alternate structures is provocative, it is not clear at this time what the significance of all these sequences might be. However, not all of them are necessary for a functional origin of replication. This conclusion was derived initially from two types of experiments which established the boundaries of the origin sequence and laid the foundation for more precise mapping of the region. One approach was the study of variant viral genomes which appear after serial passage and which have lost large amounts of the viral DNA but necessarily retain a functional origin of replication (30, 31, 32). The other approach, particularly with SV40, was the generation of viable mutants with deletions near the origin taking advantage of convenient restriction enzyme sites in the region (33, 34, 35). These studies set the origin between nucleotide numbers 5202 and 34. Recently, further deletion mutant construction and analysis has brought the boundary on the early side in to #5208 while the late side boundary is now at #29 (36). Thus, of the several unusual sequence elements described above, only the second 15 nucleotide and the 27 nucleotide palindromes, and the AT rich region are required for replication. The contribution of the repeating sequences which lie outside the late side boundary has been analyzed by Bergsma *et al.*, (37). They measured the replication of plasmids, carrying the origin region and appropriate deletions, in the Cos 1 cells of Gluzman (38) which

supply a functional T antigen. The 72 bp repeats were nonessential for replication although they are clearly involved in the regulation of early gene transcription (39). The 21 bp repeat elements, although not absolutely required for replication, appeared to enhance replication in a dose dependent fashion. Plasmids with two or three of the 21 bp elements replicated more efficiently than those with one unit, which in turn was better than the construct with none.

These advances in our understanding of the limits of the origin have been paralleled by an increasing sophistication in the description of the sequences involved in binding T antigen. The possibility that T antigen might interact directly with the origin has been tested by direct binding studies in vitro with T antigen and T antigen-related proteins (14, 40, 15, 41, 42, 17, 43). Initial experiments with partially purified preparations gave mixed results. Specific binding to the origin region of SV40 DNA was observed by electron microscopy (44). However, in filter binding assays, both origin and non-origin containing restriction fragments were bound, including a fragment with a site corresponding to the minor replication origin identified by Martin and Setlow (19, 45). A major advance was made when a T antigen-related protein, the D2 protein, became available (46, 14). This protein is produced by an Adeno-SV40 hybrid virus ($Ad2^{+}D2$) grown in HeLa cells and, while somewhat larger than authentic T antigen, it carries out many of the same functions as authentic T antigen (47). Relatively large amounts of the protein facilitated its purification and allowed precise binding experiments to be done. In DNAse I protection experiments, three origin region binding sites were demonstrated (14). The locations of these sites were confirmed by methylation protection experiments which identified the purines which make contact with the protein (47). Of the three sites, site I (5184-5209) is bound with the greatest affinity, followed by site II (5231-13) which is bound at higher protein concentrations while site III is protected at still higher protein concentrations (14) (36-61). Three binding sites were also shown with the T antigen from SV40 cells (15) and wild type T antigen produced in HeLa cells (41). Recently, a fourth minor site for the D2 protein has been located immmediately to the left of site I (48).

The interaction of T antigen and related proteins to these sites has been further characterized by binding studies with a series of deletion mutants constructed in several laboratories. The observation that T antigen binding to site II is 10-20 fold less efficient in mutants with deletions in site I than

to that site in wild type DNA suggested that the binding to the three sites is cooperative (49, 48). It is known that the protein will form oligomeric structures (42) and one possible explanation for site II and site III binding would be polymerization of the protein with nonspecific binding to sequences adjacent to site I. This possibility has been discounted in a thoughtful experiment by Tenen et al (48) who showed that specific sequences were required for binding to site II. These experiments, which demonstrate cooperative binding of T antigen and the location of the binding sites, were performed with T antigen (49) or the related D2 protein, obtained from nonpermissive cells. Experiments with T antigen obtained from lytically infected permissive cells give a somewhat different picture of the origin binding activity of the protein. DNA footprinting studies from Nathans laboratory (36) and fragment binding assays described by Tegtmeyer and his colleagues (17) do not support the conclusion that T antigen binding to the sites is cooperative. Furthermore, in DNAse protection experiments, evidence for site III binding was not found (17). In addition, there is some disagreement as to the precise position of the binding sites. These discrepancies may be due to differences between T antigens in permissive and nonpermissive cells or variability in experimental procedure (protein concentration, etc.)

Regardless of the final outcome of the origin binding experiments, the available results coupled with the origin boundary data permit some conclusions to be drawn regarding the relationship between T antigen binding and replication. The most obvious conclusion is that binding to site I and III are not required for replication. It has been suggested that site I in particular is involved in the regulation of early gene transcription (50, 51, 41, 52). Even though most of site I is not required for replication, deletions which eliminate binding of T antigen to the site tend to reduce the efficiency of replication of those mutants. The influence of point mutations in site II on replication efficiency has been studied by the laboratory of Nathans using the elegant site directed mutagenesis procedure developed by Shortle and Nathans (52a). They generated a collection of viable point mutants in site II and found that most of the mutants replicated more poorly than wild type, although one class of mutants showed a greater replication rate (53). Although T antigen binding studies with these mutants have not been described, two lines of evidence suggest that there is a direct relationship between the effects of the mutations on replication and interaction with T antigen. Many of the base changes in site II which reduce replication are replacements of guanines

identified as making contact with T antigen in the methylation protection experiments (47, 54). Direct evidence of replication-related interaction in vivo has come from the isolation of pseudorevertants of the site II point mutants. These mutants have second site alterations in the sequence of the T antigen gene itself, produce an altered T antigen and restore the replication activity of the original site II mutant (53). The studies of T antigen binding and replication suggest that although necessary, T antigen binding is probably not sufficient for replication. Mutants with deletions which eliminate all of site I and the sequence between sites I and II, but leave site II intact, do not replicate (49). At the time of this writing, the only known enzymatic function assigned to T antigen is an ATPase activity (50). It seems likely that this protein will interact with other cell proteins and enzymes at the origin, but such interactions remain to be described.

Several groups have described monomer, dimer and tetramer forms of T antigen (55, 56, 42, 57). There is some disagreement as to which form binds the origin of replication. Bradley et al. (55) and Myers et al. (42) suggest that it is the tetramer while Gidoni et al. (57) favor the monomer. When bound to the origin, the protein appears to be a tetramer (42). The protein is phosphorylated (58) although the functional consequences of this modification are unclear . A fraction of the T antigen population binds host coded "nonviral" tumor antigen of 48,000 and 55,500 daltons (59, 60, 61, 62). The role of the complex in replication is not understood at this time. Scheller et al. (63) found that T antigen, when complexed to the 53K host protein, would bind the origin fragment. On the other hand, the complex forms in non-permissive cells in which there is no viral DNA replication.

Heterogeneity in the T antigen population has also been shown through the use of monoclonal antibodies directed against T antigen (64, 65). Gurney and her colleagues isolated three monoclonal antibodies and found that one of them immunoprecipitated less than 10% of the T antigen. The other antibodies precipitated the majority of the T antigen molecules. Recently, Scheller et al. (63) found that the monoclonal antibody of Gurney, which precipitated a minor class of T antigen, precipitated the majority of the origin fragment binding activity. They suggested that a post-translational modification of T antigen might convert an inactive molecule into one which could bind specifically to the origin. If this is true, then such a modification may be necessary for a replication-competent T antigen.

In the case of polyoma virus, the origin region has not been as well defined. There is a 43 nucleotide sequence in the polyoma origin which shows a strong homology to the corresponding sequence in SV40 (27, 66). This conserved region includes the AT-rich region and the sequences homologous to the second T antigen binding site. On the early gene side of this conserved sequence, the limits of the origin have been partially defined by deletion mapping (67, 68, 69, 70, 71). Deletions which begin about 40 bases from the conserved sequence have a negative effect on replication while those about 60 bases away do not effect replication in a cis-acting fashion. Thus it appears that sequences outside the conserved sequence are necessary for efficient replication (70). The binding of the polyoma T antigen has also been studied by deletion mapping. A sequence of 16 nucleotides, 19 nucleotides from the early side of the conserved region, appears to be a high affinity binding site for T antigen (72). These studies are substantially more difficult than those with SV40 T antigen because much smaller amounts of the polyoma antigen are available.

One of the best known characteristics of papovavirus replication is its bidirectionality. The perfect twofold rotational symmetry of the 27 bp palindrome, virtually all of which lies in site II, prompted the analysis of the direction of replication in mutants which had lost the symmetry of the sequence (54). Replication was still bidirectional indicating that symmetry of the site is not necessary for bidirectional replication. Further information on this question and on the nature of the first daughter strands, which start at the origin, has been provided by the recent work of Hay and DePamphilis (73). They found that synthesis of the first daughter molecules in the early gene direction began at a number of sites in the second and third T antigen binding sites, with two predominant starts in site II. Start sites for the leading strand going into the late gene region were found within and to the early side of the first T antigen binding site. In other words, the initial replication bubble appears to be a small D loop (74). These authors propose that replication begins in the early gene direction followed by the start of the opposite strand as the template for the synthesis is exposed. Since the late side direction starts are in a region which can be deleted, it seems likely that the bidirectional character of papovavirus replication is not related to a specific sequence. Furthermore, there must be some additional feature of the origin region which assigns the early side direction of the daughter strands which start within the 27 bp palindrome. This information may be contributed by the sequences on either side of the T antigen binding site.

Another important conclusion comes from this study. There are many start sites for synthesis in both directions. It seems likely that the papovaviruses, and perhaps eukaryotic replicons, display a certain flexibility in this regard. The advantage of such a strategy for survival of the replicon is obvious.

CHAIN ELONGATION - OKAZAKI FRAGMENT FORMATION

The DNA polymerase α of mammalian cells has been shown to be responsible for the replication of SV40 DNA (75, 76, 77) and polyoma DNA (78). This enzyme has been purified and studied extensively by several groups (79, 80, 81, 82). In recent years, it has been shown, using a variety of templates and primers that the enzyme functions in a quasi-processive fashion. Polymerization of small chains is followed by dissociation of enzyme from the template and growing chain (83, 81, 84, 85). In a study with a homopolymer template and a ribo-oligonucleotide primer, it was found that the enzyme had an increasing probability of termination once six nucleotides had been added (86). Most of the new chains were 8-12 nucleotides long. Interestingly, when helix destabilizing protein HD.1 (from mouse) was added, the number of product molecules with less than 12 nucleotides was decreased and the peak of the modal distribution of molecules was shifted to 21 residues. These experiments were done under conditions in which reinitiation did not occur. Results from studies with a native DNA template (ϕX174) are consistent with the homopolymer experiments (87) and suggest that the tendency for chain termination is an inherent property of the enzyme, rather than being dictated by specific sequences. Additional factors apparently reduce the probability of termination (86) but do not eliminate it. For example, in experiments with the α DNA polymerase from Drosophila, the addition of single strand binding protein increased the average size chain from 14 to 284 nucleotides (88). Unless the enzyme's tendency for termination is completely suppressed by the entire replication complex, there will always be a probability of termination of a growing chain. If the template is on the leading side of the replication fork, then synthesis will proceed either by reinitiation with the nascent chain as the primer, or by _de novo_ synthesis of a new ribonucleotide primer and the start of a new chain.

Depending on replication factors and the efficiency of reinitiation events, replication on the leading side of the replication fork could be continuous in the sense that only one growing chain would be detected at any one time and this chain would extend from the fork back to the origin. On the

retrograde side of the fork, however, synthesis must necessarily be discontinuous. The notion of discontinuous DNA replication was advanced by Okazaki and has been demonstrated in both prokaryotic and eukaryotic systems (for review, see Ogawa and Okazaki, 89). Okazaki fragments have been traditionally identified as small daughter chains which incorporate label in brief pulses of radioactive nucleotide precursors in vivo. They are initiated by the formation of short RNA primers which are synthesized by enzymes called primases, and the characterization of the RNA primers has been the subject of many reports (89).

Okazaki fragments were first observed on replicating SV40 DNA in the studies of Fareed and Salzman (6). They characterized small molecules of 100-200 nucleotide length that were associated with replicating molecules at all stages of replication. The production of these fragments during polyoma DNA replication has been extensively studied by Reichard and his colleages (90, 91, 92, 93) while DePamphlis and his colleagues (3) have compiled a complete description of the fragments in the replication of SV40. The structure and metabolism of these molecules have been reviewed by DePamphilis and Wassarman (3) and will not be discussed here. From the studies of many labs, it is clear that the viral Okazaki fragments are initiated with short RNA primer sequences about ten bases long. The primer sequences are random, indicating that there is no highly specific initiation site for the fragment synthesis. Recently, a large number of "preferred initiation sites" have been identified suggesting that initiation of Okazaki fragments may not be completely random with respect to sequence. Nonetheless, the requirement for specific sequences is not stringent (94).

Okazaki fragment synthesis, as noted above, was proposed as a solution to the problem of retrograde strand synthesis. One of the most persistent questions about these molecules over the years has been that of the proportion of fragments copied from one side of the fork or the other. This problem has been addressed in stages as technical advances have permitted more precise definition of Okazaki fragments. From the discussion of the properties of DNA polymerase α, it is clear that if the enzyme, or enzyme complex, terminates while copying the leading side template strand, then further synthesis will proceed either by reinitiation with the nascent chain as a primer, or by de novo initiation by a primase of a new chain, unlinked to the nascent chain. If the new chain terminates after 100-200 nucleotides, and has not been ligated to the adjacent long nascent chain, an Okazaki fragment will have been formed

and will be recognized as such. Thus, the formation of an Okazaki fragment on the leading side of the fork is one of the products of a competition between several possible reactions (termination vs. continued elongation, reinitiation vs. de novo initiation). The other product is simply a longer nascent chain. The persistence of a newly formed fragment will depend on the efficiency of the primer excision, gap filling and ligation operation. A change in the balance of factors involved could easily shift the apparent concentration of fragments on the leading side.

In the early experiments, replicating SV40 molecules were isolated and the 4S fragments purified from them. These fragments were then tested for self-complementary sequences (95). A high proportion (70-90%) of the fragments self-annealed and the authors suggested that Okazaki fragments were synthesized on both sides of the replication fork. Other groups working with in vitro replication systems found both low (±20%) (96, 97) and high (±60%) (92, 91) proportions able to self-anneal. The self-annealing experiments were criticized because of the unavoidable inclusion of small cellular DNA fragments and because the Okazaki fragments were identified simply as small pulse-labeled chains. In the next series of experiments, the selection of small chains was continued, but they were hybridized against separated strands of SV40 DNA restriction fragments located on one side or the other of the origin. In experiments of this type, several groups have found with fragments labeled in vivo (98) and in vitro (99, 100) that the hybridization is asymmetric, with a 4-5 fold greater hybridization to the templates from the retrograde side of the fork. The realization that the misincorporation of dUMP into DNA was followed by repair-related excision and the generation of transient, short DNA fragments (101, 102, 103) provided another mechanism for the production of small DNA fragments. Although these repair-related fragments probably had not been a significant factor in the previous experiments, they did provoke a general dissatisfaction with the definition of Okazaki fragments as short, pulse-labeled chains. In recent work, the molecules chosen for analysis have been required to show signs of covalent linkage to RNA primers, at the expense, of course, of those molecules whose primers have been excised. Methods for demonstrating the linkage of RNA to Okazaki size fragments have been reviewed by Ogawa and Okazaki (89). Narkhammer-Meuth and colleagues have recently studied the hybridization of tritiated fragments pulse labeled in vivo to separated strands (104). They found a 2.5 fold enhancement of hybridization to the retrograde strand template. However, when the fragments were end labeled

with ^{32}P, the hybridization of the end label to the two separated strands was virtually identical. In the ^{3}H-label experiment, the hybridization of the mass average of the fragments is measured, while with the ^{32}P-end label, it is the number average that is monitored. These authors suggested that the number of molecules synthesized from either side of the fork was approximately equal, but that the chains from the lagging side were longer. In a companion paper, Narkhammer-Meuth et al., (105), performed a similar experiment with fragments synthesized in vitro and obtained similar results. It should be noted that the fragment size problem had been discussed previously by Hunter et al. (99), who concluded that size disparity could not explain their measurement of 4-5 fold enhancement of Okazaki fragments synthesized from retrograde templates. Furthermore, the DePamphilis group, using the unmasking procedure of Okazaki et al. (106) in which only chains with RNA primers are labeled, again found strong template asymmetry in the Okazaki fragment population (73).

Undoubtedly, the last word on this subject has yet to be heard. Given the view of DNA synthesis on the leading side of the fork presented above, it seems likely that the results of any given study may indeed be true and the variability of results may reflect the plasticity of the replication machinery which may have several alternative mechanisms available. The survival advantage of such versatility is readily apparent. The more fundamental questions of how Okazaki fragments are synthesized, where they are synthesized, and why they form only on initiated DNA remain to be answered. The isolation and characterization by several groups of enzymes with the properties of an initiator RNA primase would appear to be a promising approach to these problems (107, 108, 109, 110).

TERMINATION

Replication of DNA by the Cairns mode is complete when the progeny DNA exists as complete double-stranded molecules with no topological relationship between the original template strands. Termination of replication requires the polymerization of the last nucleotides of the daughter strands, ligation of the 5' and 3' ends of the new strand to one another to form a covalently closed molecule, and segregation of the interlocked progeny circles. Although these are necessary events, the order in which they take place is by no means fixed and the specific sequence of events may vary in different systems.

Termination has been analyzed in several systems, including bacterial plasmids, phage, mammalian mitochondrial DNA, eukaryotic chromosomal DNA and,

of course, papova viruses. In most studies, molecules whose structures suggest a role in the late stages of replication have been isolated. Generally, pulse–chase data supply the strongest argument for identifying the various molecules as genuine replicative intermediates.

During the replication of penicillinase plasmids in Staphylococcus aureus, there is a termination intermediate with the structure of a catenated mixed dimer with one open circular molecule interlocked with one covalently closed superhelical molecule. The conversion of this to mature covalently closed circular monomers is a rate–limiting step that requires 3-4 minutes. (Based on the rate of replication of the bacterial chromosome, the plasmid would be expected to replicate in about 30 seconds). Pulse-chase data support the identification of these molecules as replicative intermediates (111, 112). Similar results were obtained by other workers with the R6K plasmid in E. coli (113). In the initial phase of the growth cycle of phage lambda (λ), circular parental molecules are replicated bidirectionally in the Cairns mode from a fixed origin (114). As in the plasmid systems, a mixed catenated dimer accumulates as a late replicative intermediate (115). Label in these structures could be chased into monomer circles. Furthermore, the mixed catenanes appeared when both recombination–deficient host and phage were used making it unlikely that they were the product of recombinational events. It seems likely that in these three different systems that the mixed catenane is a genuine late replicative intermediate, the result of complete synthesis and ligation of one daughter strand and incomplete synthesis or ligation of the other, all without separation of the two daughter duplex circles. At the time of the description, it was not clear how they might dissociate to monomers, and some thought was given to mechanisms of dissociation (116, 115). The discovery and characterization of topoisomerases offers an apparent solution to virtually any topological problem the daughter molecules might encounter (117) although the direct involvement of these enzymes in the resolution process has yet to be demonstrated.

Mitochondrial DNA replication, in contrast to the above, is unidirectional and daughter strand synthesis is asynchronous. Nonetheless, daughter molecule segregation must occur and does so to yield one molecule whose daughter strand synthesis is essentially complete. The other molecule has a large gap about one third the genome in size in the daughter strand (118, 119). The gapped molecule is filled in and converted to a covalently closed monomer circle. Catenated molecules have been observed and characterized by several groups

(120, 121). Using BrdUr density labeling, Flory and Vinograd (121) convincingly showed that the quarter heavy molecules found in cesium chloride gradients were catenated dimers composed of one half heavy and one completely light molecule. There were no heavy:light catenanes. These results eliminated the possibility that the catenanes played a role in the replication cycle of mitochondrial DNA. The results of other studies also argue against a replicative derivation for catenanes (122, 123, 119). At the time, these several authors suggested that recombination events between monomers might be the path of catenane formation. Now, of course, the action of topoisomerase could explain the generation of these structures.

The data discussed here indicate that the necessary steps of termination may occur in different order in different replication systems. Catenated dimers may form as late and rate limiting structures in some systems, but are clearly unrelated to replication in others. These conclusions are directly relevant to the problem of termination of SV40 and polyoma DNA replication. In the system cited above, one or another termination pathway is predominant. In the case of SV40, both strategies may be employed, perhaps depending on culture conditions.

Termination of SV40 DNA replication does not appear to involve or require specific DNA sequences. This conclusion comes from the study of deletion mutants and evolutionary variants which have lost the sequences normally found in the termination region in the wild type genome but which terminate in a normal fashion (124, 34). These mutants offer a convincing argument for the sequence indifference of termination functions in permissive monkey cells. However, the defective genomes that arise after SV40 infection of human glioblastoma cells always retain the sequences in the wild type termination region (125). Given the strength of the data from the experiments with deletion mutants in permissive cells, it seems reasonable to suggest that these sequences are retained, after passage in the human cells, for reasons other than an obligate role in termination. Furthermore, the human cells have yet to be infected with the appropriate deletion mutants to directly test a putative involvement in termination. The simplest description of the termination signal is that it occurs when the opposing replicating forks meet, generally 180° away from the origin of replication.

As in the case of the staphylococcal plasmid (111), the actual replication time of the viral DNA, 10-15 min, is significantly longer than that calculated from the known rate of mammalian cellular DNA replication of 2 μ/min. Thus,

the viral DNA would be expected to replicate in less than one minute. The early workers who measured the replication time (126, 127, 128, 7) suggested that there might be a rate-limiting step and several groups have shown that there is an enrichment of late replicative intermediates in the population of replicating molecules (129, 130, 131). Although there is some variability in the precise ratio of late (60-90% replicated) vs. early (0-60%) structures, there is a 2-4 fold accumulation of the late replicating molecules relative to the early forms (127, 129, 131, 130, 132). The predominant late replicating molecule is about 90% completed (130, 131, 94, 132). These molecules have been further characterized by isolating long nascent chains and localizing the 3' end of these chains. There were a number of arrest sites, with a major stop position on either side of the termination region. The two sites were separated by about 470 base pairs with nucleotide number 2671 (from the BglI site at the origin) at the center of the gap (94).

The resolution of the late, 90%, replicating molecule appears to be the rate-limiting step in the production of monomer DNA according to the data of most of the groups that have considered this problem. As illustrated in the discussion of plasmid, phage λ and mitochondrial DNA replication, there are at least two pathways by which covalently closed monomers might be produced. In a scenario similar to that of the mitochondrial model, separation of the daughter molecules of the 90% structure without further replication would yield two gapped circular molecules with a single strand region of nearly 500 nucleotides. Such molecules are not observed. Instead, circular monomer molecules with gaps of about 50 nucleotides in the termination region have been characterized (133, 134). Pulse-chase studies indicated that the gapped molecule is a legitimate intermediate in the replication pathway and is the precursor of the mature covalently closed molecule (95). An explanation to account for the absence of the 500 nucleotide gapped molecule and the presence of the 50 nucleotide gapped intermediate has been proposed by DePamphilis and his colleages (94, 132). They suggest that the former molecule might be rapidly filled in by DNA polymerase α which is responsible for SV40 DNA synthesis (75, 77). This enzyme is inefficient at filling gaps of 20-30 nucleotides and is arrested at specific DNA sequences (135, 87). However, it has not been shown that the enzyme will actually stop 20-30 nucleotides short of completing synthesis of a much longer chain. This question should be resolved by *in vitro* experiments with purified replication components.

An alternative pathway for segregation is provided by the example set by the plasmid and phage λ studies. If the replication forks continue and complete the synthesis of one or both daughter strands without segregation, a catenated dimer would be generated. Depending on the completion of one or both nascent chains, the interlocked circles would be both relaxed, one relaxed, one supercoiled or both supercoiled. Catenated dimers in SV40 infections were described several years ago and extensively characterized (136). Pulse chase data supported the conclusion that these molecules arose from the replication process (136). These structures decay to monomers, and possibly circular dimers, which are found in infected cells as well, and have been shown to form by replication (137). Circular dimers may also arise from rolling circle replicative structures found late in SV40 and polyoma infections (137, 138, 8). An important observation was made by Jaenisch and Levine (136) who found that cycloheximide treatment of infected cells resulted in a several fold increase in the amount of catenated dimers, suggesting that the path of the late replicating molecules is sensitive to perturbations of the cells which affect protein synthesis. It should be noted that while catenated dimers do appear during SV40 and polyoma replication, they differ from those reported in the plasmid and phage systems in that they do not appear to represent a rate limiting step in the production of monomer DNA.

Recently, catenated dimers have been described in greater detail by using agarose gel electrophoresis to resolve several types of catenated structures (139, 140). These authors found catenated dimers in preparations of SV40 minichromosomes isolated from nuclei after prolonged incubation in a triton extraction buffer. Pulse label could be chased out of the molecules and the authors have suggested that they are genuine replicative intermediates. Three classes of molecules were observed: Form A which consists of interlocked relaxed circles (nicked or gapped); form B with one covalently closed supercoiled circle and one relaxed circle; and form C which has two interlocked supercoiled circles. From their data, these structures do not seem to be indicative of rate-limiting steps, that is, they do not accumulate under standard conditions of culture. However, the relative amount of form C molecules is substantially increased by hypertonic shock of the cells, a treatment which inhibits the initiation of protein synthesis (141) and an observation which recalls that made previously by Jaenisch and Levine (136). The hypertonic shock form C molecules may be interwound as many as 20 times. The catenanes decay to monomers when the cells are restored to normal medium.

The decay of the catenanes to monomers shows the versatility of the cell and suggests that segregation failures during replication are not as dangerous as once thought (130, 131). The conclusion that the catenated dimers, particularly the form C structures, arise from replication is plausible. While it is reasonable to propose that an appropriate topoisomerase will segregate the interlocked structures which might arise during replication, it is useful to recall that these enzymes can also generate these same structures (117). The relative proportion of the reaction products could easily depend on conditions in the cell or in the cell lysate during the extraction process. The exclusive use of pulse-chase experiments in experiments with the wild type virus overlooks the fact that an efficient encapsidation process will sequester monomer molecules (142) from the purview of nonreplicative catenation activities. Consequently, the monomer molecules most available for nonreplicative catenation are those which have just finished replication, and these would be labeled in short pulse-labeling protocols. The experiments with density label, which eliminated a replicative etiology for catenanes in the mitochrondrial system (121) have not been performed in the viral system.

REPLICATION IN NONPERMISSIVE CELLS

The work of the last ten years has shown that T antigen is a necessary element in the initiation of viral replication and in this capacity must interact directly with sequences at the origin of replication. However, the presence of T antigen and viral DNA in the same cell apparently is insufficient for measurable replication. Such a situation is found during the infection of mouse, or other rodent cells, by SV40 virus. After infection, T antigen is produced and a variety of cellular functions, such as cellular DNA replication, are stimulated (143). However, viral DNA replication is not detected despite the apparent presence of all the necessary components. Nonpermissive cells survive the infection and a few cells become permanently transformed, a process which requires stable chromosomal integration of viral DNA. Two features of this sequence of events are relevant to a discussion of viral DNA replication; the first is the stable integration of viral DNA, frequently in tandem duplications, in transformed cells; the second is the lack of measurable viral DNA synthesis.

The integration of viral DNA in transformed cells and the structure of the integrated viral genomes has been described by many groups (144, 145, 146, 147, 2). Recently, this subject has been studied by Chia and Rigby (148). They

found that polymeric forms of linear SV40 DNA appeared after infection of mouse cells. The polymers were formed by replication but recombination between the molecules does occur. Synthesis of the polymers of a mutant viral DNA, temperature sensitive for T antigen, was observed after infection of cells at the restrictive temperature. It seems likely that this replication does not require T antigen, at least not in the same fashion as in a permissive cell, and thus replication may be a function of cell factors exclusively. The authors suggested that the polymeric DNA would be the precursor to stably integrated tandem arrays of viral DNA in transferred cells.

The lack of a "lytic" type of replication in nonpermissive cells has been known for a long time. The block to replication is not absolute, for it is possible to treat already transformed rodent cells in such a way so as to induce viral DNA replication and, often, infectious virus formation. This can be done by fusion of the transformed cells with permissive cells (149). The replication of the integrated viral DNA and the appearance of monomeric, replicating viral DNA have been described by Sambrook, and Botchan and their colleagues (150, 151, 152). After cell fusion, there was an increase in the amount of SV40 DNA in the high molecular weight fraction of cellular DNA, followed by the appearance of increasing amounts of free monomeric viral DNA in the low molecular weight fraction (153). Substantially more monomeric DNA appeared if the original transformed cell carried tandem duplications of intact copies of the viral genome. The amplification of the viral DNA required a functional T antigen (150). To explain these results, Botchan et al. (150) formulated the well known "onion skin" model in which multiple rounds of replication of the viral sequences, driven by T antigen from a viral origin, occurs first in the integrated DNA. This leads to a local proliferation of the viral and flanking host DNA sequences. Recombination, obviously facilitated by the presence of tandem viral genomes, would produce the monomeric DNA which could be encapsidated if the genes for the late proteins were intact. The results of these experiments have prompted the conclusion that a factor or factors necessary for viral replication are supplied to the transformed rodent cell by the permissive (monkey) cell.

In an effort to identify the factors and the chromosomes carrying the necessary genes for permissivity, hybrids of permissive and nonpermissive cells have been constructed and then infected with SV40 or polyoma virus (154, 155). In the study of Huebner et al. (155), hybrids between a permissive human tumor cell line and nonpermissive mouse primary cells were made. These hybrids tend

to lose mouse chromosomes and were permissive for SV40 replication but not polyoma, even when all the mouse chromosomes were present. In hybrids which segregated human chromosomes, only polyoma replicated. Thus, in this system, it was not possible to identify chromosomes necessary for replication of either virus. In contrast to these results, Garver et al. (154), who studied SV40 replication in monkey and Chinese hamster cell hybrids, concluded that there was a strong correlation between an elevated level of viral DNA synthesis and the retention of particular monkey chromosomes in the hybrids. The reasons for the differences in the results between the two groups are not clear, although different cells were used in each study. It is important to recall that these experiments differ from the cell fusion experience cited above (150) in that the process of cell fusion was long past at the time of viral infection.

There is a striking qualitative similarity between the description of viral DNA replication in transformed rodent cells fused to permissive cells and viral DNA replication in transformed rodent cells treated with DNA damaging agents such as radiation and a variety of chemical carcinogens and mutagens. It has been known for some time that exposure of SV40 or polyoma transformed cells to such agents often leads to the production of infectious virus (156, 157, 158, 159). Although virions are not formed in all cell lines after treatment, it now seems likely that those lines which failed to yield virus probably did not have complete copies of the viral genome (160, 161). This is a possibility which could not have been properly considered with the technology available at the time of these experiments. More recently, Lavi (162, 163) has repeated many of these experiments using an SV40-transformed Chinese hamster line and a variety of DNA damaging agents. Viral DNA synthesis was monitored by hybridization rather than virion production. As before, many DNA reactive chemicals or radiation induced viral DNA synthesis in the majority of cells. A functional T antigen is required for the several hundred-fold amplication of viral DNA which follows treatment. Cellular sequences linked to SV40 DNA are also amplified. These workers suggested that the onion skin model of Botchan and Sambrook would describe the events in the induced cells. The results of these experiments and the cell fusion experiments indicate that the nonpermissive rodent cell can be converted, at least transiently, to a permissive cell. The cell fusion data have generally been explained by suggestions that some factors, necessary for replication, are supplied by the permissive cell. On the other hand, induction of virus, or replication of viral sequences by DNA-damaging agents has been ascribed to exision of viral

DNA following chromosomal breakage (161) or perturbation of cellular DNA replication by these agents (163). The recent observations of Prives and her colleagues (63) may offer an explanation for viral DNA replication in both situations. They found that a small fraction of the T antigen from lytically infected cells contained the majority of the replication origin binding activity (63). Since origin binding is certainly a requirement for initiation of replication by T antigen, they suggested that only a minority of the T antigen molecules are actually competent to support replication.

A post-translational modification might convert some portion of the T antigen population to initiation competence. If this is correct, then a simple explanation for the failure of T antigen in mouse cell to drive SV40 replication would be that post-translational modification either does not occur or the concentration of appropriately modified T antigen molecules is too low to yield measurable replication. The permissive cell, when fused to the transformed rodent cell, would either provide the compatible modification activity or stimulate the rodent activity to such levels that sufficient T antigen would be modified. In the case of chemical or radiation induced viral replication, such treatments would also induce or stimulate the necessary modification activity. In either situation, activation of T antigen by post-translational modification would be the key event. In this light, it is interesting to note that Scheller _et al_. (63) called attention to the modification of a minor amount of T antigen by poly ADP ribose, described by Goldman _et al_. (164). This is particularly intriguing because it has been well established that the enzyme poly ADP ribose polymerase is stimulated in cells treated with DNA damaging agents (165, 166). The enzyme requires free DNA ends as a cofactor and the activity of the enzyme is proportional to the number of strand breaks in DNA present in reaction mixtures. If the rodent poly (ADP ribose) polymerase had a weak affinity for SV40 T antigen, then sufficient T antigen might be modified only in cells in which the activity had been enhanced. It should be pointed out that we have emphasized the possible role of poly ADP ribosylation of T antigen because of the activation of the enzyme by DNA damage and because it is the only "minor" modification of T antigen described at this time. Obviously, there could be other modification reactions, as yet undetected, which are necessary for activation of T antigen (see, for example, 167). Another consideration from this discussion is that the putative minor modified fraction of T antigen may be present in insufficient concentration, or be unstable, in the _in vitro_ replication

systems, established by several investigators, all of which fail to support the initiation of replication.

REFERENCES

1. Challberg, M. D., and Kelly, T. J. 1982. Ann. Rev. Biochem. 51, 901-934.
2. Tooze, J. 1980. DNA Tumor Viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
3. DePamphilis, M. L. and Wassarman, P. W. 1980. Ann. Rev. Biochem. 49:627-666.
4. DePamphilis, M. L., Anderson, S., Bar-Shavit, R., Collins, E., Edenberg, H., Herman, T., Karas, B., Kaufmann, G., Krokan, H., Shelton, E., Su, R., Tapper, D. and Wassarman, P. M. 1979. CSHSQB XLIII, 679-692.
5. DePamphilis, M. L., Cusick, M. E., Hay, R. J., Pritchard, C., Tack, L. C., Wassarman, P. M. and Weaver, D. J. 1982. In: New Approaches in Eukaryotic DNA Replication (Ed. deRecondo, A. M.), Plenum Press, in press.
6. Fareed, G. C. and Salzman, N. P. 1972. Nature New Biology 238, 274-277.
7. Danna, K. J. and Nathans, D. 1972. Proc. Natl. Acad. Sci. USA 69:3097-3100.
8. Crawford, L. V., Syrett, C. and Wilde, A. 1973. J. Gen. Virol. 21, 515-520.
9. Crawford, L. V., Robbins, A. K., Nicklin, P. M. and Osborn, K. 1975. CSHSQB 39:219-225.
10. Griffin, B. E., Fried, M. and Cowie, A. 1974. Proc. Natl. Acad. Sci. USA 71:2077-2080.
11. Tegtmeyer, P. 1972. J. Virol. 10:591-597.
12. Chou, J. Y., Avila, J. and Martin, R. G. 1974. J. Virol. 14:116-119.
13. Francke, B. and Eckhart, W. 1973. Virology 55:127-131.
14. Tjian, R. 1978. Cell 13:165-179.
15. Shalloway, D., Kleinberger, T. and Livingston, D. M. 1980. Cell 20:411-422.
16. Tegtmeyer, P. and Anderson, B. 1981. Virology 115:67-74.
17. Tegtmeyer, P., Anderson, B., Shaw, S. B., and Wilson, V. G. 1981. Virology 115:75-87.
18. Fareed, G. C., Garon, C. F. and Salzman, N. P. 1972. J. Virol. 10:484-491.
19. Martin, R. G. and Setlow, V. P. 1980. Cell 20:381-391.
20. Subramanian, K. N., Dhar, R. and Weissman, S. M. 1977. J. Biol. Chem. 252:355-367.
21. Soeda, E., Kimura, G. and Miura, K.I. 1978. Proc. Natl. Acad. Sci. USA 75:162-171.
22. Dhar, R., Lai, C. J. and Khoury, G. 1978. Cell 13:345-358.
23. Van Heuverswyn, H. and Fiers, W. 1979. Eur. J. Biochem. 100:51-60.
24. Friedman, T., LaPorte, P. and Esty, A. 1978. J. Biol. Chem. 253:6561-6567.
25. Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. and Weissman, S. M. 1978. Science 200:494-502.
26. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herrweghe, J., Volckaert, G. and Ysebaert, M. 1978. Nature 273:113-120.
27. Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. and Griffin, B.E. 1980. Nature 283:445-453.
28. Seif, I., Khoury, G. and Dhar, R. 1979. Cell 18:963-977.
29. Queen, C., Lord, S. T., McCutchen, T. F. and Singer, M. F. 1981. Mol. Cell Biol. 1:1061-1068.

30. Gutai, M. and Nathans, D. 1978. J. Mol. Biol. 126:259-274.
31. Davoli, D., Ganem, D., Nussbaum, A. L., Fareed, G., Howley, P. M., Khoury, G., and Martin, M. A. 1977. Virology 77:110-124.
32. Lund, E., Griffin, B. E. and Fried, M. 1977. J. Mol. Biol. 117:497-513.
33. Cole, C. N., Landers, T., Goff, S. P., Manteuil-Brutlag and Berg, P. 1977. J. Virol. 24:277-294.
34. Lai, C. J. and Nathans, D. 1974. J. Mol. Biol. 89:179-193.
35. Subramanian, K. N. and Shenk, T. (1978). Nucl. Acids Res. 5:3635-3637.
36. DiMaio, D. and Nathans, D. 1980. J. Mol. Biol. 140:129-142.
37. Bergsma, D. J., Olive, D. M., Hartzell, S. W. and Subramanian, K. N. 1982. Proc. Natl. Acad. Sci. USA 79:381-385.
38. Gluzman, Y. 1981. Cell 23:175-182.
39. Gruss, P., Dhar, R., and Khoury, G. 1981. Proc. Natl. Acad. Sci. USA 78, 943-947.
40. Kriegler, M. P., Griffin, J. D. and Livingston, D. M. 1978. Cell 14:983-994.
41. Myers, R., Rio, D. C., Robbins, A. K. and Tjian, R. 1981. Cell 25:373-384.
42. Myers, R. M., Williams, R. C. and Tjian, R. 1981. J. Mol. Biol. 148:347-353.
43. McKay, R. and DiMaio, D. 1981. Nature 289:810-813.
44. Reed, S. I., Ferguson, J., Davis, R. W. and Stark, G. W. 1975. Proc. Natl. Acad. Sci. USA 72:1605-1609.
45. Jessel, D., Landau, T., Hudson, J., Lalor, T., Tenen, D. and Livingston, D. M. 1976. Cell 8, 535-545.
46. Hassell, J. A., Lukanidin, E., Fey, G. and Sambrook, J. 1978. J. Mol. Biol. 120:209-247.
47. Tjian, R. 1979. CSHSQB 43:655-662.
48. Tenen, D. G., Haines, L. L. and Livingston, D. M. 1982. J. Mol. Biol. 157:473-492.
49. Myers, R. M. and Tjian, R. 1980. Proc. Natl. Acad. Sci. USA 77:6491-6495.
49a. Solnick, D. 1981. Cell 24:135-143.
50. Tjian, R. 1981. Cell 26:1-2.
51. Hansen, U., Tenen, D. G., Livingston, D. M., and Sharp, P. A. 1981. Cell 27:603-612.
52. DiMaio, D. and Nathans, D. 1982. J. Mol. Biol. 156:531-548.
52a. Shortle, D. R. and Nathans, D. 1978. Proc. Natl. Acad. Sci. USA 75:2170-2174.
53. Shortle, D. R., Mangolskee, R. F. and Nathans, D. 1979. Proc. Natl. Acad. Sci. USA 76:6128-6131.
54. Shortle, D. and Nathans, D. 1979. J. Mol. Biol. 13:801-817.
55. Bradley, M. K., Griffin, J. D. and Livingston, D. M. 1982. Cell 28:125-134.
56. Fanning, E., Nowak, B. and Burger, C. 1981. J. Virol. 37:92-102.
57. Gidoni, D., Scheller, A., Barnet, B., Hantzopoulos, P., Oren, M. and Prives, C. 1982. J. of Virol. 42:456-466.
58. Tegtmeyer, P., Rundell, K. and Collins, J. K. 1977. J. Virol. 21:647-657.
59. Greenspan, D. S. and Carroll, R. B. 1981. Proc. Natl. Acad. Sci. U.S.A. 78:105-109.
60. Linzer, D. and Levine, A. J. 1979. Cell 17:43-52.
61. Lane, D. P. and Crawford, L. V. 1979. Nature 278:261-163.
62. Harlow, E., Pim, D. C. and Crawford, L. B. 1981. J. Virol. 37:564-573.
63. Scheller, A., Covey, L., Barnet, B. and Prives, C. 1982. Cell 29:375-383.
64. Gurney, F. G., Harrison, R. J. and Fenno, J. 1980. J. Virol. 34:752-763.
65. Deppert, W., Gurney, E. G. and Harrison, R. O. 1981. J. Virol. 37:478-482.
66. Soeda, E., Kimura, G. and Miura, K. I. 1978. Proc. Natl. Acad. Sci. USA 75:162-171.

67. Bendig, M. M. and Folk, W. R. 1979. J. Virol. 32:530-535.
68. Magnusson, G. and Berg, P. 1979. J. Virol. 32:523-529.
69. Wells, R. D., Hutchinson, M. A. and Eckhart, W. 1979. J. Virol. 32:517-522.
70. Bendig, M. M., Thomas, T. and Folk, W. R. 1980. Cell 20:401-409.
71. Griffin, B. E. and Maddock, C. 1979. J. Virol. 31:645-656.
72. Gaudray, P., Tyndall, C., Kamen, R. and Cuzin, F. 1981. Nucl. Acids Res. 9:5697-5710.
73. Hay, R. T. and DePamphilis, M. L. 1982. Cell 28:767-779.
74. Clayton, D. A. 1982. Cell 28:693-705.
75. Edenberg, H. J., Anderson, S. and DePamphilis, M. L. 1978. J. Biol. Chem. 253:3273-3280.
76. Krokan, H., Schaffer, P. and DePamphilis, M. 1979. Biochemistry 18:4431-4443.
77. Otto, B. and Fanning, E. 1978. Nucl. Acids Res. 5:1715-1728.
78. Gourlie, B. B., Pigiet, V., Breaux, B. C., Krauss, M.R., King, C. R. and Benbow, R. M. 1981. J. Virol. 38:826-832.
79. Chang, L. M. S., Brown, M. and Bollum, F. J. 1973. J. Mol. Biol. 74:1-8.
80. Chiu, R. W. and Baril, E. F. 1975. J. Biol. Chem. 250:7951-7957.
81. Wilson, S. H., Matsukage, A., Bohn, E. W., Chen, Y. C. and Sivarajan, M. 1977. Nucl. Acids. Res. 4:3981-3996.
82. Weissbach, A. 1977. Ann Rev. Biochem. 46:25-47.
83. Das, S. K. and Fujimura, R. K. 1979. J. Biol. Chem. 254, 1227-1232.
84. Korn, D., Fisher, P. A., Battey, J. and Wang, T. S. F. 1979. Cold Spring Harbor Symp. Quant. Biol. 43:613-624.
85. Matsukage, A., Nishizawa, M., Takahashi, T. and Hozumi, T. 1980. J. Biochem. 88:1869-1877.
86. Detera, S. D., Becerra, S. P., Swack, J. A. and Wilson, S. H. 1981. J. Biol. Chem. 256:6933-6943.
87. Weaver, D. J. and DePamphilis, M. J. 1982. J. Biol. Chem. 257:2075-2086.
88. Villani, G., Fay, P. J., Bambara, R. A. and Lehman, I. R. 1981. J. Biol. Chem. 256:8202-8207.
89. Ogawa, T. and Okazaki, T. 1980. Ann. Rev. Biochem. 49:421-457.
90. Eliasson, R. and Reichard, P. 1979. J. Mol. Biol. 129:343-409.
91. Magnusson, G., Pigiet, V., Winnacker, E. L., Abrams, R. and Reichard, P. 1973. Proc. Natl. Acad. Sci. USA. 7:412-415.
92. Pigiet, V., Winnacker, E. L., Eliasson, R. and Reichard, P. 1973. Nature New Biol. 245:203-205.
93. Pigiet, V., Eliasson, R. and Reichard, P. 1974. J. Mol. Biol. 84:197-216.
94. Tapper, D. P. and DePamphilis, M. L. 1980. Cell 22:97-108.
95. Fareed, G. C., McKerlie, L. and Salzman, N. P. 1973. J. Mol. Biol. 74:95-111.
96. Francke, B. and Hunter, T. 1974. J. Mol. Biol. 83:99-121.
97. Francke, B. and Vogt, M. 1975. Cell 5:205-211.
98. Perlman, D. and Huberman, J. 1977. Cell 12:1029-1043.
99. Hunter, T., Francke, B., Bacheler, L. 1977. Cell 12:1021-1028.
100. Kaufmann, G., BarShavit, R., and DePamphilis, M. L. 1978. Nucleic Acids Res. 5:2535-2545.
101. Tye, B. K., Chien, J., Lehman, I. R., Duncan, B. K. and Warner, H. R. 1978. Proc. Natl. Acad. Sci. USA 78:233-238.
102. Brynolf, K., Eliasson, R. and Richard, P. 1978. Cell 13:573-580.
103. Lindahl, T. 1979. In Progress in Nucleic Acid Research and Molecular Biology (Ed. Cohn, W. F.) N. Y. Academic 22:135-188.
104. Narkhammer-Meuth, M., Kowalski, J. and Denhardt, D. T. 1981. J. Virol. 39:21-30.

105. Narkhammer-Meuth, M., Eliasson, R. and Magnusson, G. 1981. J. Virol. 39, 11-20.
106. Okazaki, R., Hirose, S., Okazaki, T., Ogawa, T. and Kanosawa, Y. 1975. Biochem. Biophys. Res. Comm. 62:1018-1024.
107. Kaufmann, G. and Falk, H. H. 1982. Nucl. Acids. Res. 10:2309-2321.
108. Yara, T., Kozu, T. and Seno, T. 1982. J. Biol Chem. 257:11121-11127.
109. Conaway, R. L. and Lehman, I. R. 1982. Proc. Natl. Acad. Sci. U.S.A. 79:2523-2527.
110. Kozu, T., Yagura, T. and Seno, T. 1982. Nature 298:180-182.
111. Novick, R. P., Smith, K., Sheehy, R. J. and Murphy. E. 1973. Biochem. Biophys. Res. Comm. 54:1460-1469.
112. Sheehy, R. J. and Novick, R. P. 1975. J. Mol. Biol. 93:237-253.
113. Kupersztoch, Y. M. and Helinski, D. R. 1973. Biochem. Biophys. Res. Comm. 54:1451-1459.
114. Schnos, M. and Inman, R. B. 1970. J. Mol. Biol. 51:61-73.
115. Sogo, J. M., Greenstein, M. and Skalka, A. 1976. J. Mol. Biol. 103:537-562.
116. Gefter, M. 1975. Ann. Rev. Biochem. 44:45-78.
117. Liu, L.F., Liu, C., and Alberts, B. M. 1980. Cell 19:697-707.
118. Berk, A. J. and Clayton, D. A. 1974. J. Mol. Biol. 86:801-824.
119. Berk, A. J. and Clayton, D. A. 1976. J. Mol. Biol. 100:85-102.
120. Hudson, B. and Vinograd, J. 1967. Nature 216:647-652.
121. Flory, J. P., Jr. and Vinograd, J. 1973. J. Mol. Biol. 74:81-94.
122. Bogenhagen, D. and Clayton, D. A. 1976. J. Biol. Chem. 251:2938-2944.
123. Clayton, D. A., Doda, J. N. and Friedberg, F. C. 1974. Proc. Natl. Acad. Sci. USA 71:2777-2781.
124. Brockman, W. W., Gutai, M. W. and Nathans, D. 1975. Virology 66:36-52.
125. Carroll, D., and O'Neill, F. J. 1978. Virology 87:120-129.
126. Levine, A. J., Kang, H. S. and Billheimer, F. E. 1970. J. Mol. Biol. 50:549-568.
127. Bourgaux, P., Bourgaux-Ramoisy, D. and Seiler, P. 1971. J. Mol. Biol. 59:195-206.
128. Sebring, E. D., Kelly, J. J., Thoren, M. M. and Salzman, N. P. 1971. J. Virol 8:478-490.
129. Mayer, A. and Levine, A. J. 1972. Virology 50:328-338.
130. Tapper, D. P. and DePamphilis, M. L. 1978. J. Mol. Biol. 120:401-422.
131. Seidman, M. M. and Salzman, N. P. 1979. J. Virol. 30:600-609.
132. Tapper, D. P., Anderson, S. and DePamphilis, M. L. 1982. J. Virol. 41:877-892.
133. Chen, M. C. Y., Birkenmeier, E. and Salzman, N. P. 1976. J. Virol. 17:614-621.
134. Laipis, P. L., Sen, A., Levine, A. J. and Mulder, C. 1975. Virology 68:115-123.
135. Fisher, P. A., Wang, T. S. F. and Korn, D. 1979. J. Biol. Chem. 254: 6128-6137.
136. Jaenisch, J. and Levine, A. J. 1973. J. Mol. Biol. 73:199-212.
137. Goff, S. P. and Berg, P. 1977. J. Virol. 24:295-302.
138. Martin, M. A., Howley, P. M., Byrne, J. C. and Garon, C. F. 1976. Virology 71:28-40.
139. Sundin, O. and Varshavsky A. 1980. Cell 21:103-114.
140. Sundin, O. and Varshavsky, A. 1981. Cell 25:659-669.
141. Saborio, J. L., Pong, S. S. and Koch, G. 1974. J. Mol. Biol. 85:195-211.
142. Garber, E. A., Seidman, M. M. and Levine, A. J. 1978. Virology 90:305-316.
143. Weil, R., Salomon, C., May, E. and May, P. 1974. CSHSQB 39:381-395.
144. Botchan, M., Topp, W. C. and Sambrook, J. 1976. Cell 9:269-273.

145. Birg, F., Dulbecco, R., Fried, M. and Kamen, R. 1979. J. Virol. 29:633-648.
146. Clayton, C. E. and Rigby, P. 1981. Cell 25:547-559.
147. Kucherlapati, R., Hwang, S. P., Shumuzu, N., McDougall, J. K. and Bokhan, M. R. 1978. Proc. Natl. Acad. Sci. USA 75:4460-4463.
148. Chia, W. and Rigby, P. W. J. 1981. Proc. Natl. Acad. Sci. USA 78: 6638-6642.
149. Gerber, P. 1966. Virology 28:501-504.
150. Botchan, M. Topp, W. and Sambrook, J. 1979. CSHSQB 43:709-719.
151. Botchan, M., Stringer, J., Mitchison, T. and Sambrook, J. 1980. Cell 20:143-152.
152. Sambrook, J., Botchan, M. R., Gallimore, P., Ozanne, B., Pettersson, U., Williams, J. and Sharp, P. A. 1975. CSHSQB 39:615-632.
153. Hirt, B. 1967. J. Mol. Biol 26:365-368.
154. Garver, J. J., Pearson, P. L., Abrahams, P. J. and Vander Eb, A. J. 1980. Somatic Cell Genetics 6:443-453.
155. Huebner, K., Shander, M. and Croce, C. M. 1977. Cell 11, 25-33.
156. Rothschild, H. and Black, P. H. 1970. Virology 42:251-256.
157. Burns, W. H. and Black, P. H. 1969. Virology 39:626-634.
158. Fogel, M. and Sachs, L. 1969. Virology 37:327-334.
159. Fogel, M. and Sachs, L. 1970. Virology 40:174-177.
160. Burns, W. H. and Black, P. W. 1968. J. Virol. 2:606-609.
161. Kaplan, J. C., Wilbert, S. M., Collins, J. J., Rakusanova, T., Zamansky, G. B., and Black, P. H. 1975. Virology 68:200-214.
162. Lavi, S. 1981. Proc. Natl. Acad. Sci. USA 78:6144-6150.
163. Lavi, S., and Etkins, S. 1981. Carcinogenesis 2, 417-423.
164. Goldman, N., Brown, M. and Khoury, G. 1981. Cell 24:567-572.
165. Benjamin, R. L. and Gill, D. M. 1980. J. Biol. Chem. 255:10493-10501.
166. Benjamin, R. L. and Gill, D. M. 1980. J. Biol. Chem. 255:10502-10508.
167. Ahmad-Zadeh, C., Allet, B., Greenblatt, J. and Weil, R. 1976. Proc. Natl. Acad. Sci. USA 73:1097-1101.

3

SV40 CHROMATIN REPLICATION IN VITRO

A. RICHTER AND B. OTTO

SUMMARY

Two soluble in vitro systems for SV40 chromatin replication are presented. The first subnuclear system consists of nuclear extracts, containing replicating and nonreplicating SV40 chromosomes as well as soluble nuclear proteins, and of cytoplasmic protein extracts from mammalian cells. The second more advanced system is composed of replicating SV40 chromosomes which were purified from nuclear extracts and of protein extracts from unfertilized eggs of Xenopus laevis. In both systems elongation and completion of SV40 chromatin replication are carried out as these processes are in vivo. Both systems are useful for dissecting protein extracts and identifying replication proteins necessary for elongation and completion of SV40 chromatin replication. By analogy they also serve as a model for most of the same processes of mammalian chromatin replication.

As an example to identify replication factors with these systems we used DNA polymerases and we show by several criteria that DNA polymerase α activity is necessary for SV40 and mammalian chromatin replication.

INTRODUCTION

Identification and characterization of proteins necessary for prokaryotic DNA replication have greatly substantiated our understanding of the biochemistry of DNA replication (1). In the case of eukaryotic organisms, however, five years ago not one protein required for DNA replication had been identified. When we started to study the biochemistry of

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

mammalian chromatin replication we asked which are the proteins involved in this process. To identify some of those proteins in our mind an *in vitro* system for DNA replication with the following features was necessary:

1. The system should consist of a small, well defined DNA, organized as chromatin for which DNA synthesis had to reflect DNA replication.
2. It should be possible to omit and add proteins to the system. These manipulations should affect DNA replication and allow a characterization of the proteins necessary for the replication process.

Which small DNA would be appropriate ?

The papovaviruses, simian virus 40(SV40) and polyoma virus have small, well defined genomes, which are organized as chromatin. Their nucleosomal structure and histone composition appear to be very similar to that of the host cellular chromatin. The small circular viral chromatin (also called "circular minichromosomes") replicate in the nuclei of certain mammalian cells (2). Except for the viral gene A product, T-antigen, which is required for initiation, viral replication depends totally on the replication machinery of the mammalian host cell (3). Thus viral minichromosomes are the ideal primer templates for an *in vitro* system to characterize mammalian replication proteins.

Which *in vitro* system best allows the omission and addition of replication proteins ? Initiated by the work of M.L. De Pamphilis (4) and J. Huberman (5) we worked with a subnuclear, soluble, *in vitro* system for SV40 chromatin replication that permits the manipulations described above. This system consists of SV40 minichromosomes which have been extracted from the nuclei of SV40 infected monkey cells and thereby completely separated from cellular chromatin. Both replicating as well as nonreplicating chromosomes are present in these extracts (6). Highly concentrated protein extracts of mammalian replication factors elongate and terminate the replication of SV40 DNA thereby converting replicating into

nonreplicating chromosomes (4,5,7). We developed a second, more advanced *in vitro* system which consists of purified replicating SV40 chromosomes and of protein extracts from unfertilized eggs of *Xenopus laevis*, and in which again elongation and termination of SV40 DNA replication occurs (8).

In addition the separation of replicating and nonreplicating SV40 chromosomes allows the identification of mammalian proteins associated with replicating chromosomes.

This chapter summarizes our studies as well as those of our coworkers over the last five years at the University of Konstanz. It characterizes the two *in vitro* systems and its usefulness in identifying mammalian replication factors. As our example we use DNA polymerase α which was shown by these studies to be necessary for mammalian chromatin replication (6,7,9).

MATERIALSAND METHODS

The source of cell lines and virus as well as the preparation of SV40 chromosomes and the assay for DNA synthesis have been described earlier (6,7). Purification of and assays for DNA binding protein and DNA polymerases were done as previously described (9, 10) as well as the preparation of protein extracts from *Xenopus* eggs and all the techniques used to characterize different forms of SV40 DNA (7,8).

PROLOGUE. SV40 DNA REPLICATION IN VIVO

Semiconservative DNA replication is initiated at a unique origin in covalently closed superhelical SV40 DNA (form I-DNA) producing a pool of replicating intermediates (RI-DNA). In these DNA molecules two forks migrate bidirectionally until they reach the termination region approximately 180°C from the origin. After the separation of these molecules, DNA appears which contains a gap in the nascent DNA chain of about 50 nucleotides (form II* DNA).

When this gap is filled (creating form II DNA) the nascent DNA chain can be sealed to give form I DNA. Figure 1 summarizes the various forms of SV40 DNA during the replication cycle.

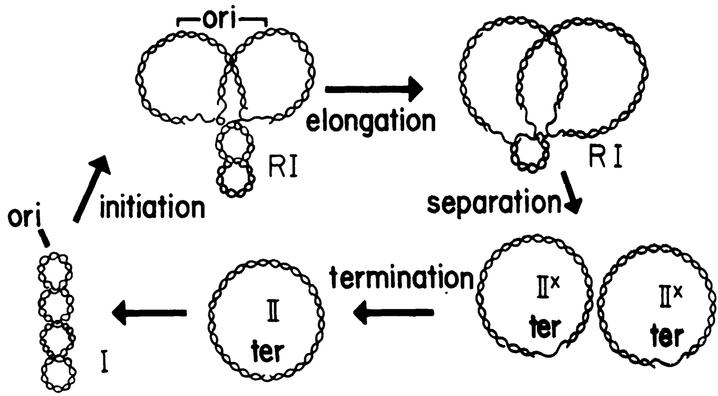


Figure 1. SV40 DNA replication cycle

DNA elongation occurs discontinously on the DNA template whose 5'-3' polarity lies in the same direction as the movement of the replication fork and continuously on the complementary strand. The discontinous synthesis leads to a transient appearance of short DNA pieces (4S, Okazaki pieces) for which synthesis is initiated at even shorter RNA primers with about 10 oligoribonucleotides in length. Once the RNA primers are excised, the resulting gaps are filled in by DNA synthesis and the elongated Okazaki pieces become joined to the growing daughter strands (11).

RESULTS

Preparation and characterization of SV40 chromosomes

SV40 chromosomes can be extracted from nuclei of SV40 infected cells prepared and kept in hypotonic conditions (4, 5, 6). Due to their small size the viral chromosomes as well

as proteins leak out of the nuclei during the extraction process whereas the 10^4 times larger host cell chromosomes remain inside the nuclei. Can both replicating and nonreplicating viral chromosomes be extracted by this procedure ? Since one replication cycle of SV40 DNA takes less than 15 minutes, a pulse label with DNA precursors of 5 minutes is mainly recovered in replicating intermediates (RI-DNA) and should appear in replicating chromosomes. A long labeling time labels nonreplicating SV40 (form I DNA) and should be recovered in nonreplicating chromosomes. Using these two labeling protocols with SV40 infected cells, both labels were found in nuclear extracts (6, 9). Moreover both labels can be separated by zone velocity sedimentation as shown in Figure 2.

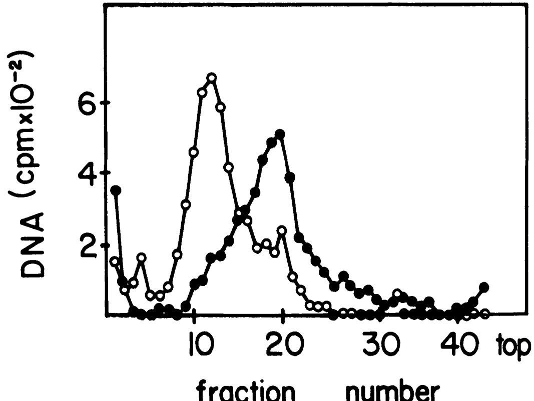


Figure 2. Separation of replicating SV40 chromosomes pulse labeled for 5 min in vivo with ^3H – thymidine (O) from nonreplicating SV40 chromosomes prelabeled in vivo for 12 hr with ^{14}C –thymidine (●) in 5-30% neutral sucrose density gradients (6).

Replicating viral chromosomes as indicated by the pulse label can be easily distinguished from slower sedimenting long labeled nonreplicating chromosomes. The chromosomes

sediment at approximately 95S and 75S respectively. In contrast pulse labeled RI-DNA as well as long labeled form I DNA, stripped of their histone and nonhistone proteins by extraction, sediment under the same conditions at 21-25S and 21S respectively.

Studies using micrococcal nuclease digestion show that SV40 chromosomes are, like host cell chromatin, organized in repeated units called nucleosomes. Replicating chromosomes containing nascent DNA are more accessible to nuclease attack than nonreplicating chromosomes indicating a change of the structural state of the chromosomes during the replication process: pulse—chase experiments _in vivo_, followed by extraction of viral chromosomes and nuclease digestion demonstrate two features of the viral DNA. Newly replicated viral chromosomes show an increased sensitivity to nuclease activity as well as a faster than usual rate of cleavage to DNA fragments of monomeric nucleosome size when compared to nonreplicating chromosomes. These alterations in replicating chromosomes are lost at characteristic times after replication and before these chromosomes are converted into nonreplicating chromosomes (12).

DNA synthesis in viral chromosomes of nuclear extracts

Does the DNA synthesis observed in viral chromosomes of this _in vitro_ system reflect SV40 DNA replication _in vivo_ ? Hypotonic extracts from nuclei of SV40 infected cells containing replicating and nonreplicating SV40 chromosomes as well as proteins _in vitro_ incorporate labeled precursors for DNA synthesis into viral DNA under suitable incubation conditions. However all the _in vitro_ labeled DNA is recovered in replicating chromosomes and characterization of the DNA shows all the label to be in short Okazaki pieces even after prolonged incubation times. Thus proteins coextracted with the chromosomes whether being soluble or associated with the chromosomes cannot join small DNA pieces into larger DNA. Addition of concentrated cytoplasmic

proteins from the same source of cells which have been used to prepare the viral chromosomal extracts enhances the in vitro DNA synthesis about two to three fold. In the presence of cytosol, small DNA pieces only appear transiently, after short incubation times, and become completely ligated during further incubations. In more than 50% of DNA molecules the nascent daughter strands are elongated to unit length of SV40 DNA, giving to the same extent form II DNA. In agreement with these results about 50% of replicating viral chromosomes are matured into nonreplicating chromosomes (7). Thus in vitro DNA synthesis occurs discontinuously; RI-DNA for which synthesis has been initiated in vivo is elongated and terminated in this in vitro system: these processes reflect SV40 DNA replication in vivo.

Initiation of SV40 DNA replication is not observed in this system. We cannot exclude completely that initiation occurs; but if it does it then must be so slight when compared to elongation that we cannot detect this process.

Salt dependence of DNA synthesis in vitro

In this in vitro system, consisting of viral chromosomes and sufficient cytosol to elongate and terminate SV40 replication, the optimal ionic strength for DNA synthesis is low. Salt concentrations close to physiological conditions (150 mM; 13) cause a decrease in DNA synthesis. This sensitivity to salt however depends on the anions used. Figure 3 shows that more than 100 mM chloride anions reduce whereas acetate anions at 50-200 mM slightly stimulate DNA synthesis. Analysis of in vitro labeled DNA indicates that in the presence of high chloride anions the joining of small DNA pieces to larger ones and the overall elongation process of RI-DNA is reduced to the same extent as the incorporation of in vitro label into viral DNA (7). This observation has been also made by others (14). We later in this chapter will demonstrate that the salt behaviour of one replication protein can explain this salt sensitivity of the replication system.

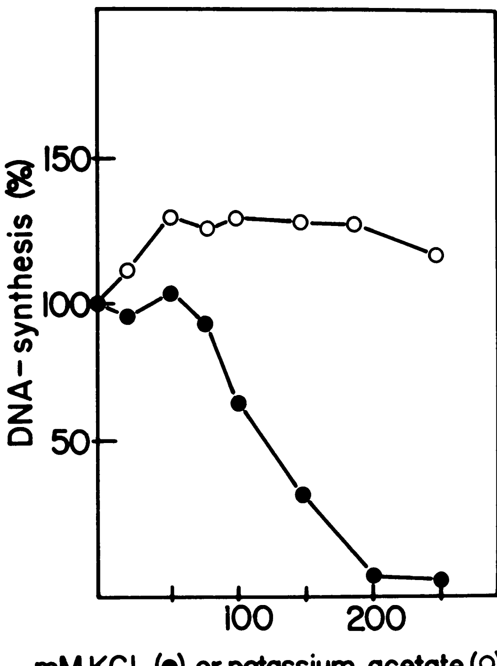


Figure 3.
Dependence of SV40 chromatin DNA synthesis on salt conditions. DNA synthesis, observed under increasing concentrations of either potassium chloride or acetate, is expressed as per cent of that DNA synthesis made in the absence of salts (7).

DNA synthesis in purified replicating chromosomes

DNA replication in replicating chromosomes of nuclear extracts can be elongated and completed by cytoplasmic protein extracts from mammalian cells. However after purification of replicating chromosomes, by which soluble proteins of nuclear extracts become separated from the chromosomes, DNA synthesis of these chromosomes can be only slightly stimulated by cytoplasmic protein extracts: synthesis of Okazaki pieces is only poorly enhanced and replication can not be completed (15, 16, 17). Why does purification of the chromosomes eliminate the ability of these highly concentrated protein extracts to complete DNA synthesis ? One reason might be physical alteration of

chromosomes during their purification. Another reason could be that these cytoplasmic extracts although highly concentrated are not by themselves but only in combination with those proteins present in nuclear extracts sufficient to complete DNA synthesis. Our data support this second explanation.

Unfertilized eggs of Xenopus laevis are programmed for rapid replication. Extracts prepared from those eggs should therefore have a higher concentration of replication factors than those from mammalian cells. This is indeed true for a DNA polymerase α -like activity (17, 18). Egg extracts stimulate elongation of SV40 DNA in replicating chromosomes. If nuclear extracts of chromosomes are used, to the same extent as mammalian cell extracts. When incubated with purified chromosomes, egg extracts enhance DNA synthesis four to six fold (Figure 4) while mammalian extracts do not (17).

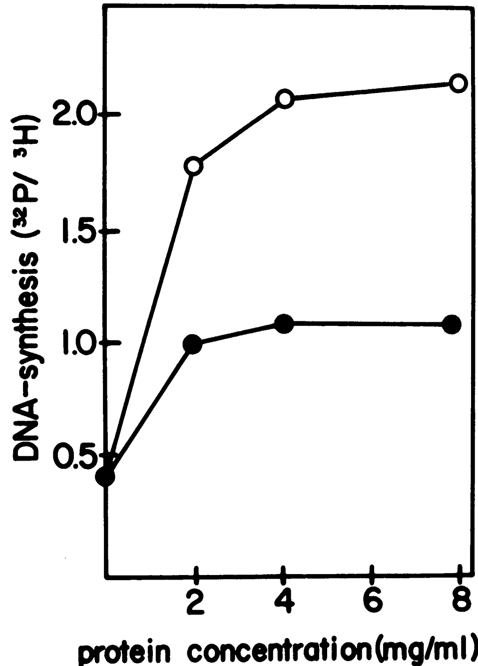


Figure 4. DNA synthesis in purified replicating SV40 chromosomes. DNA synthesis is measured as a function of increasing concentrations of protein extracts from eggs of Xenopus laevis after either 10 min (●) or 30 min (O) incubation (8).

The residual DNA synthesis of chromosomes seen without extracts does not increase between the 10 and 30 min incubation. This synthesis is due to the presence of replication proteins associated with replicating chromosomes.

DNA synthesis in the presence of egg extracts shows all the characteristics of SV40 DNA replication. It is inhibited by aphidicolin suggesting the participation of a DNA polymerase α-like activity (19). Consistent with this observation, addition of dideoxithymidine triphosphate (ddTTP) does not affect DNA synthesis. The ddTTP has been shown to inhibit DNA polymerases β and γ but not α (20). DNA synthesis requires ATP and the other ribonucleoside triphosphates; this requirement is more pronounced when low molecular weight molecules have been purified away from proteins.

Analysis of _in vitro_ labeled DNA shows in all aspects tested that DNA synthesis in this system reflects elongation and completion of SV40 DNA replication _in vitro_.

DNA is synthesized by a discontinuous mechanism: after short incubation times Okazaki pieces appear which become joined during prolonged incubation times (8).

A restriction enzyme analysis of _in vitro_ replicated DNA as can be seen in Figure 5, indicates that DNA fragments containing the terminus of replication or being close to the terminus are more _in vitro_ labeled than those fragments close to the origin. These results show that replication moves towards the terminus as one would expect for elongation and completion of replication rounds.

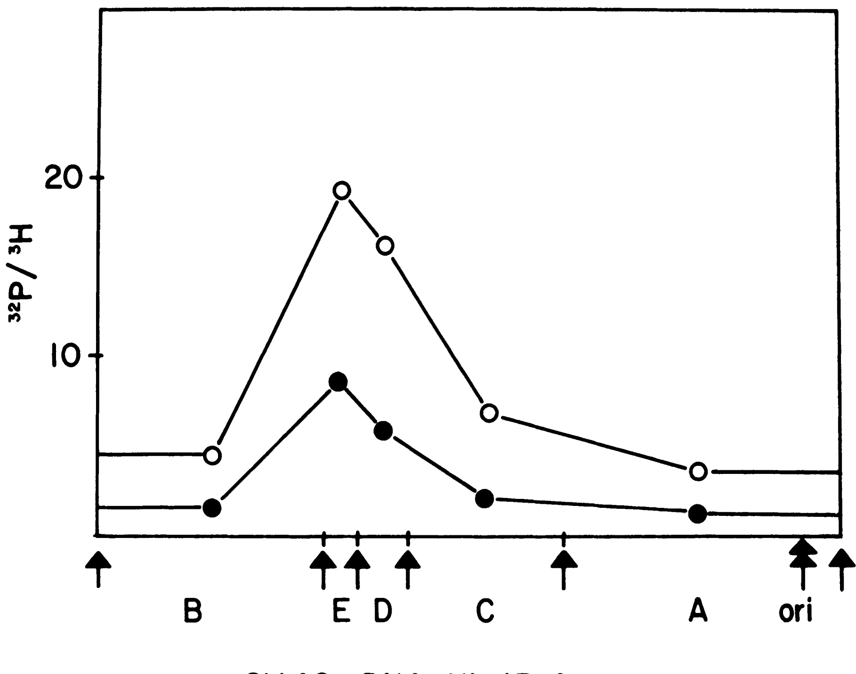


Figure 5. Restriction enzyme analysis of SV40 DNA synthesized *in vitro*. Purified replicating SV40 chromosomes of which DNA had been ^{3}H-thymidine labeled *in vivo* were incubated with egg extracts for 10 (●) and 30 (O) minutes in the presence of ^{32}P-dNTPs (8). DNA was purified and cut by HindII nuclease as described (8). The relative label ($^{32}P/^{3}H$) is shown for the five different DNA fragments.

Again in all aspects tested DNA synthesis in purified replicating SV40 chromosomes in the presence of optimal concentrations of protein extracts of unfertilized eggs from *Xenopus laevis* reflects the elongation and termination processes of SV40 DNA replication *in vivo*. For unknown reasons the only process which does not work as efficiently in this system as in an infected cell is the gap filling process of form II* DNA. About 30% of *in vitro* label accumulates in this type of DNA which contains single-

stranded DNA, 30% are recovered in form II DNA and 40% in sealed supercoiled form I DNA. This amount of form I DNA however is higher than was ever observed in studies using chromosomes in crude extracts and in similar systems described by others (4, 5, 14). Nevertheless this is to date the most useful in vitro system for dissecting protein extracts and identifying replication proteins necessary for elongation and completion of SV40 chromatin replication. By analogy it also serves as a model for most of the same processes of mammalian chromatin replication.

This system is also a prerequisite for studying initiation of replication in vitro. Initiation can be only studied if this process is followed to some extent by elongation. We have used nonreplicating SV40 chromosomes, completely free of those replicating, as a template in egg extracts to investigate the initiation process. So far these experiments are without success, but as pointed out in the introduction a viral coded protein, the T-antigen, is necessary for this process in vivo and must be included in these experiments. In addition structural alterations in nonreplicating chromosomes might be a further prerequisite.

DNA polymerases associated with viral chromosomes

Unfractionated nuclear extracts with SV40 chromosomes also contain all three DNA polymerases, α, β and γ. These extracts can be fractionated on sucrose gradients to separate replicating and nonreplicating chromosomes from each other and from soluble proteins as well. Each fraction can then be selectively assayed for the three DNA polymerases. As Figure 6 shows, 10 to 20% of the total DNA polymerase α is detected in the replicating chromosomes (95S) and 80-90% at the top of the gradient. No DNA polymerase α activity is observed in the slower sedimenting nonreplicating chromosomes (75S) suggesting a specific association of this polymerase with replicating chromosomes.

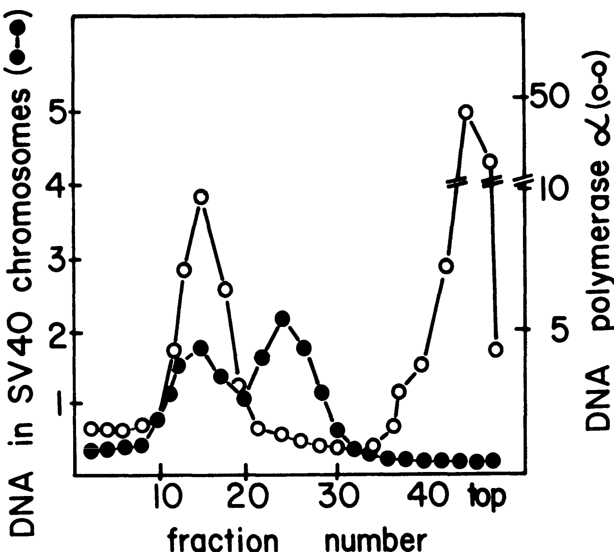


Figure 6. Association of DNA polymerase α activity with replicating viral chromosomes. Nuclear extracts were fractionated on 5-30% neutral sucrose gradients. *In vivo* labeled DNA for 25 min. (●) and DNA polymerase α activity (O) were determined in each fraction as described (6).

We also found DNA polymerase γ associated with chromosomes which sediment between the 95S and 75S viral chromosomes. In contrast DNA polymerase ß is only recovered in an unassociated form at the top of the gradient (6). Similar observations have been made by other groups (20, 21, 22).

About 80% of both types of viral chromosomes can be precipitated in nuclear extracts by the presence of 5-20 mM Mg^{++} (6). DNA polymerase α and γ activities coprecipitate with the chromosomes to the same extent as they are found to be associated with these chromosomes after fractionation of nuclear extracts on sucrose density gradients. Again DNA polymerase ß activity does not coprecipitate. The specific association of DNA polymerase α with replicating chromosomes is conserved and free polymerase activity is removed by the Mg^{++} precipitation procedure as shown in Figure 7.

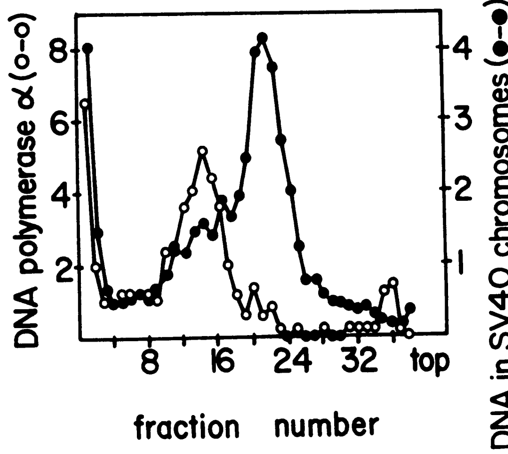


Figure 7. Specific association of DNA polymerase α with replicating chromosomes after Mg^{++} precipitation and fractionation on 5-30% neutral sucrose gradients (6).

After concentration by Mg^{++} precipitation, the DNA polymerase activities were dissociated from the complexes and further characterized. By their chromatographic behavior, sedimentation and biochemical properties the chromosome associated DNA polymerases were classified as DNA polymerases α and γ, but not as ß (6, 9).

Though the presence of DNA polymerase α in replicating chromosomes suggests that this activity may carry out SV40 DNA replication further evidence was necessary to show that this enzyme is in fact required for replication. As described before SV40 chromatin replication in the subnuclear extract is inhibited by more than 100 mM chloride, but stimulated by 100-200 mM acetate anions. We therefore investigated the dependence of DNA polymerase α activity and its association with replicating chromosomes on these concentrations of anions: we found that DNA polymerase α activity on activated DNA as a primer template depends to the same extent as SV40 chromatin replication does on chloride and acetate anions. We also found that the association of this enzyme with replicating chromosomes is conserved up to

200 mM acetate whereas more than 100 mM KCE strongly reduces the association. The same dependence on anion concentrations was found if we replaced the chromosomes by single stranded DNA immobilised to cellulose. DNA polymerase α is eluted from single-stranded DNA cellulose columns by more than 100 mM KCl but we have to use more than 200 mM acetate to elute the activity (7).

Thus the salt behaviour of DNA polymerase α can explain the salt sensitivity of the replication system and confirms our suggestion that this enzyme is required for DNA replication.

ACKNOWLEDGEMENTS

We thank Dr. R. Knippers for his encouragement during the periods of this work, his support and discussions with him. We are greatly indebted to our coworkers R. Mettke, Dr. K.-H. Klempnauer, Dr. E. Fanning, M. Baynes, R. Scheu, H. König, D. Riedel and M. Schwentorat for their help and to our colleagues Drs. E.F. Schlaeger and H. Stahl for helpful discussions. We thank Drs. F. Delamarter and D. Thatcher for their comments on the manuscript.

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 138/B4), by the University of Konstanz (Grundmittel), by a short term fellowhip to B.O. from the European Molecular Biology Organization and during the preparation of this manuscript by Biogen S.A.

REFERENCES

1. Kornberg, A., DNA replication. W.H. Freeman & Co., San Francisco, 1980.
2. Tooze, J., Molecular Biology of Tumor Viruses. Cold Spring Harbor Lab. Press, New York, 1973.
3. Tegtmeyer, P., J. Virol 10: 591-598, 1972.
4. Su, R.T. and De Pamphilis, M.L. Proc. Natl. Acad. Sci., USA. 73: 3466-3470, 1976.
5. Edenberg, H.J., Waqar, M.A. and Huberman, J.A.. Proc. Natl. Acad. Sci. USA 73: 4392-4396, 1976.
6. Otto, B. and Fanning, E. Nucleic Acids Res. 5: 1715-1728, 1978.

7. Richter, A., Scheu, R. and Otto, B. Eur. J. Biochem. 109: 67-73, 1980
8. Richter, A., Otto, B. and Knippers, R. Nucleic Acids Res. 9: 3793-3807, 1981
9. Otto, B., Fanning, E. and Richter, A. Cold Spring Harbor Symp. Quant. Biol. 43: 705-708, 1979.
10. Otto, B., Baynes, M. and Knippers, R. Eur. J. Biochem. 73: 17-24, 1977.
11. Edenberg, H.J. and Huberman, J.A. Annu. Rev. Genet. 9: 245-284, 1975.
12. Klempnauer, K.H., Fanning, E., Otto, B. and Knippers, R. J. Mol. Biol. 136: 359-374, 1980.
13. Hasselbach, W. In: Physiologie des Menschen, Verlag Urban & Schwarzenberg, München, 1971
14. Seidman, M.D., Garon, C.F. and Salzman, N.P. Nucleic Acids Res. 5: 2877-2893, 1978.
15. Edenberg, H.J., Waquar, M.A. and Huberman, J.A. Nucleic Acids Res. 4: 3038-3095, 1977.
16. Su, R.T. and De Pamphilis, M.L. J. Virol. 28: 53-65, 1978.
17. Richter, A. Thesis University of Konstanz, 1981.
18. Fox, A.M., Breaux, C.B. and Benbow, R.M. Dev. Biol. 80: 79-95, 1980.
19. Spadari, S., Sala, F. and Pedrali-Noy, G. Trends in Biochemical Sciences Jan: 29-32, 1982.
20. Edenberg, H.J., Anderson, S. and De Pamphilis, M.L. J. Biol. Chem. 253: 3778-3780, 1978.
21. De Pamphilis, M.L. et al. Cold Spring Harbor Symp. Quant. Biol. 43: 679-692, 1979.
22. Tsubota, Y. Waqar, M.A., Burke, J.F., Milavetz, B.J., Evans, M.J., Kovalski, D. and Huberman, J.A. Cold Spring Habor Symp. Quant. Biol. 43: 693-704, 1979.

4

ORGANIZATION AND REPLICATION OF CHROMOSOME-ASSOCIATED POLYOMA VIRUS DNA AND FLANKING CELLULAR SEQUENCES IN POLYOMA-TRANSFORMED RAT CELLS

HAIM MANOR, AHARON TOMER, NAVA BARAN, ADA NEER AND ELLA MENDELSOHN

SUMMARY

In this chapter we review studies on the organization and replication of polyoma virus (Py) DNA sequences which have integrated into chromosomal DNA, and of cellular DNA flanking the Py integration site, in an inducible line of polyoma-transformed rat cells designated the LPT line. We also discuss the relationship between replication of the chromosome-associated viral DNA and excision of free viral DNA molecules in LPT cells treated with virus-inducing agents. These studies have shown that LPT cells contain multiple copies of Py DNA integrated into a single chromosomal site. These copies are not evenly distributed among the cells. Instead, various Py insertions include 1, 2, 3,...whole viral genomes arranged in a direct tandem repeat. A 3.0 kilobase pairs (kb) deletion of cellular DNA was discovered next to the Py integration site. LPT cells were found to be heterozygous with respect to the viral insertions and the cellular deletion. In LPT cultures treated with mitomycin c, the most effective inducing agent in this system, replication of Py DNA is enhanced after lag periods which vary in individual cells between 9 and 20 hours. DNA synthesis is repeatedly initiated within the viral insertions and the replication forks proceed in opposite directions. The leftward moving forks were analyzed in detail and found to cross the left viral DNA-cell DNA boundary and to proceed into the flanking cellular DNA sequences. Fork movement is halted within a 0.30 kb segment which maps about 2.0 kb away from the boundary. This segment may include a termination site of a normal cellular replicon. Similar results were obtained in LPT cells exposed to bromodeoxyuridine and ultraviolet light. Extrachromosomal circular Py DNA monomers are excised and replicated in induced LPT cells after the lag phase. The viral DNA monomers are encapsidated and converted into infectious virions. Excision of free

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

viral DNA molecules and virus synthesis do not occur in an LPT subclone which contains chromosome-associated tandemly repeated viral DNA, but in which replication of the viral DNA cannot be induced. Models of papovavirus integration and induction based on studies of the LPT line and of other Py and SV40 transformants are discussed.

INTRODUCTION

The LPT line of polyoma-transformed rat cells was isolated by Fogel and Sachs (1) after infection of rat embryo muscle cells with a large plaque strain of the polyoma virus (Py). LPT cells possess the morphological characteristics of the transformed state and synthesize Py tumor antigens. A small percentage of the cells in each LPT culture (<0.20%) synthesize virus coat proteins and infectious virus. This property, designated as "spontaneous induction", is hereditarily transmitted, as indicated by the observation that cultures of 92 out of 100 clones and subclones derived from the LPT line include virus-synthesizing cells (1). The percentage of virus-producing cells can be artificially increased up to 60% by treating LPT cultures with various physical and chemical virus-inducing agents that cause DNA damage (1,2,3,4), or inhibit protein synthesis (5), and by fusing the cells with mouse cells that are permissive for polyoma virus synthesis (1).

In this review article we describe experiments which showed that in uninduced LPT cells the Py DNA is integrated into chromosomal DNA and that in induced cells extrachromosomal Py DNA monomers are generated from the integrated viral DNA. We present an analysis of the organization and replication of the chromosomally associated Py DNA and flanking cellular DNA sequences in normally growing LPT cultures and in LPT cultures treated with inducing agents. Finally, we discuss our findings in relation to current ideas on integration and replication of Py and SV40 DNA in cell lines transformed by these viruses and on the mechanism of virus induction in these transformants.

RESULTS

Quantity and state of the Py DNA in normally growing LPT cells

We have initially used RNA-DNA hybridization assays to determine the concentration of Py DNA in LPT cells. Synthetic Py-specific RNA (cRNA) labeled with ^{3}H at a high specific radioactivity ($>10^9$ dpm/μg)

was hybridized with whole cell DNA, or with high molecular weight chromosomal DNA purified from the cells by alkaline glycerol gradient centrifugation (Fig. 1). The RNA-DNA hybrids were collected on

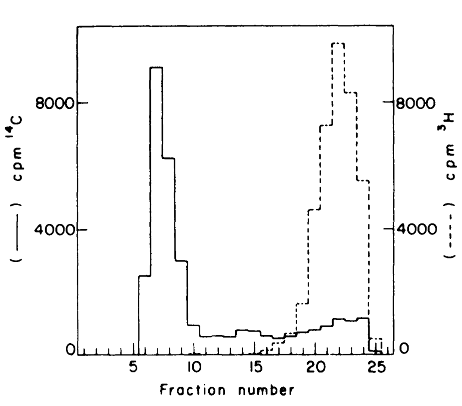


FIGURE 1. Alkaline glycerol gradient centrifugation of LPT cell DNA mixed with Py virions

LPT cultures were labeled with ^{14}C-thymidine. The cells were removed with trypsin and mixed with Py virions whose DNA had been labeled with ^{3}H-thymidine. The mixture was layered on an alkaline glycerol gradient which was then centrifuged in a preparative ultracentrifuge. Fractions were collected and counted in a scintillation counter. This figure was reproduced from a paper published in Virology (6) by permission. The paper presents further experimental details. It can be seen that the fast sedimenting chromosomal DNA peak does not contain any detectable traces of the ^{3}H-labeled viral DNA. A similar resolution was obtained in other control experiments in which purified viral DNA, rather than virions, had been mixed with the cells before centrifugation. Chromosomal DNA isolated by this method was used for the hybridization assays described in the text.

nitrocellulose filters and counted in a scintillation counter. Appropriate calibration curves were constructed by hybridizing the cRNA with mixtures of normal rat DNA and known amounts of purified Py DNA. The curves were used to estimate the amounts of viral DNA in the LPT cell DNA preparations. These studies revealed that normally growing LPT cultures contain 30-40 Py genome-equivalents (g.e.) per cell, of which 6-9 g.e. were found in the high molecular weight chromosomal DNA fraction (6).

It is to be noted that cosedimentation of viral DNA sequences with the high molecular weight chromosomal DNA could not be considered as a rigorous demonstration of covalent linkage between the viral sequences and chromosomal DNA. It was possible, for example, that LPT cells

synthesize oligomers consisting of viral DNA molecules linked in a head to tail configuration and that these molecules cosediment with the chromosomal DNA. Indeed, such oligomers of viral DNA were observed in SV40-infected cells (7). However, in a subsequent section, we shall provide a more rigorous proof for the presence of Py DNA linked to LPT chromosomal DNA.

The assays described above have indicated that most of the viral DNA in LPT cells is extrachromosomal and subsequent experiments have shown that the extrachromosomal viral DNA consists mainly of open and closed circular monomers (Baran, N. and Manor, H., unpublished results). We have used in situ hybridization analysis to find out whether these monomers are derived from the chromosomally associated viral DNA, or are propagated autonomously as plasmids. The cells were fixed on microscope slides, the intracellular DNA was denatured in situ and hybridized with Py-specific ^{3}H-cRNA. The slides were coated with a photographic emulsion and exposed for autoradiography. Fig. 2a shows a typical autoradiogram in which only cell nuclei are visible because the cytoplasms were not stained. One nucleus in the field contains about 100 grains, whereas the number of grains in the rest of the nuclei and in all the cytoplasms is equal to the background level. A quantitative analysis has indicated that each grain in this particular assay corresponds to about 200 g.e. of Py DNA and thus the single labeled nucleus contains about 20,000 molecules of Py DNA (8). A survey of many similar fields revealed that less than 0.10% of the cells contained a significant number of grains. These results have indicated that most or all of the extrachromosomal viral DNA molecules are derived from the small population of spontaneously induced cells and that, therefore, the rest of the cells contain primarily or exclusive chromosomally associated viral DNA in amounts that are below the level of detection of our in situ hybridization assays.

Induction of Py DNA replication in LPT cells treated with mitomycin C

Fig. 3 shows the time course of Py DNA replication in LPT cells treated with mitomycin C (MMC) which is the most effective inducing agent in this system. It can be seen that during the initial 9 hr which follow MMC treatment the ratio Py DNA/total cell DNA in these cells remains equal to the ratio found in uninduced LPT cells. At the end

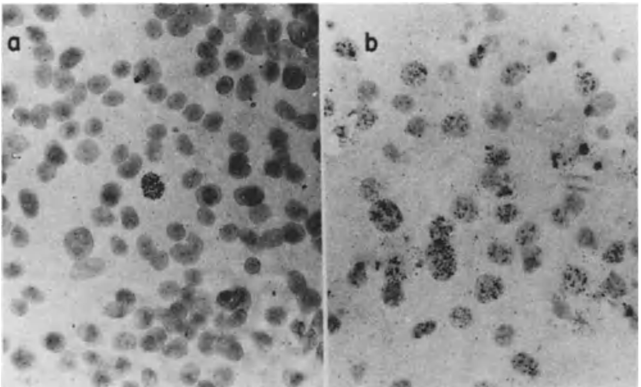


FIGURE 2. In situ hybridization analysis of untreated and MMC-treated LPT cultures. a) An LPT culture maintained under normal growth conditions was hybridized in situ with Py ^{3}H-cRNA and then exposed for autoradiography. This photograph was reproduced from a paper published in Cell (8) by permission. The paper presents further details of the experimental procedures. b) An LPT culture was treated with MMC, harvested after 33 hr and hybridized in situ with Py ^{3}H-cRNA. This photograph has also been published in cell (8) and reproduced by permission.

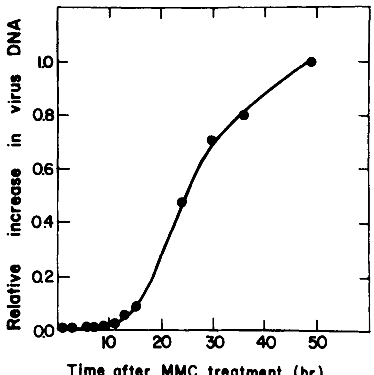


FIGURE 3. Time course of Py DNA replication in LPT cells treated with mitomycin C.

LPT cultures were exposed to 1 μg/ml of MMC for 1 hr, as described in reference No. 5, and harvested after various time periods. Whole cell DNA prepared from each culture was hybridized with synthetic Py ^{3}H-cRNA to determine the amount of Py DNA (6). The ordinate denotes the ratio Py DNA/total cell DNA relative to the maximal value which is given the value 1.0.

of this period the intracellular concentration of Py DNA increases exponentially for about 10 hr, after which time it continues to increase at a reduced rate for at least 30 hr.

The data presented in Fig. 3 have not provided information on whether the viral DNA replicates synchronously in all MMC-induced cells

or whether instead individual cells are recruited to replicate the viral DNA at different times after the initial 9 hr lag period. This question was examined by in situ hybridization analysis of the time course of Py DNA replication in MMC-treated LPT cultures. One autoradiogram of a culture harvested 33 hr after MMC treatment is shown in Fig. 2b. In this and in other autoradiograms the induced cells, which contained 30-120 grains, were easily discernible from uninduced cells which contained no grains. Table 1 represents a summary of the results obtained in this

Table 1. In situ hybridization analysis of the kinetics of Py DNA replication in LPT cultures treated with mitomycin c.[a]

Time after MMC Treatment (hr)	Percentage of Induced Cells	Average Number of Grains per Cell	Relative Increase in Virus DNA[b]
0	0.04	40	1.0
9	0.06	57	2.1
12	2.4	60	90
15	10	96	600
18	13	96	780
33	30	100	1875

[a]LPT cultures were treated with MMC, harvested at the indicated times and analysed by in situ hybridization. This table has been reproduced from an article published in Cell (8) by permission. The paper presents experimental details.

[b]Each number represents a product of the percentage of induced cells and the average number of grains per cell normalized with respect to the product found at time 0, which is given the value 1.0.

study. It can be seen that the overall kinetics of viral DNA replication deduced from these data is similar to the kinetics presented in Fig. 3. It is also evident that the accumulation of viral DNA after the lag period is mainly a consequence of an increase in the fraction of cells synthesizing viral DNA and not in the amount of virus DNA synthesized within individual cells. We conclude that different cells in each culture initiate Py DNA replication at different times after the initial lag period and synthesize most of the viral DNA which they eventually accumulate within a period of less than 6 hr.

Table 2. Amplification of chromosomally associated Py DNA in LPT cells treated with MMC[a].

Time after MMC Treatment (hr)	Py genome equivalents/cell		$\frac{(1)}{(2)} \times 100$
	Chromosomal DNA (1)	Whole Cell DNA (2)	
0	7	31	22
6	7	29	24
7	12	66	18
24	124	1380	9
6(superinfected)[b]	8	3023	0.26

[a]LPT cells were treated with MMC as described in the legend to Fig. 3 and harvested at the indicated times. Chromosomal DNA was purified by alkaline glycerol gradient centrifugation as shown in Fig. 1. The number of Py genome-equivalents in chromosomal DNA and in whole cell DNA prepared from each sample was determined by hybridization with Py ^{3}H-cRNA (6).

[b]In this control experiment LPT cells were treated with MMC, then superinfected with polyoma virions and harvested after 6 hrs. The result of this experiment confirms that chromosomal DNA purified by alkaline glycerol gradient centrifugation contains virtually no residual free Py DNA monomers (see Fig. 1).

Amplification of Py DNA cosedimenting with chromosomal DNA in MMC treated LPT cells

We have also carried out cRNA hybridization assays of high molecular weight chromosomal DNA preparations isolated by alkaline glycerol gradient centrifugation from LPT cells treated with MMC and harvested after various time periods. In parallel we determined the amounts of Py-specific sequences in whole cell DNA preparations obtained from the same cultures. Table 2 shows the results of these assays. It can be seen that the amount of chromosomally associated and total Py DNA in cells harvested 6 hr after MMC treatment remained equal to the corresponding amounts found in normally growing cells, in agreement with the data presented in Fig. 3 and in Table 1. Both quantities increased slightly after 7 hr. Twenty-four hrs after MMC treatment, not only the total amount of Py DNA increased 45-fold, but also the amount of chromosomally associated Py DNA sequences increased 17-fold. In a sub-sequent section we shall describe experiments in which, by using a more

direct approach, we proved that covalently linked viral and cellular DNA sequences are amplified in MMC-treated LPT cells. It should be noted that the rate of induction of Py DNA replication was found to be variable in different experiments, even though the same LPT clonal isolate was used for these experiments (Compare the data presented in Tables 1 and 2 and in Fig. 3). We do not know the reason for this variability.

Mapping the integration site of Py DNA in LPT cells

The development of the Southern blotting technique paved the way for detailed mapping of integrated viral genomes in transformed cells (9,10). We used this method to map the Py DNA in the LPT line (11).

Three types of restriction enzymes were used for this project. (a) Enzymes that cleave the Py DNA once (one cut enzymes), (b) enzymes that do not have a cleavage site within the Py genome (no cut enzymes) and (c) enzymes which cut the Py DNA at two or more sites (multiple cut enzymes). Digestion of LPT chromosomal DNA with each of four one cut enzymes was found to generate whole linear viral DNA molecules and two additional fragments containing both viral and cell DNA sequences. Digestion of LPT chromosomal DNA with no cut enzymes was found to generate a series of fragments whose lengths differ by increments of a whole Py genome. The pattern of integration deduced from these data was as follows: Py insertions of various lengths are located at the same chromosomal site in all LPT cells. The length heterogeneity is due to the presence of 1, 2, 3,...whole Py genomes inserted in tandem within invariable sequences of viral DNA and flanking cellular DNA. Fig. 4 shows the first three insertions in the postulated series. The arrangement of the Py DNA sequences within the insertions has been determined by double digestion of LPT chromosomal DNA with the three types of enzymes mentioned above. These data also allowed physical mapping of the flanking cellular DNA. Fig. 5 shows a comprehensive physical map of the Py integration site containing the smallest viral insertion. This map includes additional information to be discussed below.

Cloning of cell DNA sequences flanking the Py insertions

In order to carry out a more detailed analysis of the cell DNA sequences flanking the Py integration site in LPT cells, we have used the "shotgun" approach to clone the flanking sequences in the bacteriophage

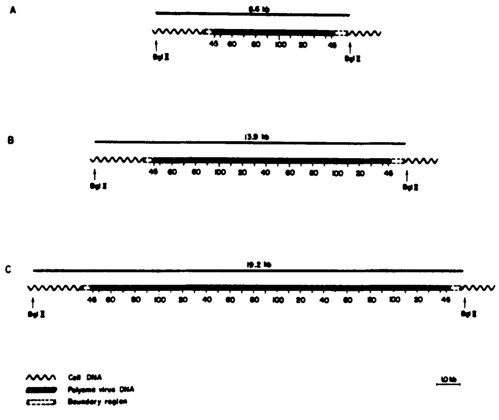


FIGURE 4. Illustration of three Py DNA insertions of varying lengths integrated into identical chromosomal sites in LPT cells

The regions marked as "polyoma virus DNA" are colinear with the standard Py physical map (12) and are divided into Py map units (1.0 Py map unit = 53 base pairs). The boundary regions represent the uncertainty limits of the positions of the cell DNA-virus DNA junctions. Fragments generated upon cleavage of each of these structures with the no cut enzyme BglII are drawn above the corresponding maps and their sizes are indicated. This figure was reproduced from a paper published in the Journal of Virology (11) by permission.

vector λgtWES (13). We prepared in this phage vector a library of LPT chromosomal DNA fragments obtained by partial digestion of the LPT DNA with the enzyme EcoRI. The library was screened by the technique of Benton and Davis (14) for recombinant phage clones containing Py DNA. Three such clones were identified and isolated out of about 1.5 x 10^6 recombinant phages that have been screened. One of these phage isolates contained two whole Py DNA monomers arranged in a head to tail configuration. The presence of this structure in the recombinant phage confirmed the integration scheme illustrated in Fig. 4. The second of these recombinant phages contained Py DNA and flanking cell DNA. A 0.90 kb segment of cellular DNA was later prepared from this phage and subcloned into the plasmid pBR322. This segment was found to map next to the Py insertions on their left side, as shown in Fig. 5. The third recombinant phage has not been characterized as yet.

LPT cells are heterozygous with respect to the Py insertions

We have used the cloned 0.90 kb flanking cell DNA segment as a hybridization probe for a Southern blot analysis of LPT and normal rat DNA digested with various restriction enzymes. This study revealed that LPT cells are heterozygous with respect to the Py insertions. Fig. 6

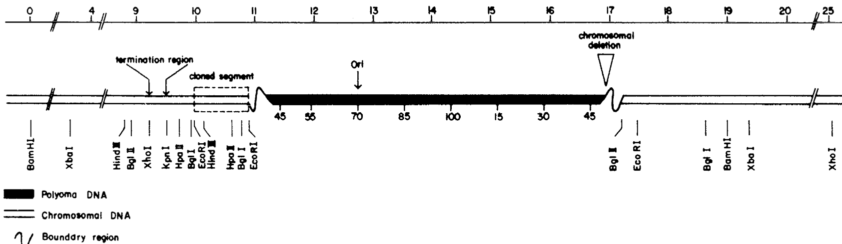


FIGURE 5. A comprehensive physical map of the Py integration site in LPT cells. This map presents the smallest Py insertion shown in Fig. 4A. It also shows the positions of the cloned cell DNA segment, the cellular deletion and the cellular termination site for replication. These elements are found next to all types of insertions and are described in the text. The map is based on data presented in references Nos. 11, 15 and 16. Ori désignates the origin of replication of the viral DNA.

illustrates the evidence which led to this conclusion. It shows two homologous LPT chromosomes of which only one contains a Py insertion but both contain the cloned flanking cell DNA segment. The chromosome which does not carry viral DNA is identical to the corresponding two chromosomes in normal rat cells which are also illustrated in Fig. 6. A restriction enzyme which has a cleavage site within the viral DNA,but has no cleavage site within the cloned flanking cell DNA segment, would therefore generate from LPT chromosomal DNA two fragments which would hybridize with the cloned cell DNA segment. One of these fragments would be identical to the single fragment generated from the normal rat DNA. Fig. 6 also shows a blot of KpnI digests of LPT and normal rat DNA hybridized with the cloned cell DNA segment. This blot displays the patterns of cleavage shown in the diagram, and described above. We carried out a similar analysis of LPT and normal rat DNA digests generated by treating the DNAs with many other restriction enzymes. These data were used to construct physical maps of the unoccupied integration site in LPT cells and the corresponding site in normal rat cells. The maps were found to be exactly identical (15).

Identification of a cellular deletion which maps next to the Py insertions

A comparison between the maps of the cellular sequences flanking

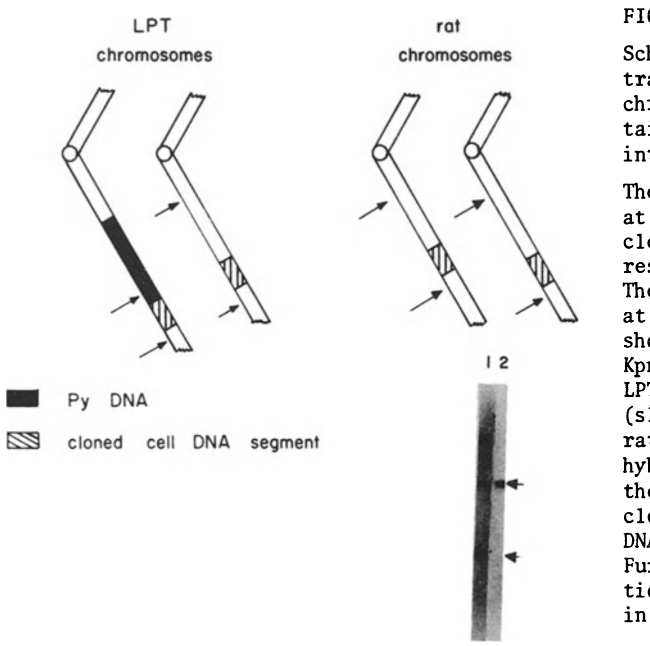


FIGURE 6.

Schematic illustrations of chromosomes containing the Py integration site

The arrows point at sites of cleavage by a restriction enzyme. The photograph at the bottom shows blots of KpnI digests of LPT chromosomal DNA (slot 1) and normal rat DNA (slot 2) hybridized with the 0.90 kb cloned cellular DNA segment. Further explanations are presented in the text.

the Py insertions in LPT cells and of the corresponding sequences in normal rat cells have indicated that a 3.0 kb segment of cellular DNA has been deleted from the LPT chromosome which carries Py DNA (15). The deletion was found to map next to the viral insertions, as indicated in Fig. 5. The homologous LPT chromosome which does not carry Py DNA does not have a deletion.

<u>Amplification of integrated Py DNA and flanking cell DNA sequences and identification of a cellular termination site for replication</u>

We have used the techniques developed for mapping the Py integration site and the information gained in these studies to reexamine the observation that the chromosomally associated viral DNA is amplified in LPT cells treated with MMC. For this purpose chromosomal DNA was selectively extracted from normally growing and MMC-treated LPT cells. Samples of these DNAs were cleaved with restriction enzymes and the cleavage products were analysed by the Southern technique. Fig. 7 shows the results of one experiment in which the DNAs were digested successively with the two enzymes HpaII and HindIII. The blot prepared

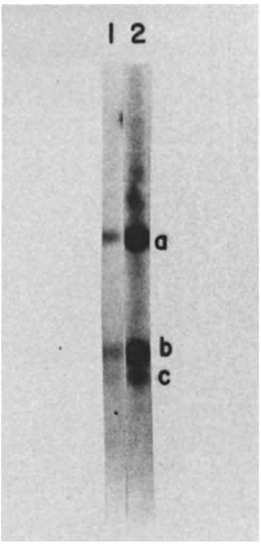


FIGURE 7. Amplification of specific fragments containing Py DNA and/or flanking cellular sequences in MMC treated LPT cells

Ten µg of chromosomal DNA from normally growing LPT cells (slot 1), and from LPT cells treated with MMC and harvested after 24 hr (slot 2), were digested with the two enzymes HpaII and HindIII. The digests were analysed by gel electrophoresis and Southern blotting. The blots were hybridized with the cloned cell DNA fragment whose map position is shown in Fig. 5, and exposed for autoradiography. Details of the experimental procedures are presented in reference No. 11 and further explanations are included in the text.

from these digests was hybridized with the 0.90 kb cloned cell DNA segment. It can be seen that the same three fragments are generated from both samples of DNA, but that their concentration is much higher in the DNA from the MMC treated cells. The fragment designated as a extends from the HindIII site which maps at position 45 on the left side of the viral insertion shown in Fig. 5 to the HpaII site within the cloned segment. This fragment therefore includes both Py and cell DNA sequences. Fragment b extends from the HpaII site to the HindIII site in the cloned segment and fragment c extends from this HindIII site to the next HpaII site beyond the left side of the cloned segment. Thus, fragments b and c contain just cell DNA sequences.

Fragments generated with many other enzymes were similarly shown to be amplified in chromosomal DNA isolated from MMC treated cells (16). Amplification of the viral sequences in fragments containing both Py and cell DNA sequences was also directly demonstrated by hybridization with Py DNA. These studies have shown that MMC treatment of LPT cells causes several rounds of replication of the chromosomally associated Py DNA and that the replication proceeds beyond the left viral DNA-cell DNA junction into the flanking cell DNA sequences.

One question raised by these findings was whether the replication ends at a specific site within the flanking cell DNA, or whether, instead,

it ends at many different sites. To examine this question we have analysed the amplification of fragments ending at more distant positions within the cell DNA. Our results have indicated that fragments whose ends map more than 3 kb away from the left viral DNA-cell DNA joint are not amplified. Therefore, we concluded that the replication ends at a site (or sites) located less than 3 kb away from the junction. A more refined analysis revealed that the replication ends within a 0.30 kb segment designated in Fig. 5 as a termination region. A similar analysis of DNA extracted from LPT cells exposed first to bromodeoxyuridine and then to ultraviolet light generated the same results, namely that replication induced within the Py insertion extends into the cell DNA sequences and ends within the segment designated in Fig. 5 as termination region. A detailed account of these studies is presented elsewhere (16).

DISCUSSION

Integration

Our studies of the patterns of integration of Py DNA in LPT cells have shown that the viral DNA is integrated into the same chromosomal site in all the cells. However, the Py insertions were found to have variable lengths due to the presence of 1, 2, 3,..whole viral genomes arranged in a direct tandem repeat (Fig. 4). Attempts to isolate LPT subclones containing just one type of insertion have failed. Of 30 subclones that we have isolated and analysed, most contained multiple insertions of the type described above and in a few subclones we observed additional rearrangements of the viral sequences (17).

One possible interpretation of these data is that each LPT cell contains the entire set of insertions described above, and that a few cells contain additional types of insertions which are not visible in the blots prepared from digests of DNA of the original LPT clone. This interpretation implies that the entire chromosomal segment including the Py insertions and the flanking cellular sequences, extending from the BamHI site on the left side of the map shown in Fig. 5 to the XhoI site on the right side of the map, i.e. a Py insertion plus 19 kb of cellular DNA, has been duplicated at least 5 times. However, a quantitative analysis has indicated that the flanking cellular DNA is unique in the rat genome (Baran, N. and Manor, H., unpublished results). Thus, it appears that each LPT cell contains just one insertion, but the lengths

of the insertions vary among the cells. Presumably, the single cells from which the subclones arose contained one type of insertion, but during the 10-15 cell generations needed to grow enough cells for DNA preparation the other insertions were generated in each of the subclones by deletion or duplication of whole Py genomes. This interpretation is supported by a recent report that duplication and deletion events involving homologous segments of chromosomally associated viral DNA are very frequent in Py transformed rat cells (18,19). Both the loss and amplification of Py sequences depend on the presence of an active big T (tumor) antigen, which is a Py encoded protein needed for initiation of replication of the viral DNA (ibid.).

Attempts to deduce the mechanism of integration of Py and SV40 into the cell DNA from data on the organization of the viral insertions in transformed cells are subject to the criticism that the insertions do not necessarily represent the structures created by the initial integration event; for, as described above, integrated viral DNA sequences may undergo rearrangements. Nevertheless, there are some common features in Py and SV40 insertions which appear to exclude certain mechanisms and to favor a recently proposed model which is discussed below. One common feature is the prevalence of partial and complete duplications in the viral insertions which are often arranged in a direct tandem repeat, like the Py insertions in LPT cells. This feature has led to the suggestion that the immediate precursor to papovavirus integration is a molecule containing tandemly joined viral monomers (20). One molecular species which could serve in this role is a rolling circle intermediate in viral DNA replication. Rolling circle intermediates were observed in both SV40- and Py-infected cells (21,22), even though the major pathway for replication of these viruses is the Cairns θ mechanism (22,23,24). Another possible candidate for the role of integration precursor is an oligomer generated by recombination of viral monomers. Both linear and circular oligomers have been observed in SV40-infected cells (7).

None of the various papovavirus insertions analysed so far has a structure which could have been generated by a single crossover between a circular viral precursor, either monomeric or oligomeric, and the cell DNA. On the other hand, many of the insertions, including those observed in LPT cells, could arise by double crossover between the cell

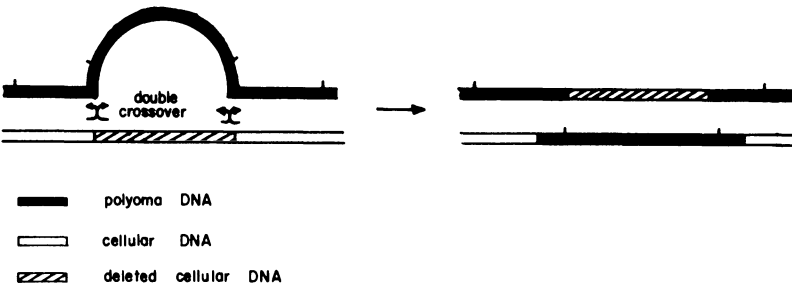


FIGURE 8. A model for papovavirus integration. The short vertical lines indicate boundaries between tandemly joined viral monomers. The viral DNA precursor, of which only a portion is shown in this figure, could be either one of two species described in the text. The length of the deletion in LPT cells is 3.0 kb and the length of the smallest Py insertion is 5.9 kb. Therefore the deletion shown here is smaller than the insertion. Other transformed lines should be examined to find out whether the presence of a deletion is a general characteristic of papovavirus transformed cells and to examine the relationship between sizes of deletions and insertions. The essential features of this model have been proposed by Botchan (see ref. 20) and by Chia and Rigby (7).

DNA and one of the two viral precursors mentioned above. Such a mechanism would necessarily generate a deletion of cellular DNA at the integration site, as shown in Fig. 8. Our discovery of the cellular deletion in LPT cells is therefore in line with this model. Cellular deletions have also been recently discovered at the SV40 sites of integration in three SV40-transformed rat cell lines (25,26).

Many studies have indicated that there are no specific sites for integration of papovavirus DNA on either the viral or the cellular genomes (20). Recently, it was demonstrated that in one SV40-transformed rat cell line, the SV40 and the rat genomes share a 5 base pair DNA sequence homology at the point of integration of the two genomes (25). Thus, the first interaction between the viral and the cell DNA, which leads to integration, may involve the matching of short stretches of homologous sequences.

Presumably, a double crossover event of the type shown in Fig. 8

happened only once in the history of the LPT line; for only one cellular integration site exists in LPT cells and only one of the two sites on homologous chromosomes harbors Py DNA. This initial insertion could have been one of the structures shown in Fig. 4. The whole series of insertions may have been generated from the initial insertion by unequal crossover between chromatids, as depicted in Fig. 9, or by excision and reintegration of whole Py genomes.

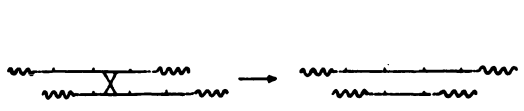


FIGURE 9. Unequal crossover between homologous Py segments on two chromatids

This process generates a deletion of a whole viral genome in one chromosome and an addition of a whole viral genome in the other chromosome.

Induction

Polyoma virus synthesis can be induced in LPT cells by various physical and chemical agents which cause damage to DNA and possess carcinogenic and mutagenic activity. These agents include ultraviolet light and X-irradiation (1,2), nitrogen mustard, N-nitrosomethylurea, mitomycin C, bromodeoxyuridine (BUdr), iododeoxyuridine and fluorodeoxyuridine (2,3). Pretreatment of LPT cells with BUdr followed by ultraviolet- or X-irradiation was more effective than either of these agents used alone (3). This result supports the notion that the activity of these agents as virus inducers is related to their ability to cause damage to DNA. Carcinogenic polycyclic hydrocarbons such as benzo(a)-pyrene, 20-methylcholanthrene and 7,12-dimethylbenz(a)anthracene were active as inducing agents only after being metabolized to reactive intermediates (4). This result suggested that virus induction depends on metabolic conversions of these hydrocarbons into the same reactive compounds that are responsible for malignant transformation and mutagenesis. Non-carcinogenic polycyclic hydrocarbons such as pyrene, chrysene and benz(a)anthracene were not active as virus inducers in LPT cells (4).

In addition to agents which cause damage to DNA, fusion of LPT cells with mouse cells which are permissive for Py infection, and exposure to

compounds which inhibit various steps in protein synthesis were also found to increase the rate of Py induction in LPT cells above the spontaneous level (1,5). The protein synthesis inhibitors may cause virus induction indirectly by inhibiting cell DNA replication. The rate of induction observed after treating LPT cells with different agents was found to vary over a wide range. Inhibitors of protein synthesis are the least efficient and cause just an 8-fold increase in the fraction of virus-synthesizing cells above the spontaneous level. The most efficient inducing agent for LPT cells is mitomycin C which causes up to a 1000-fold increase in the fraction of virus-synthesizing cells. A correlation was found between the rates of induction of viral DNA replication and of virus coat protein synthesis, determined by nucleic acid hybridization and immunofluorescence assays, respectively (Manor,H. and Fogel, M., unpublished results).

In the Results we described a detailed study of the kinetics of Py DNA replication in LPT cultures treated with MMC. We observed that there is a minimal lag period of 9 hours before the onset of virus DNA replication in these cultures. For individual cells the lag varies between 9 and 20 hours (Table 1). Exposure of MMC treated cells to cycloheximide during the lag phase inhibits viral DNA replication. The inhibition is reversible - removal of the cycloheximide from the growth medium allows resumption of viral DNA replication after an additional time period (5). It appears, therefore, that a protein (or proteins) synthesized during the lag is required for viral DNA replication. It is presently not known whether this protein is encoded in the viral genome, or by a cellular gene.

The only Py-encoded protein which is known to be required for replication of the viral DNA is the big T antigen. This protein might be synthesized in uninduced LPT cells, because it has been demonstrated that uninduced cells contain big T mRNA as well as the other "early" mRNAs (27; Neer and Manor, unpublished results), and show intense nuclear staining in immunofluorescence assays of the Py T antigens (1). However, no biochemical assays of the T antigens have been carried out in LPT cells as yet, and thus the presence of big T in uninduced cells has not been directly demonstrated. Even the presence of biochemically identifiable big T antigen would not necessarily imply that it is functional in uninduced cells. It is to be noted in this connection that newly

synthesized SV40 big T antigen molecules bind to the SV40 origin of replication more efficiently than "older" molecules and therefore only newly synthesized big T may be active in viral DNA replication (28). In the case of the LPT line functional Py big T antigen may be synthesized only after transition of an LPT cell from the uninduced to the induced state. It is interesting in this regard that during the lag period which follows MMC treatment of LPT cells no qualitative or quantitative changes occur in the population of "early" mRNAs, including the big T mRNA. Only after the onset of Py DNA replication, synthesis of the "early" mRNAs is enhanced in these cells and synthesis of "late" mRNAs is initiated (27). Apparently, the onset of Py DNA replication is not dependent on synthesis of new big T mRNA molecules after MMC treatment.

In regard to the possibility that synthesis of a cellular protein during the lag phase is required for the onset of Py DNA replication, it is to be recalled that most of the agents which induce Py DNA replication in LPT cells also induce bacteriophage λ synthesis and SOS functions in E. coli cells (29). In E. coli one of the primary effects of these agents is induction of synthesis of proteases which inactivate repressors, such as the λ repressor, by proteolytic cleavage (29,30). It would be interesting to find out whether the same agents induce synthesis of proteases in LPT cells, and whether such proteases may be involved in Py induction.

Our studies of Py induction in LPT cells treated with MMC have indicated that at the end of the lag period the chromosomally associated viral DNA replicates and is amplified at least 20-fold. In addition, extrachromosomal viral DNA molecules are generated and replicated. About 20,000 viral DNA molecules are synthesized in each cell. Botchan et al (31), who analysed SV40 DNA replication in SV40–transformed rat cells fused with monkey cells, found that chromosomally associated viral DNA sequences are amplified before the appearance of extrachromosomal viral DNA molecules. These authors suggested that in these heterokaryons DNA synthesis is first initiated at the origin of the integrated SV40 DNA, and proceeds bidirectionally into the flanking cell sequences. Repeated activation of the origin thus generates a so-called "onion-skin" structure (Fig. 10). Botchan et al also proposed that the high local concentration of duplicated SV40 DNA in this structure, and the exposure

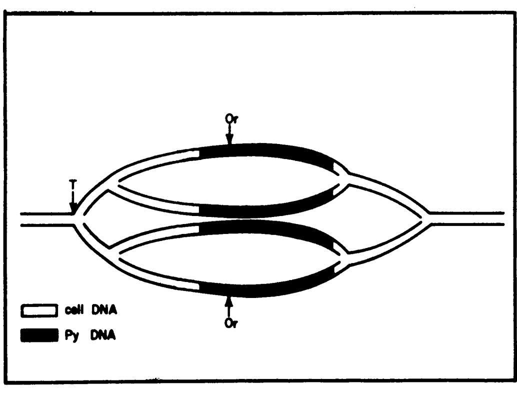


FIGURE 10 "Onion skin" model of replication

This model assumes that bidirectional replication is initiated from the origin (Or) of the integrated viral DNA and extends beyond the viral DNA-cell DNA junctions into the flanking cellular DNA sequences. In this diagram the leading leftward moving fork has reached a termination site(T).

of single-stranded regions at the replication forks, promote excision of circular viral DNA molecules by homologous recombination events. It is to be noted that replication of chromosomally associated SV40 DNA can also be induced by carcinogens and mutagens in SV40-transformed Chinese hamster cells (32).

Our data on amplification of specific restriction enzyme fragments containing viral and cellular sequences, or flanking cellular sequences alone (Fig. 7), support the "onion skin" model of replication. Kinetic analysis has indicated that in LPT cells replication of the integrated viral DNA may also precede the synthesis of extrachromosomal viral DNA. The fact that no excision of free Py DNA molecules was detected in LPT cells before the onset of viral DNA replication is consistent with the suggestion that replication of the integrated viral DNA promotes excision. The maps shown in Fig. 4 indicate that all the Py insertions found in LPT cells can readily give rise to non-defective circular Py DNA molecules by a single homologous recombination event; for even the shortest insertion contains a partially duplicated copy of the Py genome. The presence of more than one origin of replication in the longer insertions can give rise to several "onion skin" structures within one insertion, and consequently to more frequent recombination events. It is to be noted that excision of viral DNA monomers from "onion skin" structures by homologous recombination is not expected to remove all the viral DNA from the chromosomal integration site.

The data discussed above indicate that, unlike bacteriophage λ induction in E. coli, in which the primary effect of the various inducing agents on the phage metabolism is to allow transcription from the integrated phage genome by inactivation of the λ repressor (30), the primary effect of the same agents on Py metabolism in LPT cells might be to cause initiation of replication of the chromosomally associated Py DNA. The inducing agents, or effectors generated by these agents, might cause initiation of replication by interacting directly with the integrated Py DNA, or with proteins that are permanently associated with the Py DNA, e.g. histones, or with proteins which are required for initiation of Py DNA replication but are not permanently associated with the viral DNA. We hope to obtain some information on these interactions by analysing non-inducible subclones of the LPT line. About 10% of subclones isolated from LPT cells were found to be non-inducible either spontaneously or artificially. This is a rather high frequency and therefore, these subclones are probably not generated by point mutations; instead, the non-inducible subclones may arise from rearrangements of Py DNA sequences at the integration site, which occur at rather high rates, as discussed above.

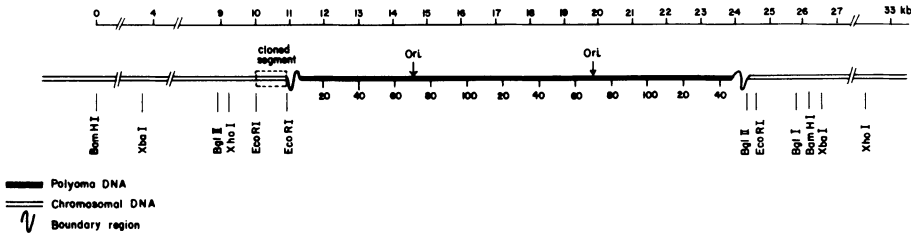


FIGURE 11 Physical map of a non-inducible subclone of the LPT line. For explanations see the text and the legends to figures 4 and 5.

We have mapped the Py insertion and flanking cellular sequences in one non-inducible LPT subclone (Fig. 11). This subclone contains a single Py insertion resembling in its structure the insertion shown in Fig. 4b, except that it includes an additional segment of Py DNA on its left side.

The flanking cellular sequences have not been altered. Neither replication of the integrated Py DNA, nor synthesis of free viral monomers can be induced in cells of this subclone. This observation implies that the presence of homologous DNA segments in the viral insertion is not by itself sufficient for production of free viral monomers. It is also interesting that this subclone contains only one type of insertion. Apparently, the generation of multiple insertions of the type shown in Fig. 4 is also dependent on replication of the chromosomally associated Py DNA. An attractive hypothesis is that the additional Py DNA sequences in this subclone have led to a change in the chromatin structure of the Py insertion and that this change blocks induction of replication of the insertion. This hypothesis can be indirectly examined by superinfecting cells of this and other non-inducible subclones with polyoma virus and by studying the replication of the superinfecting virus after treating the cells with various inducing agents. A more direct examination of the hypothesis would be an analysis of the chromatin structure of the segment including the Py origin of replication in uninduced and induced LPT cells and in cells of non-inducible LPT subclones.

Termination site for chromatin replication

We have used the LPT system to examine the question whether there are preferred termination sites for replication in mammalian chromatin. It is generally agreed that replication of mammalian chromatin occurs at many regions along the genome defined as replicons. Individual replicons are organized in clusters both spatially and with regard to timing of replication during the S-phase (33). Bidirectional replication appears to be initiated simultaneously at origins of adjacent replicons. The replication forks proceed until they meet. Eventually, the newly replicated DNA merges. It has been reported by several authors that the completion and merging of replicons represents a rate-limiting step in chromatin replication (34,35,36). Recent studies of SV40 mini-chromosome replication have suggested that the two SV40 replication forks which proceed from a common origin towards a common termination region stop at preferred DNA sequences located at this region (33,37). The retardation of replicon merging can be accounted for by the presence of similar sequences in mammalian chromatin at the junctions between adjacent replicons. The presence of DNA sequences that retard progress

of replication forks has been demonstrated in the E. coli chromosome (38,39). A terminator sequence identified in the E. coli plasmid R6K has been cloned into other plasmids and shown to arrest the progression of replication forks in these plasmids in vivo and in vitro (40,41).

The chromosomal segment including the Py insertions and flanking cellular sequences in LPT cells can be regarded as an artificially constructed replicon whose replication can be induced independently of other replicons. We have demonstrated that replication initiated within the Py insertions proceeds beyond the left cell DNA-viral DNA junction into the cellular sequences. Presumably, the replication fork proceeding to the left does not, in this case, meet another fork proceeding to the right, because the cellular replicons adjacent to the Py insertion are not activated. Nevertheless, the replication ends within a 0.30 kb segment whose position is shown in Fig. 5. This site in the chromatin could be a termination site for the cellular replicon into which the Py DNA happened to be inserted. We are now attempting to clone this chromosomal segment in a bacteriophage λ vector and hope to be able to determine its sequence soon. We also plan to carry out similar studies of replication forks which move into the cellular DNA sequences flanking the viral insertions on the right.

ACKNOWLEDGEMENTS

We thank the Leukemia Research Foundation and the United States-Israel Binational Science Foundation for financial support. We also thank Yaffa Bot for excellent technical assistance, Orit Segev for helping us to map the non-inducible LPT subclone and Jean Kenton for careful typing of the manuscript.

REFERENCES

1. Fogel, M., and Sachs, L. Virology 37:327-334, 1969.
2. Fogel, M., and Sachs, L. Virology 40:174-177, 1970.
3. Fogel, M. Virology 49:12-22, 1972.
4. Huberman, E., and Fogel, M. Int. J. Cancer 15:91-98, 1975.
5. Manor, H., and Neer, A. Cell 5: 311-318, 1975.
6. Manor, H., Fogel, M., and Sachs, L. Virology 53:174-185, 1973.
7. Chia, W., and Rigby, P.W. Proc. Natl. Acad. Sci. USA 78: 6638-6642, 1981.
8. Neer, A., Baran, N., and Manor, H. Cell 11: 65-71, 1977.
9. Botchan, M., Topp, C.W., and Sambrook, J. Cell 9, 269-287, 1976.
10. Ketner, G., and Kelly, T.J. Proc. Natl. Acad. Sci. USA 73:1102-1106, 1976.

11. Mendelsohn, E., Baran, N., Neer, A., and Manor, H. J. Virol. 41: 192-209, 1982.
12. Griffin, B.E., Soeda, E., Barrel, B.G., and Staden, R. In: DNA Tumor Viruses, Part 2, Second Edition (Ed. J. Tooze). Cold Spring Harbor Laboratory, 1980, pp.831-896.
13. Enquist, L., Tiemeier, D., Leder, P., Weisberg, R., and Sternberg, N. Nature, London 259: 596-598, 1976.
14. Benton, W.D., and Davis, R.W. Science 196: 180-182, 1977.
15. Neer, A., Baran, N., and Manor, H. Submitted for publication.
16. Baran, N., Neer, A., and Manor, H. Submitted for publication.
17. Mendelsohn, E., D. Sc. dissertation, Technion-Israel Institute of Technology, 1981.
18. Basilico, C., Gattoni, S., Zousias, D., and Della Valle, G. Cell 17: 645-649, 1979.
19. Colantuoni, V., Dailey, L., and Basilico, C. Proc. Natl. Acad. Sci, USA 77: 3850-3854, 1980.
20. Topp, W.C., Lane, D., and Pollack, R. In: DNA Tumor Viruses, Part 2, Second Edition (Ed. J. Tooze). Cold Spring Harbor Laboratory, 1980, pp. 205-296.
21. Bjursell, G. J. Virol. 26: 136-142, 1978.
22. Fareed, G.C., Garon, C.F., and Salzman, N.P. J. Virol. 10: 484-491, 1972.
23. Crawford, L.V., Robbins, A.K., and Nicklin, P.M. J. Gen. Virol. 25:133-142, 1974.
24. Danna, K.J., and Nathans, D. Proc. Natl. Acad. Sci. USA 69: 3097-3100, 1972.
25. Stringer, J.R. Nature 296: 363-366, 1982.
26. Botchan, M., Stringer, J. Mitchison, T., and Sambrook, J. Cell 20: 143-152, 1980.
27. Manor, H., and Kamen, R. J. Virol. 25: 719-729, 1978.
28. Prives, C., Oren, M., Gidoni, D., Shure, H., and Beck, Y. Cold Spring Harbor Symp. Quant. Biology 44: 123-130, 1979.
29. Witkin, E.M. Bacteriol. Rev. 40: 869-907, 1976.
30. Roberts, J.W., Roberts, C.W., Craig, N.L., and Phyzicky, E.M. Cold Spring Harbor Symp. Quant. Biology 43: 917-920, 1979.
31. Botchan, M., Topp, W., and Sambrook, J. Cold Spring Harbor Symp. Quant. Biology 43: 709-719, 1978.
32. Lavi, S. Proc. Natl. Acad Sci. USA 78: 6144-6148, 1981.
33. DePamphilis, M.L., and Wassarman, P.M. Ann. Rev. Biochem. 49: 627-666, 1980.
34. Funderud, S., Andreassen, R., Haugli, F. Nucl. Acids Res. 6: 1417-1428, 1979.
35. Kowalski, J., Cheevers, W.P. J. Mol. Biol. 104: 603-615, 1976.
36. Lanotte, M., Moerman, C. Panijel, J. Exp. Cell Res.109:191-200,1977.
37. Tapper, D.P., DePamphilis, M.L. J. Mol.Biol. 120:401-422, 1978.
38. Kuempel, P.L., Duerr, S.A., and Seeley, N.R. Proc. Natl. Acad. Sci. USA 74: 3927-3931, 1977.
39. Louarn, J., Patte, J., and Louarn, J.M. J. Mol. Biol. 115:295-314,1977.
40. Crosa, J., Luttrop, L., Heffron, L., and Falkow, S. Mol. Gen.Genet. 140:39-50, 1975.
41. Germino, J., and Bastia, D. Cell 23: 681-687, 1981.

5

REPLICATION OF ESCHERICHIA COLI CHROMOSOME

TUNEKO OKAZAKI

SUMMARY

Elongation of chromosomal DNA of Escherichia coli proceeds by a discontinuous mechanism whereby short DNA chains (nascent fragments) are synthesized and subsequently joined into long DNA. Studies on the metabolic and structural properties of the nascent fragments have elucidated the major steps of the discontinuous chain elongation: Synthesis of the nascent fragments is primed by oligoribonucleotides which are excised later with RNase H and by the **nick-translation** activity of DNA polymerase I. Resulting primer-free nascent fragments are joined into long DNA chains with DNA ligase. The biochemical basis for the discontinuous replication is provided by an elaborate replication system reconstituted in vitro for ϕX174 phage DNA.

A round of chromosome replication is controlled at the genomic origin where bidirectional replication forks are established. Thanks to the isolation of the genomic DNA segment responsible for an origin function (oriC) in the form of small plasmids, considerable progress has been made in dissecting the reactions at the origin in molecular level. RNA primed initiation sites of DNA synthesis have been shown to exist on only one of the component strands of oriC, which suggests that DNA synthesis starts unidirectionally at the oriC and then is converted to the bidirectional chain elongation.

INTRODUCTION

Chromosomal DNA of E. coli is a duplex circular molecule composed of 4 x 10^6 nucleotide pairs. Its replication initiates at a unique origin (oriC; 83.5 min on E. coli genetic map), proceeds bidirectionally with formation of a θ-type replication intermediate and terminates in a region diametrically opposed to the site oriC on the genetic map (1-5). The replication cycle of the chromosome is regulated by the initiation at the origin of the genome where two replication forks are established. Two daughter strands with antiparallel polynucleotide polarities are synthesized simultaneously and sequentially at

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

each replication fork. Discontinuous synthesis is the basic feature of the chain elongation mechanism for E. coli genome, as it is true also for many other organisms (6, 7). We proposed the discontinuous model of DNA synthesis based on the discovery of nascent DNA fragments in E. coli and other prokaryotic organisms (8-10). In recent years, we have been studying the metabolic and structural properties of nascent fragments of various organisms with special emphasis on the elucidation of the priming mechanism for the fragments (11-13). In the early part of this chapter, I will describe our work related to discontinuous replication in E. coli cells. Much success has been attained in reconstituting the process of discontinuous replication in vitro with purified protein (14-16): Many gene products categorized to function in a chain elongation step in vivo are now proved to be the essential components of the replication machinery which carries out discontinuous synthesis of DNA in vitro.

Complex biochemical events that occur at the replication origin of the E. coli are mostly unclear but major advance has been made last few years thanks to the recombinant DNA technology. The genomic region required for the start of replication has been isolated as an autonomously replicating plasmid (oriC plasmid) and the precise map position and nucleotide sequence of the origin area have been determined (17-20). The minimal essential region for ori-function has been further delimited to the 245 nucleotide region (oriC) and the effects of deletion or substitution of nucleotides within oriC have been analyzed in vivo (21-23). Recently, Fuller et al. have established a soluble enzyme system of E. coli which recognizes and replicates oriC plasmids or phage chimeras (24). Thus, the initiation events now can be dissected by biochemical procedures. Specifically what we are studying is the basic feature of the initiating DNA chain at the origin in vivo: Is the first DNA chain primed by RNA? If so, what is the feature of the priming site and the primer RNA? From where do the leading and the lagging strands for the clockwise and counterclockwise replication forks start? In the later part of the chapter, I will describe analyses related to the subject (12, 13, 25).

DISCONTINUOUS CHAIN ELONGATION OF E. COLI CHROMOSOME

Dual origin of nascent DNA fragments

Two daughter strands at each replication fork of E. coli genome are synthesized simultaneously as replication fork advances. Thus, overall polarities of chain elongation are 5' to 3' in one strand and 3' to 5' in the

other, while DNA polymerases synthesize DNA exclusively in 5' to 3' direction. To reconcile the paradox on the polarities of chain growth, the discontinuous model of DNA replication was put forward (9, 10). It proposed that either one or both strands of DNA at a replication fork be synthesized as short chains, which would be detectable as nascent fragments when the replicating portion of the chromosome is pulse-labeled with ^{3}H-thymidine. By the model, the daughter strand whose overall growth direction is 3' to 5' (the lagging strand) can be made by the repeated syntheses of short chains in the retrograde direction (5' to 3' direction) and successive joining of these short chains. Consistent with their postulated role as intermediates in DNA synthesis, the nascent fragments in the wild type strain of *E. coli* are incorporated into long DNA chains shortly after their synthesis (Fig. 1). Nascent fragments are known to be accumulated or overproduced on three occasions; under non-permissive conditions in the mutant cells defective in the DNA ligase (*lig*), and in DNA polymerase I (*polA*) or dUTPase (*dut*) activities (Fig. 1) (6). The majority of the nascent fragments which are accumulated in the *lig* or *polA* mutant cells indeed represent the intermediates of discontinuous DNA replication as will be described in the following sections, whereas the nascent fragments which are found in *dut* strain of *E. coli* are produced by a post-replication cleavage of uracil–containing nascent DNA and may be much shorter than the nascent fragments seen in wild, *polA* or *lig* strains of *E. coli* (compare Fig. A and B). Thus, at least two origins of nascent fragments are now evident, one produced by the discontinuous synthesis (so called Okazaki fragments) and the other by the repair process (pseudo-Okazaki fragments).

Function of a primer RNA and the major steps of discontinuous replication

The discontinuous mechanism requires repeated chain initiation but DNA polymerase cannot start a new chain and requires primers (15). To solve the problem of the initiation, an RNA primer model was built up which supposed a function of primer RNA made *de novo* for the synthesis of nascent fragments (26-28). According to the model, the process of chain elongation can be subdivided into the following steps, some of which may proceed simultaneously (Fig. 2): (i) unwinding of the parental strands; (ii) synthesis of an RNA primer utilizing a single-stranded template strand; (iii) extension of a DNA chain in the 5' to 3' direction from the primer terminus; (iv) removal of the primer; (v) filling in of the gaps between the DNA pieces; (vi) joining of the DNA pieces.

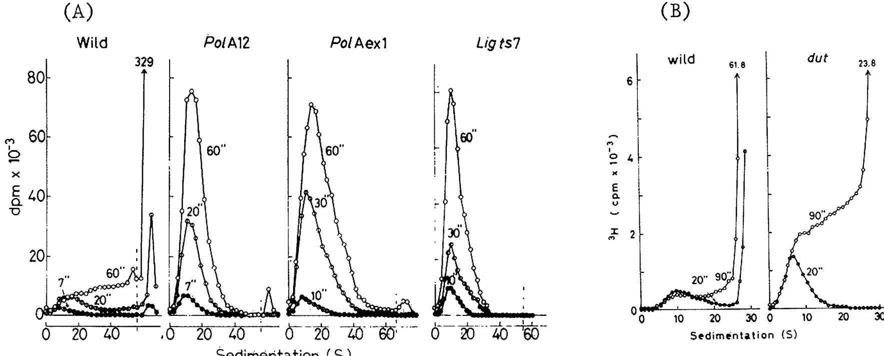


Fig. 1. Alkaline sucrose gradient sedimentation of pulse-labeled DNA from various strains of E. coli.

(A) E. coli W3110 (wild type), MM383 (polA12), RS5064 (polAex1) and KS268 (ligts7) were grown up to 6 x 10^8 cells/ml at 30°C and pulse-labeled with ^{3}H-thymidine for indicated time at 43°C, beginning at 2 min after the temperature shift. The pulse was terminated by the addition of ethanol-phenol mixture, DNA was extracted in the denatured state and sedimented through 5-20% sucrose gradients containing 0.1 M NaOH, 0.9 M NaCl and 1 mM EDTA. The vertical broken lines indicate the boundary of the 82% sucrose cushion.

(B) E. coli KS474 (wild type) and RS5087 (dut) were grown up at 30°C and pulse labeled with ^{3}H-thymidine for indicated time at 30°C. The labeled DNA was analyzed as A.

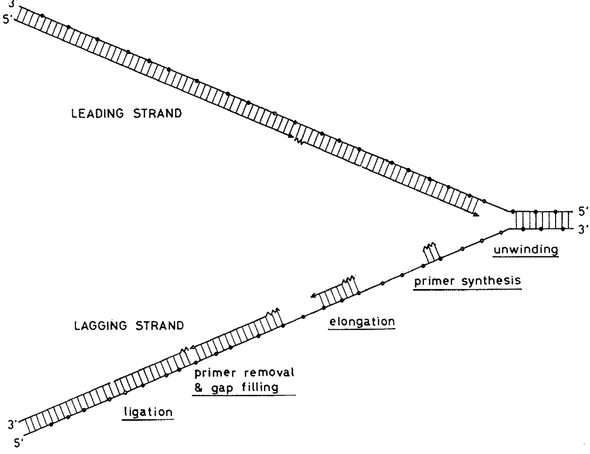


Fig. 2. Scheme for discontinuous replication:closed circles in the parental strand represent the potential initiation sites for primer synthesis.

Early findings indicated that polC(dnaE)-coded DNA polymerase III rather than DNA polymerase I functions in step iii, since nascent fragments were not made under its inhibition (29). It is now established by extensive analyses *in vitro* that the holoenzyme form of DNA polymerase III, which contains polC(α subunit), dnaZ(γ subunit), dnaX(δ subunit), dnaN(β subunit) proteins and at least three more protein factors (ε, θ and τ subunits), carries out the step (15, 30).

The function of DNA polymerase I in the steps iv and v was suspected by the earlier result that nascent fragments accumulated in the polA1 strain of *E. coli* (an amber mutant defective in polymerase activity of DNA polymerase I) (31, 32). This possibility was also supported by the following properties of DNA polymerase I: that the 5' to 3' exonuclease activity of DNA polymerase I exhibits RNase H activity and the concerted action of the exonuclease and the polymerase activities (the nick-translation activity) makes possible the simultaneous removal of the putative primer RNA and the filling in of the resulting gaps by DNA (15). We have examined the above possibility using temperature-sensitive strains of *E. coli* in the nick-translation activity (polA12), in the 5' to 3' exonuclease activity (polAex1 or 480) and in both the polymerase and exonuclease activities of DNA polymerase I (polA4113) (the latter two mutants are also temperature sensitive in cell growth). If DNA polymerase I functions in step iv, the accumulated nascent fragments shown in Fig. 1A should be enriched with RNA-linked nascent fragments to which primer RNA are covalently linked at 5' ends. To detect the RNA-linked nascent DNA fragments, we have developed an assay which measures the proportion of pulse-labeled nascent DNA hydrolyzable by spleen exonuclease after an alkaline hydrolysis (the spleen exonuclease assay of RNA-linked nascent DNA) (6, 11, 33). The assay is based on the facts that RNA-linked DNA exposes a 5' hydroxyl (5'-OH) DNA end after an alkali digestion of RNA moiety and spleen exonuclease only degrades 5'-OH terminated polynucleotides. As shown in Fig. 3, in polAex1 or polA12 cells at 43°C, 5.2 or 4 times as many RNA-linked nascent DNA fragments as in wild type cells are found, respectively. In the absence of DNA ligase activity (ligts7 at 43°C), the amount of the RNA-linked DNA pieces is similar to that in the wild type cells though RNA-free DNA fragments are accumulated. This result is expected if DNA ligase functions at step vi. The proportion of RNA-linked nascent fragments in dut strain of *E. coli* (Fig. 1B) was only a small percentage as was expected by the production of many excision repair nascent fragments in this strain (6, 11, 34). Essentially the same

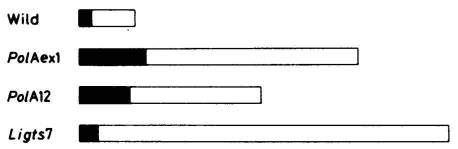


Fig. 3. Relative amount of RNA-linked and RNA-free nascent fragments in various strains of E. coli.

Alkaline treatment of a nascent DNA chain containing an RNA primer exposes a 5'-hydroxyl (5'-OH) end at the RNA-DNA junction of DNA chain. Spleen exonuclease degrades 5'-OH terminated DNA but not 5'-phosphoryl (5'-P) terminated DNA. Utilizing these two properties, proportion of RNA-linked nascent DNA has been determined by spleen exonuclease (33). E. coli W3110 (wild type), RS5064 (polAex1), MM383 (polA12), KS268 (ligts7) were grown up at 30°C and pulse labeled with ^{3}H-thymidine 40s at 43°C. Nascent fragments were purified, 5' ends were phosphorylated with polynucleotide kinase and ATP, then hydrolyzed in alkali and susceptibility of radioactivity to spleen exonuclease is determined directly and after alkaline phosphatase treatment or after phosphorylation. The relative amount of the RNA-linked (■) and RNA-free (□) nascent DNA was calculated from the ratio of 5'-OH and 5'-P DNA assuming uniform labeling of the nascent pieces.

Table 1. Number of RNA-linked DNA pieces in various E. coli strains

Strain	5'-OH ends in alkali-treated pieces (molecules/cell) 30°C	43°C
W3110 (wild type)	11	10
C-N27 (polA4113)	12	93
RS5064 (polAex1)	31	44
MM383 (polA12)	15	25
KS268 (ligts7)	13	14

Alkali-treated nascent pieces were prepared from cultures of E. coli pulse-labeled at 30°C or 43°C. They were phosphorylated with [γ-^{32}P]ATP and T4 polynucleotide kinase at 0°C, the condition to minimize the exchange of phosphate between ATP and the 5'-P terminated DNA (the selective-labeling condition) (35). The amount of 5'-OH DNA per cell was calculated from the amount of ^{32}P transferred per ^{3}H count and the ^{3}H radioactivity incorporated into the nascent pieces per cell.

result was obtained by specific determination with polynucleotide kinase of the number of the 5'-OH ends of DNA which were produced after alkaline treatment of short DNA fragments (RNA-DNA junction) (6, 11, 35). As shown in Table I, RNA-linked DNA fragments accumulate in the polA ts strains at a restrictive temperature (7). That the most prominent accumulation of RNA-linked DNA pieces is seen in polA4113 cells at the restrictive condition supports the view that primer RNA is removed by the nick-translation mechanism.

In vivo and in vitro studies showed a major role of dnaG (primase) and dnaB proteins in the step ii (14, 15, 36). Kornberg's group has established a priming system in vitro for ϕX174 viral ssDNA to RF and to successive stages (14-16). In the system, primosome, composed of seven proteins (n, n', n", i, dnaC, dnaB, dnaG) and assembled at the unique protein n' recognition sites on ϕX174 viral strand, moves processively around the viral chromosome coated with single-stranded DNA binding protein (SSB) to generate multiple primers (14). The group has proposed the function of a conserved primosome in the progress of the replication fork of the E. coli chromosome and in initiating the synthesis of nascent fragments. The ϕX174 primosome travels processively in an anti-elongation direction (5' → 3' direction on a template strand), thus synthesizing multiple primers only for a lagging daughter strand (one strand discontinuous replication in vitro). Single-stranded template DNA is required for the priming reaction in vitro, showing the requirement of the unwinding step (step i) before the primer formation. The exact mechanism by which the duplex parental strands at replication fork are opened in advance of replication has yet to be clarified. Unwinding and fork progression may be carried out by the combined actions of enzymes such as topoisomerases, helicases (rep proteins, helicases), primosome and SSB (14-16, 37). Many of these enzymes may be components of a replication fork at the replication origin of chromosome.

Structure of the primer RNA

We have analyzed the structure of primer RNA attached to the DNA fragments purified from wild type or polA strains of E. coli by the following post-labeling procedures (6, 7). 5' ends of extensively purified nascent fragments are dephosphorylated and then phosphorylated with [γ-^{32}P]ATP and T4 polynucleotide kinase. Primer RNAs are liberated (leaving one residue of deoxyribonucleotide covalently linked at their 3' ends) from the end-labeled RNA-DNA fragments by extensive digestion with pancreatic DNase and the 3' to 5'

exonuclease associated with T4 DNA polymerase (38). Primer RNA molecules thus isolated from polA mutant strains as well as from a wild type strain are surprisingly short; 80 to 90% were found to be mono- and dinucleotides, less than 10% were trinucleotide and only a small percentage are longer than trinucleotide (7). Four kinds of bases are found at the 5' ends of all chain-length primers. These results contrast with the results obtained with bacteriophage T4 or T7, in which mono- to pentaribonucleotide primers were linked to the 5' ends of DNA fragments and 5' terminal nucleotides of each chain length primer RNA showed characteristic base compositions indicative of the structure of the intact primer RNA, pppApCpN1pN2(pN3) (38, 39, 40). (In the case of T7-primer RNA, N1 is mainly C and some A, N2 is rich in A and C and a switch to DNA synthesis occurs mainly at N3 (thus the intact primer is mainly tetraribonucleotide), whereas in the case of T4-primer RNA, N1 to N3 are occupied by unspecified ribonucleotides and the intact primer is mainly pentaribonucleotide). The question is, do the very short RNA stretches linked to the E. coli DNA fragments represent partially degraded primers or do they contain a significant portion of intact-size primers? As mentioned previously, DNA polymerase I is implicated for the removal of primer RNA. Why then is the chain length of primer RNA in polA mutant cells the same as that found in a wild type strain? One explanation for this might be that DNA polymerase I degrades only the short primer RNA in the size range of mono- to trinucleotides, and the longer primers are removed by other enzyme(s) such as RNase H. It is known that RNase H hardly liberates the ribonucleotide at the RNA-DNA junction (41). We have recently isolated an E. coli mutant which has undetectable RNase H activity in vitro (less than 0.01% activity of the wild type strain) either at 30°C or at 43°C (T. Ogawa and T. Okazaki, in preparation). The mutant cell grows normally at 30°C but forms a very small colony at 43°C. DNA synthesis in the mutant cell proceeds at 43°C for at least a few hours and no accumulation of nascent fragments is seen. These results suggest either that RNase H plays auxiliary role in the removal of primer RNA in discontinuous replication or that the in vivo level of RNase H activity in the mutant cell is high enough to accomplish the excision of the primer RNA. We have examined whether a longer primer RNA exists in the RNase H-deficient E. coli cells using post-labeling procedures. In addition to the mono- to trinucleotide primers, a small amount of primer RNA of chain length around decanucleotides has been detected in the rnh mutant as well as in an rnh polA double mutant, at 43°C, whereas only the short chain-length primers have been

detected in a polA rnh^+ or wild type strains of E. coli (T. Ogawa and T. Okazaki, unpublished observation). The results show that RNase H is indeed responsible for the degradation of the long primer RNA. It seems likely that the chain length of E. coli primer RNA is heterogeneous, ranging from a few nucleotide residues to up to around a decanucleotide. The longer primers might be made so rarely that they could be excised with or without a very low level of RNase H activity. Similar to E. coli primer RNA made in vivo, size of the primer RNA made by the ϕX174 SS to RF conversion system in vitro is mono- to up to around a decanucleotide and size distribution is variable (T. Ogawa, T. Okazaki and K. Arai, in preparation). Unique chain length is a common feature of primer RNA in mammalian cells and viruses thus far reported (42). We have found that intact primers for Drosophila and sea urchin discontinuous replication are octaribonucleotides (13). E. coli primer RNA may represent an exception by its heterogeneous chain length. It would be interesting to know whether bacterial primers all share this property. The priming sites of ϕX174 phage DNA replication in vivo and in vitro are distributed quite frequently on the ϕX174 genome and no obvious recognition sequence is found around these sites (T. Ogawa, T. Okazaki and K. Arai, in preparation). The frequent priming sites and flexibility in priming reaction are general feature in discontinuous replication (43). In the case of bacteriophage T7, however, 3' $-C-T^*-G-$ 5' is the recognition sequence for the primer synthesis on a template strand (primer synthesis starts at opposite T^*) (43; Richardson, this volume). In contrast to the case of bacteriophage T7, it has been proposed that E. coli primase, dnaG protein, recognizes the secondary structure of DNA created by dnaB protein rather than a specific sequence (44).

Is DNA replication discontinuous on both strands or on one strand?

Discovery of repair fragments in dut mutant strain of E. coli (45) raised a question of what portion of the nascent fragments are replicative intermediates in dut^+ strains of E. coli (Fig. 1B). In the dut strain, nascent fragments due to uracil excision-repair are produced by the following mechanism: Because of the increased cellular level of dUTP relative to that of dTTP by the defect in dUTPase activity, dUMP is frequently incorporated into the DNA in place of dTMP. Uracil- DNA glycosylase removes the uracil bases from the DNA and the resulting apyrimidinic sites are recognized by a specific endonuclease, and incision produces DNA fragments having an apyrimidinic site. Trimming of the 5' terminus back to an intact nucleotide, gap filling and

ligation will complete the excision repair sequence. Since dUTP is a normal product of pyrimidine nucleotide metabolism, it is suspected that low levels of dUMP might be incorporated into DNA despite the presence of dUTPase (dut^+). As already mentioned, in mutants defective in either DNA ligase or DNA polymerase I, nascent fragments accumulate and little if any label is found in long DNA chain (Fig. 1A). The result was originally interpreted to suggest that both daughter strands are replicated discontinuously. The interpretation of the experiments has become clouded by the discovery of excision-repair nascent fragments in dut strain, since both enzymes function also for the dut-repair process. If DNA synthesis at the replication fork is discontinuous only on one strand and uracil repair is the only post-replication repair mechanism that contributes to the nascent fragments pool, it is expected that in a mutant strain defective in uracil-DNA glycosylase (ung) half of the label appears in high-molecular-weight DNA and the other half in short chains. In a double mutant constructed by introduction of the ung allele into dut mutant strain, the accumulation of pseudo-Okazaki fragments is no longer observed, indicating that the dut-repair process is the main cause for the production of nascent fragments in the mutant strain (46). An ung mutation in a wild-type background does not reduce the amount or increase the size of the nascent fragments (47). In a polA ung double mutant, however, nascent fragments accumulate as they do in polA single mutant strain (48). The results suggest that DNA replication is discontinuous on both daughter strands or it is discontinuous only on one daughter strand but other postreplication cleavage gives rise to DNA pieces.

We have analyzed this problem using bacteriophage P2-infected polA ung strains of E. coli (11). Replication of bacteriophage P2 DNA proceeds unidirectionally from a fixed origin, therefore overall direction of the H-daughter strand is 3' to 5' and that of the L-daughter strand is 5' to 3' (49). Chain elongation of P2 DNA is catalyzed by E. coli replication enzymes. Therefore, the P2 system is a good model to examine whether one or two daughter strands are replicated discontinuously. When P2-infected, wild-type E. coli cells were labeled by a very brief [^{3}H]thymidine pulse, the label was incorporated mainly into two classes of DNA chains: short chains of about 10S and long chains of one-genome length (Fig. 4). The short chains were composed mainly of an H-strand component but also contained a small portion of L strand and the genome length chains were composed mainly of L-strand component (49). The same results were obtained when P2-infected E. coli ung cells were analyzed. When P2-infected E. coli $polA^{ts}$ cells were pulse-labeled at a

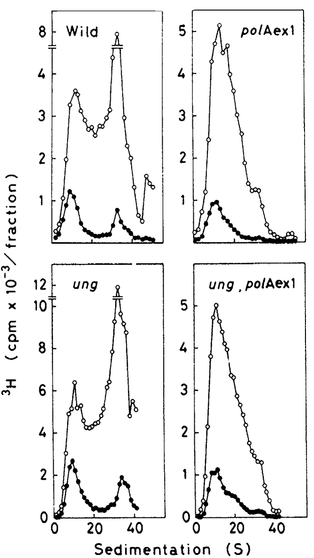


Fig. 4. Alkaline sucrose gradient sedimentation of pulse-labeled DNA from bacteriophage P2-infected E. coli cells.

Cultures of E. coli C1200 (wild), CN3 (polAex1), SN281 (ung) and SN292 (polAex1, ung) were infected with bacteriophage P2. When the rate of phage DNA synthesis was nearly at its maximum, cultures were shifted to 43°C and pulse-labeled with ^{3}H-thymidine for 10s (-•-) and 30s (-o-), beginning at 2 min after the temperature shift. DNA was extracted in a denatured state and sedimented through the alkaline sucrose gradients as described in Fig. 1. Host strains were indicated.

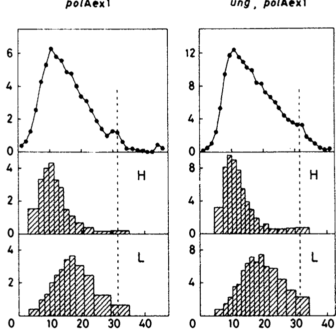


Fig. 5. Size distribution of the H- and L-strand components of radioactive DNA from bacteriophage P2-infected E. coli polAex1 or polAex1, ung cells.

Bacteriophage P2-infected E. coli CN3 (polAex1) and SN292 (polAex1, ung) cells were pulse labeled with ^{3}H-thymidine for 30s at 43°C as described in Fig. 4. The DNA samples were sedimented through the alkaline sucrose gradient. The acid-insoluble radioactive label in each fraction was determined (-•-). The amounts of the radioactivity incorporated into H- and L-strand components were determined in each fraction by DNA-DNA hybridization (49). Host strains were indicated.

restrictive temperature, virtually all the label incorporated was found in short chains (Fig. 4) which were composed of equal amounts of H- and L-strand components, but their size distributions were different (Fig. 5) (11). The average sedimentation coefficient of H-strand nascent chain is 10S, whereas that of L-strand nascent chain is 17S. The possibility that the L-strand short chains are produced by a dut-repair mechanism rather than by a discontinuous mechanism is less likely because of the following results. A polA ung double mutant accumulated nascent short chains at a restrictive temperature, and similar asymmetry in the size distributions of H- and L-strand nascent chains was observed in the accumulated short chains (Figs. 4 and 5). The spleen exonuclease assay of nascent short chains revealed that 25-50% portions of both H- and L-strand nascent chains were composed of RNA-linked DNA fragments (Table 2) (11). These results are compatible with the concepts that both H and L strands of P2 DNA are replicated discontinuously using an RNA primer but the priming reaction occurs several times less frequently on L strand (the leading strand). Whether or not RNA-linked nascent fragments derive from both daughter strands of a replication fork of E. coli chromosome is yet to be determined.

Table 2. Proportion of RNA-linked nascent DNA fragments of H- or L-strand components of bacteriophage P2 DNA

Host	Unfractionated (%)	H strand (%)	L strand (%)
C1200 (wild type)	37	43	23
CN3 (polAex1)	46	56	38
CN27 (polA4113)	48	50	46
SN292 (polAex1, ung)	41	53	41
SN262 (polA4113, ung)	50	51	55

Proportion of RNA-linked nascent DNA was determined by the spleen exonuclease assay (33). P2 phage-infected E. coli cells were pulse-labeled with ^{3}H-thymidine for 30s at 43°C as described in Fig. 5. Nascent DNA fragments of less than 1800 nucleotides were purified as described (7, 33). After the phosphorylation of pre-existing 5'-OH ends of DNA with polynucleotide kinase and ATP, the nascent fragments were incubated at 37°C for 20 hr in 0.15 M NaOH. Each sample was then divided into three portions (A, B and C). Portion A was phosphorylated with polynucleotide kinase, portion B was treated with phosphatase, and portion C was incubated under the same conditions without enzymes. The samples A, B and C were then digested with spleen exonuclease before and after strand separation. The proportion of the 5'-OH terminated DNA (RNA-linked DNA) was calculated as described (33).

INITIATION OF DNA REPLICATION AT THE ORIGIN OF THE *E. COLI* CHROMOSOME

Genetic information

Many temperature-sensitive mutants in DNA replication are classified as initiation-defective mutants because they stop DNA synthesis after completion of the ongoing replication at the time of the temperature shift (50). Belonging to this class are *dnaA*, *dnaC*, one *dnaB* (*dnaB* 252) and several other *dna* mutants (*dnaI*, K and P, etc.). Rifampicin and chloramphenicol which inhibit RNA and/or protein synthesis, respectively, also stop a new round of the chromosome replication of *E. coli*. Some mutations in *dnaA* gene are suppressed by mutation in *rpoB* gene, the cistron for the RNA polymerase β-subunit, the target polypeptide of rifampicin (51). This suggests that RNA polymerase may interact with *dnaA* protein in an initiation reaction. The suppression of *dnaAts* mutants by a number of plasmids or phage DNA integration in the bacterial chromosome (the integrative suppression) indicates the function of *dnaA* protein in the initiation stage of chromosome replication. Results of the *in vivo* studies suggest the following order of the actions of these proteins: the chloramphenicol-sensitive protein, *dnaA* and RNA polymerase in concert or in this order, and then *dnaC* protein. Data *in vivo* suggest that the *dnaB* gene product is required before or during the action of RNA polymerase, whereas the *in vitro* replication system of ϕX174 DNA show its function with *dnaC* protein in the prepriming step. The nature of the mutants and the function of required gene products has been extensively reviewed previously (52).

Replication origin of E. coli chromosome (oriC)

Earlier genetic studies have located the origin of bidirectional replication of *E. coli* chromosome near the *ilv* gene (2, 3). Marsh and Worcel physically mapped the region of the *E. coli* K12 chromosome initially replicated upon downshift of a *dnaA* and a *dnaC* ts mutants to the permissive temperature (53). They showed that replication starts within or very near a 1.3 kilobase pair (kb) *Hind*III fragment which itself is located in a 8.6 kb *Eco*RI fragment. This *Eco*RI fragment was found in an autonomously replicating plasmid (*oriC* plasmid) which was constructed by the ligation of *Eco*RI digests of total *E. coli* DNA to non-replicating DNA fragments containing β-lactamase gene and isolated from the cells transformed by the DNA to ampicillin resistance (17). The same *Eco*RI segment was contained in the F' factors (F' poh^+), which carried the genomic region left to *ilv*, and in the specialized transducing phage λ (λ-*asn*); Both of them were thought to carry the replication origin of *E. coli*

genome, since they could be maintained as plasmids in Hfr bacteria or in λ-lysogen, respectively (54, 55). Using the oriC plasmids, the replication origin was mapped precisely, the nucleotide sequence surrounding the origin was determined and the physical and genetic maps were correlated (17-20). By the mapping of Ori-function using ColE1-oriC chimera plasmids, the minimal region for initiation of DNA replication was delimited to 245 base pairs stretch, oriC (Fig. 6) (21). The oriC is located at a non-coding area between the markers unc and asn (at 83.5 min on genetic map (1)) about 2 kb away in an asn

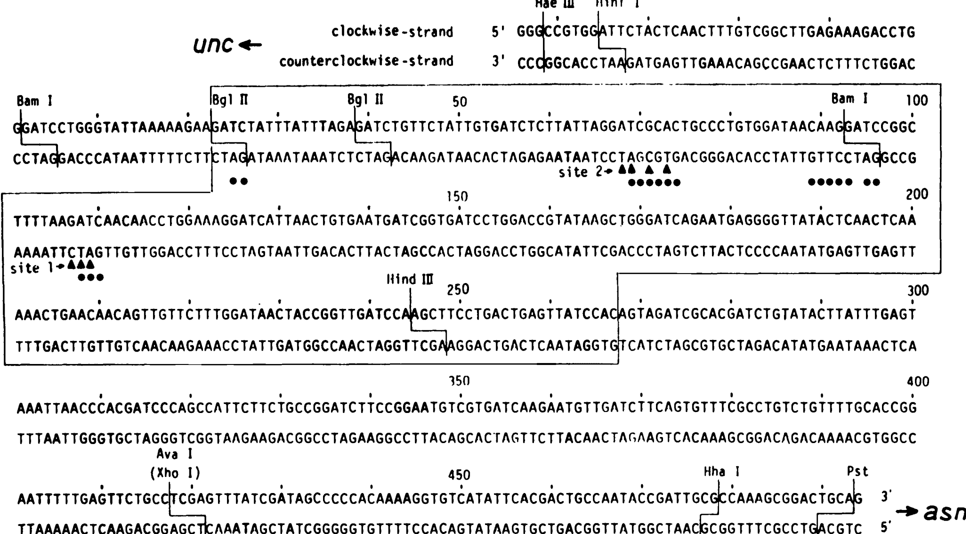


Fig. 6. RNA-DNA junction sites and the 5'-ends of RNA-linked DNA molecules mapped in the nucleotide sequence of oriC region. Nucleotide sequence of oriC area and the numbering system are by Oka et al. (21). The region essential to Ori-function (oriC) is enclosed. The clockwise-strand and the counterclockwise-strand represent the strands whose 5' to 3' polynucleotide polarity lie clockwise (rightward) and counterclockwise (leftward) on the E. coli genetic map, respectively (1). Priming sites have been located by the procedures depicted in Fig. 7.
▲, The sites where deoxyribonucleotides at the RNA-DNA junction of the RNA-linked DNA molecules have been located. Only the most frequently used junction sites in vivo (sites 1 and 2) are shown.
●, The sites where 5'-terminal ribonucleotides of RNA-linked DNA molecules have been located.

direction from the 1.3 kb Hind III fragment. It is flanked by two genes of unknown functions, genes for a 16-kdal protein (protein X) to the right and a 70-kdal protein (gid) to the left (56). Neither of these two genes are essential for growth or initiation of chromosome replication at oriC. Initiation of replication on oriC plasmid depends upon host functions known to be required for initiation of chromosome replication, such as functions of dnaA and dnaC gene products and RNA polymerase (18). Thus the oriC plasmids are ideal model substrates for the analyses of the molecular mechanism of the initiation of chromosome replication in vivo and in vitro. The oriC lack clear symmetrical sequences for a bidirectional start of DNA synthesis. Several potential secondary structures exist in the oriC sequence (57). The GATC sequence that appears eleven times in the 245-bp oriC and also many times in the sequence surrounding the oriC might play a important part in the stem formation of such potential secondary structure. Using in vitro methods generating deletions, insertions and substitutions of base pairs in the oriC region which had previously been cloned in colicin E1 plasmid vectors, effects of such change to the Ori-function were analyzed in vivo (21-23). Basing on the results, Oka et al. predicted that the oriC region is composed of two categories of sequences, one functions as recognition sites (recognition sequence) and the other as spacers (spacer sequences), and the accurate arrangements of the recognition sequences in oriC is essential for expression of Ori-function. Zyskind et al. compared the oriC sequence of five enterobacteria and demonstrated that several discrete regions with homologous sequences (consensus sequences) are separated by non-conserved sequence (58). The consensus and non-conserved sequences may correspond to the recognition and spacer sequences, respectively.

Priming sites for DNA synthesis in oriC region in vivo

One of the key reactions at the replication origin is a priming for DNA replication. Syntheses of the leading and the lagging strands for both clockwise and counterclockwise replication forks start presumably at oriC and/or its close vicinity and they will probably be primed by RNA (52). To obtain information on the nature of the priming reaction at the origin in vivo, we have located the RNA-DNA junction site and the 5' end of the primer RNA on the nucleotide sequence of both strands of oriC (12, 13, 25). We have used DNA chains purified from the dnaC ts cells which were synchronously initiating DNA replication by a temperature shift from a non-permissive to a permissive condition in the presence of arabinosyl cytosine (25, 53). A small amount of

DNA synthesized under the condition was enriched with short chains which had homology to the sequence around the origin (25). The strategy for the mapping of the priming sites is shown in Fig. 7 (13). The priming sites on the

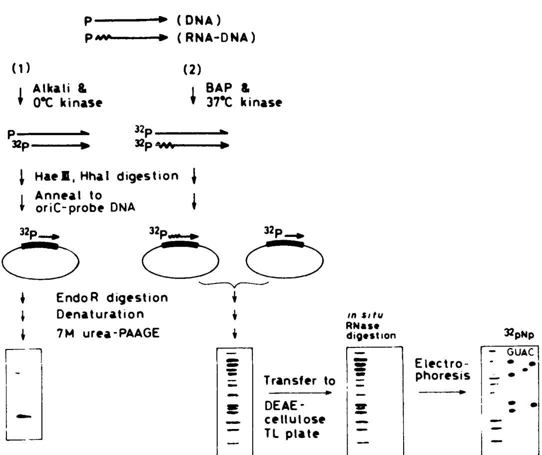


Fig. 7. Mapping procedures in oriC region for RNA-DNA junction points and 5' terminal ribonucleotide locus of the RNA-linked DNA molecules.
1. The procedure for the RNA-DNA junction points. Short DNA and RNA-linked DNA molecules are purified from dnaC cells which are synchronously initiating DNA replication. The 5'-OH DNA ends are uncovered from the RNA-linked DNA molecules by alkali treatment and the ends are selectively labeled with [γ-^{32}P]ATP and polynucleotide kinase at 0°C (35). The end-labeled molecules are then digested in single-stranded state with Hae III and Hha I, which cleave the left and right ends of the origin sequence, respectively (Fig. 6). The cleaved molecules are then annealed to the single-stranded oriC-probe DNA, composed of one each of the component strands of oriC region shown in Fig. 6 and pBR322 or fl vector DNA. The annealed molecules are digested with the second restriction endonucleases: the endo-labeled clockwise strand with Xho I or Hind III and the counterclockwise strand with Bgl II or Bam HI. The digests are denatured and electrophoresed in sequence gels.
2. The procedure for the 5'-terminal ribonucleotide locus of the RNA-linked DNA molecules. The purified DNA and RNA-linked DNA molecules are treated with alkaline phosphatase and 5'-termini of both DNA and RNA-linked DNA are ^{32}P labeled with polynucleotide kinase and [γ-^{32}P]ATP at 37°C. The end-labeled molecules are treated as described in procedure 1. After the electrophoresis in the sequence gels, the radioactive ladders, composed of 5'-^{32}P-labeled DNA and RNA-DNA, are transferred from the gels to DEAE-cellulose TL plates. After in situ RNase digestion, ^{32}P-labeled pNps derived from RNA-DNA molecules are separated by electrophoresis (13).

purified DNA chains have been labeled with highly radioactive [^{32}P]phosphate. ^{32}P-labeled DNA chains which have the homology to the oriC sequence have been isolated in hybrid form with single-stranded origin-probe DNAs carrying one each of two component strands of the oriC stretch shown in Fig. 6, and the position of the ^{32}P-labeled nucleotides in the known sequence of oriC has been determined after digestion with appropriate restriction endonucleases. We have found RNA-primed initiation sites only in one of the component strands of the oriC segment: in the strand whose 5' to 3' polynucleotide polarity lies counterclockwise on the E. coli genetic map (the counterclockwise strand, see Fig. 6). In addition to the two strong RNA-DNA junction sites shown in the figure (sites 1 and 2), many weak junction sites (not shown) have been found also in the counterclockwise strand. The mapping of the 5' ends of primer RNA has shown that the primer ends could be detected also exclusively on the counterclockwise strand and the sites were overlapped with the strong and the weak RNA-DNA junction sites (Fig. 6). A conspicuous feature of these priming sites is that they all are located at or near the GATC sequence previously mentioned. The results show that RNA primed DNA synthesis starts at one of many preferred sites in the oriC region but asymmetrically with respect to the component strands. The primer RNA molecules we have detected are all very short but there is no information about their intact size. It is not definitive at present whether these sites represent the priming sites for the leading strand of the counterclockwise replication fork or those for the lagging strand of the clockwise replication fork. Our current hypothesis for the starting mechanism of DNA chains at the replication origin of E. coli is shown in Fig. 8: After or coupled with a specific recognition and unwinding (activation) of the origin (i, ii), DNA replication first starts unidirectionally within the oriC by a counterclockwise strand initiation (iii). The strand is established as the leading strand at the counterclockwise replication fork by the aid of helix-opening enzymes. The lagging strand for the counterclockwise replication fork may be synthesized subsequently on a newly exposed template strand somewhere left to the oriC using the initiation mechanism for discontinuous replication. The first lagging strand may be established as the leading strand for the clockwise replication fork within or outside the oriC by the aid of helix-opening enzymes (iv). Finally, initiation of the lagging strand for the clockwise replication fork takes place on the newly exposed template strand to the right of the oriC by the discontinuous mechanism (v). Chloramphenicol-sensitive protein, dnaA protein and RNA

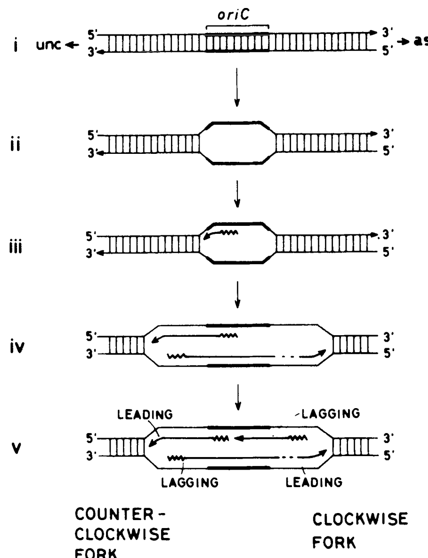


Fig. 8. A proposed mechanism for chain initiation in the E. coli replication origin. Asymmetric initiation model in which the counterclockwise-strand initiation occurs first at the oriC region (thick line) is depicted. See details in the text.

polymerase may take part in steps i and ii. The function of an RNA polymerase transcript as the primer in step iii is strongly suspected but there is no clear evidence at present. A possibility also exists that dnaG proteins might play a part in this step by some mechanism (59). Helix-opening enzymes which may be essential for the fork establishment are topoisomerases, helicases (rep protein and helicase II), primosome proteins and SSB (14-16). An oriC type primosome constituted with dnaB, dnaC and dnaG and other proteins might be assembled at the oriC and migrate to the 5' to 3' direction on the template strand to the places where primers for the lagging strands are made in steps iv and v (14). Alternatively, it might be assembled outside of oriC with only broad dependency on the nucleotide sequence.

Priming sites for DNA synthesis used by oriC enzyme system

OriC plasmids provided an ideal template for analysis of the initiation reaction at the chromosome origin in vitro. Fuller et al. discovered a soluble enzyme system which specifically replicates oriC plasmid or phage chimeras (24). The system recognizes the 245 bp intact oriC stretch, initiates replication within or near the oriC sequence and proceeds bidirectionally (60). The oriC specific replication depends on the function of dnaA protein, dnaC

protein, RNA polymerase, DNA gyrase and numerous replication proteins including dnaB and single-stranded DNA binding protein. Using the mapping method described in Fig. 7, we have located the RNA-DNA junction sites of the reaction products of the oriC enzyme system on the sequence of the two component strands of oriC. Many junction sites have been detected exclusively on the counterclockwise strand and many sites agree with the in vivo priming sites shown in Fig. 6 (unpublished results by Y. Kohara, T. Ogawa and T. Okazaki). Thus, the oriC enzyme system reaction mimics the in vivo initiation reaction. Progress in characterization of replication proteins and their function in this oriC enzyme system promises an advanced understanding of the reaction at the replication origin.

Transcription at the oriC

Some function of RNA polymerase is required for the initiation of DNA synthesis at the replication origin. It may activate the origin sequence, synthesize a primer RNA or do both. To elucidate the role of RNA polymerase, the nature of the required transcription should be clarified. Several groups have reported studies which have analyzed promoters functioning in vivo or in vitro in and close to oriC. Hansen et al. and Morita et al. have mapped in vivo functioning promoters on the cloned chromosomal DNA by using promoter-cloning vectors (56, 61). These results show that there is no high level transcription originating inside oriC and passing leftward through the boundary of the minimal essential origin. One strong leftward transcription promoted in the promoter region of protein X (15.5 kD protein) gene enters oriC across the Hind III site (at position 244 in Fig. 6) and then is efficiently terminated within oriC. This is the only transcript entering oriC. Deletion studies, however, show that this transcription is not required for efficient oriC functioning. Another transcription promoted from the outside region left to oriC, probably the promoter region of the gid gene, also proceeds in a leftward direction (therefore it does not enter oriC). Morita et al. have found that one rightward transcription is promoted from the vicinity of the right boundary of oriC with an Ori^{+} base-substitution mutant in the Hind III site. They pointed out the possibility that, unless a promotion site is created by the base-substitution, a promoter is located in the vicinity of the Hind III site. On the other hand, Lother and Messer have found by in vitro transcription experiments and RNA sequencing that oriC contains two promoters, one for the leftward and one for the rightward transcription, arranged back-to-back at the left and the right region of the Hind III site, respectively (62). The

relation between these leftward transcriptions and the RNA-DNA junction sites we have found on the counterclockwise strand of the oriC segment should be clarified by further analyses. The oriC enzyme system (24) will provide a most powerful tool to solve the problem.

CONCLUSION

Studies of structural and metabolic properties of the nascent DNA fragments of E. coli have elucidated the major steps of the discontinuous chain-elongation of E. coli chromosome. Repeated chain initiation is primed by oligoribonucleotides in heterogeneous size, ranging from a few to up to around ten nucleotide residues. A model in vitro system for the discontinuous replication of E. coli chromosomal DNA has been provided by the reconstitution studies of the conversion system of single-stranded DNA of phage ϕX174 to duplex replicative form. The priming of the DNA synthesis by this system mimics the reaction we have detected in vivo. The one discrepancy is found, however, between the in vivo and in vitro reactions: in vivo, leading daughter strands are synthesized discontinuously, although with much infrequent initiation than the lagging strand, whereas in vitro, leading strands are synthesized continuously. This difference might be caused by some structural difference between the in vivo and in vitro replication forks. At present, the least clear process of the discontinuous replication is the unwinding mechanism of the parental duplex DNA. Studies of the reaction at the replication origin of the chromosome will reveal what and how the unwinding machinery is introduced to a replication fork. They may also help to solve the problem of the leading strand synthesis. Enzymatic and molecular events at the replication origin are just about unveiled and good prospects are in sight.

REFERENCES

1. Bachmann, B.J. and Low, K.B. Micribiol. Rev. 44:1-56, 1980.
2. Masters, M. and Broda, P. Nature New Biol. 232:137-140, 1971.
3. Bird, R.E. Louarn, J., Martuscelli, J. and Caro, L.G. J. Mol. Biol. 70:549-566, 1972.
4. Prescott, D.M. and Kuempel, P.L. Proc. Natl. Acad. Sci. USA 69:2842-2845, 1972.
5. Cairns, J. Cold Spring Harbor Symp. Quant. Biol. 28:43-46, 1963.
6. Ogawa, T. and Okazaki, T. Ann. Rev. Biochem. 49:421-457, 1980.
7. Ogawa, T., Hirose, S., Okazaki, T. and Okazaki, R. J. Mol. Biol. 112:121-140, 1977.
8. Sakabe, K. and Okazaki, R. Biochim. Biophys. Acta 129:651-654, 1966.
9. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., Sugino, A. Proc. Natl. Acad. Sci. USA 59:598-605, 1968.

10. Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K., Kainuma, R., Sugino, A., and Iwatsuki, N. Cold Spring Harbor Symp. Quant. Biol. 33:129-143, 1969.
11. Okazaki, T., Kurosawa, Y., Ogawa, T., Seki, T., Shinozaki, K., Hirose, S., Fujiyama, A., Kohara, Y., Machida, Y., Tamanoi, F. and Hozumi, T. Cold Spring Harbor Symp. Quant. Biol. 43:203-219, 1979.
12. Okazaki, T., Hirose, S., Fujiyama, A. and Kohara, Y. In: Mechanistic Studies of DNA Replication and Genetic Recombination, ICN-UCLA Symp. Molec. Cell. Biol., Vol. 19 (Ed. B. Alberts), Academic Press, New York, 1980, pp. 429-447.
13. Okazaki, T., Kohara, Y., Ogawa, T., Yoda, K., Fujiyama, A., Seki, T., Hozumi, T., Kitani, T., Yasuda, H. and Miyasaka, T. In: Nucleic Acid Research: Future Development (Eds. K. Mizubuchi, I. Watanabe and D. Watson), Academic Press, Japan, in press.
14. Arai, K., Low, R. and Kornberg, A. Proc. Natl. Acad. Sci. USA 78:707-711, 1981.
15. Kornberg, A. DNA replication. W.H. Freeman & Co., San Francisco, 1980.
16. Kornberg, A. 1982 Supplement to DNA Replication. W.H. Freeman & Co., San Francisco, 1982.
17. Yasuda, S. and Hirota, Y. Proc. Natl. Acad. Sci. USA 74:5458-5462, 1977.
18. von Meyenberg, K., Hansen, F.G., Riise, E., Bergmans, H.E.N., Meijer, M. and Messer, W. Cold Spring Harbor Symp. Quant. Biol. 43:121-128, 1979.
19. Meijer, M., Beck, E., Hansen, F.G., Bergmans, H.E.N., Messer, W, W., von Meyenberg, K., and Schaller, H. Proc. Natl. Acad. Sci. USA 76:580-584, 1979.
20. Sugimoto, K., Oka, A., Sugisaki, H., Takanami, M., Nishimura, A., Yasuda, S. and Hirota, Y. Proc. Natl. Acad. Sci. USA 76:575-579, 1979.
21. Oka, A., Sugimoto, K., Takanami, M. and Hirota, Y. Molec. Gen. Genet. 178:9-20, 1980.
22. Hirota, Y., Oka, A., Sugimoto, K., Asada, K., Sasaki, H. and Takanami, M. In: The Initiation of DNA Replication, ICN-UCLA Symp. Molec. Cell. Biol. Vol. 22 (Ed. D. Ray), Academic Press, New York, 1981, pp. 1-12.
23. Oka, A., Sugimoto, K., Sasaki, H. and Takanami, M. Gene 19:59-69, 1982.
24. Fuller, R.S., Kaguni, J.M. and Kornberg, A. Proc. Natl. Acad. Sci. USA 78:7370-7374, 1981.
25. Hirose, S., Hiraga, S. and Okazaki, T. Molec. Gen. Genet. in press.
26. Brutlag, D., Schekman, R. and Kornberg, A. Proc. Natl. Acad. Sci. USA 68:2826-2829, 1971.
27. Sugino, A., Hirose, S. and Okazaki, R. Proc. Natl. Acad. Sci. USA 69:1863-1867, 1972.
28. Okazaki, R., Sugino, A., Hirose, S., Okazaki, T., Imae, Y., Kainuma-Kuroda, R., Ogawa, T., Arisawa, M. and Kurosawa, Y. In: DNA Synthesis in Vitro, ICN-UCLA Symp. Molec. Cell. Biol., Vol. 3 (Eds. R.D. Wells and R.B. Inman), University Park Press, Baltimore, 1973, pp. 83-106.
29. Gefter, M.L., Hirota, Y., Kornberg, T., Wechsler, J.A. and Barnoux, C. Proc. Natl. Acad. Sci. USA 68:3150-3153, 1971.
30. McHenry, C. and Crow, W. J. Biol. Chem. 254:1748-1753, 1979.
31. Kuempel, P.L. and Veomett, G.E. Biochem. Biophys. Res. Commun. 41:-973-980, 1980.
32. Okazaki, R., Arisawa, M. and Sugino, A. Proc. Natl. Acad. Sci. USA 68:2954-2957, 1971.
33. Kurosawa, Y., Ogawa, T., Hirose, S., Okazaki, T. and Okazaki, R. J. Mol. Biol. 96:653-664, 1975.
34. Machida, Y., Okazaki, T., Miyake, T., Ohtsuka, E. and Ikehara, M. Nucleic Acids Res. 9:4755-4766, 1981.

35. Okazaki, R., Hirose, S., Okazaki, T., Ogawa, T. and Kurosawa, Y. Biochem. Biophys. Res. Commun. 62:1018-1024, 1975.
36. Lark, K.G. Nature New Biol. 240:237-240, 1972.
37. Geider, K. and Hoffman-Berling, H. Ann. Rev. Biochem. 50:233-260, 1981.
38. Kurosawa, Y. and Okazaki, T. J. Mol. Biol. 135:841-861, 1979.
39. Seki, T. and Okazaki, T. Nucleic Acids Res. 7:1603-1619, 1979.
40. Ogawa, T. and Okazaki, T. Nucleic Acids. Res. 7:1621-1633, 1979.
41. Berkower, I., Leis, J. and Hurwitz, J. J. Biol. Chem. 248:5914-5921, 1973.
42. DePamphilis, M.L. and Wassarman, P.M. Ann. Rev. Biochem. 49:627-666, 1980.
43. Fujiyama, A., Kohara, Y. and Okazaki, T. Proc. Natl. Acad. Sci. USA 78:903-907, 1981.
44. Arai, K. and Kornberg, A. J. Biol. Chem. 256:5260-5266, 1981.
45. Tye, B.-K., Nyman, P.-O., Hochhauser, S. and Weiss, B. Proc. Natl. Acad. Sci. USA 74:154-157, 1977.
46. Tye, B.-K. and Lehman, I.R. J. Mol. Biol. 117:293-306, 1977.
47. Tamanoi, F., Machida, Y. and Okazaki, T. Cold Spring Harbor Symp. Quant. Biol. 43:239-242, 1979.
48. Tye, B.-K., Chien, J., Lehman, I.R., Duncan, B.K., Warner, H.R. Proc. Natl. Acad. Sci. USA 75:233-237, 1978.
49. Kainuma-Kuroda, R. and Okazaki, R. J. Mol. Biol. 94:213-228, 1975.
50. Wechsler, J.A. In: DNA Synthesis (Ed. I. Molineux and M. Kohiyama), Plenum Press, New York, 1978, pp. 49-70.
51. Bagdasarian, M.M., Isakowska, M. and Bagdasarian, M. J. Bacteriol. 130:577-582, 1977.
52. Tomizawa, J. and Selzer, G. Ann. Rev. Biochem. 48:999-1034, 1979.
53. Marsh, R.C. and Worcel, A. Proc. Natl. Acad. Sci. USA 74:2720-2724, 1977.
54. Hiraga, S. Proc. Natl. Acad. Sci. USA 73:198-202, 1976.
55. von Meyenberg, K., Hansen, F.G., Nielsen, L.D. and Jorgensen, P. Molec. Gen. Genet. 158:101-109, 1977.
56. Hansen, F.G., Koefoed, S. von Meyenberg, K. and Atlung, T. In: The Initiation of DNA Replication, ICN-UCLA Symp. Molec. Cell. Biol., Vol. 22 (Ed. D. Ray), Academic Press, New York, 1981, pp. 1-12.
57. Hirota, Y., Yamada, M., Nishimura, A., Oka, A., Sugimoto, K., Asada, K. and Takanami, M. In: Progress in Nucleic Acid Research and Molecular Biology, Vol. 26, Academic Press, New York, 1981, pp. 33-48.
58. Zyskind, J.W., Harding, N.E., Takeda, Y., Cleary, J.M. and Smith, D.W. In: The Initiation of DNA Replication, ICN-UCLA Symp. Molec. Cell. Biol., Vol. 22 (Ed. D. Ray), Academic Press, New York, 1981, pp. 13-28.
59. Arai, K. and Kornberg, A. Proc. Natl. Acad. Sci. USA 76:4308-4312, 1979.
60. Kaguni, J.M., Fuller, R.S. and Kornberg, A. Nature 296:623-627, 1982.
61. Morita, M., Sugimoto, K., Oka, A., Takanami, M. and Hirota, Y. In: The Initiation of DNA Replication, ICN-UCLA Symp. Molec. Cell. Biol., Vol. 22 (Ed. D. Ray), Academic Press, New York, 1981, pp. 29-35.
62. Lother, H. and Messer, W. Nature 294:376-378, 1981.

6

DNA REPLICATION IN MAMMALIAN CELLS

J. HERBERT TAYLOR

CONCEPTS OF REPLICONS AND ORIGINS

The pattern of replication in chromosomal DNA is still incompletely understood, but it is clear that there are many initiation sites per chromosome. The concept has developed that these may represent origins for replicons, i.e., units with some degree of autonomy in regulation (1,2,3). Replicons in prokaryotes may be whole genomes or chromosomes and the same terminology and concepts apply to plasmids and viral genomes. However, when there are so many sites where replication can be initiated as in the chromosomes of higher organisms, the isolation or observation of one replicon has been so difficult that the concept of defined entities with specific limits within a chromosome is still unproven. Indeed some recent work casts doubt on the idea that specific nucleotide sequences are required for initiation (4). Laskey and Harland injected SV40 and polyoma viral DNAs into unfertilized frog eggs (*Xenopus laevis*) and produced measurable amounts of replication through two rounds as demonstrated with a buoyant density label. They also obtained replication of prokaryotic DNAs, for example, the closed circular replicative forms of bacteriophage G4, M13 and the bacterial plasmid ColE1. Since they could not determine directly by electron microscopy whether the viral origin of replication in SV40 was utilized, they cut the viral genome into segments with a restriction enzyme and ligated the ends to reform circles. When these were injected some of the circles, without the known origin, replicated through two rounds in trace amounts. From these experiments they concluded that specific sequences may not be required for the initiation of DNA replication. Others have reported that closed circular DNAs replicate when injected into

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

Xenopus eggs (5,6,7), but the amount of replication of heterologous DNAs compared to a plasmid with a *Xenopus* origin is small (8). For example, when a bacterial plasmid with a 500 bp segment of *Xenopus* DNA, which had been selected as an efficient replicator, was injected into *Xenopus* eggs it replicated up to 40-fold more efficiently than a bacterial plasmid with only its prokaryotic origin. Two other recent studies, in which the origins were identified by the "eye structures" seen with the electron microscope, indicated much more efficient initiation in the eukaryotic DNA when cloned segments of *Xenopus* DNA were injected into the *Xenopus* egg. Both studies indicated that when prokaryotic DNAs were occasionally utilized for initiation in the eukaryotic system, the prokaryotic origin was very likely to be used rather than random sites. One study was based on injection experiments (7), but the other was based on initiation *in vitro* with a crude enzyme preparation from *Xenopus* eggs (9). Therefore, it is possible that origins have some common property throughout all living systems. In addition, replication may be initiated in the egg or early embryo more readily than in other cells since many more initiation sites are ordinarily utilized in these stages when divisions and DNA replication are rapid (10).

A HYPOTHESIS THAT ORIGINS ARE SPECIFIC SEQUENCES, THE MODERATELY REPEATED SEQUENCES.

I wish to develop the thesis that origins with regulatory functions do, indeed, exist not only in prokaryotic DNAs and certain eukaryotic viral DNAs, but in eukaryotic chromosomes. Initiation of DNA replication may occasionally occur without a specific origin, and origins may be as important in reducing or regulating replication as they are in initiating it. It is fairly clear in the studies with *Xenopus* eggs that the first initiation of heterologous DNAs occurs more readily than the one for a second round of replication, but homologous DNAs will frequently initiate a second time (6). Why should that be unless the first replication is in some respect defective in heterologous DNA compared with that initiated at a homologous origin?

The origins which function best in the *Xenopus* system are

the moderately repetitive sequences (8,9). We found that several plasmids with one of the Alu family of repeated sequences from the human genome functioned as well as the plasmid with the 500 bp of *Xenopus* DNA in our injection studies. The plasmid (pSW14) with the 500 bp of *Xenopus* DNA was originally selected for its efficiency in replication when injected into the *Xenopus* egg. It was then tested by hybridization to see if it contained a repetitive sequence. This hybridization test, in which random segments of genomic DNA were compared with pSW14 for their efficiency in binding a nick-translated probe prepared from the 500 bp segment, showed that a sequence in the 500 bp segment was present about every 8-12 kb on the average throughout the *Xenopus* genome. The Alu family of sequences in human cells is present in about 300,000 copies per genome i.e. one is spaced on the average about every 10 kb (11). Some of these copies from mammalian cells are transcribable by RNA polymerase III while others are not (12). In the human genome many of the copies have a single site for the restriction enzyme Alu I, hence the name Alu family. This sequence is not present in the similar repeated sequences isolated from other mammals, but Haynes et al. (12) call all of the sequences Alu-equivalent sequences. I wish to use the remainder of this review to bring together the admittedly incomplete evidence that these repetitive sequences, which are transcribable by RNA polymerase III, function as origins of replication in cells of vertebrates and similar sequences probably function in other eukaryotic cells to initiate and regulate replication. Indeed they may play a more significant role in regulating the activity of genes, i.e. the inactivation and release of genes from the inactive state during differentiation. This concept with some changes based on new information has been a long standing thesis of mine that goes back more than 20 years to my early studies of the changes in time of replication and the inactivation of X-chromosomes in mammals (13).

THE SIZE OF THE REPLICON IN CELLS OF VERTEBRATES

The replicon in vertebrate cells is generally assumed to be much larger than the spacings between Alu-equivalent sequences

along the DNA. The first attempts to estimate the size of replicons was based on measurements of rates of chain growth in replicating DNA (14,15). The problem was that one could not determine directly how long the replicating fork operated from one point of initiation, but a rough estimate was made which indicated that interruptions might occur every 7-9 kb (5-6 x 10^6 daltons of duplex DNA). This estimate was based on buoyant density of sheared DNA, partially substituted with bromodeoxyuridine (14).

About the same time a much more acceptable estimate was made by Huberman and Riggs (2) based on a technique of autoradiography of single fibers (single double helixes of DNA) originally devised by Cairns (16). Cells were pulse labeled with [^{3}H]-thymidine of high specific activity, lysed and the long chains of DNA attached to a membrane filter. A photographic emulsion was applied and after an exposure of several months autoradiographs were prepared and examined with the light microscope (later the DNA fibers were oriented by flow directly on a glass slide by Lark et al., 17). The DNA fibers were invisible, of course, but the tracks or rows of silver grains indicated the location of the labeled chains. If the pulse labeling was for 20-30 min the labeled segments were about 10-15 μm long, but if labeling was for long times the fibers could be so long that measuring them was difficult if not impossible. Huberman and Riggs (2) discovered evidence in these studies that fork movement was bidirectional from a point or origin. Fibers much longer than a unit of replication were seen with origins in tandem indicated. Measurements from center to center of the tandemly labeled segments were assumed to indicate the length of the replicon. This first estimate indicated units 10 to over 100 μm in length with a mean of about 30 μm for Chinese hamster cells in culture. However, these measurements were based on cells which had been blocked with fluorodeoxyuridine (FdU) for 12 hr to partially synchronize the cells at the beginning of S phase. This treatment which depletes thymidylate and stops, or severely retards, DNA replication decreases the distance between the origins which are utilized (18,19,20). Measurements

made on unsynchronized cells by pulse labeling without pretreatments yielded a center to center length of about 75 μm which may be a more accurate measure. However, if cells are synchronized and blocked at early S phase for 12-14 hr, the distance between origins can be reduced to as little as 4 μm with additional modal frequencies at intervals of 8, 12, 16 and intermediates up to 40-50 μm (19,20). The pulse labeling period must be reduced accordingly to as little as 4 min to resolve these sites. These experiments led to the proposal that origins might be rather evenly spaced at about 4 μm (12 kb) in mammalian cells. That would indicate that there are about 250,000 per genome which is close to the estimate of the number of Alu sequences per genome in human cells (11) and the number of repeats of the cloned origin from Xenopus (8 and unpublished results). The Alu-equivalent repeats in vertebrate cells are bracketed by short repeats suggesting transposable elements (12,21,26) and are frequently in single copies at each site. Some of these are transcribed in the same direction as the adjacent genes so that a whole cluster of the transcribable Alu sequences may be oriented in the same direction, for example, as in the β-like cluster of the human hemoglobins (22). The repeats may also rarely occur in closely adjacent pairs, and if two such copies were transcribable and had opposite orientation these might represent a bidirectional origin for a replicon (Figure 1).

← ← ← ←→ → → →

unidirectional origins — bidirectional origin — unidirectional origins

FIGURE 1. Diagram showing the hypothetical arrangement of Alu sequences which are transcribable by RNA polymerase III. The paired sequences in the center of the cluster would be the bidirectional origin of a replicon. In each direction the other sequences would be origins which are usually "fired" in sequence as the replication fork approaches them, but under some conditions as in a block of replication by depletion of thymidylate, the other sequences may "fire" more or less at random when thymidylate is restored but produce unidirectional "eyes".

The other potential origins if used would initiate unidirectionally. There is evidence that many origins do initiate

unidirectionally (17) especially if the cells are blocked with FdU for some time before a pulse label is given.

A MODEL FOR REPLICON ORGANIZATION AND CONTROL

The model which I propose then is one in which there may be complex replicons composed of many subunits, each with its own potential origin. At the center may be one with two transcribable repeats which are transcribed as the primer for initiating the leading chains for a bidirectional replicon. Replication from these origins proceeds bidirectionally, but stops at each transcribable Alu-equivalent segment where initiation normally occurs very quickly again by formation of another primer segment. Under some conditions independent initiation can occur at any of these repeats, but at all except the center pair in the replicon, fork movement will be unidirectional. The primer segments are probably ligated into the new DNA chains. If the primer segments remain in the chromosomal DNA for more than a few minutes the nucleotides must be modified to make them resistant to alkali and most ribonucleases, a well known property of DNA chains. The hypothetical modification which I propose is 2'-O-methylation of the ribose. The modified primer segments could then serve as a block to replication and thereby prevent two replications in one S phase. Removal might occur during S phase, but probably at the end of S phase and certainly before the next replication cycle. The priming of replication referred to here is quite different from that which occurs on the lagging chain which grows from the fork back toward the origin. The lagging chain presumably is replicated in short segments with primers of about 7-10 nucleotides according to the model proposed for the eukaryotic viruses, SV40 and polyoma (23,24). The section below presents evidence from our laboratory going back more than 15 yr which supports this model.

EVIDENCE FOR PRIMERS AT REPLICATION ORIGINS

The first relevant observation was a peculiar partially substituted component produced when roots of *Vicia faba* were grown in a solution containing bromodeoxyuridine (BrdU). The DNA was

first labeled with $[^3H]$-thymidine for 8 hr. After a 3.5 hr chase with unlabeled thymidine the roots were grown in $[^{14}C]$-BrdU for 10 hr. Although only a fraction of the DNA in the root was converted to a buoyant density hybrid because of the asynchrony and slow division of part of the cells, some of the $[^3H]$-labeled DNA was converted to hybrid by full substitution (25). As shown in Figure 2, a significant fraction containing ^{14}C was shifted only

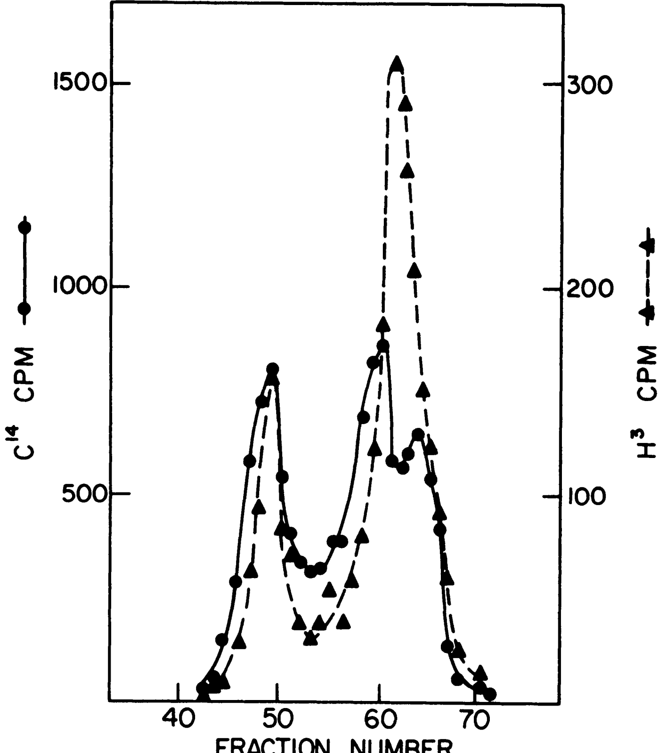


FIGURE 2. Density gradient profile of DNA banded in CsCl after extraction from roots of Vicia faba which were grown for 8 hr in a solution with $[^3H]$-thymidine to label a random sample of the DNA. The label was chased for 3.5 hr with unlabeled thymidine and then a random sample of the DNA was labeled by growing the roots in a solution of $[^{14}C]$-BrdU (10^{-4} M) along with aminopterin (10^{-6} M) and adenosine, glycine and hypoxanthine (5×10^{-6} M). Notice that besides the expected density hybrid a significant fraction of the $[^{14}C]$-BrdU is a partially substituted fraction that is banded below and overlapping the unsubstituted DNA. Density of the gradient increases to the left in the figure (redrawn from Haut and Taylor, 1967).

a small amount in density. In Vicia there were two density species of this partially substituted DNA, but when Chinese hamster cells were studied by the use of the same density label a single peak was regularly found which was partially substituted (Figure 3). Shearing to produce smaller segments increased the buoyant density of the partially substituted fraction when it was isolated and rebanded (14). In addition the fraction maintained a relatively higher density compared to unsubstituted DNA when denatured and further fragmented (25). We assumed at the time that some type of repair might be occurring in which relatively short segments

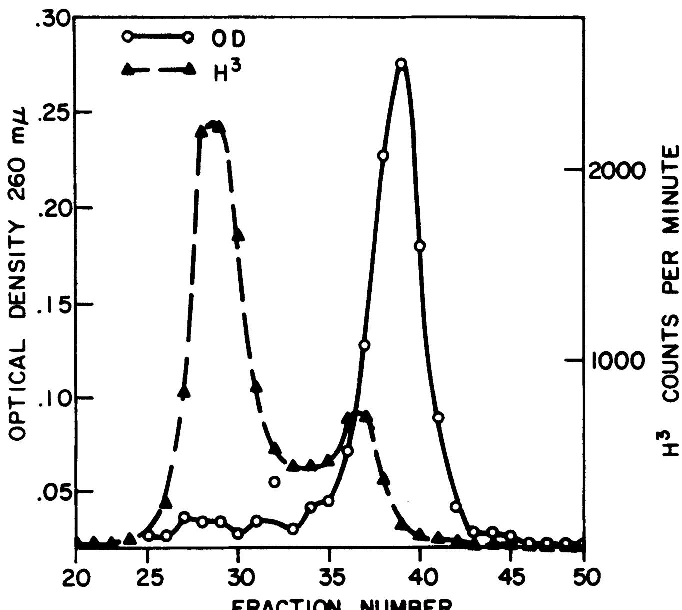


FIGURE 3. Density gradient profile of DNA banded in CsCl after extraction of DNA from Chinese hamster cells in culture which had been treated for 15 min with FdU (5×10^{-6} M) and then grown for 1 hr in [^{3}H]-BrdU (10^{-5} M) with FdU (2×10^{-6} M). DNA was extracted, sheared and banded by spinning in a Spinco SW39 rotor for 48 hr. Note that most of the BrdU is in the expected density hybrid band to the left in the figure but a distinct peak indicates a band only slightly to the heavy side of the main band of unsubstituted DNA. The cells were an asynchronous culture and therefore the minor peak could be in cells at a different stage in the cycle from those that form the hybrid bands (redrawn from the dissertation by Philip Miner submitted to the Faculty of Pure Science, Columbia University, New York in 1967).

of BrdU-substituted DNA was inserted in the newly replicated chains. The tritium-labeled chains from the previous replication cycle did not band with either the partially labeled fragments or the new chains of fully substituted DNA after denaturation (data shown by Haut and Taylor, in reference 25). This peculiar type of "repair" did not seem to make sense since no damage was known to have occurred to the DNA.

One other series of experiments yielded a large fraction of this partially substituted DNA. When unsynchronized cells were pulse labeled with [^{3}H]-BrdU at 31°C instead of 37°, a 5 min pulse produced most of the newly replicated and labeled DNA with only a partial substitution. After a 10 or 20 min pulse there were two rather narrow bands, one with the buoyant density of fully substituted DNA and another only slightly heavier than the unsubstituted DNA, with some material of intermediate density (14).

In all of the above experiments the DNA was sheared before banding in the CsCl gradients. The partially substituted material was not revealed unless the DNA was sheared. On the other hand when cells were synchronized and released into S phase with BrdU for 3 hr, two symmetrical bands were typically produced, a BrdU hydrid and the normal unsubstituted DNA. No satellite bands or partially substituted DNAs were produced with DNA from the synchronized Chinese hamster cells in the first half of S phase. In some pulse labeled cells which had been synchronized at early S phase a small satellite was evident after only a 5 min pulse, but this disappeared after longer labeling periods. The replication of a high density satellite in early S phase has been assumed to explain this change (27), but the satellite has not been demonstrated by any other means.

The interpretation of the buoyant density changes mentioned above are possibly subject to different explanations and without supplementary evidence the presence of large primer segments at origins would not be proposed. However, soon after these experiments were performed evidence of quite a different nature was obtained. Newly replicated segments (Okazaki fragments) were collected by pulse labeling Chinese hamster cells. At this time we typically blocked cells with FdU for a few minutes before or

simultaneously with pulse labeling by the use of 10^{-5} M [^{3}H]-thymidine for 30 seconds to one minute before rapid lysis. Isolation of the single-stranded, newly replicated DNA on hydroxyapatite was carried out after phenol extraction. The newly replicated Okazaki fragments appeared to be longer than some other reports indicated (34) when we measured them by sedimentation. A significant fraction of the radioactivity seemed to be associated

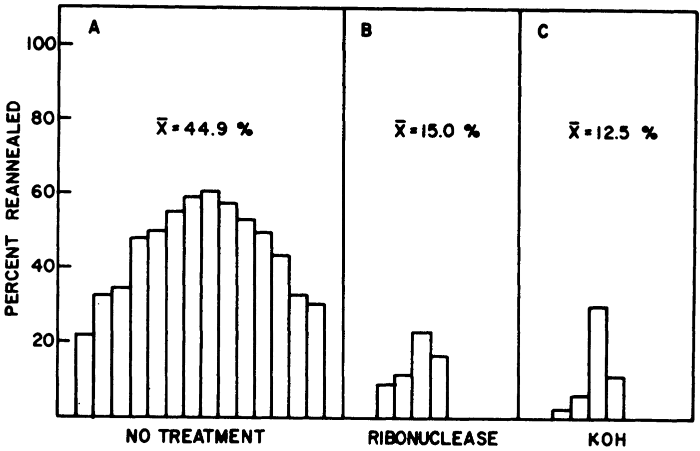


FIGURE 4. Reannealing of nascent, single-chain DNA to the intermediate repetitive DNA sequences from Chinese hamster cells. Single chains were isolated from hydroxyapatite columns in 0.12 M phosphate (pH 6.8) at 60°C after loading the pulse—labeled DNA in 0.05 M phosphate. The eluents were concentrated by evaporation in dialysis tubing and dialyzed against 0.12 M phosphate for reannealing. Intermediate repetitive DNA sequences were prepared from cells in which DNA was labeled for 2 cell cycles with [^{14}C]-thymidine (0.05 μCi/ml; 2 x 10^{-6} M). Cells were lysed with 0.15 M NaCl, 0.05 M Tris, 0.015 M EDTA, 0.3% SDS, pH 9.5. Lysates were digested for 12 hr with 1 mg/ml of pronase (Cal Biochem) and extracted into the buffer after adding phenol and chloroform. Ten μg/ml of pancreatic ribonuclease was added, and after incubation at 37°C for 1 hr, the DNA was extracted again with phenol and chloroform and dialyzed to remove traces of phenol. The DNA concentration was determined from its absorbance at 260 nm and the radioactivity per μg assayed from TCA-precipitated samples. The DNA was then sonicated until the denatured DNA was about the same size as the nascent, single-chain DNA obtained after pulse labeling. The DNA was denatured by holding in a boiling water bath for 20 min, reannealed at 65°C, and the fractions with C_0t values between 0.1 and 10.0 were separated off hydroxyapatite columns (redrawn from Taylor et al., 1972).

with fragments 500-1000 nucleotides in length. Later EM measurements indicated that the Okazaki fragments varied from 200 to about 1200 nucleotides in length (28). However, the most puzzling property of the fragments was their annealing to the middle repetitive DNA (Figure 4). Much of the newly replicated DNA (about 45%) annealed to DNA with C_0t values between 0.1 and 10.0. This property of the fragments was largely eliminated by digesting them with pancreatic ribonuclease or by treatment with alkali under conditions that would hydrolyze RNA (29).

From one minute pulse—labeled cells prepared as above we could also detect two bands of single—stranded DNA segments in CsCl equilibrium gradients (Figure 5). When these fragments were isolated from the gradients, dialyzed and examined with the electron

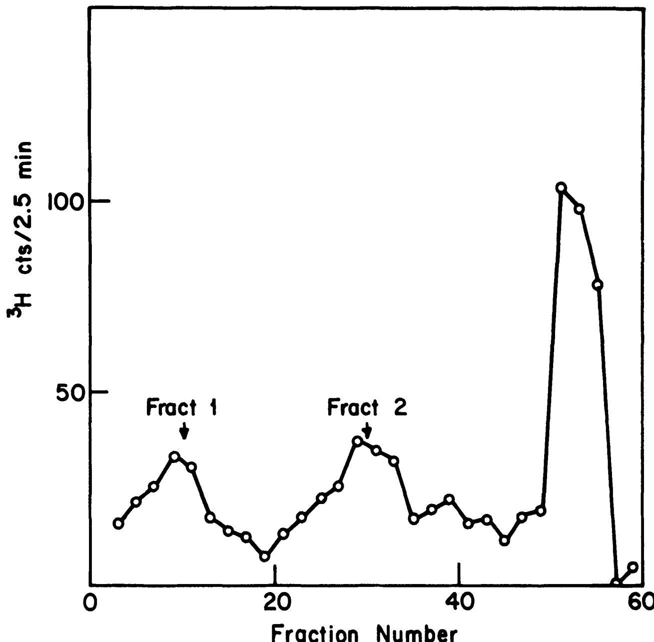


FIGURE 5. Profile showing the CsCl banding pattern of DNA from CHO pulse labeled for 1 min with [^{3}H]-thymidine following 2 min in Ham's medium with 10^{-6} M FdU. The centrifuge tube contained 7.35 g of CsCl per 6.0 g of lysate. It was spun for 48 hr at 35,000 rpm in a Spinco 65 rotor. Sixty fractions of 8 drops each were collected. The even numbered ones were refrigerated for use in electron microscopy while the alternate ones were precipitated and counted to produce this profile (from Taylor et al., 1975).

microscope, the length varied from about 200 to 1200 nucleotides. The bands were broad and the separation of the two classes of fragments was incomplete in the gradients, but a distinct difference between those of higher buoyant density and those with a density typical of single chains of DNA was noted (28). When the two types of fragments were examined in spreads for the electron microscope in which the hypophase contained 10% formamide, the fragments of higher buoyant density could be distinguished from those of the same density as regular single-stranded DNA. More than half of those of higher buoyant density retained a terminal bush which indicated that a considerable length of RNA was attached (Figure 6). From the buoyant density shift a rough estimate of 140-180 nucleotides for the length of the RNA was made. This may be compared to the length of the Alu-equivalent repeats in Chinese hamster cells which is in the same size range (12).

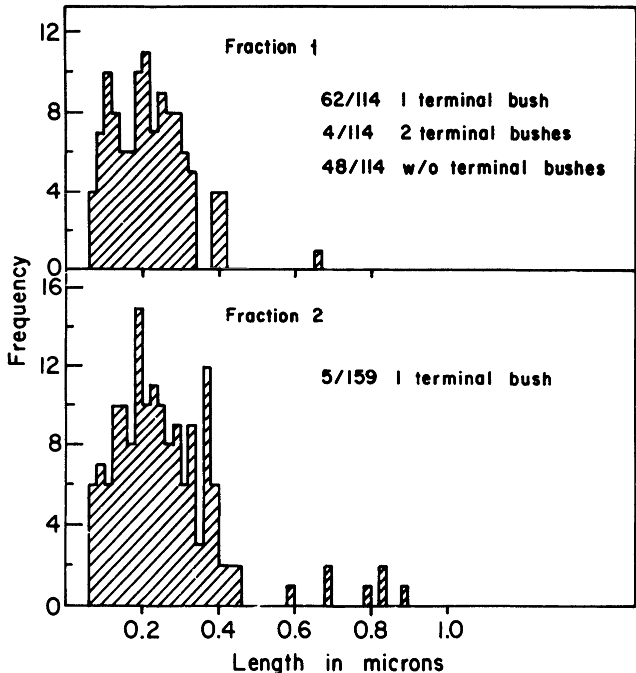


FIGURE 6. Frequency diagrams showing the sizes of Okazaki fragments as well as the occurrence of terminal bushes (presumably RNA) seen with the electron microscope from fractions 1 and 2 shown in Figure 5 (from Taylor et al., 1975).

RECENT EVIDENCE FOR A 10 kb INTERMEDIATE IN DNA REPLICATION.

One other recent bit of evidence for special sites about every 10 kb in the human complement comes from a report by Lönn (30). He detected a 10 kb replication intermediate in cultured human melanoma cells in the following way. When the DNA in cells was labeled with [^{3}H]-thymidine and then prepared for electrophoresis in a rather unusual way, a segment of 10 kb was found. The cells were lysed in 0.03 M NaOH at 0°C in the dark. After 30 min, the solution was neutralized by adding 0.067 M HCl, 0.02 M NaH_2PO_4 and 0.5% SDS. After 30 min at 25° the labeled DNA was separated by electrophoresis in 6% acrylamide gels in the cold. Soluble counts were rinsed from the gels with 5% trichloroacetic acid and the insoluble radioactivity was determined by

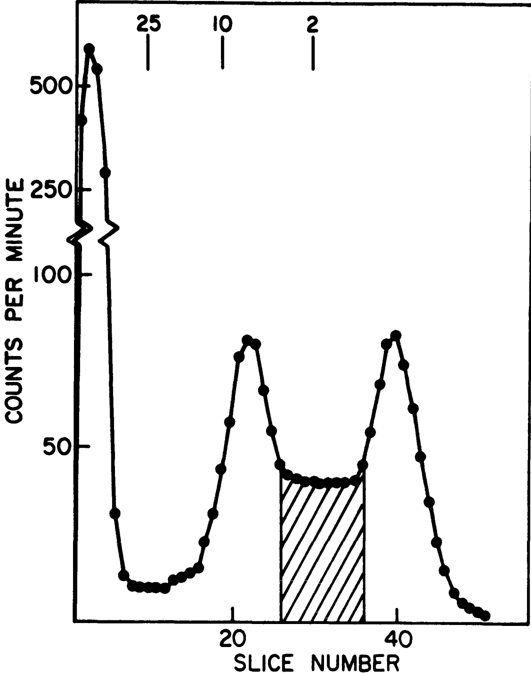


FIGURE 7. Separation of newly replicated DNA in a 0.75% agarose gel by electrophoresis. Human melanoma cells were pulse labeled for 45 sec with [^{3}H]-thymidine and the DNA prepared for electrophoresis as described in the text (redrawn from unpublished data supplied by Ulf Lönn, Karolinska Institutet and Radiumhemmet, Stockholm; further details are being published in Chromosoma).

slicing the gel and counting by liquid scintillation. Figure 7 shows the distribution of counts after a 45 sec pulse label. Much of the labeled DNA was very long, but an 8-10 kb component and the much shorter Okazaki fragments were also detected. These segments of intermediate size could represent the newly replicated DNA separated by RNA primers, gaps or nicks at 10 kb intervals. The NaOH should break the RNA, even if it were ligated into a continuous piece with DNA, and release the 10 kb single chains. However, as soon as the primer is removed or modified by 2-O-methylation as proposed in the model given here the intermediate segments would no longer be produced. During the removal of the primer by the special repair process proposed, some gaps or nicks would appear. Consistent with this was Lönn's (30) report that a very small fraction of the DNA could be isolated as 10 kb intermediates many hours after the end of the pulse labeling of the DNA.

SUMMARY AND RESERVATIONS

The evidence for the large bidirectional replicons with many smaller subunits each with a unidirectionally transcribed primer is fragmentary, but I think it is consistent with most of the observations and data we have available. However, the existence of the larger regulatory unit, the replicon implied by this model of the eukaryotic chromosome, is probably supported by less evidence than some other aspects. That bidirectional replication occurs is not in question, but that functional or regulatory replicons of 200,000-300,000 bp are a regular feature of cells is going to be difficult to establish. To clone pieces of this size is not yet feasible; therefore such large domains of DNA must be studied by segments. More extensive mapping and sequencing of the Alu-equivalent sequences is necessary and highly desirable.

The most interesting and yet the least supported aspect of the model is its possible regulatory role in differentiation. Since evidence is rapidly accumulating to support the hypothesis that methylation of cytosine in DNA is an important control mechanism in gene expression as proposed by Holliday and Pugh (31), Riggs (32) and Taylor (33), the establishment and maintenance of methylation patterns becomes of great interest. Since there is

evidence that changes in patterns of replication over the cell cycle is sometimes associated with activation or inactivation of genes, the link between control of replicons or their sub-units becomes more intriguing (35).

REFERENCES

1. Taylor, J.H. J. Cell. Comp. Physiol. 62 (Suppl. 1):73-86, 1963.
2. Huberman, J.A. and Riggs, A.D. J. Mol. Biol. 32:327-341, 1968.
3. Callan, H.G. Proc. R. Soc., Lond. B. 181:19-41, 1972.
4. Harland, R.M. and Laskey, R.A. Cell 21:761-771, 1980.
5. Laskey, R.A. and Gurdon, J.B. Eur. J. Biochem. 37:467-471, 1973.
6. Watanabe, S. and Taylor, J.H. Proc. Natl. Acad. Sci. USA 77: 5292-5296.
7. Hines, P.J. and Benbow, R.M. Cell (submitted), 1982.
8. Taylor, J.H. and Watanabe, S. In: The Initiation of DNA Replication, ICN-UCLA Symp. Mol. Cell. Biol. (Ed. D.S. Ray), Acad. Press, New York, 1981, pp. 597-606.
9. Hiraga, S., Sudo, T., Yoshida, M., Kubota, H. and Ueyama, H. Proc. Natl. Acad. Sci. (in press) 1982.
10. Blumenthal, A.B., Kriegstein, H.J., Hogness, D.S. Cold Spr. Harb. Symp. Quant. Biol. 38:205-223, 1974.
11. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. Proc. Natl. Acad. Sci. USA 77:1398-1402, 1980.
12. Haynes, S.R., Toomey, T.P., Leinwand, L. and Jelinek, W.R. J. Mol. Cell. Biol. 1:573-583, 1981.
13. Taylor, J.H. J. Biophys. Biochem. Cytol. 7:455-464, 1960.
14. Taylor, J.H. and Miner, P. Cancer Res. 28:1810-1814, 1968.
15. Taylor, J.H. J. Mol. Biol. 31:579-594, 1968.
16. Cairns, J. J. Mol. Biol. 15:372-373, 1966.
17. Lark, K.G., Consigli, R. and Toliver, A. J. Mol. Biol. 58: 873-875, 1971.
18. Ockey, C.H. and Saffhill, R. Exptl. Cell Res. 103:361-373, 1976.
19. Taylor, J.H. and Hozier, J.C. Chromosoma (Berl.) 57:341-350, 1976.
20. Taylor, J.H. Chromosoma (Berl.) 62:291-300, 1977.
21. Bell, G.I., Pictet, R. and Rutter, W.J. Nuc. Acids Res. 8: 4091-4109, 1980.
22. Fritsch, E.F., Shen, C.K.J., Lawn, R.M. and Maniatus, T. Cold Spr. Harb. Symp. Quant. Biol. 45:761-776.
23. Reichard, P., Elliasson, R. and Soderman, G. Proc. Natl. Acad. Sci. USA 71:4901-4906, 1974.
24. DePamphilis, M.L., Anderson, S., Bar-Shavit, R., Collins, E., Edenberg, H., Herman, T., Karas, B., Kaufmann, G., Krokan, H., Su, R., Tapper, D. and Wasserman, P.M. Cold Spr. Harb. Symp. Quant. Biol. 43:679-692, 1978.
25. Haut, W.F. and Taylor, J.H. J. Mol. Biol. 26:389-401, 1967.

26. Duncan, C.H., Jagadeswaran, P., Wang, R.R.C. and Weissman, S.M. Gene 13:185-196, 1981.
27. Maio, J.J. and Shildkraut, C.L. J. Mol. Biol. 40:203-216, 1969.
28. Taylor, J.H., Wu, M., Erickson, L.C. and Kurek, M.P. Chromosoma 53:175-189, 1975.
29. Taylor, J.H., Wu, M. and Erickson, L.C. Cold Spr. Harb. Symp. Quant. Biol. 38:225-231, 1973.
30. Lönn, U. Chromosoma (in press, 1982).
31. Holliday, R. and Pugh, J.E. Science 187:226-232, 1975.
32. Riggs, A.D. Cytogenet. Cell Genet. 14:9-25, 1975.
33. Taylor, J.H. In: Molecular Genetics, Part 3, pp. 89-115 (J.H. Taylor, ed.) Acad. Press, New York.
34. Endenberg, H.J. and Huberman, J.A. Ann. Rev. Genet. 9:245-284, 1975.
35. Mohandas, T., Sparkes, R.S. and Shapiro, L.J. Science 211:393-396, 1981.

7

REPLICATION OF DROSOPHILA MITOCHONDRIAL DNA

DAVID R. WOLSTENHOLME, JUDY M. GODDARD, AND CHRISTIANE M.-R. FAURON

Mitochondrial DNA (mtDNA) molecules of Drosophila include a region of exceptionally high adenine + thymine (A+T) content which, though homologously located, varies in size from 1.0 kb to 5.1 kb in different species. In all species examined, replication begins at a unique site within the A+T-rich region and proceeds in the same direction towards the molecule's ribosomal RNA genes. Most molecules are replicated by a highly asymmetrical mode in which synthesis of one strand can be up to 100% complete before synthesis of the second strand is initiated. In a minority of molecules, a more symmetrical mode of synthesis is employed. Evidence has been obtained that within A+T-rich regions extensive sequence divergences have taken place both between species and within species. The A+T-rich region shares many features with the replication origin-containing region of vertebrate mtDNA molecules, supporting the view that the two regions are derived from a common ancestral sequence.

INTRODUCTION

Metazoan mitochondrial DNA (mtDNA) shows a remarkable evolutionary conservation of size and shape. In all organisms the molecules are circular duplexes, which usually comprise approximately 16.5 kb. Recently, sequencing of the entire mitochondrial genome of human, beef, and mouse has been completed (1, 2) and much of the gene content of these mtDNAs has been determined. Each mtDNA molecule is identical and encodes the two rRNAs and 22 tRNAs of the mitochondria's distinct protein synthesizing system, and five known polypeptides (3 subunits of cytochrome c oxidase, ATPase subunit 6, and cytochrome b). In addition there are 8 presently unidentified open reading frames which are presumed to also code polypeptides. The replication origin lies within a sequence which varies somewhat in length between species.

Much less is known concerning other metazoan mtDNAs. In mtDNA of the amphibian Xenopus laevis the relative locations of the rRNA genes, tRNA genes, replication origin, and the templates for a number of polyadenylated RNA transcripts are similar to those found in mammalian mtDNAs (3, 4). Among invertebrates, Drosophila mtDNA has been the most studied. Again, it is known

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

that the rRNA genes and the origin of replication occupy similar relative positions to those found in vertebrate mtDNAs (5-8) and some transcripts have been mapped on the D. melanogaster mtDNA molecule (9, 10).

Replication of metazoan mtDNA has been studied extensively. Using electron microscopy, replicative forms were first identified in rat liver mtDNA (11). Since then studies have been carried out on replication of mtDNAs from tissues of a number of mammals (12-15), but most thoroughly on the mtDNA from mouse L-cells in culture (see 2 for references). All metazoan mtDNAs studied so far employ a unique highly asymmetrical mode of replication in which synthesis of one strand is two thirds or more complete before synthesis of the second strand is initiated. However, equally interesting is the finding that in different molecules from tissues of some animals, a variety of more nearly symmetrical modes are employed.

In this article we review studies on the replication of Drosophila mtDNAs, which differs in a number of details from replication of vertebrate mtDNAs. In Drosophila mtDNAs the region containing the replication origin is unique among metazoa, in that it has a very high (about 95%) A+T content, and shows extensive sequence divergence both within species and between species.

MATERIALS AND METHODS

The origins and derivations of all Drosophila species and different lines of a single species used in the experiments referred to in this article are given in references 16-18. All methodological details can be found in references 6, 7, 17, 19-21.

RESULTS AND DISCUSSION

The A+T-rich region of Drosophila mtDNA molecules

It was shown independently by three groups (22-24) that approximately 25% of the DNA isolated from mitochondria of D. melanogaster denatured at a temperature indicative of a higher adenine + thymine (A+T) content. A clear demonstration that most of this A+T-rich DNA occurs as a single region of constant size (5.1 kb) in each mtDNA molecule was provided from electron microscope observations of molecules which had been partially denatured by either heat (Fig. 1, 16, 19) or alkali (5). Confirmation that this segment of lower thermal stability is in fact a region with an A+T content of approximately 95% was obtained from buoyant density analyses of restriction fragments containing the entire A+T-rich region (20).

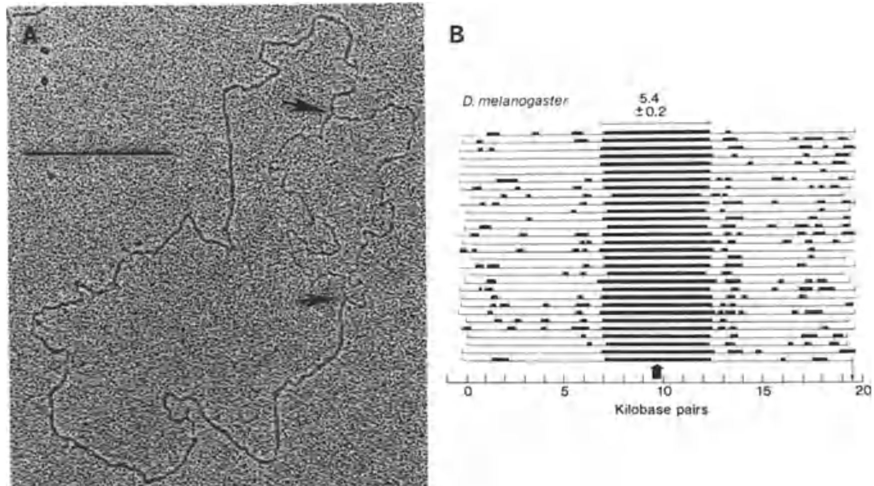


FIGURE 1. D. melanogaster circular mtDNA molecules from a preparation heated at 40°C for 10 min in 0.05 M sodium phosphate (pH 7.8) and 10% formaldehyde. A. An electron micrograph. The limits of the denatured (A+T-rich) region are shown by arrows. The bar indicates 0.5 μm. B. Denaturation maps derived from molecules such as those shown in A.

The G+C content of Drosophila mtDNA molecules outside the A+T-rich region, derived from both buoyant density and thermal melting studies, is about 25% (16, 19) which is much lower than that of other well-studied mtDNAs such as human (44.3% G+C; 1), mouse (36.7% G+C; 2), and Xenopus laevis (42.9% G+C; 25). The two complementary stands of vertebrate mtDNAs have a base bias which permits easy separation into H (heavy) and L (light) strands. In contrast, the complementary strands of Drosophila mtDNAs have little or no base bias (16, 20). This prevents comparisons of the Drosophila complementary strands and the H and L strands of vertebrate mtDNAs.

We determined that the mtDNA molecules of different species (20 examined from 9 subgroups) of the melanogaster group of the subgenus Sophophora differed in size from 15.7 to 19.5 kb. Using denaturation mapping (Fig. 1) we showed for 8 species that these differences could be accounted for by differences in size (1.0 to 5.1 kb) of an A+T-rich region. Among 20 other species representing 13 groups of the 5 subgenera of Drosophila, the size of mtDNA molecules was within the narrower range of 15.7 to 16.8 kb, and the mtDNA molecules of one species, D. virilis was shown to contain an A+T-rich region of 1.0 kb (16).

Six species of Drosophila, chosen because of their different degrees of

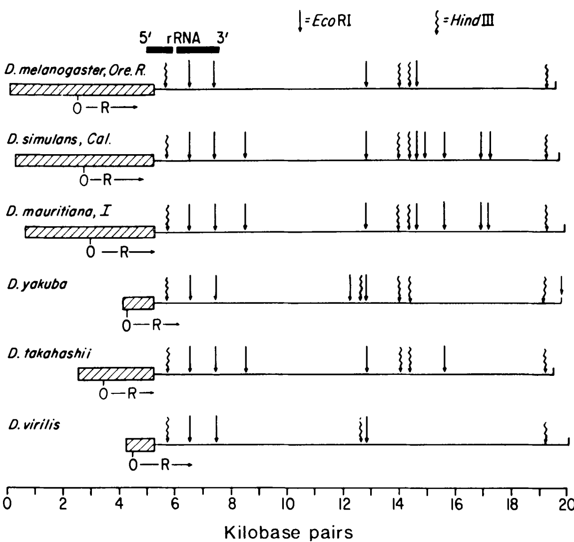


FIGURE 2. Maps of the mitochondrial genomes of the six Drosophila species indicated, showing the relative positions of the A+T-rich regions (hatched areas) and the positions at which EcoRI and HindIII cleave. The genomes have been oriented so as to maximize the coincidence of enzyme sensitive sites, and then aligned by the common EcoRI site nearest the A+T-rich region. This site defines the right end of the A+T-rich region in the figure and all molecules are linearized at what is then the left end of the A+T-rich region. The origin (O) and direction of replication (R) are shown for each genome. The position and polarity of the two rRNA molecules transcribed from the D. melanogaster mtDNA molecules are from Klukas and Dawid (5).

taxonomic relatedness, were used in further studies. D. melanogaster, D. simulans, D. mauritiana, D. yakuba, and D. takahashii are all members of the melanogaster group of the subgenus Sophophora. The first four species are members of the melanogaster subgroup. Of these D. melanogaster, D. simulans, and D. mauritiana are the most closely related and will interbreed, though the products are largely sterile. D. takahashii is in a separate subgroup. D. virilis was chosen as a taxonomically more distant species belonging to the separate subgenus Drosophila (8, 18 for references, discussion and details).

In order to gain information on the relationship of A+T-rich regions of mtDNA molecules of the six different Drosophila species, we determined in each the relative locations of sites cleaved by the restriction enzymes EcoRI and HindIII and of the A+T-rich regions. When the restriction maps were aligned, as shown in Fig. 2, and oriented so as to maximize the coincidence of

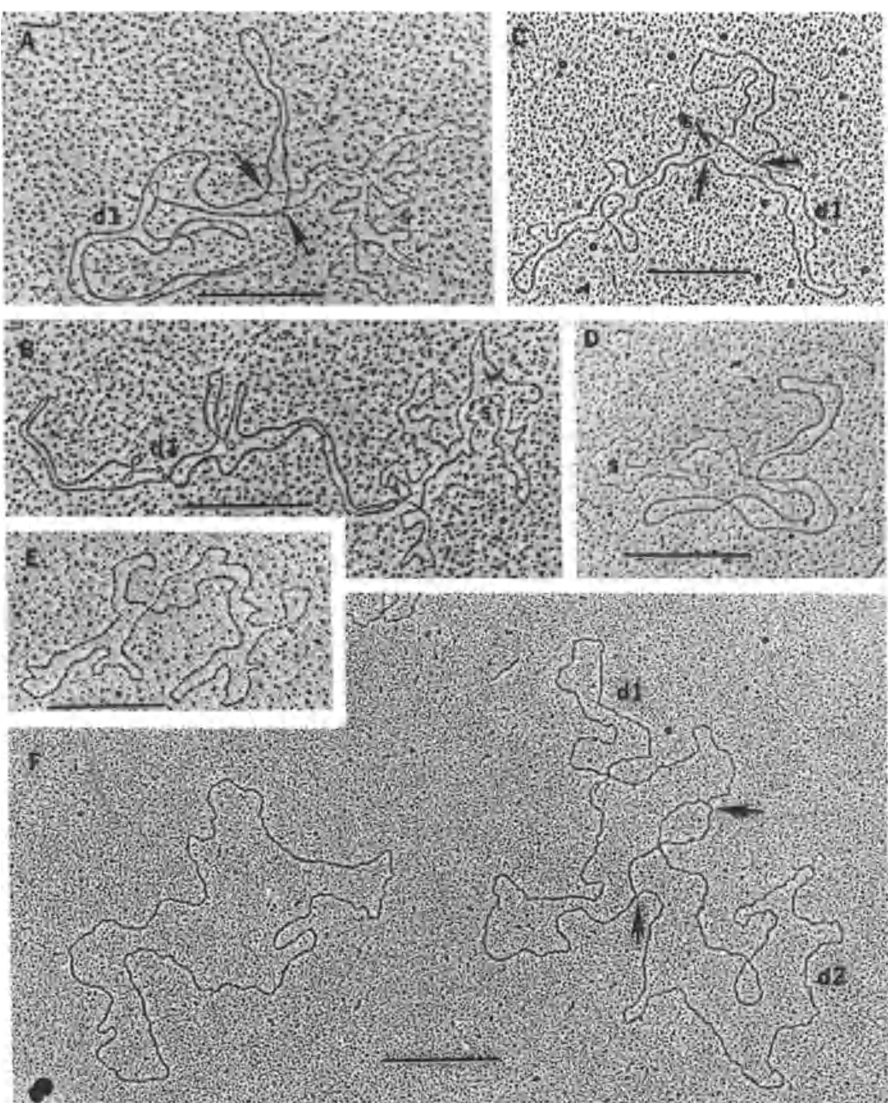


FIGURE 3. Electron micrographs of replicative intermediates of *D. melanogaster* mtDNA molecules (bar = 0.5 m). The molecules shown in A, B, C, and F (right side) are double-forked (arrows) circular molecules in which one daughter segment (d1) is totally double-stranded, and the second daughter segment is either totally double-stranded (d2 in F), totally single-stranded (s in A and B), or totally double-stranded at one fork and single-stranded (s) at the other fork (C). The replicated regions in A, B, C, and F are 77%, 99-100%, 34% and 90% of the genome length, respectively. The molecules shown in D, E, and F (left side) are simple (nonforked) circular molecules. F is completely double-

restriction sites, it appeared that all molecules have 3 EcoRI sites and 2 HindIII sites in common. Other common restriction sites between various groups of species were also found. Assuming that these inferences of common restriction sites and the corresponding homologies are correct, then it is clear that the A+T-rich region occupies an homologous position in the mtDNA molecules in all of the species studied.

Relicative forms

Approximately 1% of circular DNA molecules obtained from mitochondria of embryonated eggs of D. melanogaster (Fig. 3) had the characteristics of size and structure which had been interpreted by us (13, 14) and others (12, 15) to be replicative forms of mtDNA molecules from various other metazoa. All of these molecules (n=59) contained two forks and had a totally double-stranded daughter segment (defined as the first daughter segment) in the replicated region. The molecules differed in the structure of the other daughter segment (second daughter, Table 1). In the majority (n=49) of them, the second daughter segment was totally single-stranded (Fig. 3A and B). The length of the replicated region ranged in these molecules from 2 to 99% of the genome length. A distinct class of molecules containing D-loops [a tristrand structure formed by synthesis of the first few hundred nucleotides of the H strand in vertebrate mtDNAs (26)] was not apparent. In 6 of the replicative intermediates observed, the second daughter segment was also totally double-stranded, and the replicative region was 90 to 96% of the genome length. In 4 of the replicative intermediates, the second daughter segment (10 to 98% genome length) was single-stranded at both forks but double-stranded in a central region. In other preparations molecules were found in which the second daughter segment was single-stranded at only one fork.

We next examined mtDNAs obtained from lysed ovaries of each of the species D. simulans, D. mauritiana, D. yakuba, D. takahashii, and D. virilis. All of the structural kinds of circular replicative intermediates found in D. melanogaster mtDNA were found in mtDNAs of each of these species (Table 1).

We obtained further data on replicative intermediates of D. melanogaster mtDNA from observations on another sample of embryo mtDNA, and prepared for electron microscopy by the aqueous monolayer technique which results in

---

stranded, E is completely single-stranded, and D contains a single-stranded region (s) accounting for 40% of the genome length (in part, modified from Fig. 5 of reference 20).

Table 1. Data concerning the second daughter segment of partially replicated molecules of mtDNA from six species of Drosophila, observed in the electron microscope in foramamide protein monolayer preparations. The number of molecules observed is given by n. (Data from references 6 and 7.)

Species	Secondary structure of second daughter segment. Totally single-stranded		Totally double-stranded		Single-stranded only at one or both forks	
	n	Range of lengths (% genome)	n	Range of lengths (% genome)	n	Range of lengths (% genome)
D. melanogaster	49	2-99	6	90-96	4	10-98
D. simulans	15	5-72	2	38-91	8	17-98
D. mauritiana	17	5-71	4	50-97	3	16-96
D. takahashii	16	5-97	1	97	4	31-65
D. yakuba	22	5-83	3	24-72	1	98
D. virilis	25	2-59	1	96	1	84

collapse of single-stranded DNA into a "bush". This preparation was scored only for replicative intermediates in which the second daughter segment was either totally or partially double-stranded. Of 156 such molecules observed, 120 were totally double-stranded. Daughter segment lengths of molecules of this group ranged from 5 to 97% of the genome length. However, in 83% (n=99) of these, the daughter segment lengths fell within the narrow range of 87 to 97% of the genome length (Fig. 3F). In 34 of the remaining molecules of this sample, the second daughter segment was single-stranded at one fork only, and in two others it was single-stranded at both forks.

In formamide preparations of mtDNAs of all six species of Drosophila studied, and in the aqueous preparation of D. melanogaster mtDNA, unforked, simple circular molecules were observed in which a single region measuring from 2 to 77% of the genome length appeared to be single-stranded (Fig. 3D). In formamide preparations of Drosophila mtDNAs we have also observed circular molecules which appear to be entirely single-stranded (Fig. 3E).

Origin and Direction of Replication

We have used the relative locations of the A+T-rich regions and EcoRI sites in the mtDNA molecules of the six Drosophila species of interest to us, to locate in each molecule the site at which replication originates and the

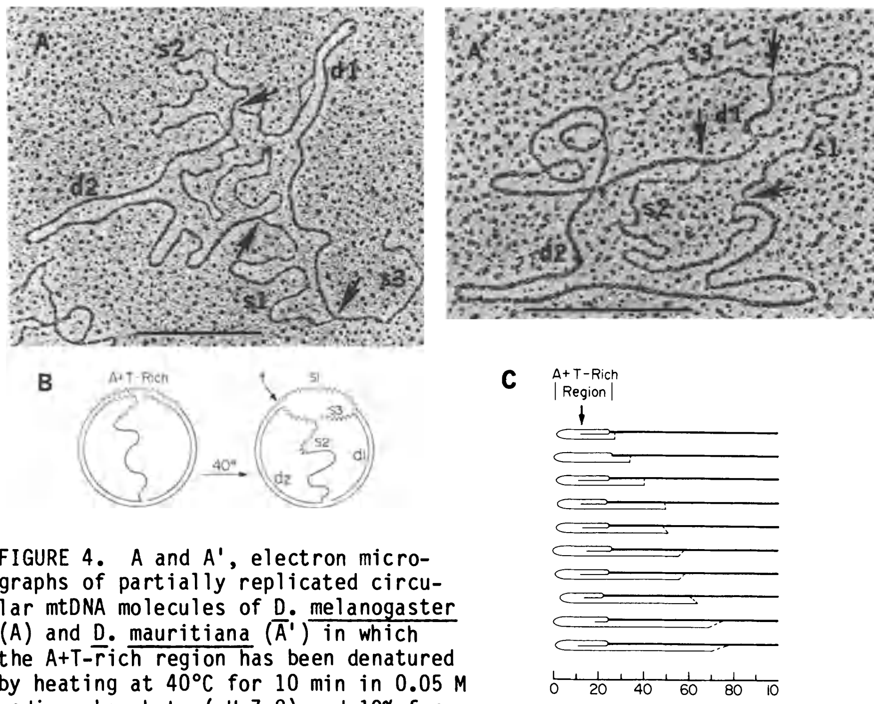


FIGURE 4. A and A', electron micrographs of partially replicated circular mtDNA molecules of D. melanogaster (A) and D. mauritiana (A') in which the A+T-rich region has been denatured by heating at 40°C for 10 min in 0.05 M sodium phosphate (pH 7.8) and 10% formaldehyde (bar = 0.05 μm). Three forks are indicated by arrows. B, a diagrammatic interpretation of the molecule shown in A. Corresponding segments of double-stranded DNA (d) and single-stranded DNA (s) in A and B are indicated by numbers. C, a comparison of 10 such molecules. For this purpose, each molecule is shown as a linear rod produced by cutting the circular molecule at the fork marked f in B and placing this fork to the left. Double-stranded and single-stranded segments are indicated by thick and thin lines, respectively (modified from Figs. 1 and 2 of references 6 and 7, respectively).

direction in which replication proceeds (6, 7). First, we examined in the electron microscope molecules of each species which had been subjected to partial heat denaturation. The large majority of circular molecules observed in each preparation were predominantly double-stranded with a major region of denaturation of a size characteristic of the respective species (Fig. 1). In addition, more complex circular molecules were observed. In each case, from a consideration of the lengths and relative arrangements of the various single-stranded and double-stranded segments, these molecules could be interpreted as replicative intermediates. Examples are shown in Fig. 4A and A'. Data for D. melanogaster mtDNA is shown in Fig. 4C. As illustrated in the diagram given in

Fig. 4B, these structures can be generated by partial denaturation of a circular molecule in which replication has originated at a single site in the A+T-rich region and proceeded unidirectionally for various distances around the molecule. These data clearly rule out any model of replication which requires either bidirectional synthesis from a unique origin, or a replicaton origin lying outside the A+T-rich region.

Next, mtDNA from each of the six species was digested with *Eco*RI and the products were examined in the electron microscope. For each species three classes of double-forked structures were observed. Examples of these from *D. melanogaster* mtDNA are shown in Fig. 5a-c, and a diagrammatic interpretation of each is shown in Fig 5a'-c'. The data obtained are summarized in Fig 5a"-c". All of the structures observed can be interpreted as the products of *Eco*RI digestion of replicative intermediates of circular molecules in which replication is unidirectional around the molecule and originates at a unique site within the A+T-rich region. In structures of the first class (Fig 5a") replication has proceeded to different positions still within the *Eco*RI A fragment. Structures of the second class (Fig 5b") are derived from molecules in which replication has proceeded to different positions within the *Eco*RI B fragment. Structures of the third class (Fig 5c") are derived from molecules in which replication has proceeded through the *Eco*RI D, B, and C fragments back to the A fragment. These data are therefore consistent with the results of the partial denaturation experiments, and further indicate that in each molecule the direction of replication is the same, towards the *Eco*RI site which delimits the *Eco*RI A and D fragments.

D. melanogaster mtDNA was also digested with *Eco*RI and the products subjected to partial denaturation. Examination of this mtDNA in the electron microscope revealed a number of complex forms which could again, in each case, be interpreted as having arisen from replicative intermediates in which replication had proceeded unidirectionally around the molecule from an origin in the center of the A+T-rich region, towards the same *Eco*RI site.

Data obtained from analysis of the structural forms observed in electron microscope preparations of *Eco*RI digested mtDNA from the other five *Drosophila* species (7) confirmed that in each of these mtDNAs, as in *D. melanogaster* mtDNA, replication originates at a unique site in the A+T-rich region and proceeds unidirectionally around the molecule towards the nearest common *Eco*RI site (Fig. 2). Our data (Fig. 6) further show that the origin of replication is located near the center of the three larger A+T-rich regions of the *D.*

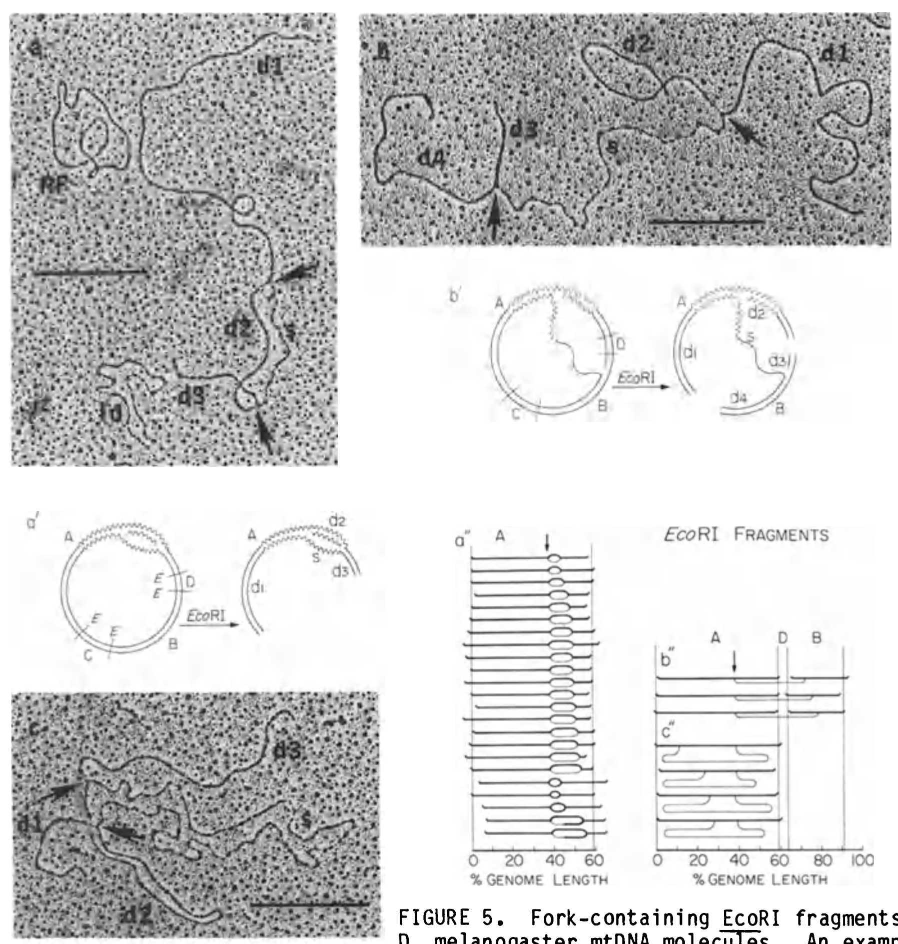


FIGURE 5. Fork-containing *EcoRI* fragments of *D. melanogaster* mtDNA molecules. An example of each of the three structural classes observed is shown in the electron micrographs a, b, and c (bar = 0.5 μm). A diagrammatic interpretation of each of the three fragments as originating from a partially replicated molecule is shown in a', b', and c'. The interpretations are based on the known positions of *EcoRI* cleavage sites (E in a'; A, B, C, and D refer to the four *EcoRI* fragments (19)) and the relative lengths of the various segments of each of the three structural classes of fragments shown in a", b", and c". In a" and b", each fragment is aligned (arrow) by the fork associated with the larger double-stranded segment, and this double-stranded segment is placed to the left. In b", the second fork of each fragment is placed to the right as shown. In c", each fragment is aligned (arrow) by the fork associated with the double-stranded segment (approximately 20% of the genome length) and this double-stranded segment is placed to the right (modified from Fig. 2 of reference 6).

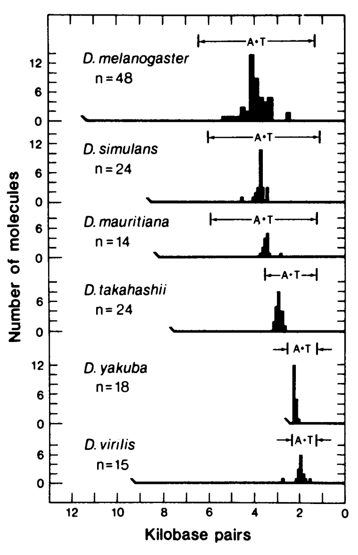


FIGURE 6. Histogram showing the position of the fork interpreted as the replication origin, within the A+T-rich region, relative to the nearest common EcoRI site. The EcoRI fragment is shown for each species by the horizontal line. The data represent the distance of the replication fork from the nearest EcoRI site. For D. melanogaster, however, the data also include the lengths of free-ended single-stranded segments of partially denatured molecules (Fig. 4A) plotted from the right boundary of the A+T-rich region. The data for D. melanogaster are plotted in 200 bp intervals compared to 100 bp intervals for each of the other species (data from Figs. 4 and 6 of references 6 and 7, respectively).

melanogaster, D. simulans, and D. mauritiana mtDNA molecules, while in mtDNA molecules of D. takahashii, D. yakuba, and D. virilis the origin is located close to the left end of the A+T-rich region (Fig. 2).

Replication Modes

Our conclusions concerning the possible modes of replication of Drosophila mtDNA are shown in Fig. 7. As in all other metazoan mtDNAs studied (27-30) replication of Drosophila mtDNA begins at a unique site and proceeds unidirectionally around the molecule (Fig. 7a). It appears that there is considerable variation in both the site of initiation of synthesis of the second strand, and the time of initiation of synthesis of the second strand relative to initiation of synthesis of the first stand. Since, in the majority of replicative intermediates, one daughter segment is completely single-stranded, and the length of this segment ranges from 2 to 99% of the genome length, it seems that most of the molecules are replicated by a highly asymmetrical mode in which considerable synthesis of one strand can take place before synthesis of the second strand is initiated (Fig. 7a-n-o+p). Evidence for such an asymmetrical mode of replication has been presented for mtDNA from a number of animal tissues (12-15). However, only in Drosophila mtDNA have molecules such as those shown in Fig. 3B and E been found, suggesting that in this case DNA synthesis of one strand can be virtually complete before

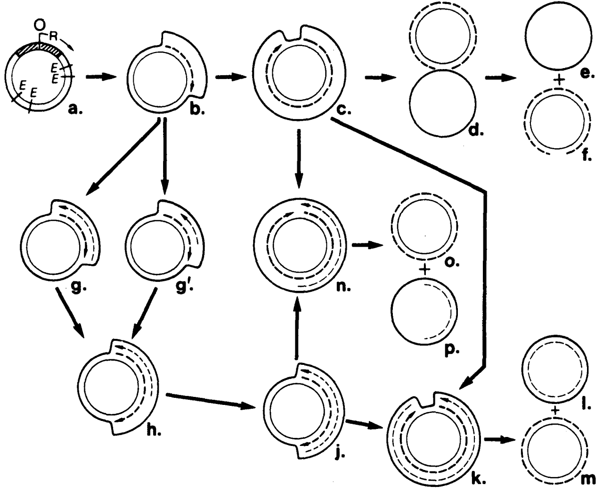


FIGURE 7. Different modes of replication of Drosophila mtDNA which account for all of the molecular forms observed in electron microscope preparations. Thin and thick continuous lines represent the complementary parental strands, and the thin and thick broken lines represent the corresponding complementary daughter strands. The arrows on daughter strands show the direction of synthesis (assuming antiparallel, 5'→3' synthesis of the two complementary strands). In a, the origin (O) and direction of replication (R) around the molecule are indicated (shown for D.melanogaster, hatched area = A+T-rich region; E = EcoRI sites). See text for discussion. (Modified from Fig. 8 of reference 8).

synthesis of the complementary strand is initiated (Fig. 7a-c), and that in some molecules daughter strand separation may occur before synthesis of the second strand commences (Fig. 7a-d-e+f).

The finding of molecules in which the replicated region measures 10% to 98% of the genome length and in which the second daughter segment is either partially or totally double-stranded (Fig. 3C and F), indicates that in some molecules synthesis of the second strand may be initiated at various positions on the molecule, even when synthesis of the first strand is as little as 10% complete (Fig. 7a-b-g(g')-h-j). The finding that, of molecules in which synthesis of the second strand had occurred, those in which replication was 87 to 98% complete predominated suggests that there is a temporal pause in replication when synthesis of both daughter segments is completed to an extent greater than 87% (Fig. 7K).

The presence of gapped (12), simple circular molecules (Fig. 3D), in which the single-stranded region measures from 2 to 77% of the genome length is con-

sistent with the view (31) that daughter molecules can separate when synthesis is incomplete over a wide range of the genome length (Fig. 7d-e+f; n-o+p).

It has been clearly demonstrated that in mouse, human and hamster tissue culture cells, synthesis of the second (L) strand is initiated only when synthesis of the first (H) strand is about 66% complete (12, 32, 33). Further, it has been shown that the site of initiation of the second (L) strand in mouse and human mtDNA is within a specific segment of 31 or 32 nucleotides which can fold into a hairpin loop (32, 34). Interestingly, however, among mtDNA molecules from tissues of a number of animals as well as Drosophila, including rat and mouse liver, and sea urchin oocytes, replicative forms occur which indicate that in some molecules initiation of synthesis on the second (L) strand occurs at many different locations within the 66% of the genome measured from the (H strand) origin of replication (13-15, 35). In fact, it appears that in mtDNA molecules of mouse liver and rat liver, this is the case in 36% and 53%, respectively, of the molecules (36).

Whether or not failure to find a distinct class of D loop-containing molecules among Drosophila mtDNAs by ourselves (6, 7, 19) and others (5, 37) indicates a real difference between the mechanism of replication operative in mtDNA of this organism and that of mtDNAs of other metazoa is not clear. It has not been ruled out that D loops are preferentially lost by branch migration due to the lower resistance to denaturation of the A+T-rich region at some time during the isolation procedure of the DNA.

Klukas and Dawid (5) have shown as illustrated in Fig. 2, that the small and large mitochondrial rRNA (mt-rRNA) genes are located respectively in tandem adjacent to the A+T-rich region. A similar location for the mt-rRNA genes of the other Drosophila species was indicated from the results of our heteroduplex studies (17). By comparing the restriction mapping data of Klukas and Dawid (5) with that of ourselves (17, 19) we determined that replication of Drosophila mtDNA molecules is always towards the mt-rRNA genes (8, Fig. 2). As mentioned above, it has been shown that the two mt-rRNA genes of rat, mouse, human, and Xenopus have a similar arrangement relative to the origin of replication to that found in Drosophila mtDNA. However, in each vertebrate mtDNA studied, the direction of replication is away from the mt-rRNA genes (3, 38-40). Up to this time, determinations of the direction of replication in other invertebrate mtDNAs have not been made.

Sequence divergence in A+T-rich regions

We have carried out denaturation-renaturation experiments to further elucidate the structure and homologies of the A+T-rich regions of the mtDNA molecules of different species. First, to gain information on the degrees of homology both outside and within A+T-rich regions, among the mtDNA molecules of the six different species, we constructed a series of heteroduplexes (17). This was done under conditions highly permissive for base pairing (35% formamide), using EcoRI fragments and whole circular molecules of the different species, in a variety of combinations. The results were quite straightforward. Complete pairing of molecules outside the A+T-rich regions was found in all heteroduplexes examined. However, in contrast, A+T-rich regions of the different species completely failed to pair in the following combinations: D. melanogaster:D. yakuba; D. melanogaster:D. takahashii; D. melanogaster:D. virilis; D. takahashii:D. yakuba; D. virilis:D. yakuba. In heterodulexes between D. melanogaster and D. simulans, and between D. melanogaster and D. mauritiana mtDNAs, up to 35% of the A+T-rich region appeared double-stranded. These data clearly indicate that between species, extensive divergence of sequences has occurred in A+T-rich regions. It is also worth noting that examination of these heteroduplexes (and various control and experimental homoduplexes) failed to reveal evidence for either heterogeneity of sequences among mtDNA molecules from flies of the individual species used in these experiments, or for tandemly repeated sequences within A+T-rich regions.

In other experiments we have obtained evidence for sequence differences between A+T-rich regions of flies derived from different strains of the same species (18). MtDNA molecules obtained from flies of two different female lines (mtDNA of Drosophila is maternally inherited (18)) originating from a single stock bottle of D. mauritiana were found to contain A+T-rich regions which differed in length by about 0.7 kb. In heteroduplexes constructed between these two kinds of mtDNA molecules, regions of strand separation were observed near the center of the A+T-rich regions. The sizes of these regions suggested that, in fact, extensive sequence differences may occur between at least 2 kb of this region. It is worth noting in this regard that the results of breeding experiments completely failed to indicate that any degree of infertility exists between flies of the two D. mauritiana lines used in these experiments.

We have also obtained evidence for sequence differences between A+T-rich regions of mtDNA molecules of three geographically separated strains of D.

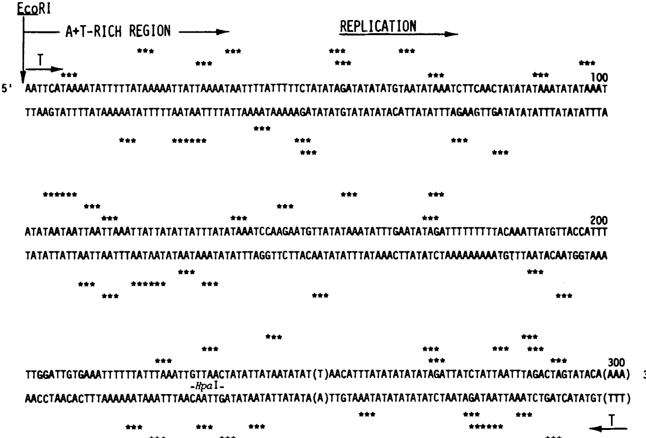


FIGURE 8. Nucleotide sequence of the left end of the A+T-rich region of the D. yakuba mtDNA molecules. The actual sequence determined was of the upper strand shown. The direction of transcription is shown by T→. Termination codons in each of the three possible reading frames of the two complementary strands are indicated by *** (Fig. 2 of reference 40).

melanogaster, and between mtDNA molecules of two geographically separated strains of D. simulans.

Sequence studies of the A+T-rich region of D. yakuba

We have begun sequence studies of Drosophila mtDNAs to determine the structure of the A+T-rich region and to gain information on the extent of sequence divergence in the A+T-rich regions of Drosophila species and on sequences associated with the origin of replication. We cloned into bacteriophage M13 the A+T-rich region-containing EcoRI fragment (the EcoRI C fragment, 2.5 kb) of D. yakuba (Fig. 2) and sequenced the ends of this fragment (41) using the Sanger procedure (42). The EcoRI site which defines one end of this fragment is located close to the left end of the A+T-rich region (1.1 kb) which in turn lies approximately 250 base pairs from the replication origin (Fig. 2).

The sequence of the 300 nucleotides at the left end of the A+T-rich region is shown in Fig.8. The first 150 nucleotides have a G+C content of 5%, but the G+C content of the remaining 160 nucleotides of this region is 14%. Inverted repeat sequences were found throughout the sequence. However, sequences homologous to sequences found close to, or associated with the origin of replication of mammalian mtDNA (2, 43, 44) were not apparent. Using the genetic code

determined for human mtDNA (45) it appears that at least 6 termination codons are present in each of the 6 possible reading frames of the complementary strands of this D. yakuba sequence, and their spacing would permit uninterrupted read-through of a maximum of 84 nucleotides.

The function of the A+T-rich region, other than the initiation of replication, is unknown at this time. Its low G+C content and notable lack of conservation of sequences make it unlikely that it contains sequences coding for proteins. This argument is supported both by two reported failures to detect RNA transcripts which hybridize to the A+T-rich region (9, 10) and our finding of frequent termination codons in the segment we have sequenced.

Potter et al. (46) found that there are five small segments in the A+T-rich region of D. melanogaster mtDNA molecules, all lying on that side of the replication origin which is the last to be replicated, which are protected in vivo from cross-linking by trioxsalen and UV light. They argued that the protected segments result from the binding of proteins which, owing to their position in the molecule, may be involved in membrane attachment or replication.

The region of mammalian mtDNA molecules which contains the replication origin and lies between $tRNA^{pro}$ and $tRNA^{phe}$ (1, 2), though of higher G+C content, shares a number of the features we have described for the A+T-rich region of the Drosophila mtDNA molecules. In mammals this region is about 1.1 kb in length, shows the greatest sequence variation, including length differences of up to several hundred base pairs between species (2, 4, 47, 48), appears not to be transcribed (except for a 230 nucleotide transcript RNA from a region near the $tRNA^{phe}$ gene of human mtDNA (49)), and lacks extensive open reading frames. It seems reasonable to argue, therefore, that this region of mammalian mtDNA and the A+T-rich region of Drosophila mtDNA have evolved from a common ancestral sequence. Inversion of a segment containing the origin of replication within this sequence may have resulted in the opposite overall direction of replication in vertebrate and Drosophila mtDNA molecules.

It seems likely from a consideration of the size of the mtDNA molecules from different Drosophila species and the results of our heteroduplex experiments, that the A+T-rich regions of most Drosophila mtDNA molecules are about 1.1 kb. The larger A+T-rich regions of the mtDNA molecules of melanogaster group species could have arisen by duplication of various segments of the smaller A+T-rich regions. However, if duplication occurred, then it must have been followed by further extensive sequence divergences as there is no evidence from our heterduplex data that any of the A+T-rich regions studied contain

duplicated sequences (17, 18, 20). A further point of interest in this regard is that the distribution of sizes of A+T-rich regions of mtDNA molecules of the different melanogaster group species was found not to follow a simple taxonomic pattern (16). Our data indicated either that increase in size of the A+T-rich region has occurred independently in separate evolutionary paths, or that decrease in size of the A+T-rich region has accompanied evolution along at least one path.

This work was supported by N.S.F. Grant No. BMS-74-21955 and N.I.H. Grant Nos. GM 18375 and K4-GM 70104.

REFERENCES

1. Anderson, S., Bankier, A.T., Barrell, B.G., DeBruijn, M.H.L. Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. Nature 290:457-405, 1981.
2. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. Cell 26:167-180, 1981.
3. Ramirez, J.L. and Dawid, I.B. J. Mol. Biol. 119:133-146, 1978.
4. Rastl, E. and Dawid, I.B. Cell 18:501-510, 1978.
5. Klukas, C.K. and Dawid, I.B. Cell 9:615-625, 1976.
6. Goddard, J.M. and Wolstenholme, D.R. Proc. Natl. Acad. Sci. U.S.A. 75:3886-3890, 1978.
7. Goddard, J.M. and Wolstenholme, D.R. Nucleic Acids Res. 8:741-758, 1980.
8. Wolstenholme, D.R., Goddard, J.M. and Fauron, C.M.-R.. In: Proc. 8th Ann. ICN-UCLA Symp. Molecular and Cellular Biology: Extrachromosomal DNA, Vol. 15 (Eds. D. Cummings, P. Borst, I.B. Dawid, S. Weismann and C.F. Fox), Academic Press, New York, 1979, pp. 409-425.
9. Bonner, J.J., Berninger, M. and Pardue, M.L. Cold Spring Harbor Symp. Quant. Biol. 42:803-814, 1978.
10. Battey, J., Rubenstein, J.L.R. and Clayton, D.A. In: Proc. 8th Ann. ICN-UCLA Symp. Molecular and Cellular Biology: Extrachromosomal DNA, Vol. 15 (Eds. D. Cummings, P. Borst, I.B. Dawid, S. Weismann and C.F. Fox), Academic Press, New York, 1979, pp. 427-442.
11. Kirschner, R.H., Wolstenholme, D.R., and Gross, N.J. Proc. Natl. Acad. Sci. U.S.A. 60:1466-1472, 1968.
12. Robberson, D.L., Kasamatsu, H. and Vinograd, J. Proc. Natl. Acad. Sci. U.S.A. 69:737-741, 1972.
13. Wolstenholme, D.R., Koike, K. and Cochran-Fouts, P. J. Cell Biol. 56:230-245, 1973.
14. Wolstenholme, D.R., Koike, K. and Cochran-Fouts, P. Cold Spring Harbor Symp. Quant. Biol. 38:267-280, 1974.
15. Matsumoto, L., Kasamatsu, H., Piko, L. and Vinograd, J. J. Cell Biol. 63:146-159, 1974.
16. Fauron, C.M.-R. and Wolstenholme, D.R. Proc. Natl. Acad. Sci. U.S.A. 73:3623-3627, 1976.
17. Fauron, C.M.-R. and Wolstenholme, D.R. Nucleic Acids Res. 8:2439-2452, 1980a.
18. Fauron, C.M.-R. and Wolstenholme, D.R. Nucleic Acids Res. 8:5391-5410, 1980b.

19. Wolstenholme, D.R. and Fauron, C.M.-R. J. Cell Biol. 71:434-448, 1976.
20. Wolstenholme, D.R., Fauron, C.M.-R. and Goddard, J.M. In: Developments in Genetics, Vol. 2. The Organization and Expression of the mitochondrial genome (Eds. A.M. Kroon and C. Saccone), Elsevier/North Holland Biomedical Press, Amsterdam, 1980, pp. 241-250.
21. Goddard, J.M., Masters, J.N., Jones, S.S., Ashworth, W.D. Jr. and Wolstenholme, D.R. Chromosoma (Berl.) 82:595-609, 1981.
22. Polan, M.L., Friedman, S., Gall, J.G. and Gehring, W. J. Cell Biol. 56:580-589, 1973.
23. Bultmann, H. and Laird, C.D. Biochim. Biophys. Acta 299:196-209, 1973.
24. Peacock, W.J., Brutlag, D., Goldring, E., Appels, R., Hinton, C. and Lindsley, D.C. Cold Spring Harbor Symp. Quant. Biol. 38:405-415, 1974.
25. Wolstenholme, D.R. and Dawid, I.B. J. Cell Biol. 39:222-228, 1968.
26. Kasamatsu, H., Robberson, D.L. and Vinograd, J. Proc. Natl. Acad. Sci. U.S.A. 68:2252-2257, 1971.
27. Robberson, D.L. and Clayton, D.A. Proc. Natl. Acad. Sci. U.S.A. 69:3810-3814, 1972.
28. Brown, W.M. and Vinograd, J. Proc. Natl. Acad. Sci. U.S.A. 71:4617-4621, 1974.
29. Robberson, D.L., Clayton, D.A. and Morrow, J.F. Proc. Natl. Acad. Sci. U.S.A. 71:4447-4451, 1974.
30. Buzzo, K., Fouts, D.L. and Wolstenholme, D.R. Proc. Natl. Acad. Sci. U.S.A. 75:909-913, 1978.
31. Berk, A.J. and Clayton, D.A. J. Mol. Biol. 86:801-824, 1974.
32. Tapper, D.P. and Clayton, D.A. J. Biol. Chem. 256:5109-5115, 1981.
33. Nass, M.M.K. J. Mol. Biol. 140:231-356, 1980.
34. Martens, P.A. and Clayton, D.A. J. Mol. Biol. 135:327-351, 1979.
35. Koike, K. and Wolstenholme, D.R. J. Cell Biol. 61:14-25, 1974.
36. Jones, S.S. and Wolstenholme, D.R. In preparation.
37. Rubenstein, J.L.R., Brutlag, D. and Clayton, D.A. Cell 12:471-482, 1977.
38. Kroon, A.M., Pepe, G., Bakker, H., Holtrup, M., Bollen, J.E., Van Bruggen, E.F.J., Cantatore, D., Terpstra, P. and Saccone, C. Biochim. Biophys. Acta 478:128-145, 1977.
39. Battey, J. and Clayton, D.A. Cell 14:143-156, 1978.
40. Ojala, D. and Attardi, G. J. Mol. Biol. 122:301-319, 1978.
41. Goddard, J.M., Fauron, C.M.-R. and Wolstenholme, D.R. In: Mitochondrial Genes (Eds. G. Attardi, P. Borst and P.P. Slonimski), Cold Spring Harbor Laboratory, 1982, in press.
42. Sanger, F., Nicklen, S. and Coulson, A.R. Proc. Natl. Acad. Sci. U.S.A. 74:5463-5467, 1977.
43. Crews, S., Ojala, D., Posakony, J., Nishiguchi, J. and Attardi, G. Nature 277:192-198, 1979.
44. Sekiya, T., Kobayashi, M., Seki, T. and Koike, K. Gene 11:53-62, 1980.
45. Barrell, B.G., Anderson, S., Bankier, A.T., DeBruijn, M.H.L., Chen, E., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. Proc. Natl. Acad. Sci. U.S.A. 77:3164-3166, 1980.
46. Potter, D.A., Fostel, J.M., Berninger, M., Pardue, M.L. and Cech, T.R. Proc. Natl. Acad. Sci. U.S.A. 77:4118-4122, 1980.
47. Upholt, W.B. and Dawid, I.B. Cell 11:571-583, 1977
48. Brown, W.M. Annals New York Acad. Sci. 1982, in press.
49. Ojala, D., Crews, S., Montoya, J., Gelfand, R. and Attardi, G. J. Mol. Biol. 150:303-314, 1981.

8

REPLICATION OF MAMMALIAN MITOCHONDRIAL DNA

DAVID A. CLAYTON

SUMMARY

The mammalian mitochondrial DNA system has proven amenable to biochemical studies of the mode of replication of this closed circular DNA. This genome contains two separate and distinct origins of replication. One origin is contained within the displacement-loop region and initiates synthesis of the heavy strand of the helix. The second origin is located 67% of the genomic length away from the heavy-strand origin and initiates light-strand synthesis in the opposite direction. Synthesis from each origin is unidirectional and initiation at the origin of light-strand synthesis is delayed until heavy-strand synthesis is greater than 67% completed. The precise start points of synthesis at each origin are known and an ordered family of replicative intermediates have been characterized. In addition to replicative functions, the displacement-loop region has been implicated as a major site of transcriptional control in mammalian mitochondrial DNA. Nucleotide sequence comparisons indicate that this region is highly species-specific. In contrast, the origin of light-strand synthesis appears to be a well-conserved functional element.

INTRODUCTION

Mammalian cells in culture typically contain several thousand copies of mitochondrial DNA (mtDNA) per cell (1). Replication of the genome is under relaxed control and mtDNA synthesis proceeds independently of nuclear DNA synthesis (2,3). Although not all molecules replicate once per doubling of tissue-culture cell populations, the net amount of mtDNA synthesis is consistent with the maintenance of a constant number of genomes (4).

The existence of oligomeric forms of closed circular DNA was first identified in mammalian mtDNA populations. Oligomers are of two types: catenanes, in which two or more monomeric units are joined as links in a chain; and unicircular dimers, in which two monomeric units exist in a head-to-tail configuration (5). The available data suggest that catenanes are formed from pre-existing monomers (6,7) while the mechanism of formation of unicircular dimers remains an enigma. However,

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

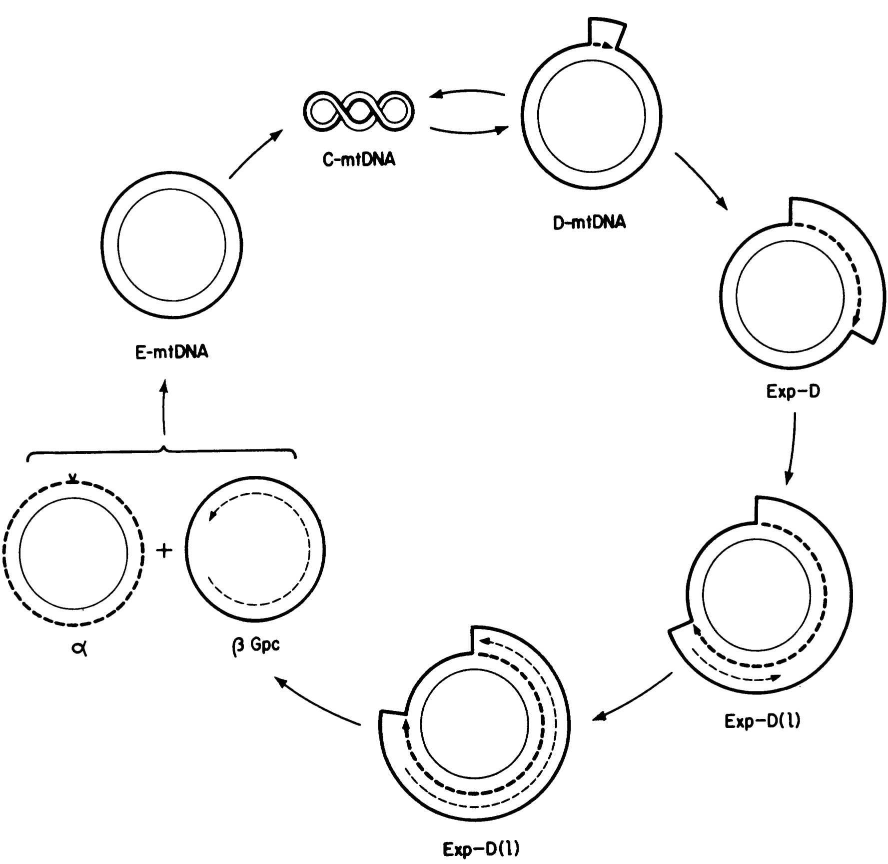


FIGURE 1. Mode of replication of monomeric mouse mtDNA. Parental H and L-strands are denoted by thick and thin lines, respectively. Daughter strands are shown as dashed lines and the order of replication is clockwise starting with D-mtDNA (D-loop mtDNA). Expanded D-loop replicative intermediates are termed Exp-D prior to initiation of L-strand synthesis and Exp-D(ℓ) after initiation of L-strand synthesis. Segregation produces two daughter molecules, α and β Gpc, and the caret indicates an interruption of at least one phosphodiester bond in the H-strand of the α daughter molecule. The gapped circle (Gpc) must complete synthesis prior to closure (E-mtDNA). The double arrows denote the metabolic instability of D-loop strands and the consequent interconversion between C-mtDNA and D-mtDNA. This figure is from (13), with permission.

the availability of a mouse L-cell line that contains exclusively unicircular dimer mtDNA (8,9) has permitted a characterization of the mode of replication of this oligomeric form (10).

MATERIALS AND METHODS

DNA sequencing was performed as described by Maxam and Gilbert (11). Electron microsopy was performed as described by Robberson and Clayton (8). 5'-end labeling and mapping of daughter light-strands (L-strands) is described by Tapper and Clayton (12).

RESULTS AND DISCUSSION

Mode of replication of monomeric mtDNA

The major form of mtDNA in mouse cells is a covalently closed circle with a displacement-loop (D-loop) at the origin of H-strand replication (Figure 1). This D-loop structure is formed by the synthesis of a short H-strand which remains associated with the parental closed circle. The entire molecule has been termed D-mtDNA. Replicative intermediates are defined as those molecules in which H-strand synthesis has extended past the D-loop region (Exp-D). Synthesis of the daughter H-strand continues in a unidirectional manner until completion of synthesis. After H-strand synthesis is 67% complete, the origin of L-strand synthesis is exposed as a single-stranded template and initiation of L-strand synthesis occurs (Exp-D(ℓ), Figure 1). The delay in initiation of L-strand synthesis results in the segregation of two types of daughter molecules (α and β), with one consisting of the parental H-strand and a partially synthesized daughter L-strand (β) (Figure 1). Both α and β daughter molecules are converted to closed circles (E-mtDNA) with few, if any, superhelical turns. Approximately 100 negative superhelical turns are then introduced into these E-mtDNA daughter molecules, resulting in the formation of C-mtDNA. This superhelical DNA (C-mtDNA) serves as the template for the formation of D-loop mtDNA which completes the replication cycle. The synthesis of full-length daughter strands requires approximately one hour and the entire cycle is completed in approximately two hours.

Replication of unicircular dimer mtDNA

Unicircular dimers contain two monomeric genomes and therefore four replication origins (Figure 2). Most unicircular dimers possesses one or two D-loops and molecules in which one of the D-loops has elongated have been used to infer the unique map position of the stationary D-loop (Figures 3,4). Productive H-strand synthesis begins by expansion of a D-loop strand beyond the D-loop region (Figure

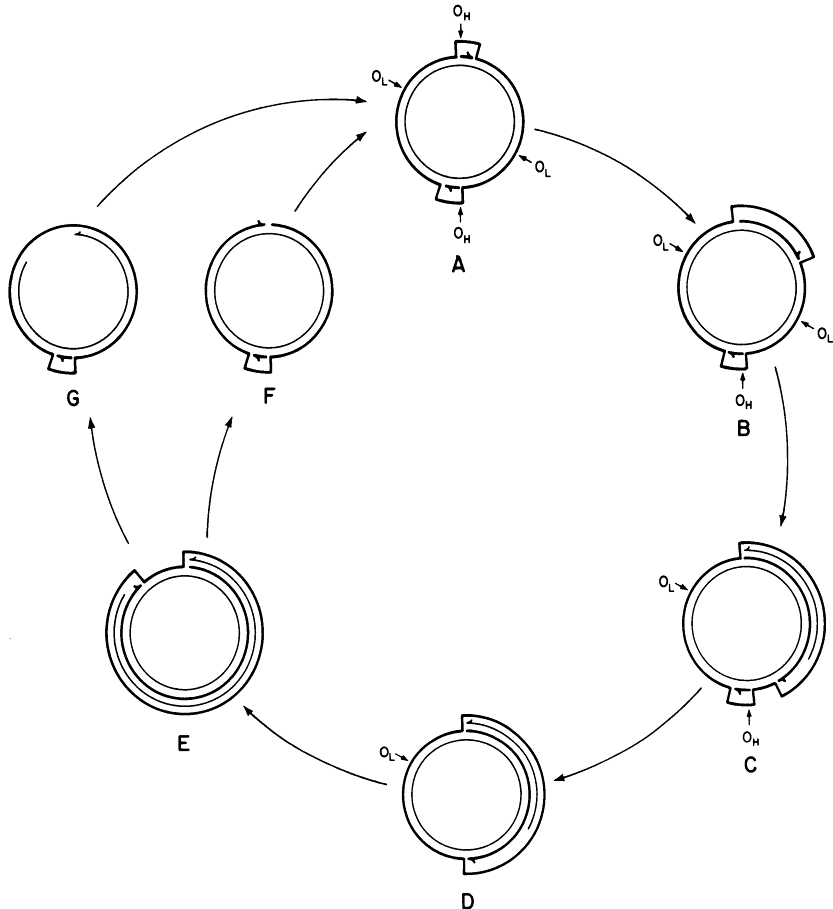


FIGURE 2. Mode of replication of unicircular dimeric mouse mtDNA. The order of replication is A through G. O_H denotes the two D-loop origins of H-strand synthesis and O_L denotes the two origins of L-strand synthesis. The major difference in the replication of unicircular dimeric mtDNA as compared to monomeric mtDNA is the synthesis of a new D-loop (F,G) in daughter molecules prior to closure. For a complete description of unicircular dimer replication, see reference (10) from which this figure is taken, with permission.

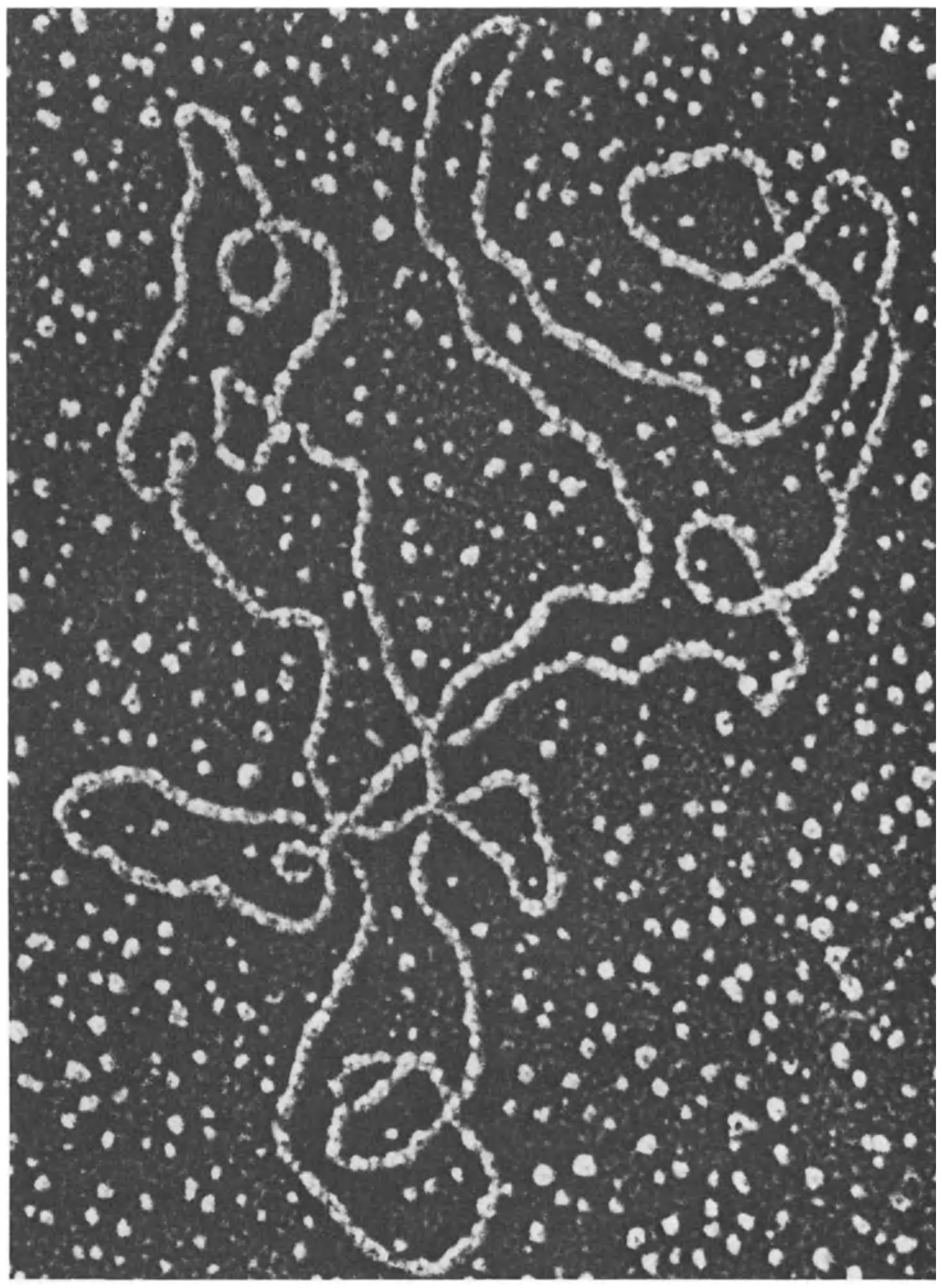


FIGURE 3. Electron micrograph of an early replicative intermediate of unicircular dimeric mtDNA. The molecule is 32.6 kb in size. One of the two D-loops (upper right) has begun to expand.

2A,B). It is not known whether pre-existing D-loop strands serve as replication primers. Elongation proceeds unidirectionally (Figure 4) past the first origin of L-strand synthesis and H-strand synthesis continues until reaching the second D-loop region (Figure 2D). Daughter H-strands 16 kb in size have been isolated from replicative intermediates which suggests that synthesis is halted for a substantial period of time (10). Whether elongation eventually progresses through this sequence or whether a second initiation event is required is unknown. H-strand synthesis in the remainder of the molecule is unimpeded until segregation of two daughter molecules (Figure 2E,F,G). In contrast, daughter L-strand synthesis begins at the first exposed origin of L-strand synthesis and proceeds unidirectionally until the first D-loop region is reached (Figure 2C). The second initiation of L-strand synthesis occurs upon displacement of the second origin of L-strand synthesis (Figure 2E). A consequence of the utilization of two origins of L-strand synthesis is that three sizes of daughter L-strands can be identified (10). The first initiation produces a 10.5 kb daughter L-strand and the second a 16 kb daughter L-strand. Subsequent joining of these two nascent L-strands results in a 27 kb strand which is maintained until after segregation (Figure 2G). The final event in L-strand synthesis is the closure of the remaining 5.4 kb gap to produce a mature closed circular species.

The topology of newly-replicated unicircular dimeric mtDNA is different from that of newly-replicated monomeric mtDNA, which lacks both superhelical turns and D-loops. Most newly-replicated dimeric mtDNA molecules contain D-loops, which may result from the presence of D-loops in late replicative molecules (Figure 2F,G). Therefore, following completion of daughter-strand synthesis and ligation, the newly-replicated molecules would be expected to contain one D-loop each. Because dimeric mtDNA molecules exist for nearly three hours as replicative intermediates, it is feasible that a D-loop H-strand may be synthesized after replication has been completed at one of the H-strand origin regions. In principle, only one L-strand and one H-strand origin could be required for unicircular dimer mtDNA synthesis. Although both origins of L-strand synthesis are functional in a single dimeric molecule, only one origin of H-strand synthesis may be utilized for replication (10). It is therefore possible that unicircular dimers contain only one D-loop region capable of initiating H-strand synthesis, although the overall frequency of replicative intermediates is such that one would not expect a significant number of doubly-initiated molecules purely on a statistical basis.

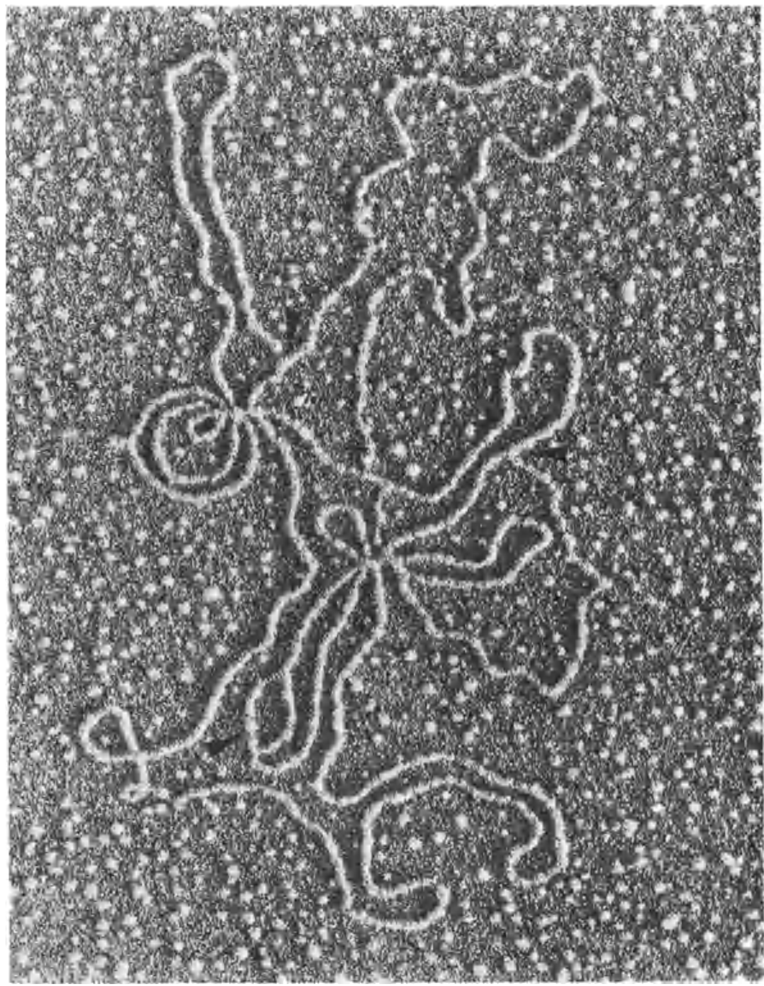


FIGURE 4. Electron micrograph of an expanded replicative intermediate of unicircular dimeric mtDNA. The double arrows denote the expanded D-loop; the single arrow denotes the stationary D-loop.

Displacement-loop region of mtDNA

In the initial studies of D-loop mtDNA it was assumed that the hydrogen-bonded H-strand within the D-loop was of unique length. That this is not the case was shown by Brown et al. (14) and Gillum and Clayton (15) who demonstrated multiple forms of D-loop strands. In the case of mouse mtDNA there are at least four species ranging from ~520-690 nucleotides in length (14-17) and the basis for the overall length difference is largely due to differences in length at the 3'-ends. In contrast, human D-loop strands are composed of three major species ~570-655 nucleotides in length, with the major contribution to length difference being at the 5'-ends of the strands (12,14,15,17).

One of the most novel features of D-loop formation is the fact that discrete stop points in synthesis are recognized. Human D-loop strands map at a single 3'-end position, while there are five discrete 3'-end map positions for mouse D-loop strands (18). There are five obvious possibilities for the cessation of synthesis of these strands. The first is that the thermodynamics of available superhelix free energy do not permit expansion beyhond a defined point (19). There are four facts which render this hypothesis unlikely: [1] If termination of D-loop H-strand synthesis were thermodynamically determined, the resultant H-strands should be of unique length or consist of a population of molecules distributed about one modal length. These fragments exist as a population of multiple, discrete lengths. [2] The termination of D-loop H-strand synthesis should be directly related to the original superhelix density of the closed circular template. It is not, because the superhelix densities of mouse and human simple closed circular mtDNAs are similar yet the lengths of D-loop H-strands differ. [3] It would be expected that elongation of the 3'-end of a D-loop strand would continue until most, if not all, superhelical turns were removed. In fact, *in vivo* synthesis proceeds until approximately two-thirds of the superhelical turns of simple closed circular mtDNA are removed. This is in contrast to the observation that D-loop strands can be extended *in vitro* until most, if not all, of the original superhelical turns of the parental strands are relieved (20). [4] In the case of unicircular dimer mtDNA, a significant fraction of the population contains only one D-loop. If the size of the single D-loop region were determined by the available free energy of superhelix formation, this D-loop should be twice the size of monomeric mtDNA D-loops because the superhelix density of unicircular dimer mtDNA is the same as that of monomer mtDNA (10). The sizes of unicircular dimeric D-loop strands are the same as those from monomer mtDNA (21). It should be cautioned that all assessments of superhelix densities of

mtDNA forms are from in vitro solution conditions and therefore do not necessarily reflect the spectrum of topological complexity which may exist in the organelle.

A second possibility, that each of these early terminated strands does not contain a proper 3'-OH end for further addition of triphosphates, is unlikely for two reasons. First, the strands can be 3'-end labeled by orthodox terminal labeling methods, such as by terminal transferase. Second, the 3'-ends of each of the human and mouse D-loop strands can be elongated in vitro by human KB cell DNA polymerase β (20,22).

In the case of mouse mtDNA, a third possibility is that the five discrete 3'-ends of mouse D-loop strands are generated by a rapid exonucleolytic process operating at the 3'-end, resulting in precise degradation to five distinct 3'-ends. This possibility has not been completely ruled out, but in vivo pulse-labeling studies have failed to reveal any obvious precursor-product relationship among mouse L-cell D-loop strands (21).

A fourth possibility is that size heterogeneity of these strands could result from actual size heterogeneity within the D-loop region in the parental duplex DNA population with each of the individual D-loop strands emanating from a unique subpopulation of parental closed circles with a single stop position. A restriction enzyme approach has shown that this region of the genome is of a single size (23).

Finally, a fifth possibility is that cessation of strand elongation is directed by a DNA sequence-specific event. A comparison of the map positions of D-loop strands and nucleotide sequences in the D-loop region of mouse and human mtDNA suggests that this intriguing concept may be correct (18). Precise localization at the nucleotide level indicates that the 3'-ends of D-loop strands of human and mouse mtDNA are discrete and map within three to five nucleotides on the complementary template strand. In the case of human mtDNA, there is a single trinucleotide stop point 51 to 53 nucleotides downstream from a 15-nucleotide template sequence (3'-TAACCCAAAAATACA-5') which is very similar to a sequence repeated four times in the mouse mtDNA D-loop region (3'-TAAPyPyAAATTACA-5'). The stop points of the five major mouse D-loop strands are from 24 to 63 nucleotides downstream from the four repeated template sequences. A productive replication event could involve initiation and synthesis past these putative stop signals (Figure 5). One would then argue that all D-loop strands represent abortive events with respect to replication and do not serve as primers. Because the four major mouse D-loop strands have a short half-life of approximately one hour (21), greater than

95% of these D-loop strands are lost by turnover and cannot serve as primers for mtDNA replication.

The nucleotide sequences of the D-loop regions of human and mouse mtDNA have been compared (17). There are limited regions of potential secondary structure in these sequences, and although the 5'-ends of D-loop strands map within or near such structures, none have as yet been directly implicated in the initiation of DNA synthesis. In addition, there are regions of extended homology between the three D-loop regions which are located in the midportion of the D-loop (Figure 5).

These data permit the delineation of a prototypic D-loop region of mammalian mtDNA (Figure 5). The short conserved sequences have been implicated as control regions for the initiation and cessation of D-loop strand synthesis (16-18) and the larger conserved sequence may be involved in anchoring the genome in situ. Given the fact that the D-loop region is the only significant amount of DNA sequence that does not code for an RNA or protein species (24), it is likely to contain the major promoters for transcription of each strand of the helix (24,25).

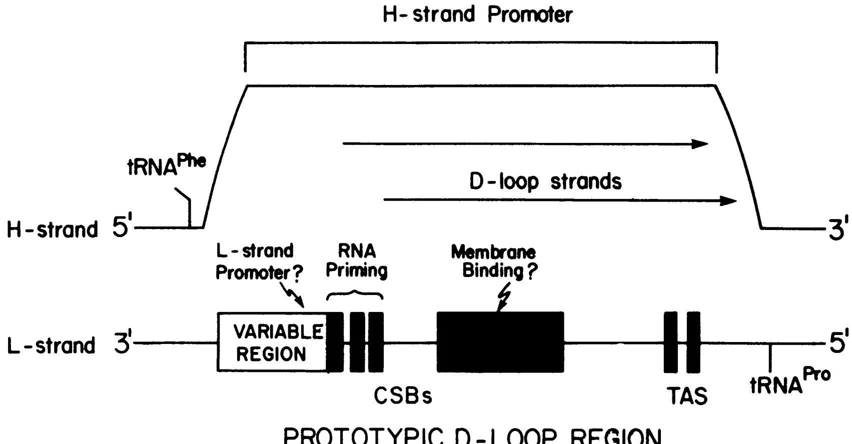


FIGURE 5. Consensus D-loop region of mammalian mtDNA. The D-loop region is defined as that portion of the genome bounded by tRNA genes Phe and Pro (24). Major promoters for transcription of each strand are thought to be located in the areas shown. The number and map position of D-loop strands are species-specific. RNA-priming of mouse D-loop strand synthesis occurs within a region of three conserved sequence blocks (CSBs) (16,17) and a large CSB in the midportion of this region could be involved in membrane binding. The termination associated sequences (TAS) map near the 3'-ends of D-loop strands (17,18).

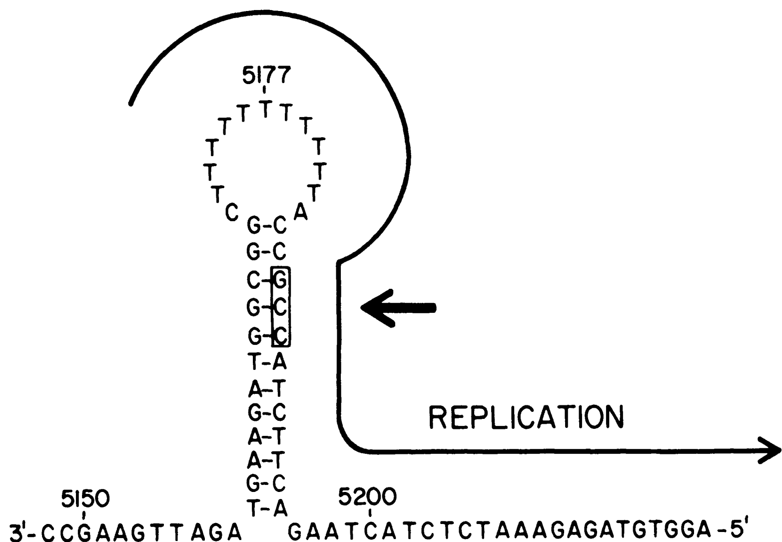


FIGURE 6. Potential secondary structure of the template strand at the origin of L-strand synthesis. The numbering is that of Bibb et al. (24). Priming of replication can begin in the T-rich loop (27) and proceeds in the direction shown. The major site of ribosubstitution in this region is boxed (26) and the major 5'-end, rG, of nascent daughter L-strands maps at the position marked by the boldface arrow. This point represents the major transition from ribo- to deoxyribonucleotide synthesis (27).

Origin of light-strand synthesis

The map positions of the human (12) and mouse (13) origins of L-strand synthesis have been precisely located. These origins have the potential to form a stem and loop structure that may be biologically significant because the template strand does not function until it is displaced as a single-stranded entity (Figure 6). For the most part, the 5'-ends of daughter L-strands are located in the stem portions of potential dyads. In particular, the larger dyad has a G+C-rich stem and T-rich loop in both systems and is predicted to be one of the most stable secondary structures assigned to a replication origin (24-26). The retention of ribonucleotides in the L-strand of closed circular mouse mtDNA at a trinucleotide site in this stem (26) may reflect incomplete excision of primer ribonucleotides (Figure 6). Tapper

and Clayton (27) have determined that the majority of 5'-ribonucleotide ends of daughter L-strands from replicative intermediates map within this same trinucleotide position in the mtDNA sequence. The flanking sequences contain tRNA genes which are more highly conserved in primary sequence (24). The origin of L-strand synthesis exhibits sequence homologies with several known prokaryotic and eukaryotic origin regions (13), and one of these is a 30-nucleotide region which is 60% homologous to the viral gene A protein cleavage site in ØX174 and G4 viral DNA. It is not recognized by gene A protein as a productive cleavage site in closed circular mtDNA (28).

Based upon the identification of a family of nascent L-strand 5'-ends (27), it is clear that RNA priming of replication can occur within the T-rich loop on the mouse template H-strand as shown in Figure 6. Because the switch to deoxyribonucleotide synthesis is most pronounced at the end of the G+C-rich portion of the stem, it is likely that the T-rich loop is recognized by the priming activity and the G+C-rich sequence effects a transition to DNA polymerase function. This view is strengthened by the fact that these primary and secondary sequence characteristics are highly conserved in human and mouse mtDNA (24).

Conclusions

The novel mode of mammalian mtDNA replication has facilitated biochemical studies of the initiation and elongation of daughter strands as well as the identification and ordering of replicative intermediates. The overall rate and sequential details of the process *in vivo* are known and documented. In contrast, our understanding of the enzymological processes involved in replicating this genome is relatively primitive. Current data suggest that DNA polymerase γ is functional in mtDNA replication and it has recently been purified to near homogeneity (29). Future studies of mammalian mtDNA replication should be directed towards identifying the factors necessary for replication by utilizing currently available, defined mtDNA primer-template systems (30).

ACKNOWLEDGEMENTS

This study was supported by a grant from the National Cancer Institute (CA-12312-12).

REFERENCES

1. Bogenhagen, D. and Clayton, D.A. J. Biol. Chem. 249:7991-7995, 1974.
2. Bogenhagen, D. and Clayton, D.A. J. Biol. Chem. 251:2938-2944, 1976.
3. Bogenhagen, D. and Clayton, D.A. Cell 11:719-727, 1977.
4. Berk, A.J. and Clayton, D.A. J. Mol. Biol. 86:801-824, 1974.
5. Clayton, D.A. and Smith, C.A. Int. Rev. Exp. Path. 14:1-67, 1975.

6. Flory, P.J., Jr. and Vinograd, J. J. Mol. Biol. 74:81-94, 1973.
7. Berk, A.J. and Clayton, D.A. J. Mol. Biol. 100:85-102, 1976.
8. Robberson, D.L. and Clayton, D.A. Proc. Nat. Acad. Sci. USA 69:3810-3814, 1972.
9. Kasamatsu, H. and Vinograd, J. Nature New Biol. 241:103-105, 1973.
10. Bogenhagen, D., Lowell, C. and Clayton, D.A. J. Mol. Biol. 148:77-93, 1981.
11. Maxam, A.M. and Gilbert, W. Meth. Enzymol. 65:499-560, 1980.
12. Tapper, D.P. and Clayton, D.A. J. Biol. Chem. 256:5109-5115, 1981.
13. Martens, P.A. and Clayton, D.A. J. Mol. Biol. 135:327-351, 1979.
14. Brown, W.M., Shine, J. and Goodman, H.M. Proc. Nat. Acad. Sci. USA 75:735-739, 1978.
15. Gillum, A.M. and Clayton, D.A. Proc. Nat. Acad. Sci. USA 75:677-681, 1978.
16. Gillum, A.M. and Clayton, D.A. J. Mol. Biol. 135:353-368, 1979.
17. Walberg, M.W. and Clayton, D.A. Nucl. Acids Res. 9:5411-5421, 1981.
18. Doda, J.N., Wright, C.T. and Clayton, D.A. Proc. Nat. Acad. Sci. USA 78:6116-6120, 1981.
19. Kasamatsu, H., Robberson, D.L. and Vinograd, J. Proc. Nat. Acad. Sci. USA 68:2252-2257, 1971.
20. Eichler, D.C., Wang, T. S.-F., Clayton, D.A. and Korn, D. J. Biol. Chem. 252:7888-7893, 1977.
21. Bogenhagen, D. and Clayton, D.A. J. Mol. Biol. 119:49-68, 1978.
22. Korn, D., Fisher, P.A., Battey, J. and Wang, T.S.-F. Cold Spring Harbor Symp. Quant. Biol. 43:613-624, 1978.
23. Bogenhagen, D., Gillum, A.M., Martens, P.A. and Clayton, D.A. Cold Spring Harbor Symp. Quant. Biol. 43:253-262, 1978.
24. Bibb, M.J., Van Etten, R.A., Wright, C.T., Walberg, M.W. and Clayton, D.A. Cell 26:167-180, 1981.
25. Clayton, D.A. Cell 28:693-705, 1982.
26. Brennicke, A. and Clayton, D.A. J. Biol. Chem. 256:10613-10617, 1981.
27. Tapper, D.P. and Clayton, D.A. J. Mol. Biol., submitted.
28. Doda, J.N. and Clayton, D.A. Plasmid 6:354-357, 1981.
29. Yamaguchi, M., Matsukage, A. and Takahashi, T. J. Biol. Chem. 255:7002-7009, 1980.
30. Kaguni, L.S. and Clayton, D.A. Proc. Nat. Acad. Sci. USA 79:983-987, 1982.

9

REPLICATION OF BACTERIOPHAGE T7 DNA

CHARLES C. RICHARDSON

SUMMARY

Bacteriophage T7 provides a model for understanding the mechanisms involved in the replication of a linear, duplex DNA molecule. Studies with purified proteins encoded by phage T7 and its host have elucidated the mechanisms by which both leading and lagging strands are synthesized, thus accounting for the movement of the replication fork along the duplex DNA molecule. Identification of the primary origin for the initiation of DNA replication on the T7 chromosome has led to an understanding of at least some of the structural and enzymatic requirements for site-specific initiation. These studies, in conjunction with those obtained with other phage and bacteria, provide considerable insight into the multiple reactions that constitute the overall process of DNA replication.

INTRODUCTION

In recent years we have been studying the replication of bacteriophage T7 DNA in order to understand the molecular and enzymatic mechanisms responsible for the replication of a linear, duplex DNA molecule (1-3). *In vitro* studies on the replication of the single-stranded, circular DNAs of phages G4, M13, and ϕX174 have defined the mechanisms by which these DNAs are converted to circular duplexes or replicative forms (4). Studies on the replication of duplex DNA molecules have made use of both circular and linear DNA molecules. There are two major classes of duplex, circular molecules: the replicative forms of the

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

single-stranded phage DNAs and a variety of plasmid DNA molecules. Among the duplex, linear DNA molecules whose replication has been studied are those isolated from phages T7, lambda, and T4 (4). Studies with duplex molecules have provided information that could not be obtained with the single-stranded DNA phages alone, such as the mechanism by which replication is initiated at a site-specific origin and the events that occur at the replication fork as both the leading and lagging strands are replicated.

Certain steps in the replication of the larger DNA molecules such as those of T7, lambda, and T4, resemble events known to occur during the replication of bacterial and eukaryotic chromosomes. Phage T7, like the more complex T4 phage, relies heavily upon its own encoded proteins for DNA replication, providing one approach to studying specific steps in replication. Of the two phages, the replication of T7 DNA appears simpler to dissect in that it has evolved an efficient and economical mechanism for the replication of its DNA. Studies with T7 have revealed the minimal requirements for the rapid and accurate replication of a duplex DNA molecule. In analogy to the replication of chromosomes of more complex organisms, T7 has multiple origins at which initiation can occur, and replication is both bidirectional and discontinuous. In addition, studies with T7 provide the opportunity to determine at least one mechanism by which a linear molecule is fully replicated without proceeding through a circular replicative intermediate.

IN VIVO DNA REPLICATION

Considerable information is available concerning the genetics and life cycle of phage T7. Kruger and Schroeder (5) have written an excellent and thorough review of phage T7. Only those topics relevant to T7 DNA replication are summarized here.

The chromosome of phage T7

The T7 virion is a polyhedral nucleocapsid 50 nm in diameter with a 20 nm long noncontractile tail. Each phage particle contains a double-stranded, linear DNA molecule, each strand consisting of an uninterrupted sequence of 40,000 nucleotides (Fig 1a). DNA molecules isolated from phage T7 particles all

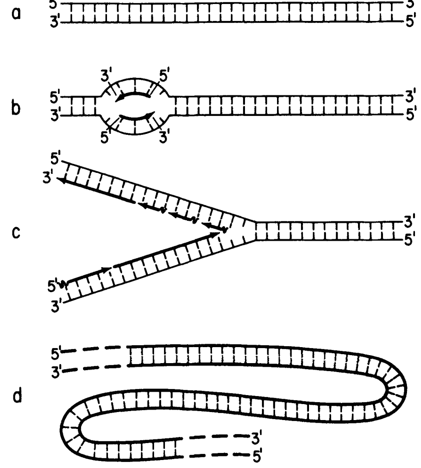


FIGURE 1. *In vivo* intermediates in T7 DNA replication. (a) Bacteriophage T7 DNA; (b) bidirectional replication initiated at the origin results in a replication bubble (eye-form); (c) Y-shaped replicating molecule showing discontinuous synthesis and RNA primers; (d) linear concatemers. From Richardson *et al.* (1).

have the same unique nucleotide sequence and a terminal redundancy of 160 nucleotides (6,7). The nucleotide sequence of the entire molecule has been determined (7, J. J. Dunn, personal communication).

Intermediates in T7 DNA replication

T7 DNA replication can be divided into three major stages: (1) initiation, (2) movement of the replication fork, and (3) formation and processing of concatemers (Fig. 1). T7 DNA replicates as a linear monomer, at least in the early stages of infection (8,9). DNA synthesis is initiated at a specific site on the viral chromosome giving rise to eye forms (Fig. 1b) in which the replication bubbles, as visualized by electron microscopy, are centered approximately 17% of the distance from the genetic left end of the DNA molecule (position 17). Bidirectional replication from this origin generates Y forms (Fig. 1c). A major portion of the newly synthesized DNA can be isolated as short fragments 1000 to 6000 nucleotides in length (10,11). The 5'-termini of these fragments bear oligoribonucleotides (12,13), suggesting a role of RNA priming in

the synthesis of these Okazaki fragments. In subsequent rounds of replication, at least 50% of the newly synthesized DNA is found in the form of linear concatemers many phage genomes in length (Fig. 1d) (14-19). Finally, maturation of replicated T7 DNA requires the cleavage and processing of these concatemers to yield the unit-size T7 DNA molecules found in the mature virion.

In addition to the structures described above, replicating T7 DNA can be isolated in more complex forms. Both parental and newly synthesized T7 DNA are found in stable association with host cell membranes (20,21). Furthermore, the newly synthesized DNA can be isolated in a form, free of membrane material, consisting of several hundred phage equivalents of DNA (19). The properties of this form of intracellular DNA suggest that it is similar to the "folded chromosome" of *E. coli* (22,23).

Genetic analysis of T7 DNA replication

Upon injection of T7 DNA into the host, *E. coli* RNA polymerase initiates transcription from three promoters located at the left end of the T7 genome, and thus accounts for the early transcription through gene 1.3 (Class I genes) (24). The product of gene 1 is an RNA polymerase (25) and it accounts for the transcription of the remainder of the genes, Class II and Class III genes. The majority of the phage replication proteins are products of the Class II genes and, as such, their synthesis commences approximately 4 min (37°C) or 6 min (30°C) after infection.

Phage mutants. Genetic analysis of phage T7 (24,26) has shown that the products of genes 1, 2, 3, 4, 5, and 6 are essential for viral DNA synthesis *in vivo* (Fig 2). The product of gene 1 is the T7 RNA polymerase; it is required for expression of most of the T7 genome, including genes 2 through 6. T7 RNA polymerase also plays a direct role in DNA replication, since T7 DNA synthesis ceases after inactivation of a temperature-sensitive T7 RNA polymerase (27). As discussed later, T7 RNA polymerase is essential for the initiation of DNA replication at the primary origin.

During infection by gene 2 mutants the block in T7 phage development (28) is at the stage of concatemer formation (29,30). Studies with purified gene 2 protein have shown

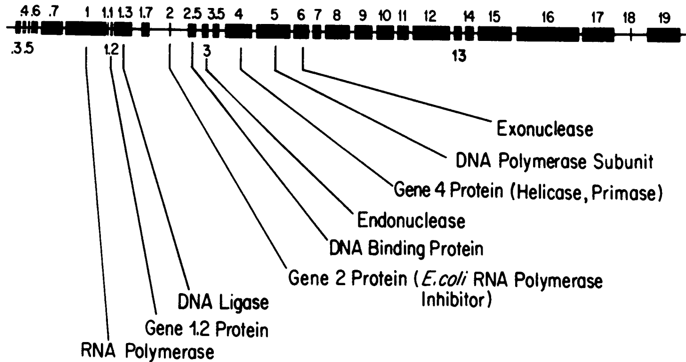


FIGURE 2. Genetic map of phage T7. Gene sizes reflect molecular weights of protein products (adapted from 7,24).

it to be an inhibitor of E. coli RNA polymerase (31-33), and E. coli mutants that have an altered RNA polymerase to which the T7 gene 2 protein cannot bind show a similar block in phage maturation when infected with wild type T7 (30).

Genes 3 and 6 code for an endonuclease and an exonuclease, respectively, and at least one of their roles is to provide nucleotide precursors for phage DNA synthesis by catalyzing the breakdown of host DNA (34-36). T7 relies almost exclusively on the breakdown of host DNA for its nucleotides. In addition, both nucleases are involved in the metabolism of concatemers (37) and the gene 6 exonuclease has been implicated in the removal of RNA primers that arise during T7 DNA replication (38).

T7 induces the synthesis of a DNA polymerase (39,40) comprised of two subunits (41). One subunit is the product of gene 5. Phage mutants defective in gene 5 synthesize an altered enzyme (39) and no phage DNA (24). The other subunit, thioredoxin, is provided by the host (see below).

E. coli infected with T7 carrying amber mutations in gene 4 synthesize only small amounts of T7 DNA (24). Studies on gene 4 mutants suggest that the gene 4 protein functions throughout replication (42-44). As will be shown, the gene 4 protein has a dual role of helicase and primase at the replication fork.

Phage T7 DNA also codes for several additional proteins that are involved in DNA replication but which can be replaced by functionally similar proteins in *E. coli*. They and their functionally similar host proteins are discussed in a later section.

E. coli mutants. Two host proteins are known to be essential for T7 DNA replication: *E. coli* RNA polymerase and thioredoxin. Chamberlin (45) has isolated three classes of *E. coli* mutants designated *tsnA*, *tsnB* and *tsnC* (T-seven negative). These mutants grow normally but are unable to support the growth of wild type T7. Whereas *tsnA* mutants are impaired in phage adsorption, infection of either *tsnB* or *tsnC* mutants is characterized by normal cell killing as well as normal patterns of phage transcription and protein synthesis. In both strains the block in phage development is in DNA replication. In *tsnB* hosts the onset of viral DNA replication is delayed, and replication terminates prematurely, while in *tsnC* hosts no T7 DNA synthesis occurs. De Wyngaert and Hinkle (46) have shown that the *tsnB* mutation alters the *E. coli* RNA polymerase so that it is resistant to inhibition by the T7 gene 2 protein. The product of the *tsnC* gene (*trxA*) is thioredoxin, the host subunit of T7 DNA polymerase (see below). As described in a later section, *E. coli* also provides several replication proteins that can replace functionally similar phage proteins.

AN *IN VITRO* REPLICATION SYSTEM

A major advance in studying DNA replication has been the development of cell-free extracts that replicate exogenous DNA (4). In addition to providing information on the proteins and intermediates of DNA replication, these systems provide complementation assays to purify replication proteins for which no conventional assay is available. Such an *in vitro* system has been developed for T7 DNA replication (47,48).

Characterization of DNA replication in vitro

Extracts of gently lysed *E. coli* cells infected with wild-type T7 carry out extensive DNA synthesis, whereas extracts prepared from uninfected cells are inactive (Fig. 3a). DNA

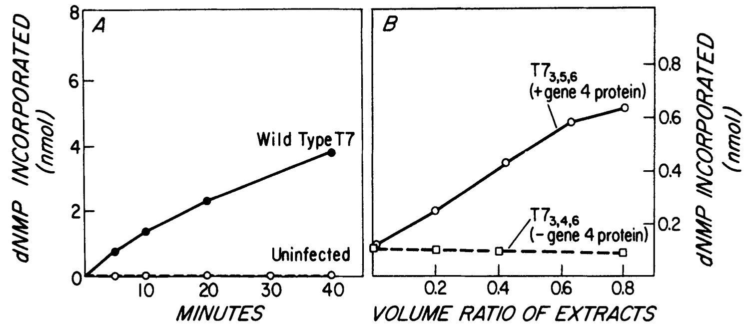


FIGURE 3. In vitro replication of phage T7 DNA. Extracts for replication of phage T7 DNA were prepared by gentle lysis of cells. Kinetics of DNA synthesis were measured by following the incorporation of dTTP into acid-insoluble DNA (48).
(A) Comparison of uninfected and T7-infected E. coli.
(B) Complementation of extracts of T7 3,4,6-infected E. coli (lacking gene 4 protein) with extracts of T7 3,4,6-infected E. coli (containing gene 4 protein).

synthesis in vitro requires the four deoxyribonucleoside triphosphates (dNTPs) and an exogenous T7 DNA template; the rate of synthesis is increased several fold by the addition of the four ribonucleoside triphosphates (rNTPs).

The initial rate of DNA synthesis is 3000 nucleotides per cell equivalent per second, and synthesis continues for at least 60 min. Under optimal conditions the amount of DNA synthesized is several fold the amount of template DNA present. Approximately 50% of the newly synthesized DNA is the same length as intact T7 DNA and is not covalently attached to the template (49). Provided the gene 6 exonuclease is present, a portion of the newly synthesized DNA is considerably larger than unit size T7 DNA, presumably reflecting the synthesis of concatemers (50). The full length product DNA molecules are fully infective and produce T7 phage when added to spheroplasts of E. coli (51). Furthermore, the newly synthesized DNA can be packaged in vitro to produce infective phage particles, the larger than phage size

T7 DNA molecules being packaged more efficiently (50). Thus the fidelity of the *in vitro* replication system is sufficient to produce biologically active DNA.

Requirement for replication proteins; In vitro complementation

DNA replication *in vitro* is dependent on several gene products necessary for DNA synthesis *in vivo*, but is independent of others. Incorporation of deoxynucleotides into DNA requires the products of phage genes 4 and 5 and the product of the host gene, *trx*. The latter two gene products, the phage gene 5 protein and *E. coli* thioredoxin, constitute T7 DNA polymerase. The products of phage genes 2, 3, and 6 are not required for synthesis *in vitro*, presumably because they affect processing of Okazaki fragments and concatemer metabolism, and hence their absence would not be detected by measuring DNA synthesis. Certainly, *in vitro*, the gene 3 and 6 nucleases are not required for the breakdown of host DNA to yield nucleotide precursors as *in vivo*, since exogenous dNTPs are present. In several cases the absence of a replication protein can be detected by examination of the product synthesized. For example, concatemer formation is greatly reduced if gene 2 protein (33) or gene 6 exonuclease (50) are absent and, in the absence of *E. coli* DNA polymerase I, fragments of T7 DNA accumulate (49).

A major use of the *in vitro* methodology is that extracts lacking a specific protein can be complemented by extracts or fractions containing the missing protein, thus providing a complementation assay to monitor its purification. As shown in Fig. 3b, extracts of *E. coli* infected with T7 gene 4 mutants do not support DNA synthesis. However, the addition of an extract containing gene 4 protein restores DNA synthesis. Thus, the requirement for the gene 4 protein *in vitro* has formed the basis of assays for its purification (42,43,52-54). Complementation assays have also been used to purify the gene 2 protein (33), the gene 5 protein (55), and *E. coli* thioredoxin (56).

THE ESSENTIAL PROTEINS OF T7 DNA REPLICATION

By either conventional enzyme assays or complementation assays, most of the proteins known to be involved in T7 DNA replication have been purified to homogeneity. Our knowledge of

Table 1. Essential proteins of phage T7 DNA replication.

Protein	Gene	Molecular weight	Reference
A. Phage proteins			
RNA polymerase	1	98,000	25,57-59
Inhibitor of *E. coli* RNA polymerase	2	7000	7,31-33,46
Endonuclease	3	17,000	35,83
Helicase, primase	4	58,000, 66,000	7,42,43,52-54
DNA polymerase subunit 3' to 5' exonuclease	5	84,000	39,41,55,67
5' to 3' exonuclease	6	31,000	36,85
B. Host proteins			
Thioredoxin	*trx*A	12,000	41,56

the overall process of DNA replication is to a large part derived from studies on the reactions carried out by these individual proteins. This section presents our current understanding of each of these proteins and the reactions they catalyze. The essential proteins of T7 DNA replication are summarized in Table 1.

T7 RNA polymerase

Phage T7 codes for its own RNA polymerase, the product of gene 1 (25). Several purification procedures have been described that lead to the preparation of a homogeneous enzyme (25,57,58). The amino acid sequence of the protein, determined from the nucleotide sequence of gene 1, reveals it to be a relatively basic protein with a molecular weight of 98,100 (59), in acceptable agreement with the molecular weight of 107,000 estimated for the purified protein (25,57,58).

T7 RNA polymerase recognizes promoters consisting of a highly conserved 23-base pair DNA sequence (60-63) occurring at 17 sites throughout the T7 DNA molecule (64). Transcription from these promoters proceeds from left to right relative to the genetic map of T7 (65,66).

T7 DNA polymerase

Properties of T7 DNA polymerasE. As shown in Fig. 4, T7 DNA polymerase, purified to near homogeneity (39,41,53,67), is composed of two subunits (41,56,68). One subunit is the 84,000 dalton polypeptide specified by gene 5 of the phage, while the other is the 12,000 dalton thioredoxin specified by the *trxA* gene of *E. coli* (69). Two lines of evidence indicate that thioredoxin functions by forming a highly stable complex with the phage-specified gene 5 protein to yield a functional DNA polymerase (41): First, thioredoxin is present in homogeneous preparations of the polymerase in a one-to-one stoichiometry with gene 5 protein (41,67). Second, reconstitution of T7 DNA polymerase from gene 5 protein and thioredoxin results in the formation of a one-to-one complex between the two proteins (41,55,67). The presence of thioredoxin in the T7 DNA polymerase has made possible the purification of the enzyme by antithioredoxin immunoadsorbent chromatography (70).

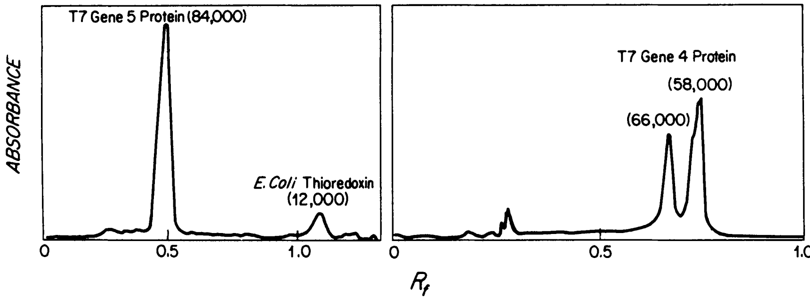


FIGURE 4. Sodium dodecyl sulfate polyacryamide gel electrophoresis of T7 DNA polymerase and T7 gene 4 protein. T7 DNA polymerase and T7 gene 4 protein were denatured and reduced, and subjected to electrophoretic analysis in the presence of sodium dodecylsulfate on a 5% polyacrylamide gel. After staining with Coomasie blue, the gels were scanned with a recording microdensitometer.

In addition to catalyzing the polymerization of nucleotides, T7 DNA polymerase has two associated exonuclease activities (67,71). Homogeneous preparations of T7 DNA polymerase possess both single-stranded and double-stranded DNA exonuclease activities, both of which hydrolyze DNA strands in a 3' to 5' direction.

Characterization of the subunits. One approach to understanding the contribution of thioredoxin and gene 5 protein to T7 DNA polymerase has been to isolate the two subunits and to determine the properties of each. Extracts of T7-infected *E. coli trxA* mutants, lacking thioredoxin, contain gene 5 protein but no T7 DNA polymerase activity. Such extracts provide the basis of assays for the purification of either thioredoxin or gene 5 protein. Thioredoxin has been purified to homogeneity from extracts of *E. coli* by monitoring the ability of thioredoxin to restore polymerase activity (56). The phage gene 5 protein has also been purified to homogeneity from extracts of T7-infected *E. coli trxA* mutants using a similar complementation assay in which gene 5 protein is measured by its ability to interact with thioredoxin to restore polymerase activity (55). A second approach to obtaining the purified subunits has been to dissociate homogeneous T7 DNA polymerase into biologically active subunits using guanidine HCl (67).

Neither the gene 5 protein nor thioredoxin itself has detectable DNA polymerase activity (55,67). However, thioredoxin and gene 5 protein together form a molecular complex indistinguishable from T7 DNA polymerase synthesized *in vivo*: the reconstituted polymerase interacts with T7 gene 4 protein to catalyze DNA synthesis on duplex DNA templates (55). Gene 5 protein alone has the 3' to 5' single-stranded DNA exonuclease activity found in T7 DNA polymerase, but not the double-stranded DNA exonuclease activity (67,71). Thioredoxin restores the double-stranded DNA exonuclease activity as well as the polymerase activity.

Using purified gene 5 protein and thioredoxin, the most efficient reconstitution occurs at a molar ratio of thioredoxin to gene 5 protein in the range of 10 to 200, depending on the

concentration of the gene 5 protein (55,67,71). Approximately 150 molecules of gene 5 protein are present in a phage-infected cell, and *E. coli* cells contain 3000-5000 molecules of thioredoxin (56). Since the apparent K_m for thioredoxin in the reconstitution reaction is 2.8 to 4.8 x 10^{-8} M (55,71) the *in vivo* assembly of T7 DNA polymerase should readily occur.

The mechanism by which thioredoxin interacts with gene 5 protein to form an active DNA polymerase is not known. However, studies with chemically and genetically altered thioredoxins should provide information on this important replication enzyme. S. Adler and P. Modrich (personal communication) have shown that reduced, but not oxidized, thioredoxin is required for both DNA polymerase and double-stranded exonuclease activity. Holmgren et al (72,73) have examined thioredoxins isolated from *trx*A mutants. One mutant, *tsn*C 7004, has no thioredoxin as measured by immunoprecipitation, suggesting a nonsense or deletion mutation in the *trx* gene, while another, *tsn*C 7007, has normal amounts of thioredoxin. Thioredoxin purified from *tsn*C 7007 has a single amino acid substitution indicative of a missense mutation (73). Although the altered thioredoxin can bind to gene 5 protein, it cannot restore polymerase activity.

Two forms of T7 DNA polymerase. T7 DNA polymerase can exist in two forms (2,74). Form I, obtained when purification is carried out in the absence of EDTA, has low levels of both the single- and double-stranded DNA exonuclease activities. Form I can initiate limited synthesis (several hundred nucleotides) at nicks in duplex DNA molecules, an event that results in strand displacement and eventually synthesis on the displaced strand via branch migration to yield a duplex branch. Form I is stimulated by the helicase activity of gene 4 protein on duplex templates. Form II, purified in the presence of 0.1 mM EDTA, cannot initiate synthesis at nicks even in the presence of gene 4 protein and has greatly increased levels of the associated exonuclease activities. The failure of Form II of T7 DNA polymerase to be stimulated by gene 4 protein is a consequence of its inability to initiate limited synthesis at a nick, a topic that will be discussed with regard to the helicase activity of the gene 4 protein. At present the physical difference between Forms I and

II is not known. It is possible that a proteolytic cleavage of Form II in the absence of EDTA yields Form I.

Gene 4 protein

The gene 4 protein has been purified by monitoring its activity using complementation assays (42,43,52-54). As shown in Fig. 4, the gene 4 protein, a single polypeptide, is found in two forms with molecular weights of 58,000 and 66,000 (24,43,52); both forms appear to have identical enzymatic activities (74). Examination of the nucleotide sequence of T7 DNA (7) reveals an AUG initiation site with a strong ribosome binding site for protein synthesis within gene 4. The separation between initiation codons (63 amino acid residues) is precisely what would be expected to give rise to the two observed polypeptides. The gene 4 protein has multiple activities; it is a single-stranded DNA dependent nucleoside 5'-triphosphatase, a helicase, and a primase. The nucleoside triphosphatase activity appears to be coupled to unidirectional translocation which is required for both primase and helicase activities.

Unidirectional 5' to 3' translocation requires the hydrolysis of NTPs. The gene 4 protein binds to single-stranded DNA, a reaction that requires a NTP. When β,γ-methylene-dTTP is substituted for the hydrolyzable NTP, a more efficient binding is observed, suggesting that the interaction of a NTP and gene 4 protein induce a conformational change that contributes to the stability of the enzyme-DNA complex. Gene 4 protein catalyzes the hydrolysis of NTPs to NDP and P_i in the presence of single-stranded DNA (75). The preferred NTP is dTTP with an apparent K_m of 4×10^{-4} M. The hydrolysis reaction is inhibited by β,γ-methylene-dTTP as are both helicase and primase activities, suggesting that the hydrolysis of NTPs is an essential component of these reactions. Studies on primer synthesis, described below, have shown that the gene 4 protein, once bound to single-stranded DNA, translocates unidirectionally 5' to 3' along the DNA strand (76). We conclude that the energy of hydrolysis of NTPs is required for the unidirectional movement of the gene 4 protein. Studies on the DNA concentration dependence of the hydrolytic activity suggest that the movement

of the gene 4 protein is highly processive. For example, the amount of single-stranded circular M13 DNA required for one-half maximal hydrolysis activity is at least 10-fold less than the amount of linear M13 DNA required, indicating that the rate-limiting step is its binding to DNA (S.W. Matson and C.C. Richardson, unpublished results).

Helicase Activity. The helicase activity of the gene 4 protein was initially inferred from its ability to stimulate T7 DNA polymerase using duplex DNA as a template (75). However, with the appropriate substrate, helicase activity can be measured directly (S.W. Matson and C.C. Richardson, unpublished results). A radioactively labeled 110 nucleotide DNA fragment is annealed to the single-stranded circular M13 DNA via a complementary 40 nucleotide sequence at the 5'-end of the fragment (Fig. 5). The remaining 70 bases on the 3'-end of the fragment are not complementary to the M13 DNA and thus provide a single-stranded tail. In the presence of gene 4 protein the radioactive fragment is displaced from the M13 DNA molecule, a reaction that is detected by analysis of the products by polyacrylamide gel electrophoresis. Presumably, the gene 4 protein binds randomly to the single-stranded M13 DNA and translocates in the 5' to 3' direction until it reaches the duplex region. Its continued movement results in displacement of the annealed fragment.

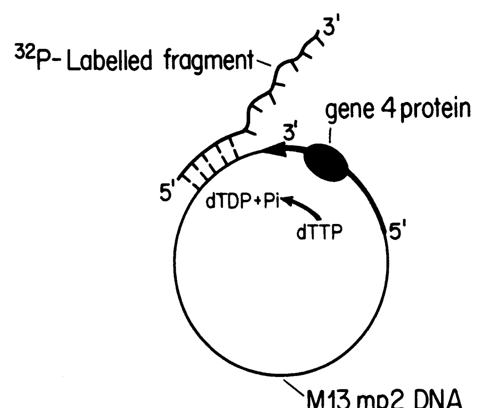


FIGURE 5. DNA substrate for assay of helicase activity of T7 gene 4 protein. A 175 nucleotide long fragment was annealed to the 7,200 nucleotide circular M13 mp2 DNA via a 40 nucleotide long sequence at the 5'-terminus of the fragment; the 3'-terminus of the fragment remains single-stranded. The gene 4 protein binds to the single-stranded circular DNA and, using the energy of hydrolysis of dTTP, translocates in a 5' to 3' direction. Helicase activity displaces the annealed fragment.

Helicase activity is dependent on the hydrolysis of a NTP. Again the preferred NTP is dTTP while β,γ-methylene dTTP inhibits the reaction. If no single-stranded 3' tail is present, the gene 4 protein cannot invade the duplex region. Using a mixture of annealed fragments with 3' tails of varying lengths, we find that 6 or 7 nucleotides are sufficient.

Helicase activity is also evidenced by the ability of the gene 4 protein to stimulate synthesis by T7 DNA polymerase on duplex DNA (75). Studies with Forms I and II of T7 DNA polymerase have shed additional light on the mechanism of action of gene 4 protein (Table 2). Form I of T7 DNA polymerase, in the absence of gene 4 protein, can initiate limited synthesis of several hundred nucleotides at a nick, resulting in the formation of a 5'-single-stranded tail (2). At some point, however, branch migration results in the displacement of the 3'-end of the newly synthesized DNA which in turn anneals to a short complementary sequence on the 5'-displaced single strand previously formed as a result of DNA synthesis. Continued synthesis produces short duplex branches (M. Engler, R. Lechner, and C.C. Richardson, unpublished results). However, when gene 4 protein is present,

Table 2. Effect of T7 gene 4 protein on the two forms of T7 DNA polymerase.

T7 DNA POLYMERASE	3' 5' (nicked circle)		3' 5' (circle with tail)	
	−gene 4 protein	+gene 4 protein	−gene 4 protein	+gene 4 protein
	(pmol dNMP Incorporated)			
Form I	10	144	14	221
Form II	0	33	6	594

Each reaction mixture (0.05 ml) contained 40 mM Tris (pH 7.5), 10 mM $MgCl_2$, 10 nM dithiothreitol, 0.3 mM each dNTP, one of which was radioactively labeled, 400 pmoles of either nicked circular pBR322 DNA or duplex circular M13 DNA bearing a 5'-single-stranded tail. T7 gene 4 protein (0.74 ug) was present as indicated. After incubation at 37° for 10 min the acid insoluble radioactivity was determined.

it can bind to the displaced single strand and translocate in the 5' to 3' direction until it reaches the replication fork where it then serves as a helicase. As synthesis occurs, it is accompanied by the hydrolysis of NTPs to NDPs and P_i. Hydrolysis is coupled to synthesis since β,γ-methylene dTTP inhibits DNA synthesis as well as NTP hydrolysis (75).

A displaced single strand to which the gene 4 protein can bind is required with Form II of T7 DNA polymerase. Form II of T7 DNA polymerase cannot initiate synthesis at a nick and gene 4 protein does not stimulate synthesis under these conditions. In order to demonstrate a direct requirement for a single-stranded tail, we have synthesized a duplex, circular DNA molecule in which one strand bears a 5'-single-stranded tail located at a unique site in the molecule (R. Lechner and C.C. Richardson, unpublished results). As seen in Table III, gene 4 protein provides a striking stimulation of DNA synthesis with Form II of T7 DNA polymerase on such a template.

Primase Activity. Gene 4 protein is also a primase. In the presence of rNTPs it catalyzes the synthesis of oligoribonucleotides on single-stranded DNAs such as ϕX174 or on the displaced strand that arises during synthesis on duplex DNA molecules (43,76-80). Primer synthesis is increased several fold by the presence of DNA-binding protein. ATP and CTP fully satisfy the requirement for rNTP.

When T7 DNA polymerase is present, it uses the oligoribonucleotides as primers to initiate DNA synthesis. When such newly synthesized chains are isolated, each chain is terminated at its 5'-end by a covalently attached tetraribonucleotide whose sequence is predominantly pppACCC and pppACCA; each of the four deoxyribonucleotides is present at the RNA-DNA junction (43,80). Since all of the RNA primers bear adenosine 5'-triphosphate residues at their 5'-termini, [γ-^{32}P]ATP is incorporated specifically into the product molecule, thus providing a sensitive and specific label for the RNA primers (79).

A possible mechanism by which the gene 4 protein catalyzes the synthesis of RNA primers is through a template mediated reaction in which the gene 4 protein first recognizes a sequence

complementary to the primer in the DNA template. We have determined the nature of the recognition signal by examining the sites of primer synthesis on the single-stranded circular DNA of ϕX174 (76) whose nucleotide sequence is known (81). With such a template DNA synthesis by T7 DNA polymerase is dependent on RNA primer synthesis catalyzed by the gene 4 protein. Thus all of the newly synthesized DNA will have a primer at its 5'-end. Analysis of the products revealed that DNA synthesis originates from 13 specific sites on the ϕX174 DNA molecule. In order to identify the precise locations on the template where each initiation event occurred, the nucleotide sequence of the 5'-terminus of DNA of each of the 13 species was determined. The template sequences complementary to the 5'-ends of all 13 predominant initiation fragments observed on ϕX174 are presented in Fig. 6. At every initiation site there is a sequence complementary to the primer sequences ACCC and ACCA.

Position	Sequence 3'		5'
886	CTG	CTGGT	CCCG
952	CAA	CTGGT	GGAT
1069	AGA	CTGGT	CGTT
1078	CTG	CTGGT	TAGA
1585	CGT	CTGGG	TATT
1789	GTG	CTGGT	CTTT
2047	TGA	CTGGG	AGTC
2736	CCG	CTGGT	AAGT
2836	AAT	CTGGT	TTGG
3650	TGA	CTGGT	CGGC
4664	ATG	CTGGT	TATA
4688	ACG	CTGGG	AGCC
4829	CTG	CTGGT	TTTA

Recognition Sequence 3'—CTGG(G/T)—5'

RNA Primer pppACC(C/A)—dN→

FIGURE 6. Recognition sites for RNA primer synthesis by gene 4 protein on ϕX174 DNA (76). The nucleotide sequences of the template ϕX174 DNA in the region of each of the thirteen gene 4 proteins initiation site is shown along with the position of the site on the template DNA where the terminal ATP of the RNA primer is base paired.

In addition, in every case there is a deoxycytidine residue adjacent to the 3'-end of the sequence complementary to the primer. Of the 65 sequences on ϕX174 DNA complementary to gene 4 protein primers, only these 13 sites are used, and they alone have a deoxycytidine residue at this position.

We conclude that the gene 4 protein recognizes the sequence, 3'-CTGGG-5' and 3'-CTGGT-5' on a single-stranded DNA template at which sites it catalyzes the synthesis of complementary pppACCC and pppACCA primers. Other sites, of the more general recognition sequence 3'-CTGGC/A-5' and 3'-CTGTN-5', can also be used for priming by the gene 4 protein, but less than 5% as efficiently as the predominant sites. Recently five *in vivo* initiation sites for Okazaki fragments isolated from newly replicated T7 DNA have been mapped (82). The same general recognition sequences (3'-CTGGN-5' or 3'-CTGTN-5') for RNA primer formation occur *in vivo*.

A comparison of the rate of utilization of primer sites suggests that the gene 4 protein binds randomly to single-stranded DNA and then translocates in a 5' to 3' direction with regard to the DNA strand until a recognition sequence is encountered. Non-hydrolyzable nucleoside triphosphate inhibit the priming activity of the gene 4 protein as well as the single-stranded DNA-dependent nucleoside triphosphatase activity. This result suggests an essential role of NTP hydrolysis and unidirectional movement of the gene 4 protein to rapidly form primers at specific recognition sites.

Gene 3 endonuclease and gene 6 exonuclease

Following infection T7 induces the synthesis of two nucleases, an endonuclease, the product of gene 3 (34), and an exonuclease, the product of gene 6 (36).

Gene 3 endonuclease. The gene 3 endonuclease, purified 2000-fold, has a molecular weight of 18,000 (35,83). The enzyme hydrolyzes phosphodiester bonds in single-stranded DNA at a rate 150-fold greater than that observed with double-stranded DNA (35,83). Furthermore, the extent of hydrolysis of bonds in single-stranded DNA is much greater than that with double-stranded DNA. The limit products with heat denatured T7 DNA are oligonucleotides of molecular weight of 3 to 6 x 10^3,

while with native T7 DNA the products, after denaturation, are 2 to 3 x 10^4 (35,84). Initial hydrolysis of duplex DNA results in the formation of single-strand breaks while more extensive hydrolysis leads to fragmentation of the molecule. In all cases phosphodiester cleavage results in the formation of 5'-phosphoryl and 3'-hydroxyl end groups. With duplex DNA the 5'-terminal nucleotides formed are predominantly deoxycytidylate (39%) and deoxythymidylate (32%).

Gene 6 exonuclease. The gene 6 exonuclease (36) has now been purified to homogeneity and shown to have a molecular weight of 31,000 (M.J. Engler and C.C. Richardson, unpublished results). The purified enzyme has a strong preference for duplex DNA (less than 0.5% activity on single-stranded DNA). It initiates hydrolysis at the 5'-terminus of a polynucleotide strand, hydrolyzing in a 5' to 3' direction, releasing nucleoside 5'-monophosphates (36,85). The enzyme ceases hydrolysis after approximately 50% of the duplex DNA is rendered acid soluble; the residual DNA is predominantly single stranded. Gene 6 exonuclease can initiate hydrolysis at either 5'-hydroxyl or 5'-phosphoryl termini. However, the initial product formed from a 5'-hydroxyl terminated substrate is a dinucleoside monophosphate. The enzyme acts distributively in its exonucleolytic attack rather than processively.

T7 gene 6 exonuclease also has an RNase H activity. The enzyme catalyzes the hydrolysis of poly(A) in a poly(A) poly(dT) hybrid duplex as well as effecting the removal of an RNA strand of a ϕX174 RNA-DNA hybrid molecule (86). Since the gene 6 exonuclease has been implicated in the removal of RNA primers during replication (38), we have examined the products of the RNase H activity resulting from the action of the enzyme on gene 4 protein primed DNA synthesis. The gene 6 exonuclease hydrolyzes the RNA primers from Okazaki fragments synthesized *in vitro*. The products arising from the hydrolysis of pppACCC and pppACCA primers are ATP, AMP, and CMP (1).

Although these two phage nucleases can account for the degradation of the entire *E. coli* chromosome, a number of questions remain unanswered. The gene 3 endonuclease exhibits a high degree of specificity for single-stranded DNA, yet repli-

cating T7 DNA would be more likely to contain such regions than would the *E. coli* chromosome. One possibility is that single-stranded regions in the supercoiled "folded chromosome" of *E. coli* provide the initial sites for attack. Once gene 6 exonuclease has rendered a portion of the *E. coli* chromosome single-stranded, the gene 3 endonuclease is ideally suited for further degradation to yield oligonucleotides that can be attacked by host exonucleases.

Gene 2 Protein

The gene 2 protein was originally purified from an *E. coli* RNA polymerase - gene 2 protein complex isolated from T7 infected cells (31,87) and subsequently shown to be the product of gene 2 (32). On the assumption that the gene 2 protein is required for maturation of T7 DNA via processing of concatemeric DNA, we have purified the gene 2 protein to homogeneity using an *in vitro* complementation assay to measure T7 DNA packaging (33). The gene 2 protein has also been purified by direct assay of its inhibitor activity against *E. coli* RNA polymerase (46).

The purified gene 2 protein has a molecular weight of 7000-9000 (31,33,46), in agreement with that of 7000 deduced from the nucleotide sequence of gene 2 (7). Gene 2 protein inhibits *E. coli* RNA polymerase but not T7 RNA polymerase (31,33,46). The gene 2 protein binds tightly to both *E. coli* RNA polymerase holoenzyme and to core polymerase. Gene 2 protein inhibits template binding and RNA synthesis by the RNA polymerase holoenzyme, but has no effect on core RNA polymerase. RNA polymerase which has initiated an RNA chain or which is bound at a promoter site is resistant to gene 2 protein (31,46). A mutant RNA polymerase, purified from *E. coli tsnB* (45), is also resistant to inhibition by the gene 2 protein (46).

THE DISPENSABLE PROTEINS OF T7 DNA REPLICATION

In addition to the essential phage proteins, other phage specified proteins are involved in T7 DNA replication but can be replaced by host proteins having functionally similar activities (Table 3). The activities of the T7 DNA ligase, the T7 DNA binding protein, and the T7 gene 1.2 protein can be replaced by the *E. coli* DNA ligase, DNA binding protein, and *optA* protein.

Table 3. Dispensable proteins of phage T7 DNA replication

Protein	Gene	Molecular weight	Reference
T7 gene 1.2 protein	1.2	12,000	97,116
E. coli optA protein	*optA*	unknown	97
T7 DNA ligase	1.3	41,000	10,88,91
E. coli DNA ligase	*lig*	77,000	88
T7 DNA binding protein	2.5	26,000	7,93,94,96
E. coli DNA binding protein	*ssb*	18,500	4,95

DNA ligase

The T7 and *E. coli* DNA ligases, both of which have been purified to homogeneity, have molecular weights of 41,000 and 77,000, respectively (see ref. 88). The catalytic properties of the two enzymes are similar with the exception that the T7 enzyme uses ATP or dATP as cofactor while the *E. coli* ligase uses NAD. Whereas the T7 enzyme can catalyze the joining of oligo(dT).poly(A) and oligo(A).poly(dT) (89), the *E. coli* enzyme can only catalyze the joining of a 5'-phosphoryl terminus of a DNA chain to the 3'-hydroxyl terminus of RNA (90).

T7 DNA ligase is the product of gene 1.3 (10,91). T7 ligase mutants grow normally in wild type *E. coli* but fail to produce progeny phage when grown on ligase-deficient *E. coli* (10,91,92); wild type T7 grows normally in such strains. No defect in T7 DNA replication is observed in wild type *E. coli* infected with T7 ligase deficient mutants. However, during infection of ligase deficient *E. coli* strains with ligase mutants, there is an accumulation of fragments of newly synthesized DNA (10). T7 DNA ligase can replace the *E. coli* DNA ligase in uninfected *E. coli* (M.J. Engler and C.C. Richardson, unpublished results), thus showing that the two enzymes are interchangable in spite of their differences in cofactor requirements.

DNA binding protein

T7 induces the synthesis of a single-stranded DNA binding protein (93,94), the product of gene 2.5 (7). The purified protein has a molecular weight of 25,000 to 31,000 (93,94), in agreement with that of 25,000 deduced from the nucleotide sequence of gene 2.5 (7). The protein binds strongly and preferentially to single-stranded DNA. It stimulates DNA synthesis on single-stranded DNA catalyzed by T7 DNA polymerase but not by the three E. coli DNA polymerases or T4 DNA polymerase (93).

T7 can grow on E. coli strains lacking the host DNA binding protein (95,96). However, T7 gene 2.5 mutants cannot grow on an E. coli strain having a defective E. coli DNA binding protein (96). DNA synthesis in the mutant infected strain is severely inhibited even in the first round of DNA replication.

Gene 1.2 protein; optA protein

In a recent search for additional proteins of T7 DNA replication we assumed that T7 duplicated some host activities in addition to the ligase and DNA binding protein. Since it was reasonable to assume that the genes for additional proteins of T7 replication would lie among the genes transcribed early in infection, we looked for E. coli mutants that would fail to support the growth of a T7 mutant lacking gene 1.2. The proximity of this small gene to the primary origin provided an additional incentive for its study. We have isolated a mutant of E. coli, optA, that is unable to support the growth of T7 phage having mutations in gene 1.2 (97). In the absence of both the phage gene 1.2 protein and the host optA protein, DNA synthesis ceases prematurely and the newly synthesized DNA is rapidly degraded (97). The optA gene is located at 3.6 min on the E. coli linkage map.

The function of the gene 1.2 and optA proteins is not yet known. Interestingly, the expression of gene 1.2 is regulated by the processing of mRNA by ribonuclease III (98). Recent insight into the function of the optA protein has come from the observation that certain mutations (antimutator phenotype) in the gene for T4 DNA polymerase or in the dda gene (helicase) of T4 lead to an inability of T4 to grow on optA strains (P. Gauss and

M. Gold, personal communication). One possibility therefore is that gene 1.2 and *optA* proteins are helicases.

Several other proteins have been implicated in T7 DNA replication. A 33,000 dalton protein has been purified to near homogeneity from uninfected *E. coli* using an assay for the stimulation of DNA synthesis by T7 DNA polymerase and gene 4 protein on duplex DNA lacking nicks (1). The purified protein has no detectable activity alone, but in the presence of T7 DNA polymerase and gene 4 protein DNA synthesis is initiated at specific sites on T7 DNA. Since the newly synthesized DNA is covalently linked to the template, synthesis must be initiated at nicks, presumably introduced by the 33,000 dalton protein. Since no *E. coli* mutants are known which are defective in the protein, its role in T7 DNA replication is not known.

The product of gene 0.3 is a protein that binds to the *Eco*B and *Eco*K restriction-modification enzymes and thus directly inhibit their activities (99,100). As a result T7 is not subject to the restriction-modification systems of these *E. coli* strains (101,102). T7 amidase, the product of gene 3.5, may also be important for DNA synthesis, perhaps to release the host DNA from membrane complex (24,103,104). A T7-induced inhibitor of *E. coli* exonuclease V, whose existence was initially postulated on the basis of its activity after T7 infection (105), has been partially purified (106). T7 also induces an endonuclease activity distinct from the gene 3 endonuclease (107,108) but no mutant phage lacking this enzyme has yet been identified. Finally, *in vivo* and *in vitro* studies indicate that *E. coli* DNA polymerase I may be important in the processing of Okazaki fragments of T7 (38,49,51).

THE REPLICATION FORK

The fundamental reactions occurring at the replication fork can be partially reconstituted using purified proteins. Based on our knowledge of the reactions catalyzed by the individual proteins and the replication intermediates and products synthesized *in vitro* it is possible to deduce the reactions occurring at the replication fork. In order to study these events it is first necessary to initiate DNA synthesis on a

duplex DNA molecule in order to simulate a replication fork. Inasmuch as the mechanism of initiation of DNA replication at the primary origin is not yet fully understood, we have initiated DNA synthesis by two different means, neither of which is dependent on site-specific initiation in contrast to DNA synthesis at the chromosomal origin.

In the first case we have made use of the ability of Form I of T7 DNA polymerase to initiate limited DNA synthesis at a nick and in the process generate a displaced 5'-single strand such as that depicted in Fig. 7. However, in the absence of other proteins DNA synthesis ceases after the polymerization of a few hundred nucleotides. Alternatively, in studies using Form II of T7 DNA polymerase we have first constructed such a replication intermediate using physico-chemical techniques.

Leading strand synthesis

The necessity for a replication intermediate such as that shown in Fig. 7 derives from the requirement of a single strand to which the gene 4 protein can bind. The gene 4 protein binds to the displaced single strand and then translocates in a 5' to 3' direction until it reaches the duplex region at which time its continued movement leads to disruption of the helix. The unidirectional translocation of the gene 4 protein requires the hydrolysis of NTPs to NDPs and P_i.

While the gene 4 protein serves as a helicase, T7 DNA polymerase catalyzes the polymerization of nucleotides and together the two proteins account for leading strand synthesis. Although the two proteins have been observed to purify together, it is not known if they physically interact at the replication

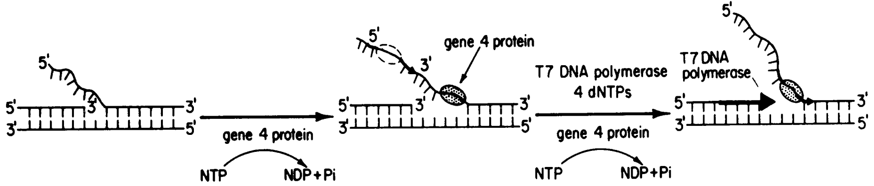


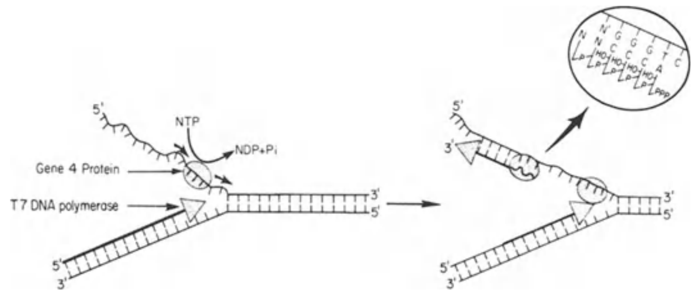
FIGURE 7. Schematic representation of leading strand synthesis catalyzed by T7 DNA polymerase and the helicase activity of the T7 gene 4 protein.

fork. As discussed earlier, the gene 4 protein alone requires a small single-stranded region on both strands in order to act as a helicase. Also unknown is whether the gene 4 protein alone can continue to invade the helix or whether DNA synthesis is required for the orderly movement of the gene 4 protein through the helix. Stoichiometric amounts of DNA-binding protein can partially replace the gene 4 protein, but the rate of synthesis is slower.

All of the DNA synthesized by T7 DNA polymerase and gene 4 protein on duplex DNAs in the absence of rNTPs is covalently attached to the template (109). Since synthesis is not specific for T7 DNA templates, we have also used duplex, circular DNA templates. Inasmuch as DNA synthesis can never proceed to the end of a circular molecule, extensive synthesis occurs by a rolling circle mechanism (110), resulting in product molecules with tails up to 20 times the length of the template molecule (109). Using duplex, circular DNA molecules in which one strand bears a 5'-single-stranded tail (see section on helicase activity of the gene 4 protein) we have measured the rate of fork movement as the T7 DNA polymerase and gene 4 protein circumnavigate the molecule. The rate of fork movement is approximately 300 nucleotides per second at 30° C.

<u>Lagging strand synthesis</u>

T7 DNA polymerase and gene 4 protein can also account for the initiation of lagging strand synthesis provided rNTPs are present (Fig. 8). As the replication fork proceeds, the gene 4



LEADING STRAND SYNTHESIS LAGGING STRAND SYNTHESIS

FIGURE 8. Schematic representation of lagging strand synthesis catalyzed by T7 DNA polymerase and the primase activity of the T7 gene 4 protein.

protein translocates along the displaced single strand until a gene 4 protein recognition sequence 3'-CTGGT-5' or 3'-CTGGG-5' is exposed (76). The gene 4 protein then catalyzes the synthesis of the tetraribonucleotides pppACCA and pppACCC which in turn serve as primers for T7 DNA polymerase (43,77-80). The synthesis of RNA primers is stimulated several fold by the presence of T7 or _E. coli_ DNA binding protein (43,79). T7 DNA polymerase is unique among DNA polymerases with regard to its ability to use such short primers effectively (111).

In contrast to leading strand synthesis, none of the DNA synthesized on the lagging strand is covalently attached to the template (80). The length of the Okazaki fragments synthesized both _in vivo_ (112) and _in vitro_ (80) range from 1000 to 6000 bases. However, on a random DNA sequence, the predominant gene 4 recognition sites occur with an average frequencyof once per 500 bases. This may indicate that the gene 4 protein synthesizes RNA primers infrequently at a replication fork, skipping over potential sites in most instances. This is clearly the case for the minor class of gene 4 protein recognition sites on ϕX174 (76). Also unanswered is the question of whether or not primer synthesis, and hence Okazaki fragment synthesis, occurs on the leading strand.

In the absence of other proteins to process the Okazaki fragments, one of two events occur depending on which form of T7 DNA polymerase is used. With Form I of T7 DNA polymerase DNA synthesis initiated at one primer site does not stop when it reaches the 5'-terminus of the next RNA-terminated DNA fragment (80). Instead, in the presence of the gene 4 protein, synthesis continues with the resulting displacement of a single-stranded RNA-terminated fragment that in turn can provide sites for primer synthesis, an event that is repeated many times. In contrast, with Form II of T7 DNA polymerase, synthesis stops once the gap between adjacent Okazaki fragments is filled.

Processing of Okazaki fragments _in vitro_ requires the removal of the RNA primers, filling of the resulting gap, and covalent joining of adjacent fragments. Using single-stranded ϕX174 DNA as a template, we have shown that either gene 6 exonuclease or the 5' to 3' exonuclease activity of _E. coli_ DNA

polymerase I can remove the RNA primers (1). If Form II of T7 DNA polymerase and DNA ligase are present, the Okazaki fragments are joined to make a continuous strand. Form I of T7 DNA polymerase, however, is so effective in promoting synthesis at a nick that even in the presence of DNA ligase, it continues the polymerization of nucleotides once the gap is filled. The resulting synthesis leads to limited strand displacement, with the consequence that adjacent fragments fail to be ligated.

INITIATION OF DNA REPLICATION AT THE T7 REPLICATION ORIGINS

Dressler *et al.* (9), using electron microscopy, showed that DNA replication *in vivo* is initiated at or near position 17, as measured by the formation of replication bubbles (eye forms, see Fig. 1). However, an exact assignment of position for the T7 origin was complicated by the existence of mutants of T7 carrying deletions within the region of positions 14.55 to 21.85 (113,114). The most likely explanation to account for the viability of these mutants was the existence of secondary origins elsewhere on the chromosome.

Primary and secondary origins

We have examined, by electron microscopy, replicating DNA molecules isolated from cells infected with both wild type T7 and T7 phage carrying deletions around position 17 (115). The experimental design was that used by Dressler *et al.* (9). *E. coli* cells growing in heavy medium are infected with light T7 phage. The replicating molecules can then be isolated on the basis of their hybrid density.

With wild-type T7, eye forms are the predominant species with the center of the bubbles located at position 17 (Fig. 9). In contrast, replicating LG37 DNA molecules carrying a deletion between positions 14.55 and 19.35 can be distinguished from the replicating T7 wild type DNA molecules. Most of the LG37 DNA molecules are Y-forms; no replication bubbles are observed at position 17 (115). Thus deletion of the region between positions 14.55 and 19.35 eliminates the primary origin normally used by wild type T7. The majority of the molecules lacking the primary origin initiate replication around position 4 although secondary origins are also present at positions 50 and 80 (Fig. 9).

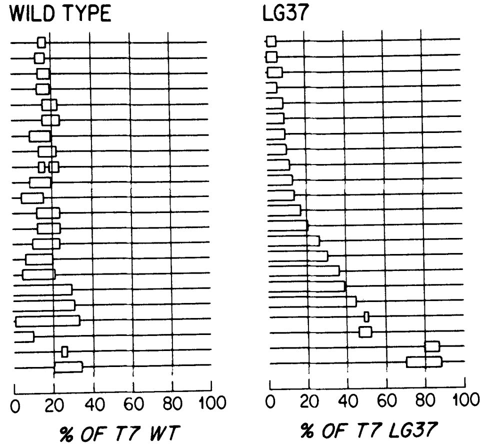


FIGURE 9. Line diagrams of partially replicated DNA molecules isolated from wild type T7 and T7 LG37 infected cells (115). E. coli 011' growing in heavy medium was infected with light T7 phage. The infected cells were harvested, lysed, and the DNA was centrifuged to equilibrium in CsCl. The replicating DNA molecules were analyzed by electron microscopy. The molecules were oriented on the basis of denaturation mapping.

Physical mapping of the primary origin

The observation that elimination of the primary origin leads to the use of secondary origins, together with the availability of T7 mutants carryingdeletions of varying sizes in this region, provides a way to map the primary origin. Using a set of six deletion mutants we found that at least a necessary portion of the primary origin is located within a 126 base pair region between positions 14.75 and 15.0 on the viral DNA molecule (115). When this region is deleted, T7 DNA replication is initiated at secondary origins of which the predominant one is located at position 4.

Nucleotide sequence and genetic organization of the primary origin

On the basis of the functional mapping of the primary origin, the sequence analysis of the region, and in vitro studies described here, we define the primary origin as the region extending from position 14.51 to position 14.95 (Fig. 10). By this definiion the primary origin lies between gene 1 (T7 RNA polymerase) on the left and gene 1.1 on the right.

The precise structure of the primary origin and its relationship to adjacent genes and regulatory elements have been obtained by determining the nucleotide sequence of this region of the T7 DNA molecule (116). A conspicuous feature of the primary

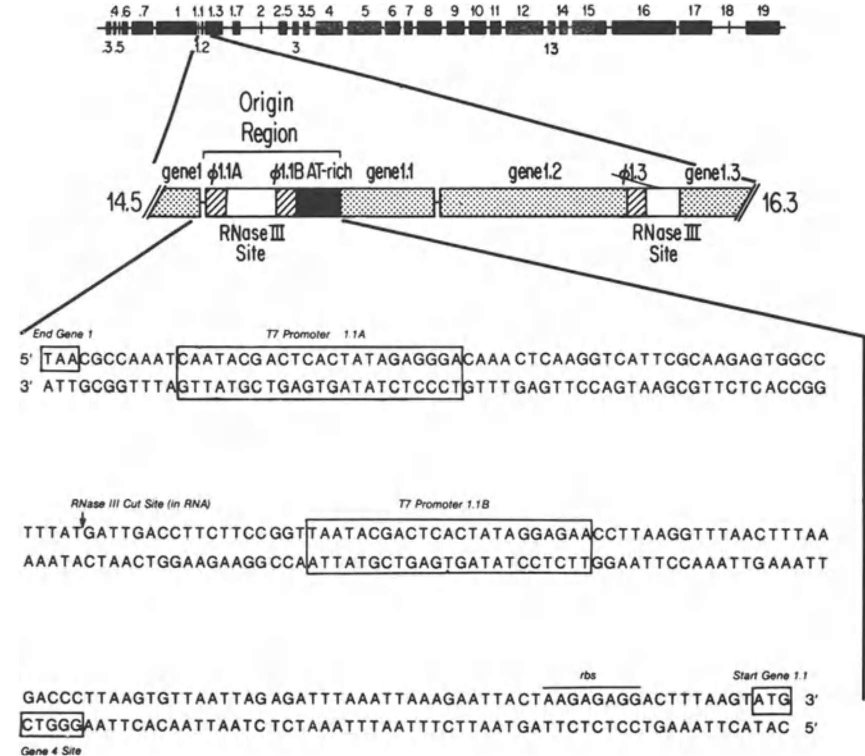


FIGURE 10. Nucleotide sequence of primary origin of T7 DNA replication. At the top is the genetic map of T7. Inserts show detailed genetic organization of the primary origin and its nucleotide sequence (116).

origin region is the presence of two tandem T7 RNA polymerase promoters, ϕ1.1A and ϕ1.1B. A single RNase III recognition site lies between the two promoters. Like all seventeen T7 RNA polymerase promoters (64), these two consist of a highly conserved 23 base-pair sequence (60-63). However, these two promoters are unique in their close proximity to each other and without interposition of any gene. Since this region of T7 is also transcribed by the E. coli RNA polymerase *in vivo*, there is no apparent requirement for these two promoters for the synthesis of messenger RNA. These considerations led us to postulate a

role of T7 RNA polymerase and transcription in the initiation of T7 DNA replication (116).

A second striking feature of the primary origin is a 61 base-pair region devoid of any genetic elements that lies between the second promoter, φ1.1B, and the ribosome binding site for gene 1.1. This region is extremely AT-rich (78% A+T); a major contribution to this high AT-content is the presence of seven copies of the palindrome, TTAA. At its center is a single gene 4 protein recognition site, 3'-CTGGG-5'. Primer synthesis at this site would direct DNA synthesis to the right.

What is the nature of the secondary origins, particularly the predominant one located near the left end? Although no precise mapping of this secondary origin has been carried out, a hitherto unknown T7 RNA polymerase promoter, φOL, at position 1.0, has recently been identified from sequence analysis (7). It is also followed by an AT-rich region.

In vitro mutagenesis of primary origin

In order to determine the role of each genetic element at the primary origin in the process of initiation, we are altering each of them by in vitro mutagenesis (S.Tabor and C.C. Richardson, unpublished results). Initially we have focused our attention on the two T7 RNA polymerase promoters, φ1.1A and φ1.1B. In the mutagenesis of these promoters we have used the plasmid, pBR322, into whose BamHI site the T7 fragment from positions 14.1 to 18.2 has been inserted. This plasmid thus contains the promoters φ1.1A, φ1.1B and φ1.3, along with the other sequences of the primary origins. Using this plasmid we have inserted, in vitro, a synthetic octamer into the center of φ1.1A and φ1.1B; the eight bases create a new restriction site. As expected from the highly conserved nature of the T7 RNA polymerase promoters, we find, using purified T7 RNA polymerase, that the insertion inactivates both promoters. Using marker rescue the T7 fragment containing the two inactivated promoters has been introduced into wild type T7 by recombination in vivo.

We have determined, using electron microscopy, the ability of T7 phage lacking the φ1.1A and φ1.1B promoters to initiate replication at the primary origin. When the two promoters are rendered inactive, DNA replication is initiated at secondary

origins, the predominant one being that located at position 4; no bubbles are observed at the primary origin. Thus, the ϕ1.1A and ϕ1.1B promoters constitute essential elements of the primary origin. Whether a single promoter at the primary origin is sufficient for initiation, as well as the role of the AT-rich region, remains to be determined.

Initiation at the primary origin by purified proteins; a requirement for T7 RNA polymerase

In view of the requirement for a functional T7 RNA polymerase promoter at the primary origin, we have examined the role of T7 RNA polymerase in initiation *in vitro* using purified proteins (117). Again we have used the formation of replication bubbles to score initiation events. When T7 DNA polymerase, T7 gene 4 protein, and DNA-binding protein are incubated with wild type T7 DNA for short periods, no eye forms are observed. However, the addition of T7 RNA polymerase results in the conversion of 17% of the DNA molecules to eye forms. Analysis of the eye forms reveals that all of the replication bubbles are located between 16 and 24% of the distance from the left end of the DNA molecule (Fig. 11). Although initiation is site specific

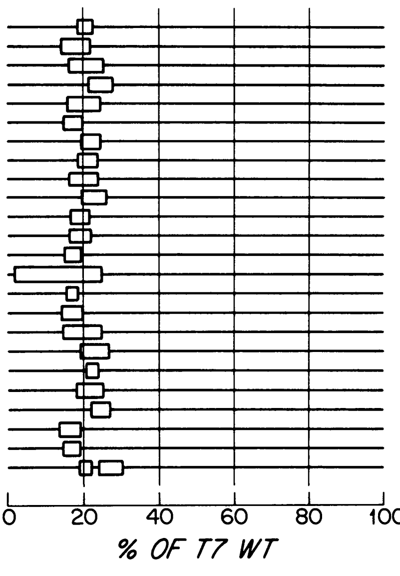


FIGURE 11. Line diagrams of partially replicated T7 DNA molecules isolated from a reaction containing T7 RNA polymerase, T7 DNA polymerase, and gene 4 protein (117). The molecules are oriented on the basis of denaturation mapping.

and the bubbles are centered near the primary origin, they do lie further to the right with their centers having a wide spread around position 19. Examination of the line diagrams of the partially replicated molecules shown in Fig. 11 reveals an interesting pattern. With only one exception the left fork of each replication bubble does not extend beyond position 15; at least a third of the left forks of the bubbles are located precisely at position 15. Thus replication _in vitro_ appears to initiate at position 15 but to proceed unidirectionally to the right, an interpretation that has been confirmed in studies using the cloned T7 primary origin.

Initiation _in vitro_ has an absolute requirement for T7 RNA polymerase and T7 DNA polymerase. The presence of gene 4 protein increases the number of initiation events 2-fold. RNA synthesis is required for initiation; no eye forms are observed in the absence of rNTPs. Furthermore, when wild type T7 DNA is replaced by the DNA of the deletion mutant, LG37, lacking the primary origin, no replication bubbles are observed.

The cloned origin of T7 DNA replication; unidirectional replication in vitro.

Assay for initiation of DNA replication by direct measurement of DNA synthesis is not feasible using the entire T7 DNA molecule. DNA synthesis catalyzed by T7 DNA polymerase and gene 4 protein can originate at nicks in the DNA, thus making it difficult to detect synthesis originating at the primary origin. In order to circumvent this problem and to have a template that can be more easily manipulated we have used plasmids containing the primary origin of T7 (117). One plasmid, pAR111, consists of pBR322 into which the T7 fragment from position 14.1 to 18.2, and thus containing the primary origin, has been inserted.

In the absence of T7 RNA polymerase, no DNA synthesis is catalyzed by T7 DNA polymerase and gene 4 protein on the linear plasmid pAR111 (Fig. 12). However, T7 RNA polymerase markedly stimulates DNA synthesis on this plasmid containing the primary origin. In a control experiment DNA synthesis on linear pDR100 was not significantly stimulated by the addition of T7 RNA polymerase (Fig. 12); pDR100 contains a fragment of T7 DNA with a

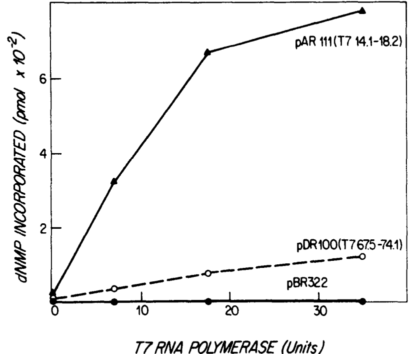


FIGURE 12. DNA synthesis on the cloned T7 primary origin (117). The plasmid DNAs were converted to linear forms with PvuII prior to use as templates. pAR111 contains the primary origin and the control plasmid pDR100 contains the T7 RNA polymerase promoter ϕ13.

single T7 RNA polymerase promoter. No synthesis is observed on the vector pBR322. The initiation specificity shown in Fig. 12 requires that the plasmid templates be first converted to linear molecules. When supercoiled templates are used T7 RNA polymerase stimulates DNA synthesis to the same extent on both the plasmid containing the origin (pAR111) and the control plasmid (pDR100), but not on the vector pBR322 (117). The elimination of specificity for initiation by supercoiling is most likely due to the known effect of supercoiling on transcription from certain promoters (118).

Electron microscopic analysis shows that 15% of the pAR111 molecules in the reaction mixture are converted to replicating forms. DNA synthesis initiates specifically within the cloned T7 DNA fragment in close proximity to the two promoters and the AT-rich region (C.W. Fuller and C.C. Richardson, unpublished results). A typical replicating pAR111 molecule is shown in Fig. 13A. More extensive synthesis results in the formation of Y forms such as the molecule shown in Fig. 13B. Restriction analysis of the Y forms observed in the electron microscope reveals that synthesis initiates near the primary origin but proceeds unidirectionally to the right in the direction of transcription from the ϕ1.1A and ϕ1.1B promoters. Thus, the predominant product molecules are Y forms with the fork of the Y located at position 15.

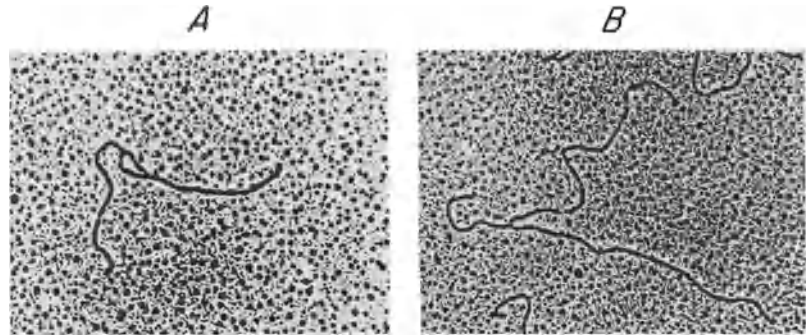


FIGURE 13. Electron micrographs of replicating linear plasmid pAR111 molecules incubated in the presence of T7 DNA polymerase, T7 RNA polymerase, and gene 4 protein. (A) Intermediate eye form with replication bubble and (B) product Y form resulting from unidirectional replication.

We suggest that leading strand synthesis proceeds rightward from the primary origin at position 15 until a gene 4 recognition site is exposed on the lagging strand (see Fig. 14). Lagging strand synthesis then proceeds leftward in a normal manner until the primary origin is reached. At this point, what was initially lagging strand synthesis now represents leading strand synthesis leftward. However, there is no displaced lagging strand at the left fork, and hence the gene 4 protein cannot bind and translocate in a 5' to 3' direction to facilitate unwinding of the duplex. Thus leftward synthesis cannot proceed past position 15. _In vivo_ this would probably represent a delay rather than a barrier since other proteins are most likely required to initiate leftward fork movement from the origin. Such a delay in movement of the fork leftward, combined with a delay in the initiation of lagging strand synthesis, may explain the observation that the replication bubbles found _in vivo_ appear at position 17 (9, 115) whereas the primary origin is located at position 15 (115,116).

Mechanisms involved in initiation at the primary origin

Earlier studies had shown that in the first round of replication, DNA synthesis is initiated at a specific site within

the left half of the T7 DNA molecule (9); a replication bubble appeared at approximately position 17. Our more recent deletion mapping of the primary origin revealed that the deletion of 126 base-pairs at position 15 led to inactivation of the primary origin (115). It seems likely, in view of our finding that DNA synthesis _in vitro_ requires additional components to proceed leftward from the origin, that the location of the midpoint of the replication bubble at position 17 _in vivo_ merely reflects more rapid movement of the right fork. DNA synthesis probably initiates near position 15 _in vivo_ as it does _in vitro_.

Knowledge of the nucleotide sequence of the region around position 15 has been particularly helpful in the analysis of precisely what constitutes an active origin (116). The conspicuous features of this region include an AT-rich region preceeded by two tandem T7 RNA polymerase promoters: a gene 4 protein recognition sequence for primer synthesis lies in the middle of the AT-rich region. The 126 base-pair region shown by deletion mapping to be an essential part of the primary origin eliminates little more than the AT-rich region and the gene 4 protein recognition sequence. The deletion that defined this region arose by a fusion of the 1.1B promoter with the 1.5 promoter further downstream. The fusion of the two promoters should, from examination of its sequence, create an active T7 RNA polymerase promoter. If this is the case, then the AT-rich region constitutes an essential part of the primary origin. Furthermore, we have now shown that inactivation of the 1.1A and 1.1B promoters also inactivates this region as a primary origin _in vivo_, a result that is supported by our _in vitro_ finding that T7 RNA polymerase is required for initiation. Earlier studies had indicated that T7 RNA polymerase is required for T7 DNA replication _in vivo_ (27).

Whereas these studies have elucidated the structure of the primary origin and shown a requirement for transcription, we do not yet know the precise mechanism by which transcription initiates DNA replication. Among the several possibilities two likely models are shown in Fig. 14. In Mechanism A, T7 RNA polymerase initiates transcription at one of the two T7 RNA polymerase promoters upstream from the AT-rich region.

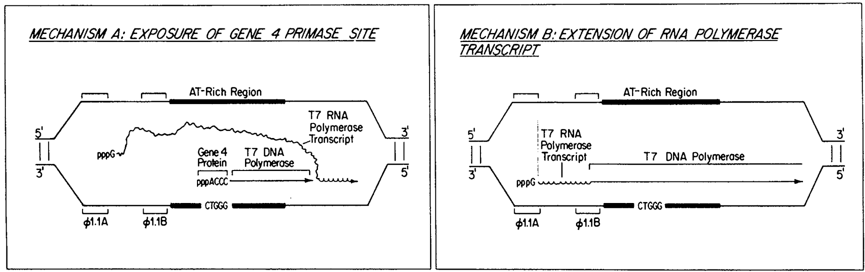


FIGURE 14. Two models for initiation of T7 DNA replication at the primary origin.

Transcription through the AT-rich region melts the DNA strands, exposing the gene 4 protein recognition site, 3'-CTGGG-5'. The gene 4 protein binds to this site and synthesizes the tetraribonucleotide primer pppACCC, which is then extended by T7 DNA polymerase, initiating DNA synthesis rightward. At a later time a gene 4 protein recognition site is exposed on the displaced strand, and lagging strand synthesis is initiated leftward. In this model, T7 RNA polymerase is serving only as an activator; its movement through the AT-rich region, and not the transcript itself, is essential. This mechanism is similar to that of transcriptional activation proposed for the initiation of phage lambda DNA replication (for review, see ref. 119).

In Mechanism B, T7 RNA polymerase initiates transcription from one of the two promoters. However, after T7 RNA polymerase terminates to make a unique RNA transcript, or the transcript is processed, T7 DNA polymerase uses the transcript itself as the primer for initiation of DNA synthesis in a rightward direction. In this mechanism the gene 4 protein is not directly involved in the initiation process, but obviously does play a role in facilitating movement of the fork and synthesizing RNA primers for lagging strand synthesis. Such a mechanism, in which an RNA polymerase transcript itself is used as the primer, is responsible for the initiation of replication of the single-stranded DNA of phage M13 (for review, see ref. 119) and of the double-stranded plasmid ColE1 (120).

At least *in vitro*, initiation of DNA synthesis appears to occur in a manner similar to that depicted by Mechanism B. Replication bubbles are observed at the primary origin in the absence of gene 4 protein (116) and either in the presence or absence of gene 4 protein newly synthesized DNA molecules bear RNA at their 5'-termini (C.W. Fuller and C.C. Richardson, unpublished results). These primer RNAs consist of two species, each having unique 5'-termini as a result of transcription from the ϕ1.1A and ϕ1.1B promoters. They are heterogeneous in length, however, their 3'-termini being determined by the transition to DNA synthesis. However, it is possible that, *in vivo*, both mechanisms depicted in Fig. 14 operate, at least under some conditions. Hopefully, current studies with T7 phage carrying mutations in each of the elements in the primary origin will provide a definitive answer to this question. Clearly, however, considerably more studies will be necessary before the complete details of this complex process are known. In fact, a requirement for additional proteins necessary to promote leftward fork movement, to process the RNA transcript, and to regulate initiation events can be easily imagined.

CONCLUSIONS

Studies on T7 DNA replication have provided a rather clear picture of the enzymatic events occuring at a replication fork. The requirement for so few proteins for this process in T7 infected-cell has simplified a determination of the requirements for both leading and lagging strand synthesis. At present the mechanism of RNA primer formation and discontinuous DNA synthesis are well characterized for the T7 system both *in vivo* and *in vitro*. Precisely what elements constitute a replication origin for the initiation of T7 DNA synthesis is now fairly well known and the prospects of identifying the enzymatic and molecular events that make for proper initiation at such an origin are good. On the other hand, little is known about the control of T7 DNA replication and the reactions that occur during the later stages of T7 DNA replication. For example, the process of concatemer formation, the control of presumably multiple

initiation events on such a large molecule, and the processing of these molecules pose intriguing questions that bear on the replication of large DNA molecules such as those found in both prokaryotic and eukaryotic cells.

REFERENCES

1. Richardson, C.C., Romano, L.J., Kolodner, R., LeClerc, J.E., Tamanoi, F., Engler, M.J., Dean, F.B., and Richardson. D.S. Cold Spring Harbor Symp. Quant. Biol. 43, 427-440, 1979.
2. Tamanoi, F., Engler, M.J., Lechner, R., Orr-Weaver, T., Romano, L.J., Saito, H., Tabor, S., and Richardson, C.C. in Mechanistic Studies of DNA Replication and Genetic Recombination. ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. XIX., ed. Alberts, B. (Academic Press, New York) pp. 411-428, 1980.
3. Tabor, S., Engler, M.J., Fuller, C.W., Lechner, R.L., Matson, S.W., Romano, L.J., Saito, H., Tamanoi, F., and Richardson, C.C. in Structure and DNA-Protein Interactions of Replication Origins. ICN-UCLA Symposia on Molecular and Cellular Biology, Vol. XXI, ed. Ray, D.S. (Academic Press, New York) pp. 387-408, 1981.
4. Kornberg, A. in DNA Replication (W. H. Freeman & Co., San Francisco), 1980.
5. Kruger, D.H., and Schroeder, C. . Microbiological Reviews 45, 9-51., 1981.
6. Thomas, C.A., Jr. J. Gen. Physiol. 49, 143-169, 1966.
7. Dunn, J.J., and Studier, F.W. . J. Mol. Biol. 148, 303-330, 1981.
8. Wolfson, J., Dressler, D., and Magazin, M. Proc. Nat. Acad. Sci. U.S.A. 69, 499-504, 1972.
9. Dressler, D., Wolfson, J., and Magazin, M. Proc. Nat. Acad. Sci. U.S.A. 69, 998-1002, 1972.
10. Masamune, Y., Frenkel, G.D., and Richardson, C.C. J. Biol. Chem. 246, 6874-6879, 1971.
11. Sternglanz, R., Wang, H.F., and Donegan, J.J. Biochemistry 15, 1838-1843, 1976.
12. Seki, T., and Okazaki, T. Nucleic Acid Res. 7, 1603-1619, 1979.
13. Ogawa, T., and Okazaki, T. Nucleic Acid Res. 7, 1621-1633, 1979.
14. Hausmann, R., and La Rue, K. J. Virol. 3, 278-281, 1969.
15. Kelly, T. J., Jr., and Thomas, C. A., Jr. J. Mol. Biol. 44, 459-475, 1969.
16. Stratling, W., Krause, E., and Knippers, R. Virology 51, 109-119, 1973.
17. Serwer, P. Virology 59, 70-88, 1974.
18. Frohlich, B., Powling, A., and Knippers, R. Virology 65, 455-468, 1975.
19. Paetkau, V., Langman, L., Bradley, R., Scraba, D., and Miller, R. C., Jr. J. Virol. 22, 130-141, 1977.
20. Center, M.S. J. Virol. 10, 115-123, 1972.

21. Hiebsch, R., and Center, M. S. J. Virol. 22, 540-547, 1977.
22. Stonington, O.G. and Pettijohn, D.E. Proc. Natl. Acad. Sci. U.S.A. 68, 6-9, 1971.
23. Worcel, A., and Burgi, E. J. Mol. Biol. 71, 127-147, 1972.
24. Studier, F.W. Science 176, 367-376, 1972.
25. Chamberlin, M., McGrath, J., and Waskell, L. Nature 228, 227-231, 1970.
26. Hausmann, R., and Gomez, B. J. Virol. 1, 779-792, 1967.
27. Hinkle, D.C. J. Virol. 34, 136-141, 1980.
28. Studier, F.W. Virology 39, 562-574, 1969.
29. Center, M.S. J. Virol. 16, 94-100, 1975.
30. De Wyngaert, M.A., and Hinkle, D.C. J. Virol. 33, 780-788, 1980.
31. Hesselbach, B.A., and Nakada, D. J. Virol. 24, 736-745, 1977.
32. Hesselbach, B.A., and Nakada, D. J. Virol. 24, 746-760, 1977.
33. LeClerc, J.E., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 76, 4852-4856, 1979.
34. Center, M.S., Studier, F.W., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 65, 242-248, 1970.
35. Sadowski, P.D. J. Biol. Chem. 246, 209-216, 1971.
36. Kerr, C., and Sadowski, P.D. J. Biol. Chem. 247, 305-310, 1972.
37. Miller, R. C., Jr. Ann. Rev. Microbiol. 29, 355-376, 1975.
38. Shinozaki, K., and Okazaki, T. Mol. Gen. Genet. 154, 263-267, 1977.
39. Grippo, P., and Richardson, C.C. J. Biol. Chem. 246, 6867-6873, 1971.
40. Oley, S.L., Stratling, W., and Knippers, R. Eur. J. Biochem. 23, 497-504, 1971.
41. Modrich, P., and Richardson, C.C. J. Biol. Chem. 250, 5515-5522, 1975.
42. Stratling, W., and Knippers, R. Nature 245, 195-197, 1973.
43. Scherzinger, E., Lanka, E., Morelli, G., Seiffert, D., and Yuki, A. Eur. J. Biochem. 72, 543-558, 1977.
44. Wolfson, J., and Dressler, D. J. Biol. Chem. 254, 10490-10495, 1979.
45. Chamberlin, M. J. Virol. 14, 509-516, 1974.
46. De Wyngaert, M.A., and Hinkle, D.C. J. Biol. Chem. 254, 11,247-11,253, 1979.
47. Stratling, W., Ferdinand, F. J., Krause, E., and Knippers, R. Eur. J. Biochem. 38, 160-169, 1973.
48. Hinkle, D. C., and Richardson, C. C. J. Biol. Chem. 249, 2974-2984, 1974.
49. Masker, W. E., and Richardson, C. C. J. Mol. Biol. 100, 543-556, 1976.
50. Masker, W. E., Kuemmerle, N. B., and Allison, D. P. J. Virol. 27, 149-163, 1978.
51. Masker, W. E., and Richardson, C. C. J. Mol. Biol. 100, 557-567, 1976.
52. Hinkle, D.C., and Richardson, C.C. J. Biol. Chem. 250, 5523-5529., 1975.
53. Scherzinger, E., and Seiffert, D. Molec. Gen. Genet. 141, 213-232, 1975.

54. Kolodner, R., Masamune, Y, LeClerc, J.E., and Richardson, C.C. J. Biol. Chem. 253, 566-573, 1978.
55. Hori, K., Mark, D.F., and Richardson, C.C. J. Biol. Chem. 254, 11591-11597, 1979.
56. Modrich, P., and Richardson, C.C. J. Biol. Chem. 250, 5508-5514, 1975.
57. Niles, E.G., Conlon, S.W., and Summers, W.C. Biochem. 13, 3904-3916, 1974.
58. Coleman, J.E. Biochem. Biophys. Res. Commun. 60, 641-648, 1974.
59. Stahl, S.J., and Zinn, K. . J. Mol. Biol. 148, 481-485, 1981.
60. Rosa, M.D. Cell 16, 815-825, 1979.
61. Panayotatos, N., and Wells, R.D. Nature 280, 35-39, 1979.
62. McAllister, W.T., and Carter, A.D. Nucleic Acid Res. 8, 4821-4837, 1980.
63. Boothroyd, J.C., and Hayward, R.S. Nucleic Acid Res. 7, 1931-1943, 1979.
64. McAllister, W.T., Morris, C., Rosenberg, A.H., and Studier, F.W. J. Mol. Biol. 153, 527-544, 1981.
65. Summers, W.C., and Szybalski, W. Virol. 34, 9-16, 1968.
66. Studier, F.W., and Maizel, J.V., Jr. Virol. 39, 575-586, 1969.
67. Adler, S., and Modrich, P. J. Biol. Chem. 254, 11605-11614, 1979.
68. Mark, D.F., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 73, 780-784, 1976.
69. Mark, D.F., Chase, J.W., and Richardson, C.C. Molec. Gen. Genet. 155, 145-152, 1977.
70. Nordstrom, B., Randahl, H., Slaby, I., and Holmgren, A. J. Biol. Chem. 256, 3112-3117, 1981.
71. Hori, K., Mark, D.F., and Richardson, C.C. J. Biol. Chem. 254, 11598-11604, 1979.
72. Holmgren, A., Ohlsson, I., and Grankvist, M.-L. J. Biol. Chem. 253, 430-436, 1978.
73. Holmgren, A., Kallis, G.-B., and Nordstrom, B. J. Biol. Chem. 256, 3118-3124, 1981.
74. Fischer, H., and Hinkle, D.C. J. Biol. Chem. 255, 7956-7964, 1980.
75. Kolodner, R., and Richardson, C.C. Proc. Nat. Acad. Sci. U.S.A. 74, 1525-1529, 1977.
76. Tabor, S., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 78, 205-209, 1981.
77. Scherzinger, E., Lanka, E. and Hillenbrand, G. Nucleic Acid Res. 4, 4151-4163, 1977.
78. Hillenbrand, G., Morelli, G., Lanka, E., and Scherzinger, E. Cold Spring Harbor Symp. Quant. Biol. 43, 449-459, 1979.
79. Romano, L.J., and Richardson, C.C. J. Biol. Chem. 254, 10476-10482, 1979.
80. Romano, L.J., and Richardson, C.C. J. Biol. Chem. 254, 10483-10489, 1979.
81. Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrell, B.G., Brown, N.L., Fiddes, J.C., Hutchison, C.A., III, Slocombe, P.M., and Smith, M. J. Mol. Biol. 125, 225-246, 1978.

82. Fujiyama, A., Kohara, Y., and Okazaki, T. Proc. Nat. Acad. Sci. U.S.A. 78, 903-907, 1981.
83. Center, M.S., and Richardson, C.C. J. Biol. Chem. 245, 6285-6291, 1970.
84. Center, M.S., and Richardson, C.C. J. Biol. Chem. 245, 6292-6299, 1970.
85. Kerr, C., and Sadowski, P.D. J. Biol. Chem. 247, 311-318, 1972.
86. Shinozaki, K., and Okazaki, T. Nucleic Acid Res. 5, 4245-4261, 1978.
87. Hesselbach, B.A., Yamada, Y., and Nakada, D. Nature 252, 71-74, 1974.
88. Engler, M.J., and Richardon, C.C. in The Enzymes, Vol. XIV, Part B, ed. Boyer, P.D., (Academic Press, New York) in press, 1982.
89. Fareed, G.C., Wilt, E.M., and Richardson, C.C. J. Biol. Chem. 246, 925-932, 1971.
90. Nath, K., and Hurwitz, J. J. Biol. Chem. 249, 3680-3688, 1974.
91. Studier, F.W. J. Mol. Biol. 79, 227-236, 1973.
92. Gellert, M., and Bullock, M.L. Proc. Nat. Acad. Sci. U.S.A. 67, 1580-1587, 1970.
93. Reuben, R.C., and Gefter, M.L. J. Biol. Chem. 249, 3843-3850, 1974.
94. Scherzinger, E., Litfin, F., and Jost, E. Molec. Gen. Genet. 123, 247-262, 1973.
95. Meyer, R.R.. Glassberg, J., and Kornberg, A. Proc. Natl. Acad. Sci. U.S.A. 76, 1702-1705, 1979.
96. Araki, H., and Ogawa, H. Molec. Gen. Genet. 183, 66-73, 1981.
97. Saito, H., and Richardson, C.C. J. Virol. 37, 343-351, 1981.
98. Saito, H., and Richardson, C.C. Cell 27, 533-542, 1981.
99. Spoerel, N.P., Herrlich, P., and Bickle, T.A. Nature 278, 30-34, 1979.
100. Mark, K.-K, and Studier, F.W. J. Biol. Chem. 256, 2573-2578, 1981.
101. Eskin, B., Lautenberger, J.A., and Linn, S. J. Virol. 11, 1020-1023, 1973.
102. Kruger, D.H., Schroeder, C., Hansen, S., and Rosenthal, H.A. Mol. Gen. Genet. 153, 99-106, 1977.
103. Inouye, M., Arnheim, N., and Sternglanz, R. J. Biol. Chem. 248, 7247-7252, 1973.
104. Jensen, H.B., and Pryme, I.F. Biochem. Biophys. Res. Commun. 59, 1117-1123, 1974.
105. Wackernagel, W., and Hermanns, U. Biochem. Biophys. Res. Commun. 60, 521-527, 1974.
106. Pacumbaba, R., and Center, M.S. J. Virol. 16, 1200-1207, 1975.
107. Center, M.S. J. Biol. Chem. 247, 146-156, 1972.
108. Sadowski, P.D. Can. J. Biochem. 50, 1016-1023, 1972.
109. Kolodner, R., and Richardson, C.C. J. Biol. Chem. 253, 574-584, 1978.
110. Gilbert, W., and Dressler, D. Cold Spring Harbor Symp. Quant. Biol. 33, 473-484, 1969.

111. Masamune, Y. and Richardson, C. C. J. Biol. Chem. 252, 8498-8503, 1977.
112. Ogawa, T., and Okazaki, T. Ann. Rev. Biochem. 49, 421-457, 1980.
113. Studier, F.W., Rosenberg, A.H., Simon, M.N., and Dunn, J.J. J. Mol. Biol. 135, 917-937, 1979.
114. Simon, M.N., and Studier, F.W. J. Mol. Biol. 79, 249-265, 1973.
115. Tamanoi, F., Saito, H., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 77, 2656-2660, 1980.
116. Saito, H., Tabor, S., Tamanoi, F., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 77, 3917-3921, 1980.
117. Romano, L.J., Tamanoi, F., and Richardson, C.C. Proc. Natl. Acad. Sci. U.S.A. 78, 4107-4111, 1981.
118. Yang, H.-L., Heller, K., Gellert, M., and Zubay, G. Proc. Nat. Acad. Sci. U.S.A. 76, 3304-3308, 1979.
119. Tomizawa, J., and Selzer, G. Ann. Rev. Biochem. 48, 999-1034, 1979.
120. Itoh, T., and Tomizawa, J. Proc. Nat. Acad. Sci. U.S.A. 77, 2450-2454, 1980.

10

POXVIRUS DNA REPLICATION

JOHN A. HOLOWCZAK

SUMMARY

Vaccinia virus DNA replication occurs primarily in the cytoplasm of infected cells. Following adsorption, penetration and uncoating, at least two "nicks" are introduced into cell-associated, cross-linked, parental DNA molecules, which can be recovered from the cytoplasm of infected cells in the form of membrane-DNA complexes (ρ = 1.175 g/cm^3). Semiconservative, discontinuous viral DNA replication, requiring continuous protein synthesis, occurs initially in association with these membrane-parental DNA complexes and involves a partially circular or "rolling-circle" intermediate. The first product of DNA synthesis appears to be a concatemer in which unit-length genomes are held together, in head-to-head and tail-to-tail arrays. As the replication of these concatemers is completed, or nears completion, they associate together into larger, complex structures termed virosomes (ρ = 1.285 g/cm^3). It is not clear if new rounds of DNA synthesis can be initiated in the virosomes or if partially replicated molecules are simply completed in these complexes. Unit length, but not yet cross-linked viral DNA molecules are excised from the virosomes and sealed, by a multi-enzyme system containing exonuclease, DNA polymerase and DNA ligase activities, present in the cytoplasm of infected cells. The mature, terminally cross-linked, 180 kilobp, progeny genomes, are then encapsulated into developing vaccinia virions.

INTRODUCTION

Viruses of the Poxviridae family (1) share a similarity

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

in size, morphology, certain antigenic properties, linear ds DNA genomes and replication in the cytoplasm of infected cells. The family has been subclassified into six genera; the orthopoxviruses, avipoxviruses, leporipoxviruses, capripoxviruses, parapoxviruses and entomopoxviruses (2). More than 40 species distributed among these genera have now been described. Electron photomicrographs of some Poxviridae are presented in Figure 1. The most striking external morphological differences are seen when negatively stained members of the parapoxvirus genus and the viruses of the other five genera are examined in the electron microscope.

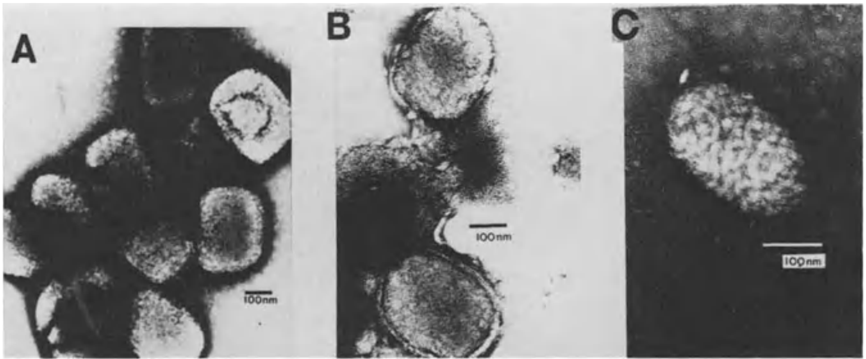


Figure 1. Electron photomicrographs of negatively stained poxviruses. Panel A. Vaccinia virus (an orthopoxvirus), Panel B. Swinepox virus. While morphologically similar to orthopoxviruses, a separate genus, the Suipoxviruses, has been proposed for this virus (3). Panel C. Milker's nodule virus (a parapoxvirus).

DNA comprises 3-5% of purified pox virions and in those cases where the nucleic acid has been analyzed, the base composition and physical properties are consistent with a ds DNA structure (4,5). Avipoxviruses (Fowlpox, canary pox virus) have a DNA genome with a molecular weight of 160-180 million (6,7), orthopoxvirus genomes vary in size from 118 million [rabbitpox virus (8)] to 145 million [cowpox (9)]. The parapoxviruses (Orf virus, stomatitis papulosa virus) contain a genome with a molecular weight estimated to be

80-90 million (10,11). This is the smallest parental poxvirus genome described thus far. The genomes of orthopoxviruses are characterized by a G+C content of 36-37 mole % (4,5,12,13) while parapoxvirus genomes have a G+C content of approximately 65% (11,14).

The majority of experimental studies in the laboratory are performed with one of the available strains of vaccinia virus. The vaccinia viruses (15,16) were among the first viruses to be successfully grown in tissue culture and the ease with which they could be propagated, purified and titrated, facilitated their use in experimental studies (17). The physical compartmentalization of host and viral genomes in the infected cell provided investigators with a major technical advantage as compared to workers who studied DNA viruses which replicated in the host-cell nucleus. To a degree this "advantage" was counterbalanced by the sheer complexity of the virus and its genome. Nevertheless, the poxviruses have and continue to provide an ideal model system for studying the organization and expression of viral genomes in a eukaryotic environment (17,18,19,20).

Recently, the use of restriction endonucleases to probe the structure of poxvirus genomes, the preparation of "libraries" of cloned restriction endonuclease fragments of vaccinia DNA and the isolation of formidable collections of temperature sensitive (ts) and host-range orthopoxvirus mutants, have opened a variety of new and exciting avenues for exploring the manner in which poxviral genomes are transcribed and replicated in the infected cell. The molecular biology of poxvirus DNA replication as exemplified by studies employing the various vaccine strains will be emphasized here. Our knowledge in this area can be best described as being incomplete but the availability of new experimental tools and systems makes it certain that this will rapidly change. This is a particularly propitious time to review the status of the problem and the directions that may be taken in the future.

STRUCTURE OF THE VACCINIA GENOME

Implications for Replication

Vaccinia virions contain a linear, ds DNA genome of approximately 180,000 base pairs (bp) which is terminally cross-linked (21,22). In addition, the genome has inverted terminal repetitions of approximately 10 kilobp (23,24) which contain direct tandem repeats of a 70 bp sequence arranged in two blocks of 13 and 17 copies, respectively (25,26). Details of the structure near the ends of vaccinia DNA have now been determined by direct nucleotide sequencing (26). Terminal hairpins were found demonstrating that the genome consisted of a single, continuous, polynucleotide chain. The 104 nucleotide apex of the hairpin contained predominantly A and T residues (92%) and was incompletely base paired. The lack of complete base pairing in this region would provide a site where the polynucleotide chain could be "nicked" by the ss DNase present in purified virions (22,24). As defined by the electrophoretic behavior, under nondenaturing conditions, of restriction endonuclease fragments which encompassed the termini, the terminal loops appeared to exist in two forms [F (fast) and S (slow) migrating] which when inverted with respect to each other were complementary in sequence. When analyzed under denaturing conditions, the mobilities of the two species was identical. It was concluded that the observed differences in electrophoretic behavior of the "terminal" restriction endonuclease fragments, under native conditions, reflected differences in nucleotide sequence and structure rather than an absolute difference in the size of the fragments (26). When Hind III digests of monkeypox and ectromelia DNA were analyzed, the terminal fragments cleaved from each end of the viral genomes were found to co-migrate upon electrophoresis in non-denaturing gels (9). Therefore, the conclusions reached in studies with DNA prepared from vaccinia virus, strain WR, cannot be generalized and applied to all the orthopoxviruses.

Nucleotide sequencing showed that a set of 13 tandem, 70 bp repeats begin 87 bp from the proximal segment of the

terminal loop of the vaccinia genome followed by a unique sequence of 325 bp and then a second set of 18 tandem, 70 bp repeats (26). The sequence of the 70 bp repeats revealed a 13 bp redundancy (26). Just beyond the second set of 70 bp repeats, approximately 600 bp of, as yet, unsequenced DNA is present, before the first transcriptionally active region of the vaccinia genome is encountered (27).

The 10 kilobp terminal repeats "bracket" a central region of the vaccinia genome. The nucleotide sequence in this central region of the genome appears to be highly conserved (as probed by a variety of restriction endonucleases) among the orthopoxviruses and appears to encode the majority of the viral specific proteins. Rapid progress is now being made in mapping this region of the genome (27,28,29,30).

Cross-linked termini have been demonstrated to be present in the genome of molluscum contagiosum virus (31) as well as the orthopoxviruses (9), avipoxviruses (7) and parapoxviruses (10), which have been examined in this regard. Preliminary evidence suggests that African swine fever virus, a member of the iridovirus family, may also contain a genome with cross-linked termini (32). This structural feature may be common to the DNA genomes of viruses which replicate in the cytoplasm of cells they infect. The biological significance of the terminal cross-links remains to be elucidated. It is clear, however, that in order to be replicated in a semi-conservative fashion, at least one cross-link must be removed or one "nick" introduced into cell-associated, parental genomes following uncoating and a mechanism for "sealing" progeny genomes or introducing terminal cross-links, must be provided for in any proposed replication scheme.

Molecular hybridization experiments have indicated that inverted terminal repeats containing sets of repetitive sequences, like those described above for vaccinia virus, are absent in variola DNA (9) and truncated (as compared to vaccinia DNA) inverted terminal repeats are present in

parapoxvirus DNA (10). Vaccinia virus isolates containing DNA with different numbers of tandem repeats and unstable isolates, in which an entire set of repeats and intervening region were amplified, have been reported (33,34). It was proposed that unequal cross-over events were responsible for multiplying and then correcting the numbers of tandem sequences which were present (34). Vaccinia virus variants and mutants (35,36,37) and rabbit poxvirus mutants (38,39) have been characterized which contain DNA with significant deletions of DNA sequences. In some cases these deletions have been mapped and shown to be within or immediately adjacent to, the inverted terminal repeat region of the genome. Variants or mutants of rabbit, monkey and cowpox viruses have also been described in which sequences have been deleted within or proximate to the inverted terminal regions of the genome but the deletion compensated for by sequences copied from the opposing terminus (40,41,42). In some cases this "transposition event" resulted in a net increase in the size of the viral DNA. These studies suggest, that in contrast to the "central portion" of the genome the sequence organization within or adjacent to the inverted terminal repeat regions of orthopoxvirus genomes can be hypervariable and subject to extensive deletions as well as complex symmetrical and asymmetrical rearrangements. Variants or mutants bearing such alterations are viable. The picture which is emerging is that such alterations appear to affect the ability of the variants or mutant viruses to replicate in certain hosts (37,39).

It is difficult to believe that a region of the genome susceptible to such drastic changes in sequence organization can play a required function during viral DNA replication. Nevertheless, the presence of the large inverted terminal repeats and repetitive sequences within this region has led some investigators to propose that these structural elements can promote annealing of opposing termini of displaced single strands at some stage during DNA replication (25,35,42). Two pieces of evidence have been put forward against such an

obligatory circular intermediate during poxvirus DNA replication. First, variola DNA does not appear to contain inverted terminal repeats (9) and rabbit poxvirus mutants have been characterized which contain DNA where a portion of the left but not the right inverted terminal repeat has been deleted (39). Since both variola virus and the mutant rabbit poxviruses can replicate their DNA in an appropriate host, it can be argued that circularization of the genome is not obligatory. Further, vaccinia virus variants containing DNA in which an entire set of tandem repeats and intervening region have been duplicated or amplified do not appear to replicate their DNA more efficiently than virus isolates containing "normal" numbers of tandem sequences (34). One might expect that duplication of these structures would promote nucleation and circularization of the genomes and their replication. The matter is complex and remains to be adequately explored. In addition to their role in promoting the circularization of the genome, the tandem repetitive sequences within the larger inverted terminal repeats have been proposed as sites where recombination between genomes could be facilitated (34,40).

EXPERIMENTAL SYSTEMS

Virus

Generally, investigators employ purified suspensions of vaccinia virus for infecting cells. This permits uniform infection of cells at a given multiplicity of infection so that the kinetics of viral DNA synthesis is highly reproducible. The rapid sedimentation rate of virions (~5000S) facilitates the isolation of virus from the cytoplasmic fraction of infected cells (43,44). By a series of differential centrifugation steps (44,45,46), followed by zonal centrifugation in preformed sucrose gradients (45,46,72), equilibrium centrifugation in cesium chloride (47) or potassium tartrate (4), highly purified virus particles can be prepared. A useful compilation of methods and physical-chemical characteristics of purified poxvirus

preparations has appeared (48). A spectrophotometric method for estimating the number of elementary bodies (EB) or virus particles in purified virus preparations has been described (45,49) and this measurement can be related to the nucleic acid and protein content of the preparation. When such measurements are coupled to an estimation of the infectivity of the virus, most often by determination of the number of plaque forming units present (pfu/ml) by titration on appropriate host cells [for other methods see (16)], it is possible to define both the "quantity" and the "quality" of the virus present in the preparation.

As reviewed above, there is evidence that the vaccinia genome is subject to a variety of sequence rearrangements. How and why this occurs is not clear. It is possible that this may be the first step in the development of defective interfering virus particles. What is clear, is that in order to appropriately interpret the structure of replicating viral DNA, one must first clearly define the structure of the parental DNA molecules. It is therefore useful to monitor for the development of virus particles containing DNA with alterations by analysis of the products obtained following restriction endonuclease digestion of DNA prepared from the virus to be used to infect cells. Virus to be employed for preparation of large stocks for use in experiments should be plaque purified at regular intervals. Finally, when preparing virus stocks, cells should be infected at low multiplicities, with inoculum which has been previously passaged a limited number of times. In this manner, the development of defective-interfering virus in stocks may be limited or avoided.

INFECTION OF CELLS

The growth cycle of poxviruses may be conveniently studied in a variety of tissue-culture systems and a one-step growth curve, with a well defined eclipse phase can be measured (50,51). L-mouse fibroblasts, HeLa cells and KB cells, adapted for growth in spinner culture, provide

particularly favorable cells for the study of vaccinia virus DNA replication (51,52,53). Uniformly infected, mass cultures can be prepared from which the appropriate number of cells can be withdrawn, at various times post-infection, to be processed for analysis.

Adsorption of vaccinia virus particles to the cell membrane is a rapid process and more than half of the inoculum may be attached within 15 min. (54). To achieve a uniform, synchronous infection, spinner-cultured cells are first concentrated to $1\text{-}2 \times 10^{7}$ cells/ml in appropriate "adsorption medium" (54) and high multiplicities of infection, on the order of 500-1000 EB/cell, are employed. Adsorption or association of virus particles to cells can occur at $4^{o}C$; penetration and uncoating of virus occurs most efficiently at $37^{o}C$. Further synchronization of the infection can therefore be obtained by allowing adsorption to proceed at $4^{o}C$ and then transferring the culture to $37^{o}C$. This is a particularly useful maneuver when infecting cells in monolayer cultures where adsorption may not proceed as efficiently or uniformly as in suspension cultures. After an appropriate adsorption period, concentrated, infected cells in suspension are rapidly diluted into medium (pre-warmed to $37^{o}C$) to a concentration of $0.5\text{-}1 \times 10^{6}$ cells/ml. This dilution is usually sufficient to prevent the subsequent adsorption of any free virus and, under these conditions, a single-step growth curve is observed.

INHIBITORS

A number of inhibitors of transcription, translation and DNA replication have been employed in experiments designed to dissect the events in viral DNA replication (17,18,19,20). There are two which appear to be particularly useful. Following the adsorption period, vaccinia virus-infected cells can be diluted into medium containing hydroxyurea (HU). At a concentration of 5×10^{-3} M, HU will inhibit both viral and host DNA replication (55,56,57). In

HU-treated cells, the uncoating of virus particles proceeds normally and early virus functions, including the viral enzymes involved in DNA synthesis and the viral products responsible for the inhibition of host-macromolecular synthesis, are expressed (55,56,57). After an appropriate period, the HU-block can be reversed by washing the drug out of the infected cells (56,57). Viral DNA replication is initiated almost immediately and can be studied against a significantly reduced background of host-macromolecular synthesis.

Rifampins are broad-spectrum antibiotics which inhibit the replication of bacteria by binding to DNA-dependent-RNA polymerase and blocking transcription (58). In the presence of rifampin, poxvirus replication is also inhibited (59). However, rather than blocking transcription, the primary effect of rifampin has been related to the assembly or morphogenesis of poxvirus particles (60,61,62,63,64,65). Viral DNA and protein synthesis are not significantly affected by the drug but assembly of mature progeny is blocked. When the drug is removed, the assembly of virions is rapidly initiated (66). The use of this drug therefore provides a means of separating the events concerned with viral DNA replication from those concerned with integrating the DNA into developing virions (63,67).

As with all drugs, one must be cognizant of the general toxic effects such inhibitors may have on the host cells being employed and the evolution of viral mutants which are resistant to the effects of the inhibitor (61). It is hoped, that the collections of ts and host-range mutants of orthopoxviruses, presently being characterized in a number of laboratories, will contain isolates blocked at various stages of DNA replication making the use of inhibitors with their potential and real limitations, unnecessary.

CELL FRACTIONATION

Vaccinia virus DNA replication occurs primarily in the cytoplasm of the infected cell. By appropriate

manipulations, the cell's plasma membrane can be ruptured or solubilized and by differential centrifugation a nuclear and cytoplasmic fraction can be prepared. This separates the host's genetic material from that of the virus. A variety of techniques can then be applied to the analysis and characterization of the viral DNA molecules present in the cytoplasmic fraction.

Mechanical homogenization in a Dounce-type ball homogenizer (68) and lysis of cells with a non-ionic detergent such as Nonidet P-4Ø [octylphenol ethylene oxide] (69) are the two methods commonly employed to "break-open" infected cells. The review of cell-fractionation techniques by Penman (7Ø) provides useful insights into the two methods and their limitations. Both methods have certain disadvantages. Dounce homogenization may subject the large vaccinia DNA molecules and DNA replication complexes to shearing forces which, potentially, may cause significant fragmentation of DNA molecules. While detergent lysis is a more gentle technique, detergents compromise the association of viral DNA with membranes and may solubilze subviral particles. Intermediates which may be present during uncoating, DNA replication or assembly of virus particles are lost or structurally altered when detergents are employed to lyse cells. To obtain a complete understanding of the nature and distribution of replicating viral DNA molecules, cytoplasmic samples prepared by mechanical and detergent methods should be analyzed.

The cytoplasmic fraction prepared from virus infected cells contains viral DNA associated with protein. It is doubtful if viral DNA completely free of protein is ever present in the infected cell. The methods which investigators selected to separate the viral DNA present in the cytoplasm of infected cells from its associated proteins was influenced by the difficulties which had been experienced in the laboratory in obtaining intact DNA from purified virus (5,17). Even after it was recognized that reducing agents, such as 2-mercaptoethanol, were required in addition to ionic

detergents to solubilize virions (71) phenol extraction procedures frequently yielded fragments of viral DNA rather than intact genomes (5,73). It was perceived that similar problems would be encountered in attempting to purify viral DNA from the cytoplasm of infected cells. When Sarov and Becker (49) showed that it was possible to obtain intact vaccinia DNA molecules by solubilizing virus particles on the surface of sucrose gradients and then sedimenting the released DNA into the gradients; modifications of this procedure were quickly applied to the analysis of viral DNA in cytoplasmic fractions (74,75). Ultimately, the manipulations required for preparation of the cytoplasmic fractions were circumvented by lysing whole, infected cells on the surface of gradients (76,77). By carefully selecting ionic conditions, host and virosome DNA remained condensed and pelleted during sedimentation analysis; viral DNA molecules released from the virosomes, as their replication was completed, were resolved and displayed in the gradients (67). Pogo and her associates (78,79) approached the problem more directly and showed that following solubilization of the cytoplasmic fraction prepared from infected cells with ionic detergents, reducing agents and protease digestion, full-length viral DNA molecules could be extracted and purified using conventional phenol extraction techniques. Organic extraction procedures for the purification of viral DNA from cytoplasmic extracts are now widely applied and have provided suitable substrates for use in experiments where restriction endonucleases are employed to probe the structure of newly replicated progeny genomes (80,81,82).

It is remarkable, that while a variety of experimental techniques have now been applied to fractionating infected cells, releasing and purifying viral DNA molecules and analyzing the replicating viral DNA species, the same general conclusions have been reached concerning the manner in which vaccinia virus DNA is replicated (see below).

ANALYSIS OF VACCINIA DNA REPLICATION

Outline of the steps in viral DNA synthesis

When infected at a multiplicity of 500 EB/cell, viral DNA replication begins in the cytoplasm of infected HeLa or L cells at about 1.5 hr post-infection (p.i.), reaches a maximum 2.5-3.5 hr p.i. and is 90% or more complete by 4-5 hr p.i. (84,85,86,87,88,89). The abrupt cessation of poxvirus DNA synthesis at a time when the assembly of progeny begins may be related to the "switch-off" of early viral protein synthesis (20).

The solubility in alkali of the cytoplasmic complexes containing replicating viral genomes permits the analysis of denatured DNA molecules labeled during short-pulses or pulse-chase experiments by sedimentation in alkaline sucrose gradients. This type of experiment coupled with molecular hybridization techniques has allowed workers to conclude that replication proceeds in a discontinuous manner (74,75,76,79,80) and that initiation and termination sites for DNA replication may be located at or near both ends of the viral chromosome (80).

The average sedimentation coefficient of material labeled during short pulses, which probably consists of growing and completed short DNA chains, was determined to be about 8-12S indicating that the "unit" of discontinuous synthesis was about 1000-2000 nucleotides. The synthesis of these "Okazaki-like" fragments has been shown to be primed by RNA (79). Pulse-chase experiments demonstrate that the 8-12S fragments are chased first into full-length but not yet cross-linked viral DNA and finally into cross-linked genomes which sediment at about 102-108S in alkaline sucrose gradients. Molecules sedimenting at ∿ 92S are also observed. The sedimentation behavior of these molecules is appropriate for denatured progeny genomes with one intact cross-link (or bearing a single "nick"). It remains to be determined if these molecules are intermediates in the formation of mature genomes.

The synthesis of full-length, but not yet cross-linked,

viral DNA molecules takes about 30 min, which would require the incorporation of 500 nucleotides/min, a rate about twice that reported for DNA synthesis in uninfected mammalian cells (91). After an initial round of viral DNA replication is completed, synthesis of a protein or proteins which function in a non-catalytic manner (18) is required in order to initiate each additional new round of DNA replication (67).

Density labeling of replicating viral DNA molecules with bromodeoxyuridine and analysis by equilibrium density centrifugation in CsCl shows that hybrid molecules (hl, ρ=1.77 g/ml) and "heavy viral DNA" (hh, ρ=1.825 g/ml) can be detected, demonstrating that vaccinia DNA replication is semi-conservative (67,75). Analysis of replicating viral DNA molecules in ethidium bromide-CsCl gradients at equilibrium failed to show the presence of circular or superhelical duplexes (75). This result and the fact that no denatured viral DNA molecules of greater than unit length were detected during long or short pulses suggested that viral DNA replication proceeded in a symmetrical fashion.

The conversion of newly replicated, full-length viral DNA strands into mature, cross-linked DNA molecules occurs late in the DNA replication cycle. The introduction of cross-links appears to be a post-replicative process (74,75,79). It has been suggested that both ends of vaccinia DNA may be regarded as a form of terminal palindrome (92). If so, introduction of internucleotide linkages at both ends of the molecules by the joint action of DNA polymerase and polynucleotide ligase would generate the covalently continuous vaccinia virus genome (93). Pogo (81) has described a multiple enzyme system, present in the cytoplasmic fraction prepared from vaccinia virus infected cells, which catalyzed the introduction of cross-links into full-length but not yet terminally cross-linked progeny genomes.

Viral DNA molecules whose replication is completed or is near completion are released from the cytoplasmic complexes or factories. Continuous protein synthesis is required for

this orderly dissociation of DNA molecules but release can be demonstrated to occur independently of virion assembly (67,77).

The experimental results described above provided a general scheme for vaccinia virus DNA replication. We will now examine in detail the fate of parental viral DNA molecules after they are uncoated in the infected cell, the manner in which DNA replication may be initiated and the nature of the initial product of viral DNA replication.

Detailed Consideration of the Fate of Parental Genomes and Viral DNA Replication.

Intracellular vaccinia virus DNA has been shown to be associated with characteristic cytoplasmic inclusion complexes which have been termed "factories" or virosomes ("viral body" in analogy to the term chromosomes). Studies employing autoradiographic techniques (9) staining with fluorescent antibody specific for viral proteins (95,96) and electron microscopy (97) have shown that viral DNA replication occurs in conjunction with the virosomes. Joklik and Becker (86) were the first to study the nature of intracellular viral DNA at various times after infection. They observed that newly replicated viral DNA molecules could be recovered as large aggregates which were heterogeneous in size. Their experimental findings suggested that both magnesium ions and proteins were responsible for maintaining the newly replicated DNA in the form of large aggregated complexes. Dahl and Kates (98) developed a sucrose density gradient technique which proved useful for the partial purification and analysis of poxvirus DNA complexes from infected cells (57,82,99). Using these experimental conditions, virosomes were isolated from vaccinia virus-infected HeLa cells in the form of rapidly sedimenting complexes with a density of 1.285 g/cm^3 (98). Two major proteins with molecular weights of 30,000 and 35,000 are associated with such virosomes (100,101,102) and *in vitro* experiments have demonstrated that the complexes contain

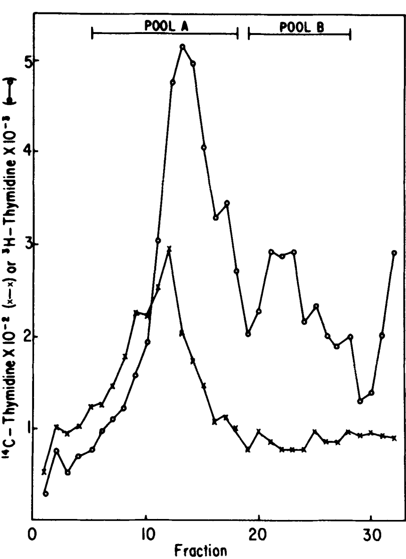
enzymes capable of transcribing both "early" and "late" viral mRNA (1Ø3).

A significant aspect of the work of Dahl and Kates (98) was the observation that parental and progeny genomes underwent dramatic alterations in density as a function of time after infection. It was suggested that some viral DNA components become membrane associated during their replication and assembly into virions.

The seminal studies of Joklik (54,1Ø4) first examined the uncoating of cell-associated poxviral genomes. Uncoating is now defined as a two step process. Virions are first converted to a subviral form termed a "core." At this stage, active synthesis of mRNA under the direction of the DNA-dependent RNA polymerase encapsulated within the virions is initiated (1Ø5,1Ø6). It has been speculated that the proteins required for the second step of uncoating may be made under the direction of these "pre-early" mRNAs. The second stage of uncoating involves the release of the parental DNA from the core structure. Pogo (78,1Ø7) has demonstrated that during the uncoating process, the cross-links are removed or, alternatively, at least two "nicks" are introduced into the parental DNA molecules. This "nicking" or removal of cross-links may be mediated by nucleases which are found enclosed within the viral cores prepared from mature virions (83).

When the techniques of Dahl and Kates (98) are applied to fractionating cell cytoplasm prepared from infected cells, virus and(or) viral cores are recovered in association with membranes (ρ=1.157 g/cm^3 in sucrose) [Figure 2, Pool B; Figure 3, Panel A]. Coincidentally uncoated and "coated" parental genomes are found associated with a membrane-rich, subcellular fraction banding at a density of 1.175 g/cm^3 [Figure 2, Pool A; Figure 3, Panels B,C,D]. When replicating progeny genomes are pulse-labeled with [^{14}C]-thymidine, in cells which have been infected with virus containing DNA labeled with [^{3}H]-thymidine, labeled progeny are first detected in association with the parental genomes at a

Figure 2. Analysis of the parental and pulse-labeled progeny genomes present in the cytoplasm of infected HeLa cells. HeLa cells were infected with vaccinia virus strain WR containing DNA labeled with [^{3}H]-thymidine (1000 EB/cell). At 1.45 hr post-infection 6x10^7 infected cells were pulse-labeled with [^{14}C]-thymidine (0.5 µCi/1x10^6 cells) and 15 min later, the cells were harvested, the cytoplasmic fraction prepared and analyzed as described by Dahl and Kates (98). No 80% sucrose cushion was employed so that virosomes were recovered in the pellet fraction.



density of 1.175 g/cm^3 and are subsequently converted to a form which sediments with the virosomes (ρ=1.285 g/cm^3). Reproducibly, the pulse-labeled replicating DNA can be shown to be slightly denser than the parental species (1.180 g/cm^3 vs. 1.175 g/cm^3, respectively) [Figure 2] (98). As noted by Dahl and Kates (98) and confirmed in this laboratory, the dual peaks containing parental DNA, which can be resolved after density centrifugation analysis of cell cytoplasm prepared at various times after infection, persist late into the infection cycle. While progeny DNA molecules rapidly become part of the virosome fraction, the uncoated parental DNA molecules remain largely associated with the subcellular species banding at a density of 1.175 g/cm^3. Between 3-5 hr after infection of HeLa cells with vaccinia virus, 90% or more of the progeny DNA molecules are recovered as part of the virosomes (98); 20% or less of the parental genomes were found in the virosome fraction at this time (unpublished observations).

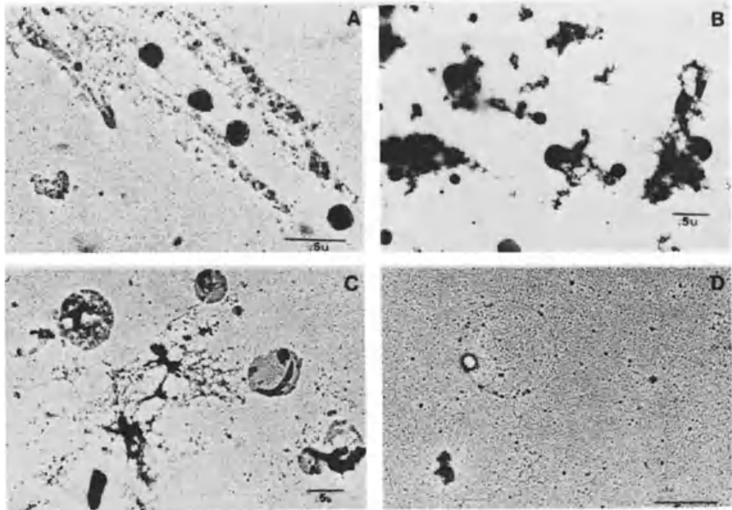


Figure 3. Electron photomicrographs of subviral particles and viral DNA molecules present in the cytoplasm of vaccinia virus infected cells. The cytoplasmic fraction from vaccinia virus infected HeLa cells was prepared and fractionated as described by Dahl and Kates (98) [Figure 2]. Appropriate fractions were combined as indicated [Figure 2, Pools A and B], dialyzed (108) and the samples processed and prepared for electron microscopy after spreading on polylysine coated grids as described by Williams (117). Panel A. Material recovered from Pool B. Virus and viral cores in association with membranes were observed. Panels B and C. Material recovered from Pool A. Viral genomes in the process of being released from subviral particles and DNA-protein complexes released from cores were observed. Panel D. Material recovered from Pool A. Fractions containing [^{14}C]-thymidine labeled, replicating viral DNA [Figure 2] were enriched with these partially circular DNA species. This kind of intermediate would be expected if viral DNA was replicated by a rolling-circle mechanism (108). The marker in each panel equals 0.5 μ.

While proteins and magnesium ions contribute to maintaining viral DNA in a condensed form in virosomes, there is evidence that the DNA molecules may also be linked together in concatemeric forms. Viral DNA can be released from virosomes by treatment with alkali or by digestion with S-1 nuclease but not RNase or pronase (67). This suggests that ss DNA regions *per se* or proteins having affinity for ss DNA may bind the progeny genomes together in virosomes.

Moyer and Graves (82) prepared virosomes from rabbitpox virus-infected cells and digested the virosomal DNA with restriction endonucleases. Analysis of the products provided evidence that the virosomal DNA consisted primarily of concatemers containing unit length molecules joined by fusion of two left (LL) or two right (RR) ends. These concatemeric, head-to-head and tail-to-tail mirror image arrays may be the substrate from which appropriate size mature genomes are excised and packaged. It was estimated that on the average, the basic concatemeric structure in the virosomes was four genomes in length (82). The fact that cytoplasmic viral DNA can be completely solubilized in alkali suggests that the molecules in the virosomal concatemers may not be bound together covalently. Alternatively, it can be postulated that nucleases present in the virosome fraction "nick" the DNA, so that upon denaturation of the complexes, fragments of DNA or unit length viral DNA molecules are released. A fundamental question which remains to be answered is whether any parental genomes associate as concatemers to provide a template for DNA replication.

The experimental findings summarized above provide evidence that within the cytoplasm of poxvirus infected cells, viral DNA replication may be compartmentalized; early after infection (1-2 hr p.i.) replication appears to occur in conjunction with membrane-associated parental genomes. Progeny are then rapidly converted to a form which sediments as part of the virosomes. Later in the infection cycle (2-3 hr p.i.) pulse-labeling experiments show that newly replicated viral DNA is associated with both the upper-band containing membrane associated parental genomes and the virosomes (98). These experimental results suggest that there are two possible pathways for viral DNA replication. Viral DNA replication may occur only in conjunction with the membrane-associated parental genomes. Subsequently, the progeny are rapidly and continuously transferred into the virosome fraction. Electron microscopic evidence has been presented that partially circular molecules are present in

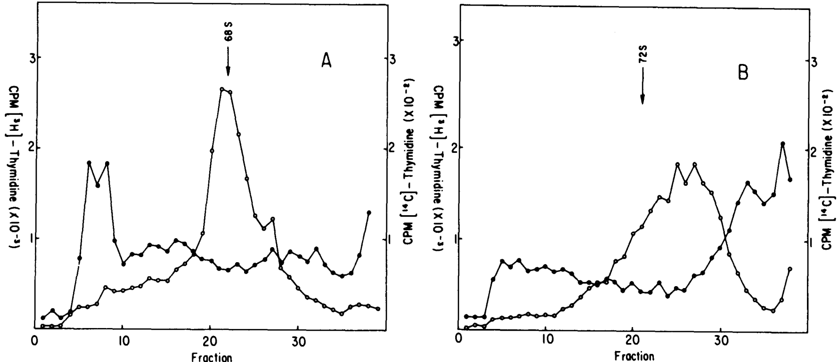


Figure 4. Analysis of parental and replicating viral DNA molecules by sedimentation in neutral and alkaline sucrose gradients. The cytoplasmic fraction prepared from vaccinia virus-infected Hela cells was prepared and analyzed as described in Figure 2. The [^{3}H]-thymidine labeled parental genomes and [^{14}C]-thymidine, pulse-labeled replicating viral DNA molecules, recovered from Pool A [Figure 2], were analyzed by sedimentation in 15-3Ø% neutral (Panel A) or alkaline (Panel B) sucrose gradients. Conditions for centrifugation, fractionation of gradients and determination of acid-insoluble radioactivity present in the recovered fractions was as described previously (75). O--O [^{3}H]-thymidine labeled parental DNA; O--O [^{14}C]-thymidine, pulse-labeled progeny DNA.

the fraction containing membrane associated parental genomes (ρ=1.175 g/cm^{3}) (1Ø8) [Figure 3, Panel D]. When the parental genomes labeled with [^{3}H]-thymidine and pulse-labeled progeny labeled with [^{14}C]-thymidine recovered from the appropriate gradient fractions (Figure 2, Pool A) were analyzed in neutral and alkaline sucrose gradients, the results shown in Figure 4 were obtained. Under both conditions, newly replicated genomes sedimenting more rapidly than full-length, cross-linked genomes are detected. Under denaturing conditions (Figure 4, Panel B), molecules sedimenting at >1Ø2-1Ø6S (appropriate for mature, cross-linked parental genomes) were detected suggesting that

covalently-linked complexes, larger than mature viral genomes are synthesized. Molecules sedimenting at 12-20S were also resolved indicating that discontinuous DNA synthesis was occurring [Figure 4, Panel B]. Taken together, these results provide support for the idea that viral DNA molecules may replicate via a "rolling circle" mechanism (108). The product may be the concatemeric DNA molecules described by Moyer and Graves (82). It can be argued that the labeled material which is localized in virosomes after very short pulses with radiolabeled thymidine (98) does not reflect viral DNA replication but is the result of repair, recombination or the formation of cross-links as progeny genomes are excised from concatomers and mature. Alternatively, viral DNA may be replicated in two stages. In the initial stages (1-1.75 hr p.i.) replication could occur in association with the membrane associated parental genomes as described above. The initial product of DNA replication would then enter the virosomes and there serve as the template for the second phase of viral DNA replication (2-5 hr p.i.). Replication may continue at both sites but with time progeny would accumulate in the virosomes and this would then appear to be the dominant site of DNA replication. Such a two-stage scheme is supported, in part, by the pulse-labeling experiments carried out by Dahl and Kates (98). It can be postulated that during the first phase of viral DNA replication, a circular intermediate may be involved while during the second phase only linear forms participate. Examination of the data presented by Joklik (54,104) concerning the uncoating of viral DNA, reveals that 50% or less of the cell-associated genomes are uncoated at the time that DNA replication is initiated. This suggests that there is a continuous uncoating of genomes during the period of DNA replication. As the parental genomes are uncoated they may associate with membranes and initiate a new round of DNA replication. Uncoating of viral genomes may therefore be closely linked to the initiation of viral DNA synthesis.

A number of experimental approaches could be applied to obtain experimental data to answer some of the questions posed above. Pulse-chase experiments would be particularly valuable in this regard. However, in our hands, we find that the fractionation procedures described by Dahl and Kates (98) are imperfect, and can produce variable results. While a "virosome fraction" can be readily isolated, the resolution of membrane-associated viral DNA species varies and there is considerable cross-contamination between the various subcellular fractions resolved by this method. Dahl and Kates (98) were cognizant of these problems and cautioned investigators that their techniques were applicable only to vaccinia virus-infected HeLa cells and that their conditions must be carefully followed in order to reproduce their results. Clearly, there is a need to explore new methods for fractionating cytoplasmic extracts prepared from infected cells and to more fully define the association of parental and replicating viral DNA molecules with membranes. Poxviruses are unique in that they direct the synthesis of new membranes in the cytoplasm of infected cells (109). Whether the putative DNA-membrane associations described above occur between cellular or viral-specified membranes is unknown.

The host-cell nucleus and vaccinia DNA replication.

Although the cytoplasm appears to be the major site of vaccinia virus replication and assembly, there is now evidence that a functional host-cell nucleus is required for production of infectious progeny. Viral DNA, RNA and protein are transported from the cytoplasm into the nucleus (110,111,112,113,114,115). The relationship between the viral DNA which enters the nucleus, and the viral DNA which remains localized in the cytoplasm remains to be defined. It is not clearly established whether the viral DNA present in the nuclear fraction replicates and, if such replication occurs, whether viral and(or) host enzymes are involved. Recent experimental results have shown that active

participation of a portion of the host-cell nuclear transcriptive system is obligatory in the vaccinia virus replicative cycle. It has been proposed that the host-cell DNA-dependent RNA polymerase II, or a subunit of this enzyme may be necessary for transcription of late virus functions (116). Normal vaccinia DNA replication kinetics and yield of DNA could be measured in enucleated cells or infected cells in which the host-nuclear transcriptive machinery (and possibly other functions) had been rendered non-functional. However, the packaging of this progeny viral DNA into a DNase-insensitive form, as occurs in cells with a functional nucleus, did not occur (112). It remains to be determined whether the viral DNA synthesized when host-cell nuclear functions are interrupted has some aberrant characteristics that preclude its incorporation into developing particles.

CONCLUSIONS

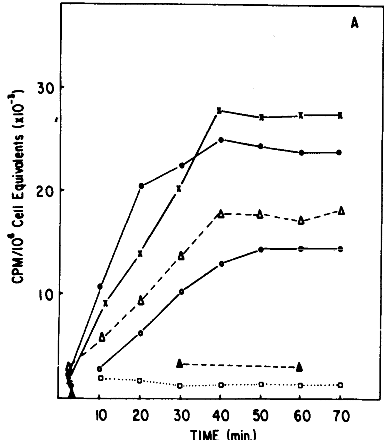
Directions for the Future

As our knowledge has increased, similarities between the structure and replication of vaccinia virus DNA and the genetic material from eukaryotic cells have become apparent. It is now reasonable to predict that an understanding of vaccinia virus DNA replication will provide important insights into the manner in which eukaryotes replicate their chromosomal DNA and how this process is controlled.

While substantial progress has been made in understanding poxvirus DNA replication, it is obvious from the experimental results reviewed above, that much remains to be done. For example, the exact origin of vaccinia virus DNA replication remains to be critically defined. Hybridization experiments (80) as well as structural considerations argue for a site at one or both ends of the genome, within or immediately adjacent to the large, inverted terminal repeats. The sequence organization within the inverted terminal repeats has now been defined (26) and restriction enzyme fragments which encompass this structure have been cloned (24). If DNA synthesis is initiated in this region of

the genome both the information and molecular reagents required to define the origin of DNA replication are now available and experiments to obtain an answer to this problem should soon be undertaken.

An aim of a number of investigators is to reproduce the replication of vaccinia virus DNA in vitro (118,119). These experiments have been only partially successful; while DNA chains were completed in vitro, the initiation of new rounds of DNA replication could not be measured. Investigators have thus far concentrated on employing a virosomal-like fraction prepared from infected cells for in vitro DNA synthesis. While DNA chains may be completed in association with the virosomal fraction, a membrane-DNA complex may be required if initiation events are to be studied in vitro. Recognizing that the topological arrangement of viral DNA genomes within the cytoplasm of infected cells may be critical in their functioning during DNA replication we have begun to develop an in vitro system where the integrity of the structure of infected cells is partially preserved and yet the incorporation of labeled, deoxytriphosphates into replicating viral DNA can be studied. Vaccinia virus—infected HeLa cells, permeabilized with lysolecithin and placed in an appropriate in vitro reaction mixture (120) actively incorporate labeled deoxytriphosphates into DNA being synthesized in their cytoplasmic fraction [Figure 5, Panel A]. The synchrony of the infection is maintained by including HU ($5x10^{-3}$M) in the culture medium and the buffers employed during cell processing. Synthesis of DNA begins as soon as the permeabilized cells are resuspended in the in vitro reaction mixture (120) and the kinetics of synthesis are not affected by inhibitors of protein synthesis [Figure 5, Panel A]. In Figure 5, Panel B are shown the results obtained when viral DNA synthesis was measured in vivo in the presence of an inhibitor of protein synthesis. Only one round of DNA replication occurred (77), demonstrating the need for continuous protein synthesis, if additional rounds of DNA replication are to be initiated.



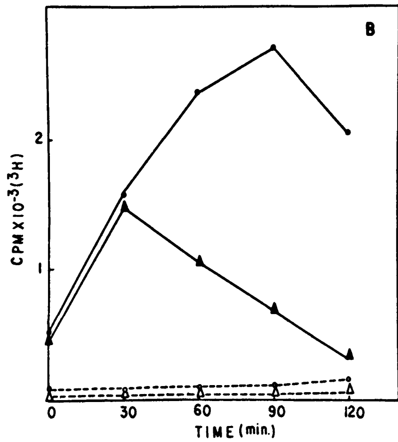


Figure 5. Vaccinia virus DNA synthesis in permeabilized cells *in vitro* compared to viral DNA synthesis *in vivo* in the presence and absence of actidione. Panel A. HeLa cells were infected with vaccinia virus (500 EB/cell). At the end of the adsorption period, the cells were diluted into medium containing 5 mM HU. Four hr later, the cells were processed, permeabilized with lysolecithin and $2x10^7$ permeable cells added to an *in vitro* DNA synthesis reaction mixture as described previously (120). The incorporation of [^{3}H]-TTP into acid-insoluble product with time after *in vitro* synthesis was initiated, is indicated. □--□, no cells added to reaction mixture; x--x unfractionated, infected cells, in the presence of actidione (100 μg/ml) ●--● unfractionated, infected cells, no actidione; △--△ cytoplasmic fraction of infected cells, with or without actidione; O--O unfractionated, mock-infected cells ▲---▲, cytoplasmic fraction of mock-infected cells. Panel B. Vaccinia virus infected HeLa cells were diluted into medium containing 5 mM HU. Four hr later the cells were collected by centrifugation, washed to remove the HU and then resuspended in fresh medium containing [^{3}H]-thymidine (2 μCi/ml), with or without actidione (100 μg/ml). At the times indicated, $2x10^5$ cells were removed and the incorporation of [^{3}H]-thymidine into acid-insoluble product assayed. ●--● infected cells, no actidione; ▲--▲ infected cells in the presence of actidione. O--O mock-infected cells, no actidione; △--△ mock-infected cells in the presence of actidione.

The kinetics and time period of viral DNA synthesis measured in the permeabilized cells *in vitro*, were similar to these events measured *in vivo*, in the presence of inhibitors of protein synthesis. This suggests that only one round of viral DNA replication occurs in the permeabilized cells under the *in vitro* conditions described (unpublished observations). We are presently attempting to supplement our system with the precursors and reagents required for *in vitro* protein synthesis (120) as well as those required for DNA synthesis to determine if the period of active viral DNA synthesis can be extended in the permeabilized cell system. Even with the limitation of being able to analyze only a single-round of viral DNA replication, the ability to incorporate [^{32}P]-labeled deoxy- and ribotriphosphates into replicating viral DNA molecules should permit a detailed analysis of the replication intermediates involved in this process. This is an important first step in developing a cell-free system with which to study viral DNA replication.

With the world-wide eradication of smallpox (121) has come the call to destroy, or severely limit the availability, of stocks of variola virus. It was possible that important questions concerning the origins of variola virus and its relationship to other orthopoxviruses, the pathogenesis of the disease caused by this virus and factors which determine host-range, would never be addressed. Fortunately, Dr. J. Esposito (Communicable Disease Center, Atlanta, Georgia) has undertaken the task of cloning restriction endonuclease fragments of variola virus. This will preserve the genetic material for study. In particular it would be useful to sequence the terminal endonuclease fragments obtained from variola virus DNA and compare the sequence arrangement with that reported for vaccinia virus (26). While some members of the orthopoxvirus genus may replicate their genome in a unique fashion, it is hoped that studies with vaccinia virus will provide information which will be applicable to all members of this genus and perhaps all the *Poxviridae*. Variola virus DNA does not appear to contain large inverted

terminal repeats (9) and the requirement for circularization of orthopoxvirus genomes during replication has been challenged, in part, because of this exception. Sequence analysis of "terminal fragments" from variola DNA would resolve this issue and could provide information which would contribute to our understanding the unique host range of this virus.

Finally, as noted above, there is a need to develop new methods for fractionating infected cell cytoplasm in a manner which will preserve the association of viral DNA with membranes. Experiments should be undertaken to further purify virosomes and to define the enzymatic components of this complex. Methods which have proven useful for the isolation and analysis of nuclear matrices (122,123) could prove productive for this purpose. Similarly, the membrane viral DNA complexes present in infected cell cytoplasm should be purified and their structural as well as enzymatic components carefully analyzed. Understanding the nature of these membrane DNA complexes and how they participate in viral DNA replication appears to be essential if we are to define how viral DNA synthesis is initiated and controlled.

ACKNOWLEDGMENTS

Work described in this review from the authors' laboratory was supported by a U.S. Public Health Service Research Grant (CA-11027) and by a National Research Service Award Institutional Grant (CA-09069) from the National Cancer Institute. The author was a recipient of a U.S. Public Health Service Research Career Development Award (CA-70458) from the National Cancer Institute. The author gratefully acknowledges the technical assistance of M. Bucolo in the laboratory, V.L. Thomas and L. Flores for providing technical expertise in the electron microscopy studies, and M. Icangelo for her help in preparing this manuscript.

REFERENCES

1. Fenner, F. Intervirology 7: 27-31 (1967).
2. Fenner, F., Pereira, H.G., Porterfield, J.S., Joklik, W.K., and Downie, A.W. Intervirology 3: 193-198 (1974).
3. Matthews, R.E.F. Intervirology 12: 160-164 (1979).
4. Pfau, C.J. and McCrea, J.F. Virology 21: 425-435 (1963).
5. Joklik, W.K. J. Mol. Biol. 5: 265-274 (1962).
6. Muller, H.K., Wittek, R., Schaffner, W., Schumperli, D., Menna, A., Wyler, R. J. Gen. Virol. 38: 135-145 (1977).
7. Gafford, L.G., Mitchell, B.E., and Randall, C.C. Virology 89: 229-239 (1978).
8. Wittek, R., Menna, A., Schumperli, D., Stoffel, S., Muller, H.K., and Wyler, R. J. Virol. 23: 669-678 (1977).
9. Mackett, M., and Archard, L.C. J. Gen. Virol. 45: 683-701 (1979).
10. Menna, A., Wittek, R., Bachmann, P.A., Mayr, A., and Wyler, R. Arch. Virol. 59: 145-156 (1979).
11. Thomas, V., Flores, L., and Holowczak, J.A. J. Virol. 34: 244-255 (1980).
12. Subak-Sharpe, H., Burke, R.R., Crawford, L.V., Morrison, J.M., Hay. J., Keir, H.M. Cold Spring Harbor Symp. Quant. Biology 31: 734-748 (1966).
13. Langridge, W.H.R., Bozarth, R.F., and Roberts, D.W. Virology 76: 616-620 (1977).
14. Wittek, R., Kuenzle, C.C., and Wyler, R. J. Gen. Virol. 43: 231-234 (1979).
15. Fenner, F., and Burnet, F.M. Virology 4: 305-314 (1957).
16. Fenner, F. Virology 5: 502-529 (1958).
17. Joklik, W. K. Bacteriol. Rev. 30: 33-66 (1966).
18. McAuslan, B.R. In The Biochemistry of Viruses (Ed. H.B. Levy), Marcel Dekker, New York, 1969, pp. 361-413.
19. Moss, B. In Comprehensive Virology, Vol. 3 (Eds. H. Frankel-Conrat, R.R. Wagner), Plenum Press, New York, 1974, pp. 405-474.
20. Moss, B., In The Molecular Biology of Animal Viruses, II, (Ed. D.P. Nayak), Marcel Dekker, Inc., New York, 1978, pp. 849-890.
21. Berns, K.I., and Silverman, C. J. Virol. 5: 299-304 (1970).
22. Geshelin, P., and Berns, K.I. J. Mol. Biol. 88: 785-796 (1974).
23. Wittek, R., Menna, A., Muller, H.K., Schumperli, D., Bosely, P.G., and Wyler, R. J. Virol. 28: 171-181 (1978).
24. Garon, C.F., Barbosa, E., Moss, B. Proc. Natl. Acad. Sci. USA, 75: 4863-4867 (1978).
25. Wittek, R., and Moss, B. Cell 21: 277-284 (1980).
26. Bahige, M.B., Sundararajan, V., and Moss, B. Cell 28: 315-324 (1982).
27. Venkatesan, S., and Moss, B. J. Virol. 37: 738-747 (1981).

28. Wittek, R., Cooper, J., and Moss, B. J. Virol. 39: 722-733 (1981).
29. Cooper, J.A., Wittek, R., and Moss, B. J. Virol. 39: 733-745 (1981).
30. Moss, B., Winters, E., and Cooper, J.A. J. Virol. 40: 387-395 (1981).
31. Parr, R.P., Burnett, J.W., and Garon, C.F. Virology 81: 247-256 (1977).
32. Ortin, J., Enjuanes, L., and Vinuela, E. J. Virol. 31: 579-583 (1979).
33. Wittek, R., Muller, H.K., Menna, A., and Wyler, R. FEBS Lett. 90: 41-46 (1978).
34. Moss, B., Winters, E., and Cooper, N. Proc. Natl. Acad. Sci. USA 78: 1614-1618 (1981).
35. McFadden, G., and Dales, S. Cell 18: 101-108 (1979).
36. Panicali, D., Davis, S.W., Mercer, S.R., and Paoletti, E. J. Virol. 37: 1000-1010 (1981).
37. Drillien, R., Koehren, F., and Kirn, A. Virology 111: 488-499 (1981).
38. Lake, J.R., and Cooper, P.D. J. Gen. Virol. 48: 135-147 (1980).
39. Moyer, R.W., and Rothe, C.T. Virology 102: 119-132 (1980).
40. Moyer, R.W., Graves, R.L., and Rothe, C.T. Cell 22: 545-553 (1980).
41. Esposito, J.J., Cabradilla, C.D., Nakano, J.H., and Obijeski, J.F. Virology 109: 231-243 (1981).
42. Wittek, R. Experientia 38: 285-410 (1982).
43. Craigie, J. Br. J. Exptl. Pathol. 13: 259-264 (1932).
44. Smadel, J.E., and Hoagland, C.L. Bacteriol. Rev. 6: 79-110 (1942).
45. Joklik, W.K. Biochim. Biophys. Acta 61: 290-301 (1962).
46. Zwartouw, H.T., Westwood, J.C.N., Appleyard, G. J. Gen. Microbiol. 29: 523-529 (1962).
47. Planterose, D.N.C., Nishimura, C., and Salzman, N.P. Virology 18: 294-301 (1962).
48. Pirie, G.D., Bishop, P.M., Burke, D.C., and Postlethwaite, R. J. Gen. Virol. 13: 311-320 (1971).
49. Sarov, I., and Becker, Y. Virology 33: 364-373 (1967).
50. Smith, K.O., and Sharp, D.G. Virology 11: 519-528 (1960).
51. Easterbrook, K.B. Virology 15: 404-416 (1961).
52. Woodson, B., and Joklik, W.K. Proc. Natl. Acad. Sci. USA 54: 946-953 (1965).
53. Galasso, G.J., and Sharp, D.G. Proc. Soc. Exptl. Biol. Med. 113: 43-47 (1961).
54. Joklik, W.K. J. Mol. Biol. 8: 263-276 (1964).
55. Rosenkranz, H.S., Rose, H.M., Morgan, C., and Hsu, K.C. Virology 28: 510-519 (1966).
56. Pogo, B.G.T., and Dales, S. Virology 43: 144-151 (1971).
57. Fil, W., Holowczak, J.A., Flores, L., and Thomas, V. Virology 61: 376-396 (1974).
58. Wehrli, W., and Staehelin, M. Bacteriol. Rev. 35: 290-309 (1971).

59. Heller E., Argaman, M., Levy, H., and Goldblum, N. Nature (London) 222: 273-274 (1969).
60. Moss, B., Katz, E., and Rosenblum, E.M. Biochem. Biophys. Res. Commun. 36: 858-865 (1969).
61. Moss, B., Rosenblum, E.N., Katz, E., and Grimley, P.M. Nature (London) 224: 1280-1284 (1969).
62. Grimley, P.M., Rosenblum, E.N., Mims, S.J., and Moss, B. J. Virol. 6: 519-533 (1970).
63. Katz, E., Grimley, P.M., and Moss, B. Nature (London) 227: 1050-1051 (1970).
64. Pennington, T.H., Follet, E.A., and Szilagyi, J.F. J. Gen. Virol. 9: 225-237 (1970).
65. Tan, K.B., and McAuslan, B.R. J. Virol. 6: 326-332 (1970).
66. Moss, B., Rosenblum, E.N., and Grimley, P.M. Virology 45: 123-134 (1971).
67. Esteban, M., and Holowczak, J.A. Virology 86: 376-390 (1978).
68. Penman, S., Scherrer, K., Becker, Y., and Darnell, J.E. Proc. Natl. Acad. Sci. USA 49: 654-662 (1963).
69. Borun, T.W., Scharff, M.D., and Robbins, E. Biochim. Biophys. Acta 149: 302-304 (1967).
70. Penman, S. In Fundamental Techniques in Virology (Eds. K. Habel and N.P. Salzman), Academic Press, New York, 1969, pp. 35-48.
71. Pfau, C.J., and McCrea, J.F. Nature (London) 194: 894:895 (1962).
72. Joklik, W.K. Virology 18: 9-18 (1962).
73. Holowczak, J.A. Virology 72: 121-133 (1976).
74. Holowczak, J.A., and Diamond, L. Virology 72: 134-146 (1976).
75. Esteban, M., and Holowczak, J.A. Virology 78: 57-75 (1977).
76. Esteban, M. J. Virol. 21: 796-801 (1977).
77. Esteban, M., and Holowczak, J.A. Virology 82: 163-181 (1977).
78. Pogo, B.G.T. Proc. Natl. Acad. Sci. USA 74: 1739-1742 (1977).
79. Pogo, B.G.T., and O'Shea, M.T. Virology 84: 1-8 (1978).
80. Pogo, B.G.T., O'Shea, M., Freimuth, P. Virology 108: 241-248 (1981).
81. Pogo, B.G.T. Virology 100: 339-347 (1980).
82. Moyer, R.W., and Graves, R.L. Cell 27: 391-401 (1981).
83. Pogo, B.G.T., and Dales, S. Proc. Natl. Acad. Sci. USA 63: 820-827 (1969).
84. Green, M. Cold Spring Harbor Symp. Quant. Biol. 27: 219-233 (1963).
85. Hanafusa, T. Biken J. 4: 97-110 (1961).
86. Joklik, W.K., and Becker, Y. J. Mol. Biol. 10: 452-474 (1964).
87. Magee, W.E., Shuk, M.R., and Burrows, M.J. Virology 11: 296-299 (1960).
88. Salzman, N.P. Virology 10: 150-152 (1960).
89. Shatkin, A.J., and Salzman, N.P. Virology 19: 551-560 (1963).

90. Magee, W.E., and Levine, S. Ann. N.Y. Acad. Sci. 173: 362-378 (1970).
91. Huberman, J.A., and Riggs, A.D. J. Mol. Biol. 32: 327-341 (1968).
92. Tattersall, P., and Ward, D.C. Nature (London) 263: 106-109 (1976).
93. Esteban, M., Flores, L. and Holowczak, J.A. Virology 83: 467-473 (1977).
94. Cairns, J. Virology 11: 603-623 (1966).
95. Kato, S., Ogawa, M., and Myamoto, H. Biken J. 7: 45-56 (1964).
96. Kato, S., Takahashi, M., Kameyama, S., and Kamahora, J. Biken J. 2: 353-363 (1959).
97. Dales, S., and Siminovitch, L. J. Biophys. Biochim. Cytol. 10: 475-502 (1961).
98. Dahl, R., and Kates, J. Virology 42: 453-462 (1970).
99. Minocha, H.C., and Sahu, S.P. Am. J. Vet. Res. 36: 1789- 1792 (1975).
100. Polisky, B., and Kates, J. Virology 49: 168-179 (1972).
101. Polisky, B., and Kates, J. Virology 66: 128-139 (1975).
102. Sarov, I., and Joklik, W.K. Virology 52: 223-233 (1973).
103. Dahl, R., and Kates, J. Virology 42: 463-472 (1970).
104. Joklik, W.K. J. Mol. Biol. 8: 277-288 (1964).
105. Kates, J.R., and McAuslan, B.R. Proc. Natl. Acad. Sci. USA 58: 131-141 (1967).
106. Munyon, W., Paoletti, E., and Grace, J.T. Proc. Natl. Acad. Sci. USA 58: 2280-2287 (1967).
107. Pogo, B.G.T. Virology 101: 520-524 (1980).
108. Holowczak, J.A. Current Topics in Micro. and Immunol. 97: 28-79 (1982).
109. Dales, S., and Mosbach, E.H. Virology 35: 564-583 (1968).
110. Gafford, L.G., and Randall, C. Virology 69: 1-14 (1976).
111. Hardy, C.L., Randall, C.C., and Gafford, L.G. Virology 84: 236-241 (1976).
112. Hruby, D.E., Lynn, D.L., and Kates, J.R. Proc. Natl. Acad. Sci. USA 76: 1887-1890 (1979).
113. LaColla, P., and Weisbach, A. J. Virology 15: 305-315 (1975).
114. Taylor, J.L. and Rouhandeh, H. Biochem. Biophys. Acta 478: 59-67 (1977).
115. Walen, K.H. Proc. Natl. Acad. Sci. USA 68: 165-168 (1971).
116. Silver, M., and Dales, S. Virology 118: 214-218 (1982).
117. Williams, R. Proc. Natl. Acad. Sci. USA 74: 2311-2315 (1977).
118. Esteban, M. , and Holowczak, J.A. Virology 78: 76-86 (1977).
119. Lambert, D.M., and Magee, W.E. Virology 79: 342-359 (1977).

120. Miller, M.R., Castellot, J.J., and Pardee, A.B. Biochemistry 17: 1073-1080 (1978).
121. Arita, I. Nature (London) 279: 293-298 (1979).
122. Berezney, R., and Coffey, D.S. J. Cell Biol. 73: 616-637 (1977).
123. Berezney, R., and Buchholtz, L.A. Biochemistry 20: 4995-5002 (1981).

11

REPLICATION OF ADENOVIRUS DNA

J.S. SUSSENBACH AND P.C. VAN DER VLIET

1. SUMMARY

The replication of linear double-stranded adenovirus DNA starts at the molecular termini and employs a unique priming mechanism. During the elongation of the newly initiated daughter strands displacement of the parental strands takes place. We have investigated the stages of initiation and elongation in an *in vitro* DNA replication system and have analyzed the functions of two viral-coded proteins (terminal protein and DNA-binding protein) and a host factor (DNA polymerase). Evidence is presented that in the initiation step the first nucleotide of the new DNA strand is covalently linked to a precursor of the terminal protein. This step does not require the presence of DNA-binding protein. However, DNA-binding protein as well as DNA polymerase are required for the elongation step. Analysis of the gene coding for the DNA-binding protein of some mutants and revertants indicates that in particular the C-terminal part of the DNA-binding protein molecule is involved in DNA replication. These results are discussed in relation to a model for adenovirus DNA replication.

2. INTRODUCTION

The group of human adenoviruses consists of 33 different serotypes, some of which can induce tumors in newborn rodents. In their natural host these viruses cause respiratory diseases (1). Most research on the molecular biology of adenoviruses has been performed with the serotypes 2 and 5 (Ad2 and Ad5).

The adenovirion has an icosahedral structure and contains a linear double-stranded DNA molecule with a molecular weight of 23 million daltons (2, 3). The 5'-termini of the two complementary strands are covalently linked to a protein with a molecular weight of 55 kD (terminal protein, TP) (4, 5). Ad2 and Ad5 DNA contain an inverted terminal repetition 102-103 nucleotides long (6, 7).

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

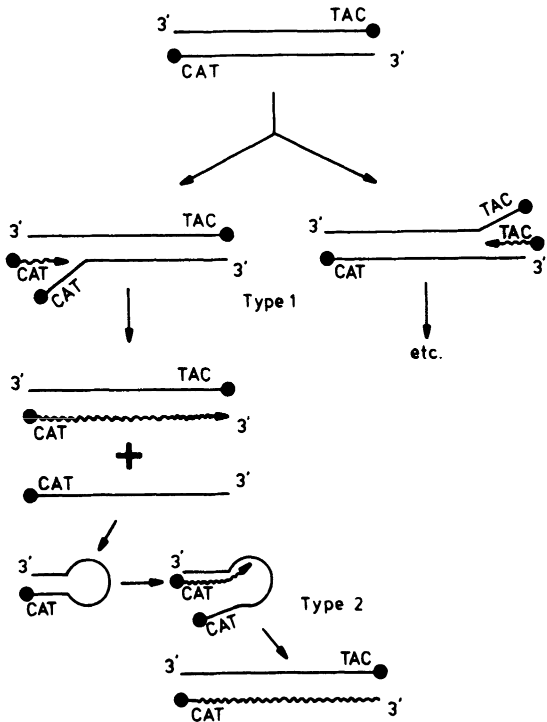


FIGURE 1. Schematic representation of adenovirus DNA replication. Parental strands are indicated by drawn lines and newly synthesized strands by wavy lines. The black dots represent the terminal protein. The inverted terminal repetition is symbolically indicated by the first nucleotides of the terminal repetition CAT. Replication might start at either molecular end. Only the complete replication cycle for initiation at the left-hand end has been drawn. Replication which starts at the right-hand end proceeds similarly. The branched intermediates are called Type 1 molecules, while unbranched intermediates are designated as Type 2 molecules.

The adenovirus genome may code for 30-50 different proteins. A major part of the organization of the viral genome has already been elucidated (1).

A few hours after infection of cells by adenovirus, five to six specific regions of the viral genome are expressed before the onset of viral DNA replication. These regions code for a number of proteins, at least two of which are involved in viral DNA replication (DNA-binding protein (DBP) and TP) (8, 9, 10). About 6 hr pi replication of viral DNA starts and proceeds for at least 20 hr. At the end of the infection cycle the infected cell

contains about 10^5-10^6 copies of the viral genome.

Biochemical and electron-microscopic investigations of replicating adenovirus DNA have revealed that adenovirus DNA replicates via a displacement mechanism (1, 11-14). Replication may start at each of the two molecular termini and proceeds by displacement of the parental strands. It has been shown that replication starts within the inverted terminal repetition at the very ends of the DNA molecule (10, 15). After completion of the displacement synthesis, the displaced parental strands are converted into duplex molecules via complementary strand synthesis (Fig. 1).

One of the most intriguing stages of the DNA replication cycle is the initiation step. All known DNA polymerases require a free 3'-OH group to start DNA synthesis. In many systems these 3'-OH groups are provided by short RNA molecules, which are used as primers of DNA replication. Due to the circular structure of many DNA molecules the replication fork passes the origin of DNA replication after completion of each round of replication and the RNA primers can be replaced by DNA during this event.

However, since adenovirus DNA replication proceeds via linear replicative intermediates, an RNA priming mechanism cannot be operational in this case. Several models for the initiation of adenovirus DNA replication have been proposed, including a self-priming mechanism, but none of these appeared to be correct (16, 17). In 1977 Rekosh *et al.* (15) proposed that adenovirus DNA replication might start with the terminal protein as primer molecule. At that time no examples of protein priming were known. However, the exploration of an *in vitro* DNA replication system has provided evidence that this hypothetical priming mechanism might be correct.

In this report further evidence for a protein-priming mechanism will be presented. In addition, new information on the functions of a number of proteins involved in adenovirus DNA replication will be provided, especially on the DNA-binding protein.

3. ADENOVIRUS DNA SYNTHESIS *IN VITRO*

Three systems have been developed for *in vitro* DNA replication. These are, in decreasing order of complexity (3.1.) isolated nuclei, (3.2.) replication complexes and (3.3.) nuclear extracts.

3.1. Isolated nuclei

Isolated nuclei of Ad5 infected human cells can elongate endogenous

replicative intermediates that have initiated viral DNA replication in intact cells (18). This type of synthesis seems to reflect *in vivo* DNA replication in many respects but isolated nuclei are defective in initiation.

Biochemical and electron-microscopical characterization of replicative intermediates from isolated nuclei have been of great use in unravelling the mechanism of elongation (11). Moreover, since isolated nuclei are permeable to immunoglobulins it is possible to investigate the effect of antibodies to replication proteins on viral DNA synthesis. Using this approach we have observed an inhibiting effect of anti-DNA binding protein. This result provided the first evidence for a role of DBP in elongation (9).

Recently, we have studied the effect of ATP on Ad5 DNA synthesis in isolated nuclei. In contrast to cellular DNA synthesis, Ad5 DNA synthesis appeared only weakly sensitive to low ATP concentrations. In the presence of endogenous (10^{-7} M) ATP levels normal chain elongation was observed and only a two-fold stimulation was observed at 2 mM ATP or dATP. GTP, which can replace ATP in initiation, was not effective in elongation. This may indicate that extensive ATP hydrolysis as performed by DNA-dependent ATPases (19) is not required for adenovirus DNA chain elongation.

3.2. Replication complexes

Replication complexes have been extracted from infected cell nuclei by extraction with 150-200 mM ammonium sulfate (20, 21) or with sodium heparin (22). These complexes are composed in part of replicative intermediates complexed with various proteins and can perform an elongation reaction.

3.3. Nuclear extracts

In 1979 Challberg and Kelly (10) prepared a nuclear extract in which initiation and elongation on exogenous templates could occur. True replication starts were only observed when DNA complexed with terminal protein (DNA-TP) was used. About 6% of the input template could complete one round of displacement synthesis.

In subsequent studies it was demonstrated that newly replicated DNA contains a 80 kD precursor of TP (pTP) which is covalently bound to the new DNA by a serine-dCMP phosphodiester bond(23, 24). The pTP is processed to mature TP only late in infection (25). The presence of protein on replica-

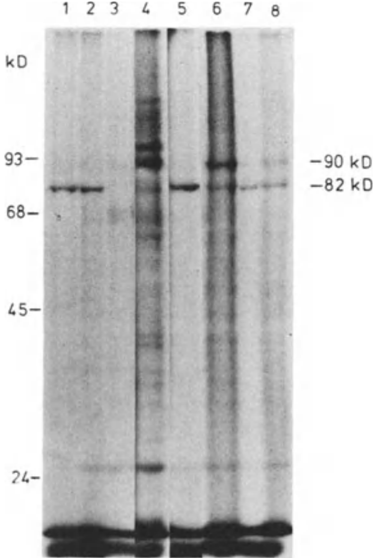


FIGURE 2. Formation of a covalent complex between dCMP and preterminal protein is required for initiation of DNA synthesis.
Nuclear extracts were prepared from Ad5-infected HeLa cells as described (10). A reaction mixture of 0.1 ml, containing 40 mM Hepes-KOH pH 7.5 - 4 mM $MgCl_2$ - 2 mM ATP - 0.8 mM dithiothreitol - 4 mM creatine phosphate - 4 μg/ml creatine kinase - 40 μM ddATP - 1 μM α-^{32}P-dCTP (100 Ci/mmole) - 0.1 mM aphidicolin, 0.2 μg Ad5 DNA-TP and 23 μl nuclear extract was incubated for 30 min at 37° C. In some cases ddATP was substituted by 40 μM dATP, 40 μM dTTP and 40 μM ddGTP to permit a limited elongation reaction (elongation conditions). The reaction was stopped by addition of 25 μl 0.25 M sodium pyrophosphate and 5 μl 0.2 M EDTA. TCA-insoluble material was electrophoresed on 10% polyacrylamide - 0.8% SDS gels in the presence of appropriate markers, stained, dried and autoradiographed. Lanes 1-5, initiation conditions. Lane 1, standard incubation, lane 2, 40 μM ddCTP added; lane 3, 0.2 μg Ad5 DNA replacing Ad5 DNA-TP; lane 4, minus ATP, minus energy-regenerating system; lane 5, as lane 4, + 2 mM GTP; lane 6, standard elongation conditions; lane 7, initiation, 15 min pulse; lane 8, pulse chase: After a 15 min pulse dATP, dTTP, dCTP (100 μM) and ddGTP were added and the incubation was continued for 60 min.

tive intermediates was previously observed *in vivo* (26-29).

We have used this system to study the requirements for initiation and elongation separately. The formation of the pTP-dCMP initiation complex can be demonstrated directly by incubation with α-^{32}P-dCTP in the presence of DNA-TP as a cofactor (24, 30, 31). Fig. 2, lane 1 shows an electropherogram of the products formed. One prominent labeled band at 82 kD,

representing pTP-dCMP, can be observed. A weaker band at 25 kD is also present but is not dependent on the presence of DNA-TP. This product has not been further characterized. The formation of the pTP-dCMP complex is dependent on the presence of ATP (lane 4) and is not inhibited by ddCTP, an inhibitor of chain elongation (lane 2). ATP can be substituted by dATP, GTP (lane 5) or dGTP and to a less extent by pyrimidine triphosphates.

Is the formation of a pTP-dCMP complex an essential step in the DNA synthesis? To determine this we have performed the following pulse-chase experiment. A nuclear extract was incubated with DNA-TP and α-^{32}P-dCTP for 15 min. Then an excess of dCTP (100 μM), dATP, dTTP and ddGTP was added and the mixture incubated another 30 min. Under these conditions (Fig. 2, lane 6) limited DNA replication can occur until nucleotide 26, the first G in the 5'-terminated strand (6). This leads to the formation of a 90 kD product consisting of the pTP and a 26 nucleotide long oligonucleotide (24, Van der Vliet,unpublished). As can be seen from Fig. 2 (lanes 7 and 8) the 82 kD initiation complex can be chased in part to the 90 kD partial elongation complex. This result provides strong evidence for a role of the pTP-dCMP complex as primer for adenovirus DNA synthesis.

4. ADENOVIRUS DNA-BINDING PROTEIN

4.1. Role of the adenovirus DNA-binding protein in replication

At early times after infection large amounts of the 72 kD DBP are synthesized. This protein has an anisometric structure and is rapidly phosphorylated after synthesis (32-34). It binds co-operatively to single-stranded DNA and less efficiently to the ends of double-stranded DNA (33, 35). Detailed analysis of the phenotype of a number of DBP mutants has revealed that the DBP is multifunctional. Analysis of H5ts107, H5ts125 and H2ND$_1$ts23 has shown that these temperature-sensitive mutants are defective in initiation of DNA replication at the non-permissive temperature (8, 63). In addition a role of the DBP in chain elongation has been assumed based upon inhibition of DNA synthesis by anti-DBP IgG (9) and temperature-sensitive elongation in replication complexes (36).

At which step in replication is DBP essential? To determine this we analyzed the formation of an initiation complex and a partial elongation complex in extracts from H5ts125 infected cells. The extracts were obtained from cells infected at 32° C and were preincubated at 32° C or 37° C. As shown in Fig. 3, lanes 1-4, the synthesis of the 82 kD pTP-

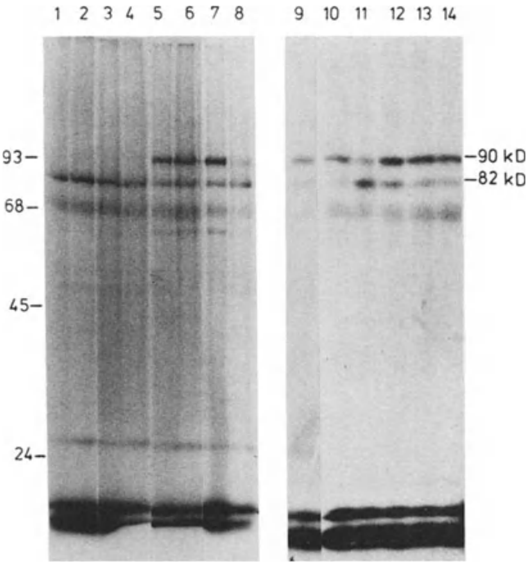


FIGURE 3. DBP is not required for the formation of a covalent pTP-dCMP complex *in vitro*.
HeLa cells were infected with wild type (wt) or H5ts125 at a moi of 10,000 particles per cell. The cells were grown for 40 hr at 32° C in the presence of 5 mM hydroxyurea present from 4-40 hr pi. Nuclei were isolated and the nuclear extract was preincubated for 15 min at 32° C or 37° C after which the reaction was initiated by addition of DNA-TP. Lanes 1-4, initiation conditions, lanes 5-14, elongation conditions. Lanes 1-8 and 9-14 represent two different experiments.
1, 5 wt, 32°   9 wt, 37°
2, 6 wt, 37°   10 H5ts125, 32°
3, 7 H5ts125, 32°   11 H5ts125, 37°
4, 8 H5ts125, 37°   12-14 H5ts125, 37° to which had been added 1.2 µg, 2.4 µg and 3.6 µg purified DBP during preincubation and incubation, respectively.

dCMP complex was not temperature-sensitive. However, the formation of the 90 kD product was strongly inhibited at 37° C in H5ts125 extracts (lanes 4-8). This inhibition could be suppressed by incubation of the extracts at 37° C in the presence of purified wt-DBP (lanes 9-14). Similar results were recently reported by Challberg *et al.* (30). This indicates that the DBP is not required, or only required at a very low concentration, for the covalent attachment of dCMP to the preterminal protein but is essential already in the early stages of elongation.

4.2. The relation between the structure and function of the DNA-binding protein

In a first step to unravel the relation between the structure and the multiple functions of the DBP molecule we have determined the complete nucleotide sequence of the Ad5 DBP gene and have examined the structure of DBP mRNA species (37). From the nucleotide sequence of the mRNA molecules the complete amino acid sequence of DBP could be derived (Fig. 4). DBP appears to be 529 amino acid residues long and has a molecular weight of 59 kD. The overall amino acid composition of DBP indicates that the molecule has a neutral character. This is in agreement with its isoelectric point of 7.5, after removal of the phosphate groups (38).

Recently we have also determined the complete nucleotide sequence of the Ad2 DBP gene and have derived the amino acid sequence of Ad2 DBP from the nucleotide sequence. In comparison with the amino acid sequence of Ad5 DBP the Ad2 sequence displays nine amino acid differences (Fig. 4). All these differences are located in the N-terminal region of the DBP molecule. Since no differences are found in the C-terminal part of the molecule, this suggests that relatively less alterations are allowed in this region of the molecule without loss of functions of DBP.

In order to obtain further information on the functional aspects of the different regions of the DBP molecule, we have mapped at the nucleotide level the positions of the mutations in a number of DBP mutants and revertants (Fig. 4).

The only mutant located in the N-terminal part of DBP is H5hr404 (amino acid residue 130). This mutant has normal DNA replication, but displays aberrant late gene expression resulting in the ability to grow in monkey cells (39). The position of the H5hr404 mutation and the phenotype of this mutant suggests that the N-terminal part of the DBP molecule is not directly involved in DNA replication, but rather in late gene expression.

Three mutants with temperature-sensitive DNA replication (H5ts107, H5ts125 and H2ND$_1$ts23) (personal communication: H.S. Ginsberg and J.F. Williams; 40, 63) are all located in the C-terminal part of the DBP molecule. H5ts107 and H5ts125 appear to carry the same mutation at amino acid position 413. In two revertants [r(ts125)2 and r(ts107)127] the amino acid residue at position 413 has been reverted to the original wild type amino acid at that position. Since no other changes in the amino acid sequence have been detected, this proves that the amino acid alteration at position

```
                    20                  40                  60                  80                 100
(I)  MASREEEQRETTPERGRGAA RRPPTMEDVSSPSPSPPPPR APPKKRMRRRIESEDEEDSS QDALVPRTPSPRPSTSAADL AIAPKKKKKRPSPKPERPPS
                                                    ↑   ↑                           ↑     ↑
                               Y←hr404              L   L                           T     S
                    120          ↓       140                 160                 180                 200
     PEVIVDSEEEREDVALQMVG FSNPPVLIKHGKGGKRTVRR LNEDDPVARGMRTQEEEEEP SEAESEITVMNPLSVPIVSA WEKGMEAARALMDKYHVDND
                                                               ↑ ↑    ↑  ↑   ↑
                                                               K S    S  I   L      F←ts23
                    220                  240                 260                 280   ↓            300
     LKANFKLLPDQVEALAAVCK TWLNEEHRGLQLTFTSNKTF VTMMGRFLQAYLQSFAEVTY KHHEPTGCALWLHRCAEIEG ELKCLHGSIMINKEHVIEMD

                                                        D←r202
                    320                  340              ↓  360                 380                 400
     VTSENGQRALKEQSSKAKIV KNRWGRNVVQISNTDARCCV HDAACPANQFSGKSCGMFFS EGAKAQVAFKQIKAFMQALY PNAQTGHGHLLMPLRCECNS

     ts107,ts125→S←r13,r202
                ↓   420                  440                 460                 480                 500
     KPGHAPFLGRQLPKLTPFAL SNAEDLDADLISDKSVLASV HHPALIVFQCCNPVYRNSRA QGGGPNCDFKISAPDLLNAL VMVRSLWSENFTELPRMVVP
                ↑
       r13→Y    P←r2,r127
         ↓          520
     EFKWSTKHQYRNVSLPVAHS DARQNPFDF
```

```
                                                                             r13,r202
                      hr404                      ts23            r202      ts107,ts125      r13
                        ↓                          ↓               ↓            ↓             ↓
(II) H2N----------------------------------------------------------------------------------------COOH
                           ████                                                 ↑
                                                                             r2,r127
            25 kD          chymotryptic              45 kD
                           cleavage site
```

FIGURE 4. (I) Complete amino acid sequence of Ad5 DBP. The first amino acid residue represents the amino-terminus. The differences in the Ad2 DBP sequence are indicated below the Ad5 DBP sequence. The amino acid alterations in the temperature-sensitive mutants H2ND_1ts23, H5ts107 and H5ts125, the revertants r(ts125)2, r(ts125)13, r(ts107)127 and r(ts107)202 and the host-range mutant H5hr404 are indicated in the Ad5 sequence. (II) A schematic representation of the Ad5 DBP sequence and the positions of mutants and revertants. The black box indicates the region, within which DBP is supposed to be cleaved by chymotrypsin. The resulting 25 and 45 kD fragments are indicated.

413 in H5ts107 and H5ts125 is indeed responsible for the temperature-sensitive viral DNA replication of these mutants. Mapping of the revertants r(ts125)13 and r(ts107)202 shows that these revertants still contain the original H5ts125 and H5ts107 mutations, but have, in addition, second-site mutations at positions 508 and 352, respectively.

It has been shown before that DBP of H5ts125 is thermolabile in its binding to single-stranded DNA (41). The positions of the mutants and revertants indicate that in particular the C-terminal part of DBP is involved in DNA replication, probably by binding to single-stranded DNA. Since replicating adenovirus DNA contains extended regions of single-stranded DNA, a role of DBP in adenovirus DNA replication is not unexpected (11). DBP might protect these single-stranded regions against nucleolytic degradation or stabilize the single-stranded regions in replicating DNA, thus facilitating migration of the replication fork. Since DBP molecules bind to single-stranded DNA in a co-operative way and therefore have to interact simultaneously with DNA and other DBP molecules, it can be expected that a relatively large part of the DBP molecule is essential for DNA replication. The fact, that the different mutations are scattered over a broad region of the amino acid sequence is in agreement with this notion. A distinct role of the C-terminal part of DBP in DNA replication is also suggested by the observation that the chymotryptic C-terminal 45 kD fragment of wild type DBP is able to complement the defective *in vitro* DNA replication system of H5ts125 (42).

Characterization of the phenotypes of some of the revertants has revealed in addition, that the second-site mutations cause unexpected new properties in these virus strains. R(ts107)202 which grows normally in HeLa cells at the non-permissive temperature does not grow in the adenovirus-transformed 293 cells under these conditions (43). Furthermore, r(ts125)13 exhibits an aberrantly high overproduction of DBP (J.C. Nicolas, personal communication). This illustrates very clearly that, although we have distinguished two different domains in the DBP molecule, the C-terminal part involved in viral DNA replication and the N-terminal part involved in late gene expression, this view is probably oversimplified.

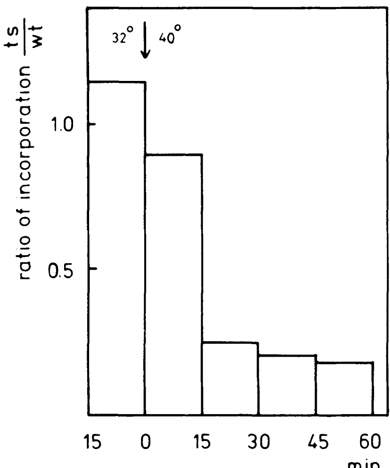
5. OTHER POSSIBLE VIRUS-CODED REPLICATION PROTEINS

In addition to pTP and DBP other viral proteins may be required. One of these proteins could be the product of H5ts36. The H5ts36 mutation, like

pTP, is located in the early region E2B and is involved in viral DNA replication (8, 61). The corresponding protein has not been characterized.

Recently a new class of mutants affected in viral DNA replication was discovered, suggesting that beside DBP and pTP another virus-coded protein might be involved in viral DNA replication. The first representative of this class of mutants is H2ts111, which was originally characterized by d'Halluin *et al.* (44). They concluded that this mutant is affected in the initiation and elongation of viral DNA replication and that viral DNA is degraded at the non-permissive temperature.

In order to investigate this mutant further, we have measured the rate of viral DNA replication of wild type and H2ts111 in shift-up experiments (Fig. 5). If infected cells are grown at the permissive temperature and the temperature is shifted to 40° C a rapid decrease in the rate of DNA



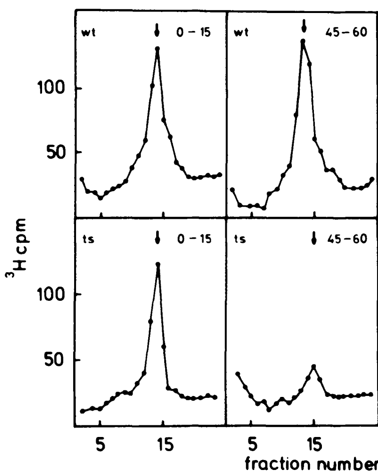


FIGURE 5. Time course of viral DNA synthesis in KB cells infected with H2ts111 or wt virus (moi 2000 particles per cell).
After 40 hr of incubation at 32° C the temperature was shifted to 40° C. DNA synthesis was measured by the incorporation of ^{3}H-thymidine during 15 min pulses ranging from 15 min to 45 min after shift-up. Viral DNA was isolated as described elsewhere (18). The left-hand panel represents the ratio of incorporation of ^{3}H-thymidine in H2ts111- and wt-infected cells. The isolated viral DNA, labeled from 0-15 and 45-60 min, respectively, after the shift to 40° C, was subjected to sucrose gradient centrifugation (right-hand panel). Sedimentation is from right to left. The 31 S position is indicated by the arrow.

synthesis is observed. Sucrose gradient centrifugation of the labeled DNA indicates that the sedimentation profile of replicating H2ts111 DNA labeled 45-60 min after the shift-up is similar to the wild type profile, indicating that no accumulation of replicating intermediates with higher sedimentation co-efficients takes place (Fig. 5). Pulse-chase experiments also indicate that in the first 20 min after shift-up the maturation of replicative intermediates in H2ts111-infected cells is normal (Fig. 6). However, 120 min after shift-up an extensive degradation of labeled DNA can be detected.

Since degradation of viral DNA is not detected before 60 min after shift-up, while the rate of DNA synthesis is already affected after 15 min,

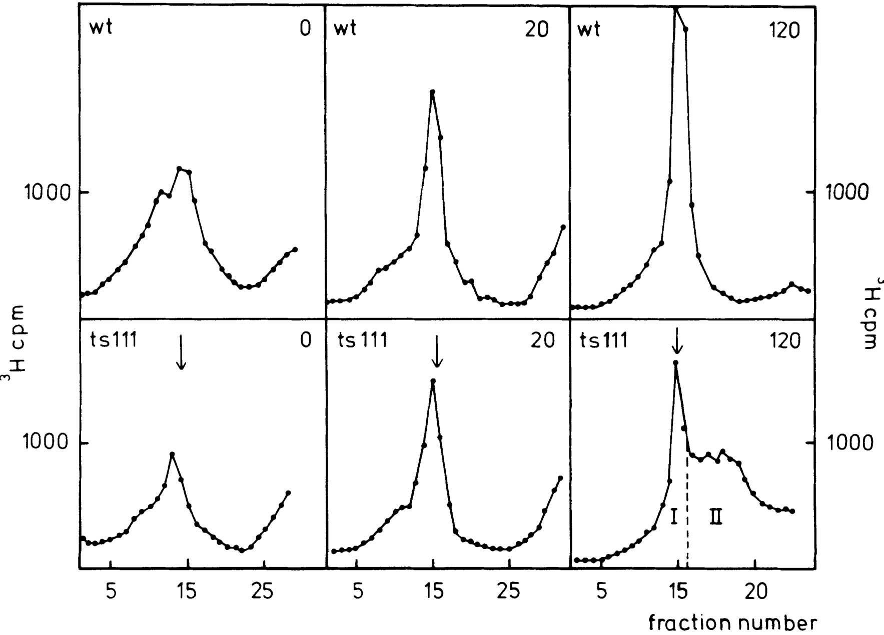


FIGURE 6. Neutral sucrose gradient centrifugation of viral DNA isolated from wt- and H2ts111-infected cells, which were incubated for 40 hr at 32° C and subsequently pulse-labeled for 5 min with ^{3}H-thymidine. The radioactivity incorporated was chased for 0, 20 and 120 min, respectively at the non-permissive temperature. Viral DNA was isolated as described elsewhere (18). Centrifugation is from right to left. The arrows indicate the position of 31 S marker DNA. Note that in H2ts111-infected cells a considerable amount of labeled DNA (region II) is degraded after a 120 min chase.

it is likely that degradation of DNA is a secondary effect. The observation that 30 min after shift-up the rate of viral DNA synthesis is considerably reduced, while the sedimentation profile of replicating DNA is comparable to that of wild type, suggests that in H2ts111-infected cells at the non-permissive temperature elongation of growing strands is normal. Probably only initiation of DNA replication is affected.

However, it should be emphasized that this type of analysis might not be adequate to distinguish between elongation and initiation. Other experiments are required to pin-point precisely the defect in H2ts111. Nevertheless, the rapid decrease of the rate of DNA replication in H2ts111-infected cells after shift-up suggests that probably another virus-coded protein exists, which plays a role in adenovirus DNA replication.

6. ELONGATION AND THE ROLE OF DNA POLYMERASE γ

Adenovirus does not code for its own DNA polymerase and major changes in DNA polymerases have not been observed during infection (45). This suggests that a cellular DNA polymerase is required for adenovirus DNA replication. In endogenous replication complexes both DNA polymerases α and γ, but not DNA polymerase β have been detected (20, 46-49).

To determine which DNA polymerase is involved specific inhibitors have been employed. We found that Ad5 DNA synthesis in isolated nuclei is very sensitive to low concentrations of ddNTP's in contrast to cellular or SV40 DNA replication (50-52, see Fig. 7). A similar sensitivity was observed in replication complexes (47-49) and in nuclear extracts (53, Van der Vliet, unpublished observations).

These results have been taken as evidence for the participation of DNA polymerase γ. This enzyme is more sensitive to ddNTP's than the other DNA polymerases. Moreover, DNA polymerase γ can recognize ddTTP as substrate in contrast to DNA polymerase α (52).

To study a possible additional role of DNA polymerase α another series of experiments were performed employing aphidicolin, a specific inhibitor of this enzyme (54, 55). *In vivo*, aphidicolin at low concentrations inhibits cellular DNA synthesis or SV40 DNA synthesis which agrees well with a functional role of DNA polymerase α in these processes. Adenovirus DNA replication is also inhibited but a 300 to 500-fold higher drug concentration is required to obtain a similar effect as in uninfected cells (56, 57). This difference is not easily explained by different intra-

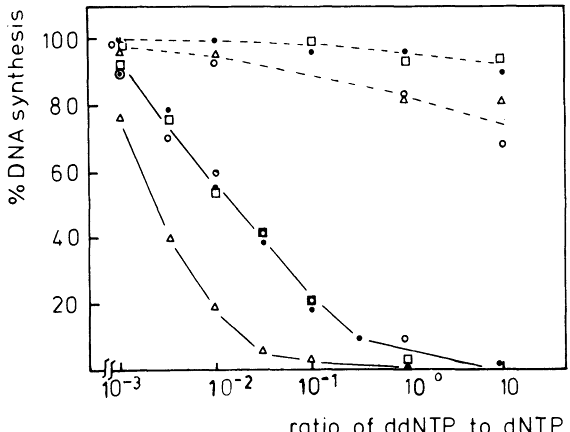


FIGURE 7. Specific inhibition of adenovirus DNA synthesis by 2'-3'-dideoxy-nucleoside 5'-triphosphates.
Nuclei from uninfected (---) or Ad5 infected (——) HeLa cells were incubated in a reaction mixture containing 50 μM dNTP's, 5 μM ^{3}H-dTTP and various concentrations of ddATP (o), ddTTP (•), ddGTP (□) or ddCTP (Δ). When ddTTP was added 50 μM dTTP and 5 μM ^{3}H-dGTP were present. The nuclei were incubated for 30 min and DNA synthesis was assayed as described (9).

cellular pools or metabolic conversion of aphidicolin since single cells, doubly infected with both Ad5 and SV40, displayed also a differential sensitivity (57). This suggests that the intracellular targets for the drug would be partly different in uninfected or adenovirus-infected cells. Moreover, inhibition of DNA polymerase α is competitive with dCTP and this is not observed for Ad5 DNA synthesis in isolated nuclei.

We have interpreted these data as evidence that DNA polymerase α as such is not involved in Ad5 DNA replication. The precise mechanism of inhibition of Ad5 DNA chain elongation by aphidicolin remains unknown. Possibly, a modified DNA polymerase is required, or other replication factors are also sensitive to the drug. Also, a cellular DNA polymerase could acquire a modified sensitivity towards aphidicolin due to a strong association with other proteins.

Recently, a DNA polymerase activity was isolated which co-purified with pTP (58). This enzyme displayed sensitivity to ddNTP's and N-ethyl maleimide and resistance to aphidicolin just like DNA polymerase γ, but could not use poly(rA)•oligo(dT) as template.

At high concentrations an effect of aphidicolin on chain growth could

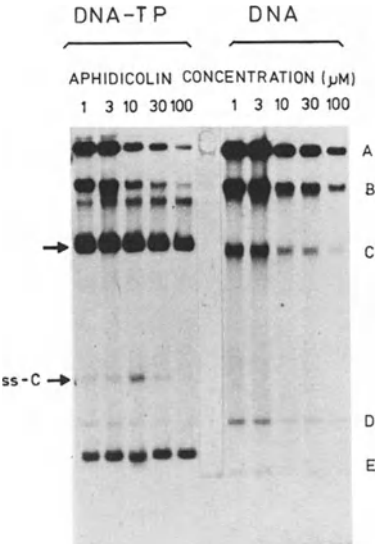


FIGURE 8. Replication of adenovirus DNA *in vitro* and the effect of aphidicolin.
Ad5 DNA or DNA-TP were digested with *Xba*I and the reaction products were incubated with nuclear extracts of Ad5-infected HeLa cells in the presence of increasing concentrations of aphidicolin. The reaction conditions were, in 30 μl, 50 mM Hepes-KOH pH 7.5 - 5 mM $MgCl_2$ - 0.5 mM DTT - 2 mM ATP - 20 μM each of dATP, dTTP and dGTP, 10 μM α-^{32}P-dCTP (5 Ci/mmol) - 5 mM creatine phosphate - 5 μg/ml creatine kinase, 0.04 μg digested complex and 7 μl nuclear extract. After 60 min at 37° C the reaction was stopped by addition of 3 μl 0.1% SDS - 0.1 M EDTA - 0.1% bromophenolblue and the complete mixture was electrophoresed in 1% agarose gels containing 40 mM Tris-HCl pH 7.8 - 1 mM EDTA - 5 mM sodium acetate - 0.1% SDS. After electrophoresis the gels were dried and autoradiographed. The terminal fragments C and E are indicated with arrows. C and E from DNA-TP have a reduced mobility due to the presence of protein. ssC = single-stranded C, resulting from multiple rounds of initiation and displacement synthesis (59).

also be observed *in vitro* (31). At low concentrations aphidicolin can be used to inhibit repair-like processes. When *Xba*I-digested DNA-TP or DNA was incubated with nuclear extracts under DNA synthesis conditions labeling of internal fragments, or terminal fragments without TP, was severely suppressed by aphidicolin (Fig. 8). However, replication of terminal fragments containing TP was rather resistant, in agreement with adenovirus DNA replication *in vivo*.

7. GENERAL DISCUSSION

7.1. The role of DBP

Adenovirus DNA replication is a very efficient nuclear process in which at least two viral proteins, pTP and DBP, are involved. The DBP, in particular the carboxy-terminal domain, is required already early during elongation but not for the formation of a pTP-dCMP complex (Fig. 9). These results, obtained *in vitro*, differ from studies on the phenotype of H5ts125 *in vivo*. In intact H5ts125-infected cells elongation continues unabated after a shift to the non-permissive temperature but new replication rounds are blocked (8). This result indicated a role of DBP in initiation. An explanation for this discrepancy could be that DBP is essential for the formation of a functional elongation complex after initiation. Once present in such a complex, the H5ts125 DBP might be protected against thermal inactivation. During the subsequent elongation process the main function of free DBP could be the protection of single-stranded displaced DNA. This role might be taken over in part by cellular DNA-binding proteins. This hypothesis would explain that H5ts125 behaves as an initiation mutant *in vivo*, while the DBP is required in the early steps of elongation.

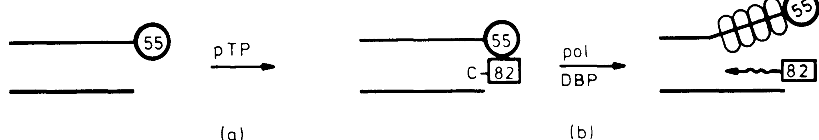


FIGURE 9. Requirements for initiation and elongation of Ad5 DNA.
a. Initiation proceeds by formation of a covalent pTP-dCMP complex. This reaction requires, besides pTP, dCMP, DNA-TP, and possible other cellular factors and ATP. ATP can be substituted by GTP, dATP or dGTP. The reaction is resistant to aphidicolin and ddCTP. b. Elongation requires DBP, dNTP's and a DNA polymerase of the γ-type. The reaction is sensitive to ddNTP's and to high concentrations of aphidicolin. (55), 55kD TP; [82], 82 kD TP; 0, DBP; pol, DNA polymerase.

7.2. The priming mechanism

The use of a protein priming mechanism, which is now strongly indicated by a vast amount of experimental evidence, has not been described before in eukaryotic cells. A similar mechanism may occur in B. subtilis phage Φ29 (60). The precise role of the parental TP is not yet clear. Protein-protein interactions between pTP and TP may be sufficient to position the

pTP correctly with respect to the 3'-terminus of the template strand. Alternatively, the TP may function to facilitate the exposure of a single-stranded nucleotide sequence in the origin. Possibly this sequence recognizes and binds pTP and thus keeps the priming protein in the correct initiation position. The conserved sequence 9-22 may be involved in this process (6).

7.3. Complementary strand synthesis

The detailed mechanism of complementary strand synthesis remains obscure. Formation of a panhandle structure in displaced single-stranded DNA is a very attractive hypothesis and provides a function for the inverted terminal repetition which is essential for viral replication. However, no such panhandles have been observed directly. *In vitro*, type 2 replicative intermediates have been observed only at low frequencies (9). This shows that complementary strand synthesis does not occur easily in nuclear extracts. Such a failure may be caused by the absence of a suitable single-stranded template, rather than by an intrinsic incapability of the extracts to perform the reaction. Studies with single-stranded templates containing TP will be required to resolve this point.

7.4. DNA polymerase

The sensitivity of adenovirus DNA chain elongation to ddNTP's has suggested that a γ-type DNA polymerase is required. Our results employing aphidicolin indicate that DNA polymerase α, if functional in Ad5 DNA replication, does not behave similarly as during cellular or SV40 DNA replication. A DNA polymerase with the properties of DNA polymerase γ but the template specificity of DNA polymerase α is found in strong association with pTP (58). It will be interesting to know whether this enzyme, when uncoupled from pTP, is identical to any of the cellular DNA polymerases. The similarities between adenovirus DNA replication and mitochondrial DNA replication are striking as has been pointed out previously (57). DNA polymerase γ is found in mitochondria and mitochondrial DNA synthesis is sensitive to ddNTP's and resistant to aphidicolin (62). Also, both replicons duplicate by a displacement mechanism. Whether these similarities are purely accidental remains to be seen.

8. ACKNOWLEDGEMENTS

We wish to thank M.G. Kuijk, M.M. Kwant, F.M.A. van Schaik, and W. van Driel for expert technical assistance and W. Kruijer, B.G.M. van Bergen and P.J. de Jong for making their unpublished results available to us. This research was supported in part by the Netherlands Organization for Chemical Research (S.O.N.) with financial support from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

9. REFERENCES

1. Cold Spring Harbor Monographs 10B. Molecular Biology of Tumor Viruses. Part 2. DNA Tumor Viruses. Ed. J. Tooze. Cold Spring Harbor, 1980.
2. Van der Eb, A.J., Van Kesteren, L.W., Van Bruggen, E.F.J. Biochim. Biophys. Acta 182:530-541, 1969.
3. Green, M., Pina, M., Kimes, R., Wensink, P.C., Machattie, L.A. and Thomas, C.A. Proc. Natl. Acad. Sci. USA 57:1302-1309, 1967.
4. Robinson, A.J., Younghusband, H.D. and Bellett, A.J.D. Virology 56: 54-69, 1973.
5. Rekosh, D.M.K., Russell, W.C., Bellett, A.J.D. and Robinson, A.J. Cell 11: 283-295, 1977.
6. Steenbergh, P.H., Maat, J., Van Ormondt, H. and Sussenbach, J.S. Nucleic Acids Res. 4:4371-4390, 1977.
7. Arrand, J.R. and Roberts, R.J. J. Mol. Biol. 128:577-594, 1979.
8. Van der Vliet, P.C. and Sussenbach, J.S. Virology 67:415-525, 1975.
9. Van der Vliet, P.C., Zandberg, J., and Jansz, H.S. Virology 80:98-110, 1977.
10. Challberg, M. and Kelly, T.J. Proc. Natl. Acad. Sci. USA 76:655-659, 1979.
11. Sussenbach, J.S., Van der Vliet, P.C., Ellens, D.J. and Jansz. H.S. Nature New Biology 239:47-49, 1972.
12. Ellens, D.J., Sussenbach, J.S. and Jansz, H.S. Virology 64:427-442, 1974.
13. Lechner, R.L. and Kelly, T.J. Cell 12:1007-1020, 1977.
14. Winnacker, E.L. Cell 14:761-773, 1978.
15. Sussenbach, J.S. and Kuijk, M.G. Virology 84:509-517, 1975.
16. Stillman, B.W., Bellett, A.J.D. and Robinson, A.J. Nature 269:723-725, 1977.
17. Sussenbach, J.S. and Kuijk, M.G. Nucleic Acids Res. 5:1289-1295, 1978.
18. Van der Vliet, P.C. and Sussenbach, J.S. Eur. J. Biochem. 30:584-592, 1972.
19. De Jong, P.J., Tommassen, J.P.M., Van der Vliet, P.C. and Jansz, H.S. Eur. J. Biochem. 117:179-186, 1981.
20. Brison, O., Kedinger, C. and Wilhelm, J. J. Virol. 24:423-443, 1977.
21. Kaplan, L.M., Kleinman, R.E. and Horwitz, M.S. Proc. Natl. Acad. Sci. USA 74:4425-4429, 1977.
22. Yamashita, T., Arens, M. and Green, M. J. Biol. Chem. 252: 7940-7949, 1977.
23. Challberg, M.D., Desiderio, S.V. and Kelly, T.J. Proc. Natl. Acad. Sci. USA 77:5105-5109, 1980.
24. Lichy, J.H., Horwitz, M.S. and Hurwitz, J. Proc. Natl. Acad. Sci. USA 78:2678-2682, 1981.

25. Challberg, M.D. and Kelly, T.J. J. Virol. 38:272-277, 1981.
26. Robinson, A.J., Bodnar, J.W., Coombs,D.H. and Pearson, G.D. Virology 96:143-158, 1977.
27. Stillman, B.W. and Bellett, A.J.D. Virology 93:69-79, 1979.
28. Van Wielink, P.S., Naaktgeboren, N. and Sussenbach, J.S. Biochim. Biophys. Acta 563:89-99, 1979.
29. Reiter, T., Fütterer, J., Weingärtner, B. and Winnacker, E.L. J. Virol. 35:662-671, 1980.
30. Challberg, M.D., Ostrove, J.M. and Kelly, T.J. J. Virol. 41:265-270, 1982.
31. Pincus, S., Robertson, N. and Rekosh, D. Nucleic Acids Res. 9:4919-4938, 1981.
32. Van der Vliet, P.C. and Levine, A.J. Nature New Biology 246:170-173, 1973.
33. Van der Vliet, P.C., Keegstra, W. and Jansz, H.S. Eur. J. Biochem. 86:389-398, 1978.
34. Linné, T., Jörnvall, H. and Philipson, L. Eur. J. Biochem. 76:481-490, 1977.
35. Fowlkes, D.M., Lord, S.T., Linné, T., Pettersson, U. and Philipson, L. J. Mol. Biol. 132:163-180, 1979.
36. Horwitz, M.S. Proc. Natl. Acad. Sci. USA 75:4291-4295, 1978.
37. Kruijer, W., Van Schaik, F.M.A. and Sussenbach, J.S. Nucleic Acids Res. 9:4439-4457, 1981.
38. Klein, H., Maltzman, W. and Levine, A.J. J. Biol. Chem. 254:11051-11060, 1979.
39. Klessig, D.F. and Grodzicker, T. Cell 17:957-966, 1979.
40. Ensinger, M.J. and Ginsberg, H.S. J. Virol. 10:328-339, 1972.
41. Van der Vliet, P.C., Levine, A.J., Ensinger, M.J. and Ginsberg, H.S. J. Virol. 15:348-354, 1975.
42. Ariga, H., Klein, H., Levine, A.J. and Horwitz, M.S. Virology 101: 307-310, 1978.
43. Nicolas, J.C., Suarez, F., Levine, A.J. and Girard, M. Virology 108: 521-524, 1981.
44. d'Halluin, J.C., Allart, C., Cousin, C., Boulanger, P.A. and Martin, G.R. J. Virol. 32:61-71, 1979.
45. De Jong, A., Van der Vliet, P.C. and Jansz, H.S. Biochim. Biophys. Acta 476:156-165, 1977.
46. Arens, M., Yamashita, T., Padmanaban, R., Tsurno, T. and Green, M. J. Biol. Chem. 252:7947-7954, 1977.
47. Abboud, M.M. and Horwitz, M.S. Nucleic Acids Res. 6:1025-1039, 1979.
48. Shaw, C.H., Rekosh, D.M. and Russell, W.C. J. Gen. Virol. 48:231-236, 1980.
49. Van der Werf, S., Bouché, J.-P., Méchali, M. and Girard, M. Virology 104:56-72, 1980.
50. Van der Vliet, P.C. and Kwant, M.M. Nature 276:532-534, 1978.
51. Krokan, H., Schaffer, P. and De Pamphilis, M.L. Biochemistry 18:4431-4443, 1979.
52. Van der Vliet, P.C. and Kwant, M.M. Biochemistry 20:2628-2632, 1981.
53. Kaplan, L.M., Ariga, H., Hurwitz, J. and Horwitz, M.S. Proc. Natl. Acad. Sci. USA 76:5534-5538, 1979.
54. Ikegami, S., Taguchi, T., Ohashi, M., Oguro, M., Nagano, H. and Mano, Y. Nature 275:458-460, 1978.
55. Pedrali-Noy, G. and Spadari, S. Biochem. Biophys. Res. Commun. 88:1194-1202, 1979.

56. Longiaru, M., Ikeda, J.-E., Jarkovsky, Z., Horwitz, S.B. and Horwitz, M.S. Nucleic Acids Res. 6:3369-3386, 1979.
57. Kwant, M.M. and Van der Vliet, P.C. Nucleic Acids Res. 8:3993-4007, 1980.
58. Enomoto, T., Lichy, J.H., Ikeda, J.-E. and Hurwitz, J. Proc. Natl. Acad. Sci. USA 78:6779-6783, 1981.
59. Horwitz, M.S. and Ariga, H. Proc. Natl. Acad. Sci. USA 78:1476-1480, 1981.
60. Mellado, R.P., Penalva, M.A., Inciarte, M.R. and Salas, M. Virology 104:84-96, 1980.
61. Galos, R.S., Williams, J.F., Binger, M.-H. and Flint, S.J. Cell 17: 945-956, 1979.
62. Zimmermann, W., Chen, S.M., Bolden, A. and Weisbach, A. J. Biol. Chem. 255:11847-11852, 1980.
63. Van der Vliet, P.C., Blanken, W.M., Zandberg, J. and Jansz, H.S. In: DNA Synthesis, Present and Future (Eds. I. Molineux and M. Kohiyama) Plenum Press, New York, 1977, pp. 869-885.

12

REPLICATION OF PARVOVIRUS DNA

LOIS A. SALZMAN, RITA MITRA, and DALE BROWN

In 1959, Kilham and Oliver characterized a virus that they isolated from several strains of metastasizing rat liver sarcomas and a transplantable leukemia. The virus named Kilham rat virus (KRV) caused a cytopathic effect (CPE) in rat embryo tissue, agglutinated guinea pig red blood cells (HA), was heat stable and ether resistant (1). Acridine orange and Feulgen stains of intranuclear inclusions in infected cells established that the virus contained DNA (2). KRV was the first characterized member of what is now a large group of viruses called parvoviruses. Parvoviruses are among the smallest eukaryotic viruses. They are ubiquitous and have been isolated from a wide variety of hosts. The parvoviruses that have been studied most include the H-1 virus from a human transplantable tumor (3), bovine hemadsorbing enteric virus (Haden; 4), bovine parvovirus (BPV; 5), feline panleukopenia virus (FPV; 6), minute virus of mice (MVM; 7), KRV (1), and human adeno-associated viruses (AAV; 8, 9).

The family Parvoviridae is divided into two genera based on the ability of the virions to replicate autonomously in cell culture. Genus A, the larger group, consists of those autonomous or non-defective parvoviruses capable of replicating independently in cells although this replication may be cell specific. Parvoviruses in Genus A include KRV, BPV, H-1, MVM, and FPV. Parvoviruses like AAV, in genus B, are defective. They depend for complete or partial replication on coinfection of the host cell with an additional "helper virus" such as adenovirus or herpes virus (10). A third genera C has been proposed for inclusion in the parvovirus family. This group would contain the autonomous viruses of arthropods which may replicate their DNA in a manner analogous to the AAV. However, there is some debate about their inclusion in the Parvoviridae and we will not include them in this discussion (11). All of the parvoviruses seem to share an affinity for the dividing cell. Because of their small size and limited coding capacity, the virus is dependent

on the cell for most of the enzymes and substrates for virion multiplication. Diseases associated with the viral infection often seem to involve tissues which at the time of infection are undergoing changes involving cell proliferation (10). Parvoviruses such as FPV, mink encephalitis (MEV) and canine parvovirus (CPV) cause naturally occurring diseases, sometimes rapidly fatal. The rest of the parvoviruses elicit a spectrum of pathological conditions, particularly in fetal or neonatal animals (10). Because parvoviruses replicate in the nucleus of infected cells, viral DNA replication presents an attractive model system for studying molecular mechanisms of eukaryotic DNA replication.

In this chapter, we will discuss the results of experiments involved in trying to understand DNA replication of the parvoviruses. We will present mainly the results found in our laboratory with the autonomous parvovirus, KRV. We will supplement this work with some of the fine research carried out in other laboratories studying DNA replication of both the autonomous and defective parvoviruses.

AUTONOMOUS PARVOVIRUS DNA

The characterization of parvovirus DNA has presented technical problems due to the compact structure of the nucleocapsid. The virus is resistant to treatments for DNA extraction. It is also difficult to grow large quantities of virions. We have found alkaline extraction of DNA from the virion and separation of the DNA and protein by cesium chloride or alkaline sucrose centrifugation most satisfactory (12, 13). The extracted DNA has been shown to be single-stranded by staining techniques, reaction with formaldehyde, digestion with single-stranded specific nucleases, nearest neighbor analysis, and the noncomplementary ratios of adenine-thymine and guanine-cytosine (10, 14. 15). Electron microscope examination of KRV revealed that the single-stranded (ss) genome was also linear. Circular forms were not seen (16). However, digestion of the ss DNA by single-stranded nucleases resulted in only 70 to 80% digestion of the genome. The remaining DNA pieces resistant to hydrolysis eluted from benzoyl-naphthoyl-DEAE (BND)-cellulose as double-stranded (ds) DNA (17, 18, 19). The double-stranded regions are the 3' and 5' termini of autonomous virus DNA. Since the extracted virion DNA did not form duplex molecules, selective packaging must occur predominantly if not exclusively on one strand of viral DNA and not on its complement. This strand is the thymine-rich or

minus strand. The properties of linearity, single-strandedness and selective packaging of one DNA strand are common for all the autonomous parvoviruses (10) with the possible exception of LuIII which was isolated from the human cell line Lu106 (20). This autonomous parvovirus may package both (+) and (-) strands.

Since probably the origin and terminus of DNA replication occurs in the small ds terminal regions (21, 22, 23), we decided to further investigate their structure. We determined the nucleotide sequence of the 3' terminus of KRV and proposed a configuration for its most highly base paired configuration (Fig.1). The nucleotides can be folded over in a Y-shaped hairpin configuration. Nucleotides 1 to 44 are base paired with nucleotides 71 to 114 with the exception of nucleotides 25, 26, 41, 88 and 89. Nucleotides 45 to 50 are base paired with nucleotides 54 to 59 and nucleotides 60 to 63 are hydrogen bonded with nucleotides 67 to 70. Only nine of the 114 nucleotides could not be base paired in this arrangement. Other reports confirmed this sequence for KRV and other rodent parvoviruses [MVM, H-1 and H-3 (24)]. All four have the same Y-shaped terminus. Although there are some minor differences, the sequences are very homologous, indicating conservation or close evolutionary relationship of the 3' terminal sequences. The autonomous parvoviruses (BPV) and human (LuIII) also possess similar 3' terminal hairpin structures, although each nucleotide sequence is unique (24).

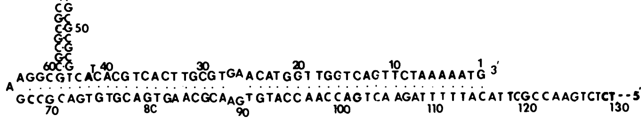


Fig. 1. Monolayer rat nephroma cells were infected with KRV (multiplicity 5:1) and incubated for 48 hr at 37°C. Virus was freed from contamination by 2 or 3 isopycnic centrifugations in cesium chloride (density 1.41 g/cm^3). DNA was extracted from KRV by sedimentation at pH 13.5 in a 5 to 20% sucrose gradient, 0.4M NaCl, 0.6M NaOH, 1 mM EDTA at 20°C in a Beckman SW41 rotor at 27,000 rpm for 16 hr. The 16S fractions were pooled, neutralized and precipitated in ethanol at -20°C. The 3' terminus of the DNA was labeled using T4 DNA polymerase and a ^{32}P deoxynucleotide triphosphate. After passage through a Sephadex G-50 column, the DNA was centrifuged again in a pH 13.5, 5-20% sucrose gradient, neutralized and precipitated in ethanol as above (19). The molecule was sequenced by the procedure of Maxam and Gilbert (25).

Sequencing studies of the 5' terminus of the KRV genome have been more difficult than with the 3' terminus. We have experienced great difficulty in labeling this 5' end of the KRV DNA with ^{32}P using polynucleotide kinase and other enzymes. This is probably because the 5' terminus has some interfering secondary structure or a terminal blocking group such as an amino acid. However, information of the 5' nucleotide sequence derived from the cloned terminal segments of the parvovirus H-1 have been published (26). Because of the similarity in the 3' terminus between these two viruses, it is likely that their 5' termini are also similar. The authors have found that there are two orientations of the 5' terminal palindrome. One form is inverted with respect to the other (flip-flop relationship). The palindrome may exist in the sequenced (5') 1→242 or (5')242→1. This dual orientation is not found in the 3' end. Adjacent to the 5' terminal palindrome is an A-T rich region that is non-coding and contains a 55 base pair tandem repeat (26).

Many of the advances in studying the fine structure of the parvovirus DNA has resulted from the use of restriction enzyme fragments. Because of their 3' hairpins, parvoviral DNAs are excellent template primers for prokaryotic DNA polymerases (18, 27). Double-stranded viral DNAs can be synthesized _in vitro_ using the 3' OH hairpin terminus as a primer; fragments could be observed synthesized in sequential order.

Comparison of the DS DNA synthesized _in vitro_ (given in the legend of figure 2) and from _in vivo_ synthesized DS KRV DNA shows identical patterns except for the size of the terminal 3' fragments (hairpin) which is shorter by about 100 nucleotides in the _in vitro_ synthesized DNA. Other restriction fragments have been mapped using a variety of methods (28).

DEFECTIVE (AAV) PARVOVIRUS DNA

Some intriguing differences have emerged between the virion DNAs of the autonomous and defective parvoviruses. These differences may require some changes in the steps in replication of the two groups of virions. Finding out more about their differences may help in our understanding of AAV defectiveness and the scheme of DNA replication. The first apparent difference between the two groups involves the state of the DNA. A controversy existed for several years over the strandedness of the genome of the defective virus AAV. When extracted, the DNA was double stranded with a molecular weight of 3 X 10^6 (10). However, a virion of the reported size

Fig. 2. Physical map of KRV DNA strain 308 with the sites of hydrolysis of restriction endonucleases HhaI, EcoRI, HaeII, TaqI, BglII, HaeIII, PstI, HindIII and HpaI. KRV ss DNA was isolated from virions by alkaline hydrolysis and sedimentation in sucrose (pH 13.5) as described in Fig. 1. After ethanol precipitation, the ss DNA was incubated at 37°C with DNA polymerase "large fragment", two units per mg of DNA in 0.2 ml of 20 mM Tris (pH 7.4), 2 mM $MgCl_2$, 0.3 mM β-mercaptoethanol, dCTP, dGTP, dTTP, and dATP, 20 mM each (27). Incubation times varied and enzyme activity was halted by addition of 20 mM EDTA and 0.5M NaCl, or heating to 70°C for 3 min. The DNA was precipitated in 2.5 volumes of ethanol at -20°C for 16 hr. After centrifugation for 30 min at 30,000 rpm and -5°C in a Beckman SW60 Ti rotor, the precipitate was resuspended in the appropriate restriction enzyme digestion mixture (100 µl) and digested with 10 to 20 units of enzyme at 37°C for 4 hr. Glycerol (20%) and the tracking dye bromophenol blue were added to each sample and they were electrophoresed in a 1.4% agarose gel in 0.04M Tris pH 8.0, 0.02M sodium acetate buffer for 2 hr at 150 volts. The HindIII digest of lambda DNA was added as a marker for molecular weight determination. The gels were stained with ethidium bromide (1.5 X 10^{-6}g/ml) and the DNA fragment bands visualized under UV light. Further fine structure analysis can be accomplished by sequential digestion with two endonucleases. The fragments can also be isolated from gels (as above) and digested with other restriction endonucleases. Restriction enzymes AvaI, SalI, SstI, SstII, XhoI, KpnI, BamHI and SmaI do not appear to cleave KRV DNA.

could not hold that much DNA (15). The controversy was resolved when AAV grown in the presence of bromodeoxyuridine (BUdR) was mixed with AAV grown in the absence of BUdR and the DNA extracted from the virions.

Double-stranded DNA of hybrid density was found after CsCl isopycnic centrifugation. One strand of the hybrid contained BUdR and the other did not (29). The result suggested that although each virion contains a single-stranded DNA molecule, it can anneal with another molecule after extraction to form a double-stranded molecule. The AAV virions, unlike most of the autonomous parvoviruses, contain either (+) or (-) DNA strands in approximately equal amounts (22).

Another difference between defective and autonomous parvoviruses was realized when the DNA molecules were observed under the electron microscope and later when the nucleotides were sequenced. Koczot *et al.* (13)

found that both ends of the purified (+) and (-) DNA strands of AAV-2 contain terminal self-complementary sequences (inverted terminal repetitions). A strand can form a single-stranded circle closed by short hydrogen-bonded duplex segments or panhandles in an anti-parallel configuration. The exact structure of the AAV DNA termini was resolved by nucleotide sequencing (30). As in the case of the 3' terminus of the autonomous parvovirus, AAV has a terminal Y-shaped palindrome of 145 bases. The terminal 145 nucleotides at the 3' end of the DNA are perfectly complementary to the 145 nucleotides at the 5' end of the genome. However, the palindrome at the 3' and 5' ends can exist in two orientations which are identical inverted sequences (flip-flop) of each other. This enables them to form the double-stranded panhandle circle (30). The nucleotide sequence of the AAV termini has very little similarity to the sequences found in either the 3' or the nonidentical 5' termini (20) of the autonomous parvoviruses.

PARVOVIRUS REPLICATION

The small parvovirus genome described above is capable of directing the synthesis of few, probably less than three, proteins. Its capability may be greater if a parvovirus DNA sequence can code for more than one protein by means of RNA splicing or if they are coded for in different reading frames. This limited ability to specify protein synthesis suggests that the replication of the viral DNA will be carried out largely by cellular enzymes and co-factors. The parvoviruses thus provide a small and more easy to characterize replicon for the analysis of the mechanisms of eukaryotic DNA replication.

Aside from the species specificity of the autonomous parvoviruses, replication of the virus does not occur unless the cell itself is in the proper stage of growth. MVM can bind to and enter a cell early in G phase. However, the viral DNA strand remains single stranded until the cell enters S phase. Only then does viral DNA synthesis start (31). The S phase requirement could be a cellular enzyme, a modified cellular enzyme, a viral or cellular gene product. A DNA polymerase associated with KRV which may be a modified cellular polymerase has been reported (32). It varies in some properties from the uninfected cellular DNA polymerases (33). Synthesis of replicative forms (Rf) and viral DNA from Rf may require the synthesis of viral proteins. Both temperature sensitive

mutants of H-1 and defective interfering viruses are defective in DNA synthesis at nonpermissive temperatures or without wild type helper. The defective viruses contain all the information and signals needed for replication but cannot synthesize a required viral protein (34, 35).

The steps in the infection process by which a parvovirus takes over a eukaryotic cell, turns it into a virus production factory and disintegrates its host has turned out to be rapid and interesting. KRV is absorbed and penetrates a rat nephroma cell within one hour after infection (36). If KRV whose DNA is radiolabeled with ^{3}H thymidine is used for infection, the fate of the viral DNA can be followed. Within one hour after infection of a permissive cell, the parental viral DNA can be converted into a new form. On the basis of sedimentation properties in sucrose gradients, elution from hydroxylapatite, buoyant density in equilibrium CsCl density gradients, and appearance in the electron microscope, this form was double-stranded and linear (37). This replicative form is called the Rf.

Structural analysis of intermediates in the replication and production of virion DNA have yielded some information to help in understanding the sequence of events in virus replication. On the basis of retention on hydroxylapatite, a portion of _in vivo_ labeled MVM DNA was also determined to be in a duplex monomer. About 35% of this duplex MVM DNA underwent spontaneous unimolecular renaturation, suggesting that some of the duplex molecules existed as covalently linked viral (v) and complementary (c) strands (38). The double-stranded replicative form (Rf) self-annealed (snapped back) on the viral 3' end suggesting a covalent hairpin-like turn between v and c strands. This hairpin structure may result from the first step in DNA replication, the covalent synthesis of a complementary strand from the self-priming viral strand 3' end hairpin.

Rf synthesis can be completely inhibited by the inhibitors of RNA synthesis, actinomycin D and α-amanitin and the inhibitor of protein synthesis cycloheximide if the inhibitors are added before the initiation of viral DNA synthesis (39). However, after viral DNA and protein synthesis are started, addition of α-amanitin no longer inhibits Rf synthesis. This indicates that Rf replication is not primed by polyribonucleotide sequences produced by the α-amanitin sensitive RNA polymerase II (39).

Electron microscope evidence has been reported that after the synthesis of the double-stranded Rf DNA, further viral DNA synthesis occurs from the 3' end of the complementary strand (5' end of the viral strand). Replication of the autonomous parvovirus H-1 occurs via a double-stranded linear Y-shaped replicative intermediate (23). The origin of Rf replication H-1 was localized within approximately 300 base pairs of the right end (3' complementary, 5' viral) of the Rf DNA as determined by partial denaturation mapping, restriction endonuclease digestion and electron microscopy (23). The right 3' end of the complementary viral genome as shown before is capable of folding back in a hairpin and self-priming DNA synthesis.

Other structural forms of KRV DNA have been isolated from the infected cell. If DNA from KRV infected cells is extracted, purified by isopycnic centrifugation in CsCl and then analyzed by velocity sedimentation in a 5 to 30% sucrose gradient (pH 7.4), oligomers of KRV Rf DNA can be detected (37). The most prominent oligomer is a dimer (up to 20% of the KRV DNA). Higher molecular weight oligomers are seen in smaller amounts (37, 40, 41). The oligomers may also play a role in parvovirus replication.

The only report of significant numbers of circular forms of autonomous parvoviruses comes from electron microscopic analysis of MVM-infected cells (42). The authors reported the presence of DNA in the form of duplex lariates (circles with tail). The total contour lengths of the lariates are equal to that of monomer duplex Rf. The significance of the forms and their part in DNA replication, if any, is unknown.

Intracellular DNA intermediates of KRV (43) and H-1 (44) are usually found associated with proteins with a capsid-like structure forming nucleocapsids. Since mutations in virion structural proteins have a profound effect on H-1 DNA synthesis (34, 45), the replication of virion DNA is thought to be closely associated with virion assembly. Revie _et al_. (44) have reported a covalent association of H-1 DNA with one or more proteins in the nucleocapsid complex. Hirt extracts of cells infected with H-1 contained viral DNA which banded in CsCl remained stable in strong alkali and electrophoresed in agarose as though the DNA was covalently linked to a protein. Treatment of the complex with protease released DNA with the properties of virion DNA. We have found similar results with a protein linked to KRV DNA.

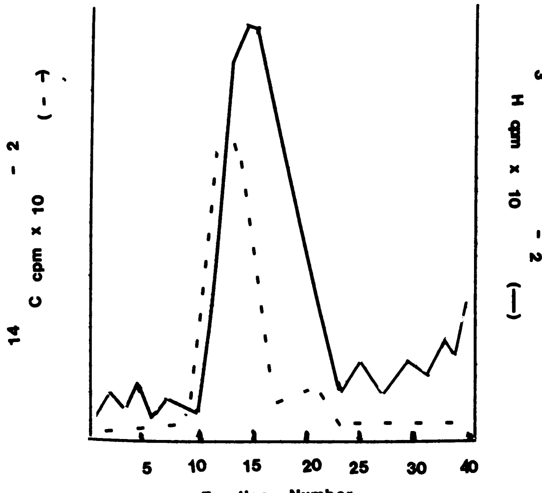


Fig. 3. Alkaline CsCl centrifugation of KRV DNA protein complex. Rat nephroma cells were grown in monolayers and infected with KRV (five infectious units per cell). At 16 hr after infection, when Rf DNA synthesis is most rapid, the intracellular DNA was labeled with (^{3}H)-deoxythymidine at 10 μC/ml or (^{14}C)-deoxythymidine at 0.5 μC/ml for two hr at 37°C. The plates were scraped and the cells sedimented by centrifugation at 3,000 rpm for 10 min. Intracellular DNA intermediates were separated from host DNA by incubating the cell pellet in 0.6% SDS, 0.02 M Tris-HCl pH 8.0, 0.001 M EDTA at room temperature for 90 min followed by incubation in 1 M NaCl overnight at 4°C. The small molecular weight DNA remains in the supernatant after centrifugation at 10,000 rpm in a Sorvall centrifuge for 30 min at 4°C. The supernatant from the cells to which (^{14}C)-deoxythymidine was added was treated with proteinase K at a concentration of 1 mg/ml for 1 hr at 60°C. Two additions of proteinase K at 1 mg/ml were made at 1 hr intervals. Proteinase K was then added at a concentration of 2 mg/ml and the solution incubated at 37°C for 5 hr. The DNA was precipitated in ethanol, 0.2 M NaCl overnight at -20°C and then centrifuged at 12,500 rpm for 10 min in an Eppendorf microcentrifuge. The precipitate was combined with the supernatant containing (^{3}H)-thymidine and held at 37°C for 30 min in 1M NaOH. It was then centrifuged to equilibrium in a 6 ml alkaline CsCl gradient (density 1.71 g/ml) in 50 mM Na_3PO_4 pH 12.0, 25 mM EDTA at 4°C in a Beckman 65 rotor at 40,000 rpm for 42 hr.

As can be seen in this figure, the majority of the radioactive DNA containing (^{3}H) has a lower density in CsCl (density 1.69 g/ml) when compared with the proteinase K treated sample (^{14}C) with a density the same as KRV DNA (density 1.71 g/ml). If the peak fractions from both isotope peaks were hybridized with double-stranded KRV DNA immobilized in 10 fold

excess to a nitrocellulose filter (46) in 4 X SSC, 0.1% SDS at 67° for 16 hr, over 90% of the sample DNA hybridized to the filter. The symmetry of the (^3H) peak suggests that both strands, viral and complementary, of the Rf are complexed with protein. The resistance of the DNA-protein linkage to alkali suggests that the linkage is covalent. The linkage of the DNA and protein on the 5' ends of the Rf can be demonstrated by decreased mobility during electrophoresis of the terminal fragments after digestion with restriction enzymes.

As shown by Revie _et al_. (44), the difference in neutral CsCl density between the proteinase K-treated and non-proteinase K-treated parvovirus DNA can be used to determine the molecular weight of the terminal protein. Using this method, the KRV terminal protein has a molecular weight of 60,000 to 65,000, close to the molecular weight of the major KRV capsid protein. Preliminary results indicate the KRV DNA is linked to the protein by a phosphodiester linkage involving serine and thymine or deoxycytosine (data unpublished). Adenovirus also contains a 5' DNA protein linkage between the β-OH of a serine residue in the protein and the 5' OH of the terminal deoxycytidine residue of adenovirus DNA (47). The adenovirus terminal protein is proposed to serve as a primer for daughter strand synthesis (47). The source (cellular or viral) and function of the autonomous parvovirus terminal proteins is not known. Because of the terminal hairpins at the 3' and 5' termini, hairpin priming in DNA replication has been favored by most authors. The protein coded by gene 5 of M13 phage is essential for production of viral single strands from the duplex replicative intermediate (48). Other functions that the protein could have include protection of the viral strand against nucleases, helping the DNA assume the configuration required for packaging, prevention of other DNA binding proteins or a role in RNA transcription.

In order to study the exact requirements of virus DNA replication, some studies have been conducted _in vitro_ in cell-free systems. Synthesis of viral DNA in intact nuclei have been studied by Kollek _et al_ (49), using H-1 virus, and in nuclear lysates by Pritchard _et al_. (50), using KRV and BPV. Both groups studied the influence of inhibitors of mammalian DNA polymerases on viral DNA synthesis. The inhibitors used were aphidicolin specific for DNA polymerase α and 2' 3' dideoxythymidine 5' triphosphate (ddTTP) which inhibits cellular polymerases in the order $\gamma>\beta>\alpha$.

Kollek *et al*. (49) found that there was an initial period of rapid DNA synthesis (20 min) involving polymerase α followed by a period of sustained synthesis (120 min) carried out by a ddTTP sensitive enzyme, probably γ. They suggest that polymerase α is involved in strand-dependent synthesis on H-1 Rf and the ddTTP sensitive synthesis may correspond to conversion of displaced single strands to Rf. Pritchard *et al*. (50) found that only inhibitors of DNA polymerase α stopped viral DNA synthesis. They found that polyamines were required to protect the template and the newly synthesized viral strands from degradative DNA synthesis in the nuclear lysate involved only elongation of strands initiated *in vivo*.

DEFECTIVE PARVOVIRUS (AAV) REPLICATION

AAV infection of cells follows a course similar to the autonomous parvovirus infection but its defectiveness makes additional requirements on the cell and helper virus. Unlike the species specific autonomous parvoviruses, human AAV will replicate in any cell so far tested if the adenovirus helper can replicate in this cell (51). Thus, if host gene products are required directly for AAV replication, they are common components of many species of animal cells. Possible candidates are DNA binding proteins, DNA polymerases or nucleases. The contribution that a helper virus such as adenovirus must make to AAV replication has received much attention. Understanding this contribution may lead to a better understanding of adenovirus replication and perhaps broader cellular application. We will mention some of the experimental results on the relationship between adenovirus and AAV.

The temporal sequence of events in AAV infection is dependent on whether AAV and adenovirus are simultaneously coinfected or if there is a delay between first adenovirus infection and then AAV infection. In the first instance, AAV DNA synthesis is delayed several hours as compared to separate infections, suggesting a need for an early adenovirus protein. Early adeno function may be all that is required since it has been shown that microinfection of Ad2 early mRNA into Vero cells was sufficient to support complete AAV replication (52). However, extensive use of adenovirus early temperature sensitive mutants deficient in their own DNA synthesis permits the synthesis, albeit inefficient, of AAV DNA on coinfection (53). Thus, the expression of multiple adenovirus genes may be a prere-

quisite for AAV production. In a series of experiments involving cloned segments of AAV and adenovirus, Janik _et al_. (54, and personal communication) have concluded that the adenovirus DNA binding protein augments the level of AAV cytoplasmic transcripts while translation of these transcripts requires adenovirus genes. This is an area for further investigation.

The early events of AAV infection such as penetration, migration to the nucleus, uncoating of the DNA, and perhaps conversion to replicative intermediates, are believed to be independent of helper virus (55) and similar to the autonomous parvoviruses. The AAV DNA can at this point, in the presence of helper virus, proceed to the production of intact virions, produce defective particles, or become integrated into the host chromosome (56, 57, 58). AAV DNA replication studies in the production of intact virions favor the 3' hairpin initiation of DNA synthesis and the hairpin transfer to replicate 5' complementary strand termini (Fig. 4). Studies on the production of the mature AAV genome strand that can be packaged into virions or recycled as an Rf molecule have also been reported.

Hauswirth and Berns (22, 59) analyzed the specific activities of restriction enzyme fragments of pulse-labeled intracellular duplex Rf of AAV DNA. Most of the radioisotope label resided in the portion of the genome that replicated last. The radioisotope profiles for the complementary strands showed the most label at the 3' end. They concluded that synthesis is a unidirectional mechanism which terminates at each progeny 3' end and involves a single-stranded displacement mechanism. We will leave the wealth of information available on integration and defective particles to other reviewers.

PARVOVIRUS REPLICATION MODEL

The parvovirus genome is linear. All described DNA polymerases require a primer with a free 3' OH group. The presence of a primer presents a problem in the complete replication of the progeny 5' terminus. The primer may be a small polyribonucleotide as found in most prokaryotes but this must be replaced by deoxynucleotides or the genome will lose some of its terminal sequences every time replication occurs. As discussed above, the parvoviruses have a 5' terminal protein which may be involved in DNA replication but they also have both 3' and 5' palindromes capable of forming

hairpin primers. Experimental evidence indicates that the virion 3' terminus both in vivo and in vitro is probably self-priming to form the replicative intermediate (Rf).

A mechanism involving hairpin structures in DNA replication was first proposed by Cavalier-Smith for the replication of linear eukaryotic chromosomal DNA (60). He proposed that the 5' terminus of the progeny DNA contained an RNA primer which could be removed by RNAse H-like activity. The 5' terminus would then have a cap and the 3' end a single strand overhang. If the 3' overhang contained a palindrome (self-complementary end), it could fold back on itself in a hairpin configuration and any gap between the 3' hairpin and the 5' end of the strand would be filled in by DNA polymerase and sealed with DNA ligase. If the parental strand was specifically nicked at the point opposite the original 3' end of the hairpin, a 5' progeny strand overhang would be created. The 3' OH on the shortened parental strand would serve as a primer for DNA polymerase to fill in the end by repair-like synthesis. The result of this transfer would be an inverted 3' terminus.

An inverted 3' terminus (flip-flop) is found in the Rf of the defective parvovirus AAV but not in the autonomous parvoviruses. Since the 3' terminus of the progeny AAV strand is identical to the 5' terminus, the progeny 3' terminus could also form a hairpin and initiate DNA synthesis and cause a flip-flop to occur at that end also. As a result, AAV molecules would be expected in four configurations, two classes of 3' (flip or flop) and two classes of 5' terminal sequences (flip or flop) (Fig. 4). All four molecules have been found experimentally (K. Berns, personal communication).

The replication of viral DNA from the autonomous parvoviruses presents some additional problems to the Cavalier-Smith replication model. In the autonomous parvovirus genome, as in the case of AAV, the 5' terminal sequence exists in two forms called "flip" and "flop". However, at the 3' terminus of autonomous parvoviruses, unlike AAV, only a single unique sequence has been found. It is possible that there is some selective process for the unique 3' sequence which occurs during encapsidation but it is also possible that there is an asymmetric step which occurs during DNA replication. This is an area for future study. Several replication models have been proposed for the replication of autonomous parvovirus DNA.

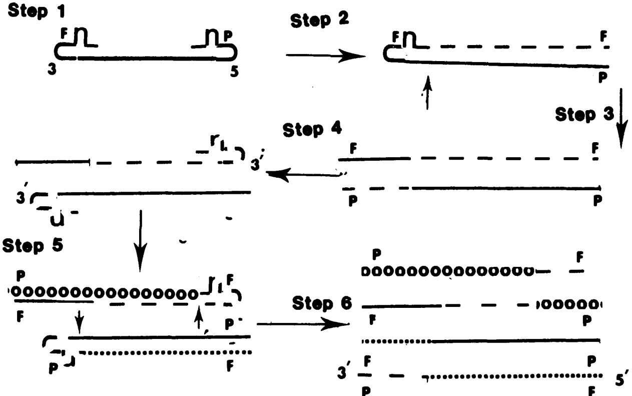


Fig. 4. Model for the synthesis of AAV DNA. Step 1. Spontaneous formation of 3' and 5' hairpin structures to serve as primers for DNA polymerase activity. If flip is the terminal palindrome sequence 1 to 145, then flop would be the sequence 145 to 1, as sequenced in the 3' to 5' direction. The four possible configurations present in equal amounts are 3' (1) flip----flip, (2) flop----flop, (3) flip----flop and (4) flop----flip. We have arbitrarily chosen flip (F) ----flop (P) to use as an illustration. Step 2. Elongation of primer to synthesize the complementary (C) strand and the duplex replication form (Rf). Endonuclease nicking (arrow) would then occur opposite the 3' terminus of the parental strand resulting in hairpin transfer to the complementary strand. The enzyme has not been isolated. If the molecule is not nicked, it is possible to open up the molecule and form a double length concatamer which could self-prime at the 3' terminus. Replicating pools containing dimer length concatameric single-stranded DNA have been reported as replicative intermediates (62, 63, 64). Step 3. DNA polymerase fills in the gap at 3' terminus of parental molecule probably using normal repair synthesis. A terminal protein could initiate DNA synthesis at this point or at step 6. Step 4. The 3' terminal hairpin is reformed and displacement synthesis occurs, resulting in the synthesis of a new duplex molecule as in Step 2. The formation of the hairpin loop at this step requires the melting out of base paired nucleotides and may involve a protein. The configuration of the termini will depend on which of the terminal configurations initiated synthesis. No attempt has been made to draw the DNA molecule to scale. The middle portion of the molecule is about 4500 nucleotides in length.

Rhodes has proposed a model for H-1 replication (26). This is a unidirectional "semi-discontinuous" model. It utilizes his electron microscope data, the observed dimer concatamers and his radioisotope labeling experiments. It does, however, require RNA primers for which, at present, there is no evidence. The rolling hairpin model has also been proposed (61). It utilizes a 3' hairpin loop to make a double-stranded Rf molecule. By unfolding and refolding the terminal palindromic sequences, a concatamer is produced that continuously doubles in length. Thus, dimers and longer concatamers

are required. Nuclease cleavage at any palindromic site would release a genome-length molecule for encapsidation. As discussed above, dimer and longer concatamers have been reported both during autonomous and AAV infections. In AAV infections, there is some evidence that dimeric concatamers can be chased into unit length strands (62, 63, 64). However, the involvement of dimer molecules as obligatory intermediates has not been established. The large pre-existing pool of monomeric molecules and the possibility of interconversion between monomers and dimers in radioisotope labeled experiments has clouded the confirmation of the dimer intermediate. There are many questions left unanswered and many experiments to be conducted. More details on the replication of these viruses, so dependent on the cell machinery for replication, may aid in understanding both viral and cellular replication.

REFERENCES

1. Kilham, L., and Oliver, L. J. Virology 7:428-437, 1959.
2. Rabson, A. S., Kilham, L., and Kirschstein, R. L. J. Natl. Cancer Inst. 27:1217-1223, 1961.
3. Toolan, H. W., Dalldorf, G., Barclay, M., Chandra, S., and Moore, A. E. Proc. Natl. Acad. Sci. U.S.A. 46:1256-1258-, 1960.
4. Abinanti, F. R., and Warfield, M. S. Virology 14:288-289, 1961.
5. Bates, R. C., Storz, J., and Reed, D. E. J. Inf. Dis. 126:531-536, 1972.
6. Johnson, R. H. Res. Vet. Sci. 6:466-471, 1965.
7. Crawford, L. V. Virology 29:605-612, 1966.
8. Atchison, R. W., Castro, B. C., and Hammond, M. W. Science 194:754-756, 1965.
9. Hoggan, M. D., Blacklow, N. R., and Rowe, W. P. Proc. Natl. Acad. Sci. U.S.A. 55:1467-1471, 1966.
10. Salzman, L. A. In: Molecular Biology of Animal Viruses, Vol. 2 (Ed. D. P. Nayak) Marcel Dekker, New York, 1978, pp. 539-587.
11. Harrap, K. A., and Payne, C. C. Advances in Virus Research 25:273-355, 1979.
12. McGeoch, D. J., Crawford, L. V., and Follet, E. A. C. J. Gen. Virol. 6:33-40, 1970.
13. Koczot, F. J., Carter, B. J., Garon, C. F., Rose, J. A. Proc. Natl. Acad. Sci. U.S.A. 70:215-219, 1973.
14. Salzman, L. A., Jori, L. A. J. Virol. 5:114-122, 1970.
15. Crawford, L. V., Follet, E. A. C., Burton, M. G., and McGeoch, P. J. J. Gen. Virol. 4:37-46, 1969.
16. Salzman, L. A., White, W. L., and Kakefuda, T. J. Virol. 7:830-835, 1971.
17. Bourguignon, G. J., Tattersall, P. J., and Ward, D. C. J. Virol. 20:290-306, 1976.
18. Chow, M. B., and Ward, D. C. In: Replication of Mammalian Parvoviruses. (Eds. D. C. Ward and P. Tattersall), Cold Spring Harbor, New York, 1978, pp. 205-217.

19. Salzman, L. A., and Fabisch, P. J. Virol. 30:946-950, 1979.
20. Majaniemi, I., Tratschin, J. D., and Siegel, G. Fifth International Congress of Virology, Strasburg, August, 1981 (abstracts) p. 366, 1981.
21. Rhode, S. L. J. Virol. 21:694-712, 1977.
22. Hauswirth, W. W., Berns, K. I. Virology 78:488-499, 1977.
23. Singer, I. I., Rhodes, S. L. J. Virol. 21:724-731, 1977.
24. Astell, C. R., Smith, M., Chow, M. B., and Ward, D. C. Cell 17:691-703, 1979.
25. Maxam, A. M., and Gilbert, W. Proc. Natl. Acad. Sci. U.S.A. 74:560-563, 1977.
26. Rhode, S. L., III, and Klaassen, B. J. Virol. 41:990-999, 1982.
27. Salzman, L. A., Fabisch, P., Parr, R., Garon, C., Wali, T. J. Virol. 27:784-790, 1978.
28. Banerjee, P. T., Rothrock, R., and Mitra, S. J. Virol. 40:118-125, 1981.
29. Rose, J. A., Berns, K. I., Hoggan, M. D., and Koczot, F. J. Proc. Natl. Acad. Sci. U.S.A. 64:863-869, 1969.
30. Lusby, E., Fife, K. H., and Berns, K. I. J. Virol. 34:402-409, 1980.
31. Wolter, S., Richards, R., and Armentrout, R. W. Biochem. Biophys. Acta 607:420-431, 1980.
32. Salzman, L. A., and McKerlie, L. J. Biol. Chem. 250:5583-5588, 1975.
33. Salzman, L. A., and McKerlie, L. J. Biol. Chem. 250:5589-5595, 1975.
34. Rhode, S. L. J. Virol. 17:659-667, 1976.
35. Rhode, W. L. J. Virol. 25:215-223, 1978.
36. Salzman, L. A., White, W. L., and McKerlie, L. J. Virol. 10:573-577, 1972.
37. Salzman, L. A., and Fabisch, P. J. Gen. Virol. 39:571-574, 1978.
38. Tattersall, P., Crawford, L. V., and Shatkin, A. J. Virol. 12:1446-1456, 1973.
39. Rhode, S. L. J. Virol. 14, 791-801, 1974.
40. Gunther, M., and May, P. J. Virol. 20:86-95, 1976.
41. Lavelle, G. A., and Li, A. T. Virology 76:464-467, 1977.
42. Bratosin, S., Laub, D., Tal, J., and Aloni, V. Proc. Natl. Acad. Sci. U.S.A. 76:42-89, 1969.
43. Gunther, M., and Revet, B. In: Replication of Mammalian Parvoviruses (D. C. Ward and P. Tattersall) Cold Spring Harbor, New York, 1978, pp. 469-477.
44. Revie, D., Tseng, B. Y., Grafstrom, R. H., and Goulian, M. Proc. Natl. Acad. Sci. U.S.A. 76:5539-5593, 1979.
45. Rhode, S. L. J. Virol. 25:215-223, 1978.
46. Lavelle, G., Patch, C., Khoury, G., and Rose, J. J. Virol. 16:775-782, 1975
47. Challberg, M. D., Desidero, S. V., and Kelly, T. J., Jr., Proc. Natl. Acad. Sci. U.S.A. 77:5105-5109, 1980.
48. Pratt, D., Laws, P., and Griffith, J. J. Mol. Biol. 82:425-439, 1974.
49. Kollek, R., Tseng, B. V., and Goulian, M. J. Virol. 41:982-989, 1982.
50. Pritchard, C., Stout, E. R., and Bates, R. C. J. Virol. 37:352-362, 1981.
51. Castro, B. C., Atchison, R. W., and Hammon, W. M. Virology 3252-59, 1967.
52. Richardson, W. D., Carter, B. J., and Wesphal, H. Proc. Natl. Acad. Sci. U.S.A. 77:931-935, 1980.

53. Myers, M. W., Laughlin, C. A., Jay, F. T., and Carter, B. J. J. Virol. 35:65-75, 1980.
54. Janik, J. E., Huston, M. M., and Rose, J. A. Proc. Natl. Acad. Sci. U.S.A. 78:1925-1929, 1981.
55. Rose, J. A. In: Comprehensive Virology (Eds. H. Conrat and R. R. Wagner) Plenum, New York, Vol. 3, 1974, pp. 1-61.
56. Hoggan, M. D., Thomas, G. F., and Johnson, F. B. In: Possible Episomes in Eukaryotes, (Ed. L. G. Silvestri), North Holland Publishing Co., Amsterdam, 1973, pp. 243-249.
57. Cheung, A. K.-M., Hoggan, M. D., Hauswirth, W. W., and Berns, K. I. J. Virol. 33:739-748, 1980.
58. Handa, H., Sheroki, K., and Shimojo, H. Virology 82:84-92, 1977.
59. Hauswirth, W. W., Berns, K. I. In: Replication of Mammalian Parvoviruses (Eds. D. C. Ward and P. Tattersall) Cold Spring Harbor, New York, 1978, 257-268.
60. Cavalier-Smith, T. Nature 250:467-478, 1974.
61. Tattersall, P., and Ward, D. C. Nature (London) 263:106-109, 1976.
62. Straus, S. E., Sebring, E., and Rose, J. Proc. Natl. Acad. Sci. U.S.A. 73:742-746, 1976.
63. Handa, H., Shimojo, H., and Yamaguchi, K. Virology 74:1-15, 1976.
64. Hauswirth, W. W., and Berns, K. I. Virology 93:57-68, 1979.

13

ORGANIZATION AND STRUCTURE OF RETROVIRUS GENOMES

Inder M. Verma

INTRODUCTION

Retroviruses are a group of viruses with single-stranded RNA genomes which require the obligatory conversion to double-stranded DNA intermediates in order to propagate. This common feature is characteristic of a very large and diverse group of agents present across the evolutionary species from the Northern pike to humans. The pathological mainifestations of infection by retroviruses include a wide variety of tumors and other diseases. Retroviruses can be transmitted both horizontally and vertically. Occasionally, these viruses can also transduce portions of host cellular sequences (1). The structure of retroviral genomes, their mode of replication and pathogenesis have been the subject of numerous reviews (2-9). It is not my intent to review the structure and organization of various viral genomes. Instead, I plan to utilize Moloney murine leukemia virus (Mo-MLV) and Moloney murine sarcoma virus (Mo-MSV) as model systems to illustrate the salient features of retroviruses. For a more extensive and well documented study of retroviruses, the interested reader should consult the recently published monograph on RNA tumor viruses (1).

GENERAL STRUCTURE

The history of retroviruses stretches over seven decades

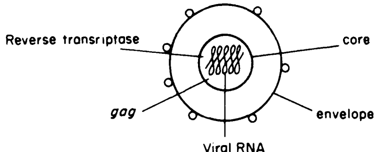
Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

even though much of the knowledge about their structure, replication and genome organization has been gathered only in the last decade. The first retrovirus was isolated in 1910 by Dr. Peyton Rous as a filterable infectious agent from a connective tissue tumor (sarcoma) in a domestic chicken (10). Among the other early isolates first noted as oncogenic agents were classes of viruses capable of inducing mammary carcinomas in laboratory mice (11) and leukemias in newborn mice (12).

All retroviruses are approximately 80-100 nm in diameter, containing nucleoprotein cores of ca. 40-70 nm in diameter and are enveloped by plasma membrane containing virally coded glycoprotein derived from the host cell (13). They have a buoyant density of about 1.16-1.18 in CsCl and are sensitive to heat, lipid solvents and detergents. However, they are relatively resistant to UV- and X-irradiation compared to other animal viruses containing RNA genomes. A common classification of the retroviruses is based upon their electron microscopic appearance (13,14). By far the most common form of retroviruses are referred to as C-type particles which assemble their nucleocapsids just beneath the plasma membrane at the site of budding. Thus the mature form of the virus has a dense, centrally located nucleoid and poorly visible spikes. In contrast, the B-type particles [represented primarily by the mouse mammary tumor virus (M-MTV)] have an eccentrically located core with conspicuous spikes. Noninfectious, nonenveloped forms of retroviruses present either in the cytoplasm or intracisternal spaces have been identified and are referred to as A-type particles, which may be precursors of B- or C-type particles. Primate retroviruses have been classified as D-type particles. They exhibit properties of both B- and C-type particles (16). All retroviruses isolated to date contain a complex of two subunits of single-stranded RNA and low molecular weight RNAs including tRNAs (17,18) Replication of retroviruses requires the participationn of three viral gene products and the genomic RNA. Below, I

shall briefly review the properties of the viral proteins and the structure of a typical replication competent retrovirus.

1) *Viral proteins:* The three viral genes referred to as *gag*, *pol* and *env* encode three viral structural proteins (19; Fig. 1):



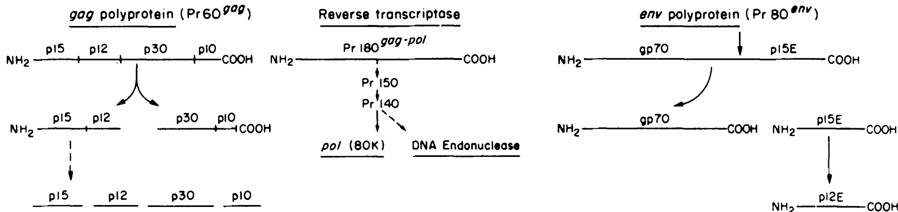


FIGURE 1. Retrovirus structural proteins. Diagrammatic sketch of a retrovirus indicating location of viral genome and structural proteins. The schematic pathway of murine leukemia virus *gag*-polyprotein, *gag*-*pol* polyprotein and *env* polyprotein is shown. The broken lines with arrows indicate the proposed pathway.

a) *Group specific antigen (gag)*: The *gag* polyprotein of approximately 60,000 daltons is encoded by a genome length 38S mRNA species (20, 21). The precursor polyprotein is later (during viral maturation) cleaved to form four viral proteins which make up most of the core of the virus particle. The biochemical events involved in the processing of the polyprotein are not well understood, but in the case of Mo-MLV *gag* Pr60, the cleavage occurs without loss of any amino acid (22). However, in the case of avian sarcoma virus, the cleavage of *gag* Pr76 to mature polypeptides involves loss of amino acids (23.24). The role of the *gag* proteins in the life cycle of retroviruses has yet to be fully appreciated.

b) *Reverse transcriptase (pol)*: Much of the interest in

retroviruses centers around the product of the _pol_ gene region, the reverse transcriptase. It was the discovery of reverse transcriptase by Baltimore (25) and Temin and Mizutani (26) that supplied the major missing link in the life cycle of retroviruses. The transcription of viral genomic RNA into double-stranded DNA is carried out by reverse transcriptase. Temperature- sensitive mutants of avian sarcoma viruses or murine leukemia viruses with lesions in the _pol_ gene fail to establish infection (27-30). Purified reverse transcriptase manifests two synthetic and two degradative activities (3). The synthetic activities are represented by RNA-dependent DNA polymerase, which transcribes ribopolymers into deoxyribopolymers and DNA-dependent DNA polymerase which converts single-stranded DNA polymers to double-stranded DNA polymers. Evidence to date indicates that the two enzymatic activities share a common active site. The degradative activities associated with reverse transcriptase are characterized by ribonuclease H (RNase H) and DNA endonucleases (3,31). RNase H specifically degrades the RNA moiety of an RNA-DNA polymer whereas DNA endonuclease activity introduces nicks in supercoiled DNA. In the case of avian reverse transcriptase, both the synthetic and the RNase H activities reside on the same polypeptide (polypeptide, 70,000 daltons), whereas the DNA endonuclease activity is associated with a 32,000 dalton polypeptide which is cleaved from the larger polypeptide (, mol. wt. 100,000 daltons) (32,33). The and polypeptides are structurally related, as shown by peptide analysis (34). By limited proteolytic digestion of avian reverse transcriptase a 24K polypeptide exhibiting only RNase H activity can be identified (35). Efforts to generate a DNA polymerase activity free of RNase H activity have not been successful. The murine reverse transcriptase consists of a single polypeptide of an average size of 80,000 daltons, which manifests the RNA-dependent DNA polymerase, DNA-dependent DNA polymerase and RNase H activities (36,37). A 40,000 dalton polypeptide with DNA

endonuclease activity has been identified in the virions of murine leukemia viruses (33).

The mature reverse transcriptase is synthesized from a _gag_-_pol_ polyprotein precursor (20,38,39). Since the reading frames for _gag_ and _pol_ genes are different in the, it is likely that splicing events produce the mRNA required for synthesis of _gag_-_pol_ polyproteins. The precise mechanism of cleavage or synthesis of polymerase from _gag_-_pol_ protein precursor is not known, but it appears that in the infected cell the amount of _gag_ polyprotein is nearly ten times that of the reverse transcriptase. Furthermore, no _pol_ mRNA has yet been identified. In the RSV genome, the termination codon for _gag_ gene product is followed by in-frame terminators, and start of the coding sequence for mature reverse transcriptase is 20 nucleotides downstream in the -1 reading frame (40). By contrast, in the Mo-MLV genome, the open reading frame for _pol_ directly follows the single _amber_ termination codon for _gag_ (41). The C-terminal end of the _pol_ gene product in both Rous sarcoma virus (RSV) and Mo-MLV overlaps the amino terminal of _env_ gene product, but in a different reading frame.

c) _Envelope protein (env)_: The envelope protein determines the host range of retroviruses. It is synthesized as a polyprotein of an approximate molecular weight of 60-70,000. A signal peptide is cleaved from the primary translation product (20,22). In the mature Mo-MLV virions, the _env_ polyprotein appears to be further cleaved to generate a peptide of 12,000 daltons (p15E), which remains covalently linked to the remainder of the envelope protein by disulfide bridges. The Mo-MLV _env_ protein is glycosylated and appears to have seven carbo- hydrate chains (42). The _env_ protein of retroviruses is also the target for neutralizing antibodies. The _env_ polyprotein is encoded by a subgenomic mRNA which has viral 5-'sequences spliced on its 5'-end (43-45). The locations of the splice acceptor sites for _env_ mRNAs of avian sarcoma viruses and Mo-MLV have been determined (46).

2) Viral genome: A single-stranded piece of RNA, ranging in length from approximately 3.5-9.5 kb among various virus strains, contains all the genetic information carried in a retrovirus particle (18). It contains a 5'-cap and a 3'-poly(A) tail typical of a eukaryotic mRNA. The viral genome in virions normally exists as a 60-70S dimer complex which, upon heating, results into two monomers of 30-35S and a number of low molecular weight RNAs of host origin (18). Thus, the genome of retroviruses is diploid (47,48). Although the organization of the two subunits in the viral genome complex is not well understood, electron microscopy studies indicate a dimer linkage by hydrogen bonds, involving sequences located near the 5'-end of each subunit (49-51). The retroviral genomic RNA appears to have an extensive secondary structure. With the exception of the primer tRNA. the role and nature of the host low molecular weight RNA is not known. Some species of tRNAs are tightly bound to the genomic RNA near its 5'-end (5,18). There is an exquisite selection for the host tRNA depending on the type of retrovirus. For instance, in the case of avian sarcoma viruses, only $tRNA^{trp}$ is hydrogen bonded to the viral genomic RNA (5), whereas in the case of Mo-MLV it is $tRNA^{pro}$ (52). The tRNA serves as primer to initiate the synthesis of the first strand of viral DNA (see section on Viral DNA Replication). The tRNA also has tight binding affinity to reverse transcriptase (53).

The organization of a typical retroviral subunit RNA and its salient features are shown in Fig. 2. Briefly, starting from the 5'-terminus of Mo-MLV genomic RNA, one encounters the following landmarks:

a) 5'-cap: Like most eukaryotic mRNAs, the retrovirus genome is endowed with a 7-methyl guanyl nucleotide. In the case of Mo-MLV, the exact 5'-cap structure is $m^7G^{5'}ppp^{5'}GmpCp$ (54).

b) Terminal redundancy (R): The genome of RNA tumor viruses is terminally redundant (55-57). The exact size of the terminal redundancy varies from about 10-80 nucleotides

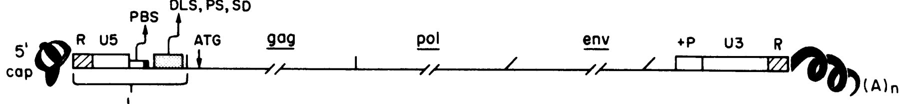


Figure 2. Salient features of retroviral genomes. Schematic diagram of the viral genome indicating important structural features. Starting from the 5'-end: a) 5'-cap; b) R (genomic terminal redundancy); c) U5 (unique 5' sequence); d) PBS (primer-binding site); e) L (the leader region), DLS (dimer linkage site), PS (packaging signal), SA (splice acceptor site); f) gag (group-specific antigen) gene; g) pol (reverse transcriptase) gene; h) env (envelope) gene; i) +P [(+) strand primer region); j) U3 (unique 3'-end); k) R (genomic terminal redundancy); l) poly(A) tail.

among various retroviruses. In the case of Mo-MLV the exact size of the terminal redundancy has not yet been established but appears to be between 44-66 nucleotides (57). The functional significance of genomic terminal redundancy will be discussed during the description of the mechanism of reverse transcription.

c) U5 sequence: Following the terminal redundancy (R) at the 5'-end of the genome, there is a sequence of about 80-100 nucleotides which is referred to as U5 (unique 5') since it occurs only once in the genomic RNA. The junction of R and the U5 sequence contains the presumptive poly(A) addition signal sequence.

d) Primer binding site (PBS): The primer binding site lies adjacent to the U5 sequence (5). In the case of Mo-MLV 18 base pairs at the 3'-end of proline tRNA are hydrogen bonded to the genomic RNA, but the terminal 3'A-OH is free and serves to form the first phosphodiester bond during the initiation of cDNA synthesis.

e) The leader region (L): Between the PBS and the first ATG of the gag polyprotein, there are, in the case of Mo-MSV, 596 nucleotides (58). Dr. John Coffin has

designated this region as the leader region (18), though the leader sequence actually contains the R, U5 and PBS sequences. The length of the L region varies in different viruses. This region most likely contains the following four signals:

i) dimer linkage site (DLS): the sequence by which two subunits of retroviruses are linked.

ii) packaging signal (PS): The region encompassing sequences which allow the packaging of viral genomic RNA in the virions; for Mo-MLV this region appears to be present between restriction sites *Pvu* I [position 419;(41)] and *Pst* I [position 563;(41); R.Mulligan and D. Baltimore,; S. Goff; and C. Van Beveren, personal communication)]. A mutant of RSV, SE21Q1b, lacking viral genomic RNA has a small deletion in the L region (59,60). The deleted region may be analogous to the packaging signal.

iii) splice doner site (SD).

iv) splice acceptor site (SA).

Finally in the case of Mo-MLV, the initiation codon for the glycosylated surface membrane $gP85^{gag}$ may also lie in the L domain (H. Fan, personal communication).

f) *gag*: In Mo-MLV, the coding region for *gag* gene product begins at position 621 and terminates at an amber codon at position 2235. This region could encode a polyprotein of 538 amino acids.

g) *pol*: The *pol* gene product initiates at position 2238 and terminates at 5831, which overlaps with the N- terminal region of the *env* gene. The *pol* gene is capable of encoding a protein of 1199 amino acids but so far only a 80,000 dalton polypeptide (i.e., approximately 700 amino acids) manifesting reverse transcriptase activity has been identified.

h) *env*: The last coding domain of Mo-MLV and other replication competent retroviruses encodes the *env* protein. The amino terminal of *env* gene product actually overlaps with C-terminal of *pol* gene, but it is translated in a different reading frame. The *env* gene product of Mo-MLV

begins at position 5777 and terminates at position 7774, just before the region purported to be involved in priming the synthesis of second strand (+) of DNA. There are seven canonical glycosylation sites (asn-x-ser or thr) present in the gp70 moiety of the *env* polyprotein of Mo-MLV and three in RSV.

A novel protein R has been reported to be present between *env* polyprotein and the (+) strand primer region (61). It now appears that this R region represents only the 16 amino acid C-terminus of the *env* polypeptide.

i) *± strand primer region (+P)*: The specific sequence in the region distal to the 3' end of the *env* gene varies considerably among various retroviruses. In general it has no coding capacity and has also been called 3' noncoding region (3'-NT). For the purposes of this review, I have designated this region as (+P) which is confined between position 7775 (end of *env* gene) and position 7815 (begining of U3 region) in the case of Mo-MLV. What is conserved in this region is a purine-rich tract (58,62,63), which is often well conserved among retroviruses. This region has been implicated in the initiation of the synthesis of the (+) strands of DNA. There is no proof, however, since the primer for the (+) strand remains elusive.

j) *U3 sequence (U3)*: This is an unique sequence near the 3'-end of viral RNA which contains important regulatory elements for transcription. In the case of Mo-MLV,the first eleven nucleotides are repeated inversely at the 3'-end of the U5 sequence, thus forming inverted repeats within each long terminal repeat (LTR)(58,62-65). An extensive variation of size (ranging from 5 to 28 nucleotides) and sequence among various retroviruses occurs in this region (66-68). The size of the U3 region of Mo-MLV is about 370-450 nucleotides, whereas in the case of avian sarcoma or leukemia viruses it is only 150-170 nucleotides. The U3 region of M-MTV appears to be well over 1,200 nucleotides (67). There is an extensive variation in the U3 sequences of retroviruses within the same species. For instance, the

U3 regions of Mo-MLV and AKR-MLV show considerable differences, but the sequences in the control elements are essentially identical (68). The U3 region contains the transcriptional control signals (discussed in more detail in the LTR section). In most retroviruses, the U3 region does not appear to encode a polypeptide but in M-MTV, a large open reading frame capable of encoding a 36,000 dalton polypeptide can be identified (67,69,70). However, to date no polypeptide encoded by the U3 region has been identified in infected cells.

k) Terminal redundancy (R): Since the genome of retroviruses is terminally redundant, the last segment of the viral genomic RNA at the 3'-end contains the same R sequences present at the 5'-end.

l) Poly(A): Like most eukaryotic mRNAs, the retro- viral genomic RNA as well as subgenomic mRNAs, has a poly(A) tail at the 3'-end (71). The poly(A) addition is not template directed and its size varies in different viruses.

STRATEGIES OF VIRAL GENOMES

The genomes of retroviruses are highly promiscuous. They recombine with other viral or cellular sequences with high frequency (17). Although RNA tumor viruses require the function of their three structural genes for replication, these functions can also be supplied in *trans*. Furthermore, the viral genomes can be rescued in the envelope of a helper retrovirus or other envelope virus (e.g., vesicular stomatitis virus), a phenomenon referred to as pseudotyping (72). Thus defective viral genomes can be propagated in the presence of competent helper viruses. The most common form of replication-deficient retroviruses studied in detail are the ones which have acquired normal cellular sequences. Often acquisition of cellular sequences imparts to these viruses the ability to induce neoplasia and transform cells in vitro (8,73). Acqusition of cellular sequences, however, usually occurs at the expense of viral structural genes (18,73,74). Not surprisingly, therefore, most acutely oncogenic viruses are deficient for replication.

What is the minimal amount of retroviral information that is indispensable for its propagation? It appears that U_3RU_5 [collectively referred to as long terminal repeat)], the primer binding site, the presumptive leader sequence, and (+) strand priming site are essential. Genomes bearing lesions in these regions are nonconditionally defective. There are, however, reports that some LTR functions can be supplied in *trans* (75).

During the last few years, largely due to the advent of recombinant DNA technology, extensive information on the structure of retroviral genomes has been obtained. Basically there are three major viral genomic organizations:

a) Replication competent viruses containing intact *gag*, *pol*, and *env genes.*

b) Replication-defective viruses containing cellular sequences and portions of viral structural genes. This class also includes viruses which have undergone recombination with endogenous viral sequences as well as deletion of portions of structural genes.

c) Replication-competent viruses containing recombinant viral genomes.

Some examples of each class of viral genome are shown in Fig. 3. The structure and properties of replication-competent typical murine leukemia viral genomes have already been described above. A major oddity in this class is RSV, which is replication competent and is capable of transforming chick embryo fibroblasts in vitro and induce neoplasia in vivo. The transforming gene of RSV referred to as *src* has been acquired from the host cellular genome (18,73). A spliced subgenomic 21S mRNA encodes the transforming protein, $pp60^{src}$ (46). In RSV the cellular sequences are located between the end of the *env* gene and the (+) strand priming site, without affecting the function of any structural or control element.

The enormous diversity in the genome organization of retroviruses is manifested by acutely transforming oncogenic viruses (18,73). The oncogenic sequences (*onc*) are acquired

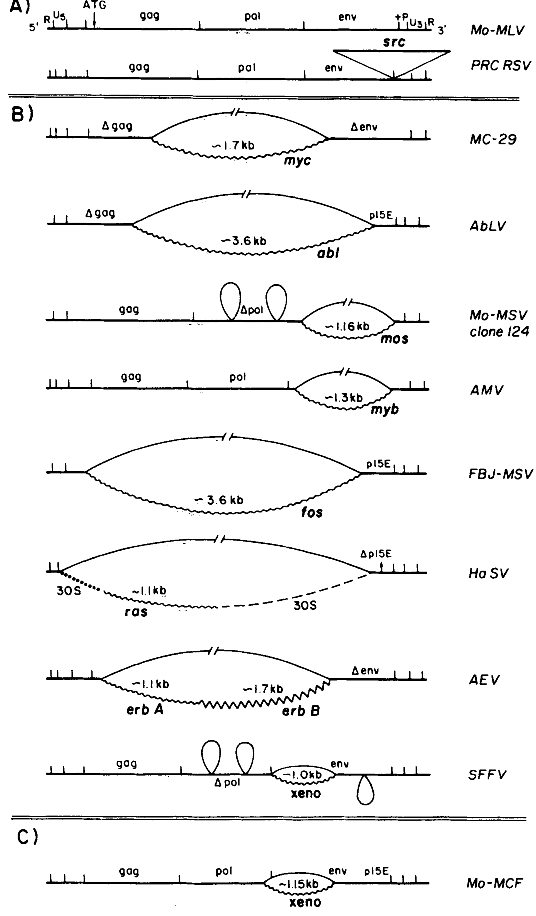


FIGURE 3. Strategies of viral genomes. Schematic diagram depicting the organization of various viral genomes. A) Replication competent viruses: Mo-MLV, PR-C RSV (Prague-C Rous sarcoma virus); the symbols on the genome are as described in Fig. 2; ATG indicates the initiation codon for the gag polyprotein; src, transforming gene of PR-C RSV. B) Replication defective oncogenic viruses: The salient features of the viral genome shown in Fig. 3A are indicated by lines. gag, pol and env indicate truncated viral genes. ( ) indicates deletions, ( ) myc (oncogene of avian myelocytomatosis virus, MC-29); ( ) abl [oncogene of Abelson leukemia virus (AbLV)]; ( ) mos [oncogene of Moloney mouse sarcoma virus (Mo-MSV)]; ( ) myb [oncogene of avian myeloblastosis virus (AMV)]; ( ) fos [oncogene of FBJ murine osteosarcoma virus (FBJ-MSV); ( ) 30S sequence in the Harvey sarcoma virus (HaSV); ( ) ras (oncogene of HaSV); ( ) erb A and ( ) erb B [oncogenes of erythroblastosis virus (AEV)]; ( ) xenotropic sequences. C) Replication

competent viruses containing recombinant viral genomes: Mo-MCF (Moloney mink cell focus-forming virus) ( ) xenotropic sequences. Note that recombination actually takes place in 3'-terminus of the *pol* gene and extends into the *env* gene.

from the host cellular genome by some unknown mechanism. Several oncogenic sequences with counterparts in the normal cellular genome have been isolated and described in detail (73). The host cellular sequences can recombine in every structural gene. The most frequent recombination occurs in the *gag* gene, often making a *gag*-onc hybrid-transforming protein. Examples in this class include Abelson murine leukemia virus (AbLV)(76), MC-29 (77), Y-73 (78), URI (79), PRCII (80), Fujinami sarcoma virus (81), and several feline sarcoma viruses (82). The hybrid *gag*-*onc* protein is presumably synthesized from an mRNA structurally equivalent to genomic RNA. It is not apparent what role the *gag* portion of the hybrid protein plays in the induction of neoplasia. Preliminary indications from work done with AbLV suggest that oncogenic properties of the transforming protein are maintained in the absence of large portions of *gag*-related sequences (J. Wang and D. Baltimore, personal communication). In other cases, perhaps the *gag*-portion of the hybrid protein is required for its proper anchorage in the cell or proper initiation of the protein.

In the case of avian myeloblastosis virus (AMV) and Mo-MSV, the viral genome displays a versatility of accommodating cellular sequences at the expense of *env* genes. In the case of AMV, the *gag* and *pol* genes are essentially intact (83), while in Mo-MSV the entire *gag* but only portions of the *pol* gene are retained (58). In another oncogenic virus, E26, containing the same *myb* oncogene (the transforming gene of AMV or E26) present in AMV, a *gag*-*onc* fusion protein has been reported (84). The biogenesis of the transforming gene of Mo-MSV has been extensively analyzed. Fig. 4 shows the nucleotide sequences at the

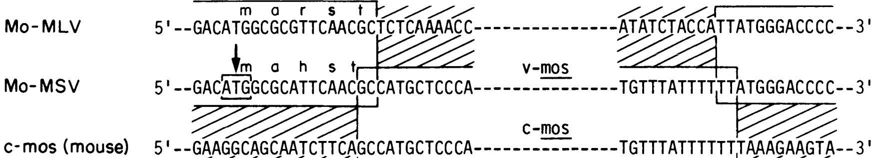


FIGURE 4. The 5' and 3' recombination junctions between Mo-MLV, Mo-MSV and c-*mos*. The 5' and 3' recombination junctions between Mo-MLV and c-*mos* leading to the formation of the v-*mos* gene sequence are shown. The open boxies ([]) denote homology whereas the hatched bars ( ) indicate nonhomology. The ATG (below the arrow) designates the position of the predicted first amino acid (methionine) of the v-*mos* and *env* gene product of amino acid. Code: m, methionine; a, alanine; r, arginine; s, serine; t, threonine, and h, histidine.

junction of viral and acquired cellular sequences. The Mo-MSV transforming gene (*mos*) is 1125 nucleotides long and encodes a protein of 374 amino acids. Four out of five NH_2-terminal residues are contributed by the helper viral *env* gene. Thus the Mo-MSV transforming gene product is a hybrid between *env* and cellular *mos* sequences (74). It is interesting to note that neither the expression nor any product of c-*mos* (cellular homologue of *mos*) sequences has been observed in a variety of different tissues and cell lines. The murine c-*mos* sequence can, however, be transcribed if a viral promoter is attached to its 5'-end (85). Perhaps the c-*mos* is a pseudogene and requires viral control elements for expression. The substitution of the viral *env* gene with c-*mos* sequences apparently allows the synthesis of subgenomic mRNA species using the splice signals of the *env* gene.

The genome organization of FBJ murine osteosarcoma (FBJ-MSV) virus presents yet another novel recombination event (86). In this case the *onc* sequences (referred to as *fos*) are acquired at the expense of the entire *gag* and *pol* genes and the gp70 portion of the *env* gene. Only the LTRs and a portion of the *env* gene, that code for p15E poly-

peptide of the parental virus, are retained (C. Van Beveren, personal communication). A single species of RNA of an average size of 3.4 kb, presumably encoding the *fos* protein has been identified (86).

Another set of complex recombinant events involving at least three different genes led to the formation of Harvey and Kirsten strains of murine sarcoma viruses. The helper virus strains of HaLV and KiLV recombined with both c-*ras*^{Ha} and c-*ras*^{Ki} gene and rat endogenous virus-like 30S RNA (VL-30S RNA) sequences to generate the Harvey or Kirsten oncogenic viruses (87,88). The oncogenic *ras* sequences are apparently located on the 5'-side of any residual viral coding domain and are expressed as nonfused protein (89).

In the genome of avian erythroblastosis virus reside two *onc* genes, one of which, *erb* A, is expressed as a fused *gag*-*onc* protein while the other domain, *erb* B, which is downstream from *erb* A, is translated independently from a spliced subgenomic mRNA (90). The two oncogenes are not related and manifest different biological activities.

The spleen focus-forming virus (SFFV) is defective for replication and causes erythroleukemia in adult mice. The genome of SFFV has been extensively analyzed by heteroduplex analysis and more recently by restriction endonucleases (91, 92). A portion of the *env* gene (ca 0.8-0.9 kb) of SFFV is substituted with *env* gene sequences of an endogenous MuLV. These sequences are generally referred to as mink cell focus-forming (MCF) specific sequences. The SFFV genome has, in addition, undergone a deletion of 600 nucleotides in the region of the *env* gene coding for p15E, two small deletions in the *pol* gene and a small deletion in the *gag* gene. The *env* gene product of SFFV is a 55 kd glycoprotein (gp55) as compared to the 70 kd glycoprotein of the helper MuLV (93,94). The gp55 is most likely encoded by a subgenomic spliced mRNA analogous to normal *env* mRNA and is implicated in transformation by SFFV.

A novel type of genome organization is observed in dual-tropic MCF MuLV's (95). These viruses are replication

competent but have an expanded host range due to substitution of portions of the env gene sequences (96-98). In the case of Moloney MCF (Mo-MCF) the substitution encompases at most 1159 bp, beginning in the carboxy terminus of the pol gene and extending to the middle of the env gene. The Mo-MCF env gene product is 29 amino acids shorter than the parental Mo-MuLV env gene product (99). The precise source of the substituted sequence is not known, but is believed to have originated from the env gene of a xenotropic MuLV (97).

The various examples of the strategies of viral genomes bespeak of the plasticity of the retoviral genome. These viruses do not appear to be fastidious about the size or the location of foreign sequences present amidst them. This remarkable adaptability perhaps allows the retroviruses to entrap such a wide variety of oncogenic sequences.

VIRAL DNA REPLICATION

The conversion of viral genomic RNA into single- and double-stranded viral DNA is carried out by the reverse transcriptase. Within hours after infection of recipient host cells, viral DNA complementary [(-) strand] to the genomic RNA can be detected. Small amounts of integrated viral DNA (proviral DNA) have been reported 8-12 hours after infection. The amount of unintegrated, and perhaps integrated viral DNA, increases for at least 24 to 48 hrs after infection (4,100). A single provirus encoding one subunit of a replication-competent genome is sufficient to direct the production of high titers of infectious virus in appropriate host cells. Several forms of viral DNA have been identified in the cytoplasm of the infected cells. They include covalently closed supercoiled DNA, double-stranded circular DNA, double-stranded linear DNA with two genome length strands, double-stranded DNA where the (-) strand is full-length and the (+) strand is in small pieces, and single-strand (-) DNA. Furthermore, restriction endonuclease data obtained with molecularly cloned unintegrated viral DNAs indicate the presence of circular DNA with

one or two long terminal repeats (101). Other forms of viral DNA with rearrangements have also been found. For a detailed and thorough description of the nature, structure and kinetics of synthesis of viral DNA, I refer the reader to the RNA tumor virus monograph (101). I do, however, wish to describe briefly the fascinating and extremely complex strategy the retroviruses employ to synthesize viral DNA from the RNA template. The precise details of the mechanism of reverse transcription, in particular the structure of the intermediate molecules, remain largely unknown. However, over the last few years enough experimental evidence has emerged to give broad outlines of this obligatory step in the life cycle of retroviruses. The starting template and the end product of reverse transcription are shown in Fig. 5

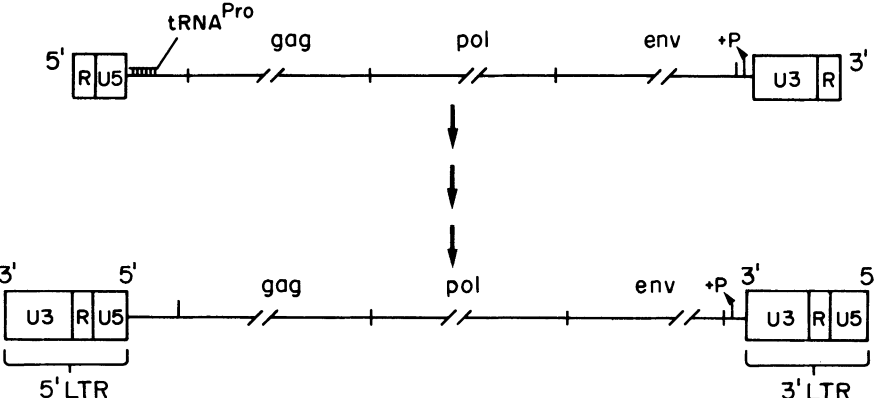


FIGURE 5. Structure of the end product of reverse transcription. The symbols have been described in Fig. 2.

while the various steps involved in this process are shown in Fig. 6. In any model of reverse transcription, the following two cardinal requirements need to be taken into account:

1) The integrated and some unintegrated forms of viral DNA are 500-600 nucleotides larger than the genomic RNA (102-106);

2) The termini of the double-stranded viral DNAs are redundant (102). The seven steps shown in Fig. 6 fulfill the

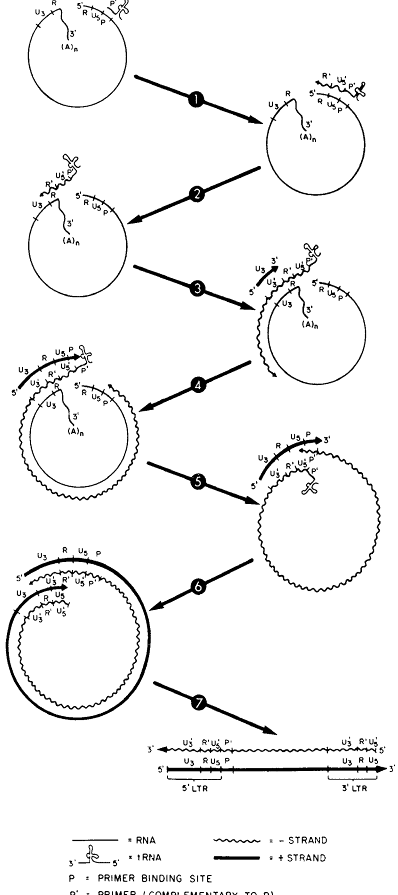


FIGURE 6. Proposed model for the mechanism of reverse transcriptase.

requirement of being able to generate these crucial features. Reverse transcription of Mo-MLV genomic RNA is used in this model, but the general outline should also apply to other retroviral genomes. This model is very similar to that proposed by Gilboa et al. (107).

<u>Step 1</u>: Viral DNA synthesis initiates at the 3'-OH end of the tRNA primer (P') located near the 5'-end of the genomic RNA (5). In the case of Mo-MLV, 18 nucleotides at the 3'-end of primer $tRNA^{Pro}$ are hydrogen bonded to the genomic RNA (region P). The 3'-A-OH of the primer tRNA is

not hydrogen-bonded and is located 146 nucleotides from the 5'-cap nucleotide of the genomic RNA (64). It forms a phosphodiester bond with the first deoxynucleotide triphosphate (dATP in the case of Mo-MLV) and the synthesis of the complementary DNA [cDNA or (-) strand DNA] proceeds until it reaches the 5' nucleotide of the genomic RNA.

Step 2: The 145 nucleotide long cDNA [also referred to as "strong stop" DNA (108), R+U5 in Fig. 5] covalently linked to the primer $tRNA^{Pro}$ dissociates and hybridizes to terminally redundant sequences (R) at the 3'-end of the genomic RNA. The synthesis of (-) strand then ensues from the 3'-end of the RNA toward its 5'-end.

Steps 3 and 4: A 600 bp DNA fragment (300 nucleotides in the case of avian retrovirus and ,1200 nucleotides in the case of Mo-MTV) of opposite polarity (+) strand can be observed following the synthesis of 0.5-1.0 kb of the (-) strand (100,109,110). The (+) strand DNA transcript spans the U3 and R regions of the viral genome. In addition, it includes the U5 region and the primer (P') sequence of the $tRNA^{Pro}$. Several laboratories have observed the synthesis of the (+) strand transcript during reverse transcription carried out in vivo or in vitro. Gilboa et al. (107) and Taylor and Hsu (111) have shown that the 18 nucleotide long tRNA primer region (P') is reverse transcribed. The (-) strand cDNA transcripts continue to elongate and transcribe the primer binding site (P) located on the genomic RNA.

Steps 5-7: The region P [located on the (+) strand] transcribed from tRNA primer can base pair with the P' sequences [present on (-) strand] as shown in step 5 of the model (fig. 6). the (-) strand then continues to be copied from the (+) strand and the (+) strand is elongated by using the (-) strand as the template leading to the formation of double-stranded viral dna molecules shown in step 7. the double-stranded viral dna in the case of mo-mlv is about 535 nucleotides longer than the viral genomic RNA (68). Furthermore, the final viral DNA transcript shown in the step 7 has a terminal redundancy of 589 nucleotides.

Despite the elegance and substantial experimental support of the proposed model of reverse transcription, several lingering uncertainties persist. For instance a) How does the strong stop DNA (U5 + R) dissociate from the viral genome to allow the synthesis of cDNA to continue at the 3'-end of the genomic RNA? b) What is the signal or primer for the synthesis of the (+) strand DNA? c) What is the role of the diploid structure of the viral genome during reverse transcription? Does each subunit get transcribed independently? d) How are the primer tRNA and poly(A) residues removed?

ANATOMY OF A LONG TERMINAL REPEAT (LTR):

An unique and fundamentally very important feature of the structure of unintegrated viral DNA and proviral DNA was unraveled by Dr. J. Taylor, Dr. H. Varmus, and their colleagues (102-105), when they observed terminal redundancy. Soon the acronym long terminal repeat (LTR) was coined, largely to distinguish it from genomic terminal redundancy (R) (112). As described in Figure 5, the end product of reverse transcription of Mo-MLV genomic RNA is double-stranded DNA with a terminal redundancy of 589 nucleotides. Over the last couple of years the complete nucleotide sequence of several LTRs from different retroviruses has been deduced (see Appendix, ref. 1). Despite the size variation, the overall salient features of all LTRs are largely identical. The structure of a typical Mo-MLV LTR displaying the important landmarks is shown in Fig. 7. They are:

a) An LTR consists of U5, R and U3 regions of the genome.

b) The 5'-capped nucleotide (nucleotide No. 1 in ,Fig. 2) is located 145 nucleotides from the 3'-boundary of the LTR.

c) The proposed RNA polymerase II initiation site ("Hogness-Goldberg box" or "TATA box") is located at positions -25 to -31.

d) A sequence CCAAT ("CAT box") found to be associated

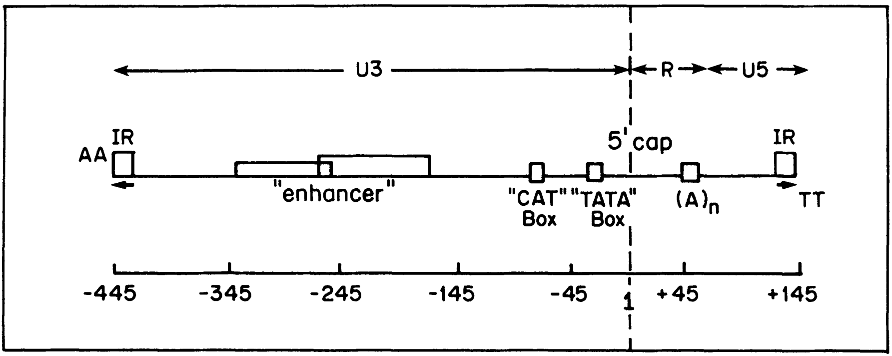


FIGURE 7. Salient features of long terminal repeat. The numbers refer to Mo-MLV LTR (68).

with the RNA polymerase II initiation site is located at around position -80, in analogy to other eukaryotic genes (113).

e) The purported eukaryotic poly(A) addition signal (AATAAA) is located at positions +45 to +52 from the 5'-cap nucleotide.

f) The Mo-MLV LTR contains an 85 bp duplication with a 10 bp overlap starting at positions -183 to -267 and -257 to -342 (Fig. 7). The size of the duplication varies from 50-110 bp and some Mo-MLV LTRs have only one copy of the repeat sequence (68). The precise role of this internal duplication is not understood, but these sequences are akin to the "enhancer" sequences observed in other systems.

g) The LTR of unintegrated Mo-MLV DNA has a 13 bp inverted repeat at its termini. In the proviral DNA the inverted repeat is only 11 nucleotides since the ultimate and penultimate 2A and 2T residues are presumably lost during integration.

h) No viral proteins encoded in the LTR have yet been observed, though cRNA to M-MTV LTR has been shown to direct the synthesis of a 36K protein, in vitro (69).

i) The 5'-end of the 5'-LTR and the 3'-end of the 3'-LTR of proviral DNA (integrated viral DNA) are bounded by a direct repeat of host cellular sequences. The number of

nucleotides in the direct cellular repeat differs with different viruses, but generally the number for each class of retroviruses is constant.

j) The 3' boundary of the 5' LTR is flanked by the primer binding site (PBS) while the 5' end of the 3' LTR is adjoined by the polypurine rich purported site of origin of (+) strand DNA synthesis (62).

Though the precise role of the LTRs in the life cycle of retroviruses remains unknown, clearly their structure suggests important biological functions. The LTR contains the control elements for both initiation and termination of transcription. Its termini are almost certainly involved in the process of integration into host chromosomes. Studies with site-directed mutations in the LTR should delineate the precise role of the LTRs in the viral life cycle.

INTEGRATION

Retroviruses integrate into the host chromosomal DNA via a DNA intermediate. Unlike other RNA- or DNA-containing animal viruses, chronically infected cells produce retroviruses without cell death. Once the viral DNA is integrated in the host genome, it behaves like a cellular gene. Despite an extensive knowledge of the structure of the viral DNA, little insight has been gained regarding the mechanism of integration into the host chromosome. There are, however, a few tenets by which the integration of retroviral DNA is governed. They are:

1) The integrated viral DNA (provirus) is colinear with the viral genome.

2) The proviral DNA is terminally redundant and the structure is U3RU5....U3RU5.

3) Two A residues at the 5'-end and two T residues at the 3'-end of the unintegrated viral DNA are missing from the proviral DNA (9,63,68,101).

4) The proviral DNA is flanked by a direct repeat of 4-6 nucleotides of cellular DNA. The duplicated cellular sequences are however present only _once_ in the pre-integration site (68,114-116).

5) The integration of retroviral DNAs appears to be random. There are, however, exceptions where the viral DNA has a selected site of integration in the host chromosome, as in the activation of the cellular oncogene, c-*myc* (117-119), and c-*erb* (H. J. Kung, personal communication).

What are the nature and structure of the viral DNA molecule prior to integration? The three most likely candidates are: a) linear double-stranded DNA; b) circular DNA with two LTRs blunt end ligated at the termini; and c) circular DNA with only *one* LTR. All three forms of DNA can be identified in virus-infected cells. In addition, other kinds of molecules with frequent rearragements can also be found (9). A mechanistically interesting set of viral DNAs found in the infected cells contain inverted LTRs. Shoemaker et al. (120) first reported variants of molecularly cloned Mo-MLV DNA in which one of the LTRs integrated into viral DNA, thus creating a molecule in which the two LTRs are not adjacent to each other. Van Beveren et al. (68) have reported encountering a similar situation in molecularly cloned Mo-MSV DNA. Fig. 8 describes the structure and nucleotide sequence analysis of one such molecule. The two LTRs are separated by a stretch of 321 nucleotides (Fig. 8, structure II). Several points can be noted from the structure of inverted LTR in the molecularly cloned Mo-MSV DNA:

i) The 5'-LTR and 321 nucleotides downstream from it are inverted (structure II) with respect to their expected orientation (structure I);

ii) The 321 nucleotides between the two LTRs correspond to sequences representing the tRNA binding site and downstream viral sequences. The inversion presumably occurred during viral DNA synthesis and is not an artefact of cloning;

iii) Two terminal A residues from the U3 region of 5'-LTR and two terminal T residues from the U5 region of 3'-LTR are lost;

iv) A 4 bp inverted repeat (CTCG...CGAG, Fig. 8) is

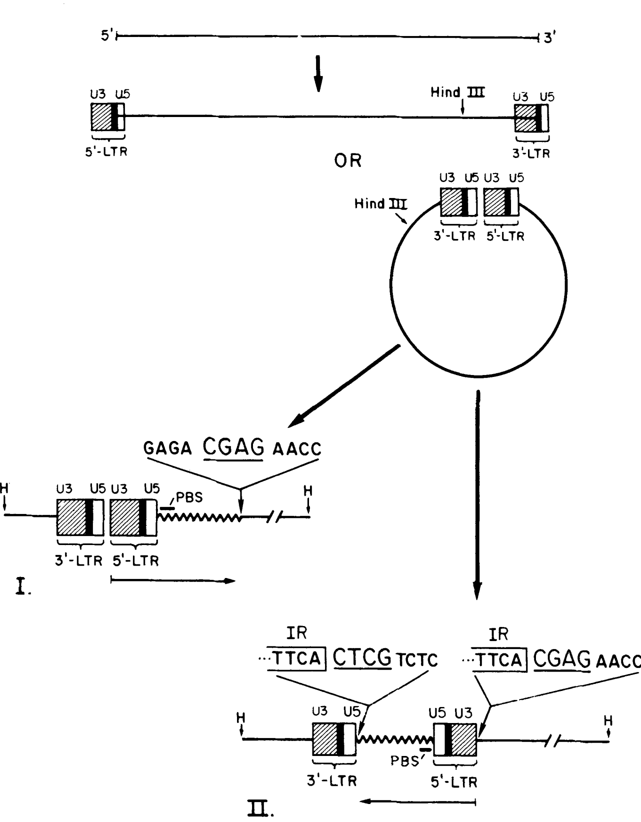


FIGURE 8. Structure of recombinant clones of Mo-MSV clone 124. The genomic RNA gives rise to double-stranded DNA having one (not shown) or two LTRs in either linear or recombinant clones of supercoils, isolated from productively infected cells and cloned at the unique *Hind* III site, is generally that shown in structure I. In one recombinant clone, pMSV-12, the region indicated by the arrow, encompassing the 5' LTR and the 321 nucleotides downstream from it [including the tRNA primer binding site (PBS)], has been inverted to give the arrangement shown in structure II. The target tetranucleotide CGAG, found once in structure I, is found repeated at the end of the inversion in structure II.

observed at the U3 terminus of the inverted 5'-LTR and at the junction of the additional 321 bp sequences and U5 terminus of the 3'-LTR. Thus in contrast to the proviral DNA, where the flanking cellular sequences at the site of integration are present as a direct repeat, in this case the flanking viral sequences are inverted because the molecule

has integrated into itself; and

v) Finally, the 4 bp inverted repeat is present once in the parental DNA at the site of inversion.

Taking into account various salient features of the viral DNA integration into the host chromosome, Shoemaker et al. (120) and Varmus (9) have proposed models for integration analogous to the one suggested by Shapiro for transposition and replication of bacteriophage Mu and other transposable elements (121). These models suggest staggered cuts in the viral DNA and chromosomal DNA at the integration site. Fig. 9 illustrates how the three major forms of unintegrated AKR viral DNA may integrate into the host chromosome.

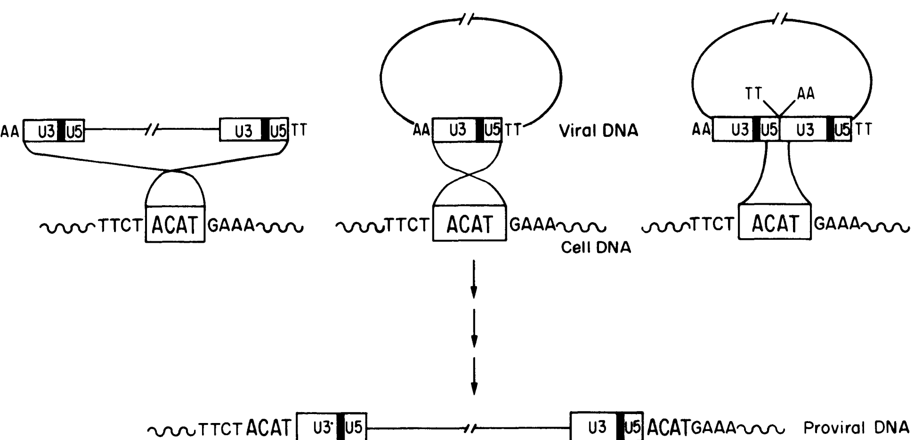


FIGURE 9. Possible mode of integration of retroviral DNA into host chromosomes. The example of AKR-MuLV into host chromosome is illustrated (68). The 4 bp direct cellular repeat in the case of AKR-MuLV clone 623 is not perfect since one base pair has undergone a change in the 5' LTR (ACAA instead of ACAT); for the sake of clarity I have not shown this change.

Though the unique structure of the unintegrated DNA is perhaps the most crucial aspect of integration, it is apparently not sufficient for creating colinearity of the genome in the proviral DNA. Linear or circular forms of

proviral DNA transfected into eukaryotic cells do not integrate in the typical U3RU5...U3RU5 comformation (122, 123) Instead, the viral DNA appears to integrate in a permuted form. Only in the retrovirus-infected cell does the proviral DNA appear to have its typical diagnostic structure (123). It is likely that some viral protein participates in the process of integration. The most likely candidate is the DNA endonuclease associated with the reverse transcriptase which may be required to introduce the staggered cuts in the viral DNA and perhaps in the cellular target DNA. Maturation of viral proteins may also be required, since proviral DNA transfection of producer cells does not guarantee integration in a proper orientation. It would be interesting to perform experiments where viral DNA containing one or two LTRs is introduced into the cells together with virus.

ANALOGY TO TRANSPOSONS

The organization of retroviral DNAs immediately suggests a parallel with the structure of some transposable genetic elements. A simple inspection of the structural features of various transposable elements suggests that retroviruses belong to a family of mobile genetic elements which includes Tn9 in bacteria (124), Ty-1 in yeast (125), and copia elements in *Drosophila* (126-128). All of these share the following general features shown in Fig. 10A:

i) An internal coding domain of one to several kbp,

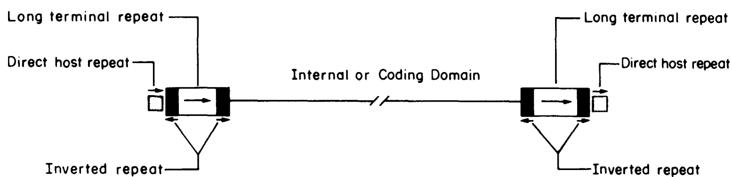
ii) A large, terminal direct repeat usually of a couple of hundred bp,

iii) Small inverted repeats at the two ends of the large terminal repeats,

iv) A short direct repeat of host DNA sequences surrounding the element, and;

v) Only one copy of the duplicated host cellular sequences at the integration site prior to integration.

There are other groups of mobile elements such as classes that include Tn3 of *E. coli*, P factors of *Drosophila*, the variable surface glycoprotein genes of



Mobile Genetic Element	Internal Domain (kb)	Inverted Repeat	Long Terminal Repeat	Host Duplication
Retroviruses				
Murine leukemia virus	7.6	13/13	517	4
AKR leukemia virus	7.6	13/13	626	4 (5)
Murine sarcoma virus	4.6	13/13	589	4
FBJ murine osteosarcoma virus	2.8	13/13	617	4
Mouse mammary tumor virus	6.5	6/ 6	1330	6
Avian sarcoma virus	8.5	10/13	270-350	6
Avian leukosis virus	7.8	12/15	270-350	6
Reticuloendotheliosis virus	7.5	5/ 5	569	5
Drosophila				
Copia	4.5	13/17	276	5
297	5.7	3/ 3	412	4
412	6.5	5/ 6	481	4
Mdg 1	6.3	8/ 6	444	4
Yeast				
Ty-1	5.1	2/ 2	338	5
E. coli				
Tn9	1.1	18/23	768	9

FIGURE 10. General features of a retrovirus-like transposable element; table indicating the characteristics of various retrovirus-like transposable genetic elements.

trypanosomes, mating type cassettes of yeast, immunoglobulin genes, the Mu-1 phage, elements of maize, etc. which have features distinguishable from the retroviral type of transposable elements.

The characteristics of transposable genetic elements related to retroviruses are tabulated in Fig. 10B. The size of the direct terminal repeat is variable, ranging from approximately 270 bp in the case of copia and avian sarcoma virus to 1330 bp in Mo-MTV. The LTRs of murine retroviruses are generally 500-600 bp nucleotides in length. It almost seems that two avian LTRs combined to make <u>one</u> murine C-type retrovirus LTR and two murine LTRs combined to make <u>one</u>

mammary tumor virus LTR. The size of the inverted repeats at the end of the direct repeats varies from as much as 2/2 nucleotides in the case of yeast Ty-1 element to 13/13 nucleotides of murine LTR and 18/23 nucleotides in Tn9. As pointed out by several investigators, the 5' and 3' termini of all LTRs and related mobile genetic elements share a minimum of TG...CA (115,124). In the case of unintegrated viral DNAs, however, the 5' terminus is AATG and the 3' terminus is CATT. Since no analogous stage exists in other elements, it is not possible to show if they also contained additional AA or TT sequences prior to integration. The number of bases in the duplicated sequences that flank the different elements also vary. It ranges from 4 in the case of murine LTRs to 9 nucleotides in Tn9. It appears that the number of direct repeats is dictated by the viral DNA since the virus grown in heterologous cells has the same number of duplicated cellular sequences. There is, however, one exception where the same virus can generate either a 4 or 5 bp direct cellular repeat (68). The size of the internal coding domain is extremely variable, spanning from 1.1 kb in Tn9 to nearly 8.0 kb in the case of ASV. The remarkable diversity of the size of retroviral genomes is consistent with the idea of a variable internal coding region.

Despite the obvious structural homology with transposable genetic elements, there is no clear-cut case of transposition of retroviruses from one site in the chromosome to another site. Perhaps the best example is the integration of LTR in the viral DNA shown in Fig. 8. It is possible that the large number of endogenous proviruses in mice were generated either by transposition or infection of the germ line. The retroviruses should be very useful to study the mechanism of transposition, since they offer an intermediary step prior to integration. In contrast the bacterial transposon Tn9 is found only as an integrated moiety.

LIFE CYCLE OF RETROVIRUSES

Prior to the discovery of the reverse transcriptase, it was generally acknowledged that retroviruses did not behave like other animal viruses with RNA genomes since their propagation was hindered by agents affecting DNA synthesis. In 1964, Howard Temin had proposed the "provirus hypothesis" which postulated that the RNA genome was converted to a DNA provirus, which then served as a template for synthesis of viral RNA (129). The biological experiments in which a form of viral genome appeared to be labeled with BudR combined with the more decisive finding of reverse transcriptase have borne out the major tenets of the "provirus hypothesis" (130). The preoccupation with reverse transcriptase and DNA replication have obscured the role of other viral proteins in the life cycle of the retroviruses.

Fig. 11 displays a diagrammatic sketch of the major events involved in the establishment of infection and propagation of retroviruses. They can be divided into six categories:

a) Adsorption and penetration: Retroviruses are believed to be introduced into cells via plasma membrane receptors which specifically interact with the glycoproteins in the viral envelope. Neither the biochemical nature of the receptors nor the precise mode of entry (plasma membrane fusion or endocytosis) of the virus particles is known. Data on the early events involved in the infection of retroviruses are woefully lacking.

b) Reverse transcription: Early after infection (one to two hours) viral genomic RNA is transcribed into DNA by reverse transcriptase. The mechanism of reverse transcription and the forms of viral DNA synthesized in the infected cell have been discussed in detail in the section on viral DNA replication.

c) Integration: Little is known about the transport of the viral DNA into the nucleus or the precise form of viral DNA involved in integration. Integrated viral DNA (provirus) can be found in the cells 10-12 hours following

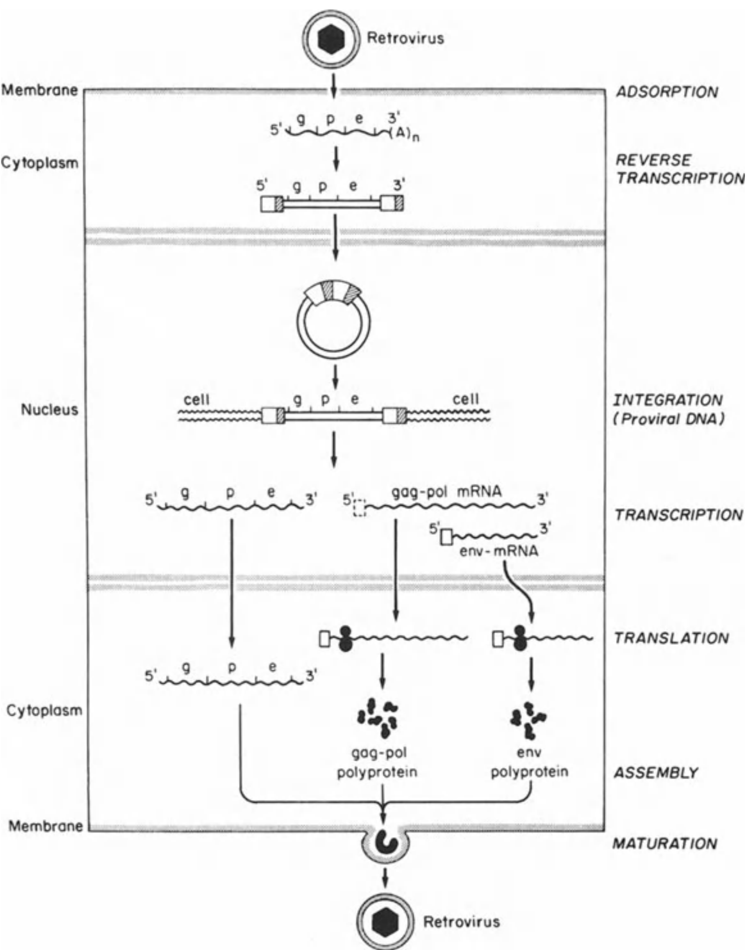


FIGURE 11. Life cycle of a retrovirus.

infection. Details of the structure of the viral DNA or possible mechanisms of integration have also been described in other sections.

d) Transcription: The proviral DNA behaves like a eukaryotic transcriptional unit. The host DNA-dependent RNA polymerase II appears to be involved in catalyzing the transcription, presumably recognizing the canonical sequence TATAAA present 25 to 31 nucleotides upstream from the 5'-cap nucleotide (131,132). In vitro transcription of molecularly

cloned retroviral DNA, in soluble systems has identified the correct initiation sites (133-135). The proviral DNA is transcribed in its entirety and only 38S RNA molecules can be detected in the nucleus. The processing of the viral RNA to generate spliced subgenomic mRNAs presumably occurs prior to transport in the cytoplasm. Transcriptionally active proviruses are usually hypomethylated and appear to be present in the DNase I-sensitive region of the chromosome (136). Hypermethylated proviruses can be made transcriptionally active by addition of 5-azacytidine which prevents methylaton of CpG dinucleotides (137). The propagation or spread of retroviruses must depend on the strength of its transcriptional promoter located in the LTR. This step in the life cycle of retroviruses is likely to be responsible for the amplification of viral genomes in the infected cell.

e) Translation: Polyadenylated transcripts of genomic and subgenomic length have been identified in the infected cell (138). It is not clear if the 35-38S viral *gag-pol* mRNA is spliced. Generally the *gag-pol* mRNA species is the most abundant form of viral mRNA, followed by subgenomic *env* or *src* mRNA species. No *pol* mRNA has yet been identified. The viral mRNAs encode polyproteins which are subsequently processed. Details of viral protein processing have been described in the section on General Structure.

f) Virion assembly and maturation: Like other enveloped viruses, the retroviruses assemble at the plasma membrane and are released from the cell by budding (139). The 35S viral RNA is encapsulated presumably by the *gag* proteins and the nucleoprotein core migrates to the cell surface. The immature virions are believed to contain the viral RNA subunits, *gag* polyprotein, *gag-pol* polyprotein, *gag* proteins, reverse transcriptase and tRNA primer. Small amounts of viral mRNAs may also be packaged in the virion. Recent experiments by several investigators suggest that the viral RNA contains a packaging signal which is presumably removed from the mRNAs by splicing events. The exact

details of the assembly of the various viral components remains largely a mystery.

The released virions undergo intracellular maturation. The *gag-pol* protein is processed to form mature viral structural proteins. The viral RNA subunits, presumably after budding, dimerize and anneal to the tRNA primer. Again, little is known about the sequence of events involved in the maturation of infectious virion particles. The release of infectious virus completes the retrovirus life cycle.

FUTURE RESEARCH TRENDS

The multifaceted studies of retroviruses remind me of the old Indian tale of the four blind men and the elephant. Each man touches a different part of the elephant and describes it as a wall, or a big rope or perhaps giant columns. Cancer biologists view a retrovirus as a tool to induce and study oncogenesis. Molecular biologists use it as a model to study the expression and regulation of eukaryotic genes. Recombinant DNA researchers consider it as an indispensable tool to synthesize cDNA transcripts. It is fast becoming the choice vector to deliver foreign genes into eukaryotic cells. I perceive progress on all fronts in the next few years, but certain questions will most likely receive greater attention, for instance: i) the precise mechanism of reverse transcription with particular emphasis on determining the nature of the primer for (+) strand DNA; ii) what form of viral DNA is integrated; iii) the biochemistry and enzymology of integration; iv) Is integration random- random or is some site specificity (maybe a short sequence of 4-10 nucleotides) involved? In this context, we will perhaps witness more c-*onc* genes activated by "promoter-insertion"; v) the role of host cellular sequences on regulation of viral gene expression; vi) the search for novel *oncogenes.* Here I would like, however, to interject my personal bias that the number of *novel* oncogenes will be limited. We are already witnessing the isolation of the same *oncogene* in a number of different independent isolates;

vii) intensive investigation of both the product and the function of *onco*genes; viii) expression of c-*onc* genes in a variety of tissues and cell types; ix) role of c-*onc* genes in human malignancy, and x) utility as vectors to carry genetic information into somatic and perhaps germ-line cells. With such marvelous prospects, retroviruses should continue to dominate and generate excitement in modern biology.

ACKNOWLEDGMENTS

I am very grateful to Drs. Tom Curran, Renato Dulbecco, Walter Eckhart, Tony Hunter, Rolf Müller, Jack Rose and Charles Van Beveren for their comments and constructive criticims. I thank Carolyn Goller and Bart Sefton for help in preparing the manuscript and S. Kronz and J. T. Simon for illustrations. Support was from grants from the National Cancer Institute and the American Cancer Society.

REFERENCES

1. The Molecular Biology of Tumor Viruses, Vol. III (Eds. R. A. Weiss, N. Teich, H. E. Varmus and J. M. Coffin), Cold Spring Harbor Laboratory,1982.
2. Temin, H. M. and Baltimore, D. In: Adv. Virus Res., Vol. 17 (Eds. K. M. Smith and M. A. Lauffer) Academic Press, Inc., New York, 1972, pp. 129-186.
3. Verma, I. M., Biochim. Biophys. Acta **473**:1-38,1977.
4. Weinberg, R. A., Biochim. Biophys. Acta **473**:39-51,1977.
5. Taylor, J., Biochim. Biophys. Acta **473**:57-71,1977.
6. Bishop, J. M. and Varmus, H. E. In Cancer, Vol. 2 (Ed. F. F. Becker), Plenum Publishing Corp., New York, 1975, pp. 3-48.
7. Weinberg, R. A., Ann. Rev. Biochem. **49**:197-226,1980.
8. Bishop, J. M., Scientific American **246**:80-92,1982.
9. Varmus, H. E., In Mobile Genetic Elements (Ed. J. A. Shapiro), Academic Press, New York,1982.
10. Rous, P., J. Exp. Med. **13**:397-413,1911.
11. Bittner, J. J., Science **84**:162-164,1936.
12. Gross, L., Proc. Soc. Exp. Biol. Med. **76**:27-32,1951.
13. Teich, N., In Molecular Biology of Tumor Viruses, Part III (Eds. R. A. Weiss, N. Teich, H. E. Varmus and J. M. Coffin) Cold Spring Harbor Laboratory,1982.
14. Bernhard, W., Cancer Res. **20**:712-727,1960.
15. Sarkar, N. H., Moore, D. and Nowinski, R., In RNA Viruses and Host Genomes in Oncogenesis (Eds. P. Emmelot and P. Bentvelzen), North Holland, Amsterdam, 1972, pp. 71-79.
16. Fine, D., and Schochetman, G., Cancer Res. **38**:3123-3139,1978.
17. Coffin, J. M., J. Gen. Virol. **42**:1-26,1979.
18. Coffin, J. M., In Molecular Biology of Tumor Viruses, Part III (Eds. R. A. Weiss, N. Teich, H. E. Varmus and J. M. Coffin) Cold Spring Harbor Laboratory, 1982, Chapter 4.
19. Baltimore, D., Cold Spring Harbor Symp. Quant. Biol. **39**:1187-1200,1974.
20. Dickson, C., Eisenman, R., Fan, H., Hunter, E., and Teich, N., In Molecular Biology of Tumor Viruses, Part III (Eds. R. A. Weiss, N. Teich, H. E. Varmus and J. M. Coffin) Cold Spring Harbor Laboratory, 1982, Chapter 6.
21. Arcement, L. J., Karshin, W. I., Naso, R. B., and Arlinghaus, R.B., Virology **81**:284-297,1977.
22. Oroszlan, S. and Gilden, R. V., In Molecular Biology of RNA Tumor Viruses (Ed. J. R. Stephenson), Academic Press, New York, 1980, pp. 299-344.
23. Eisenman, R. N. and Vogt, V.M., Biochim. Biophys. Acta **473**:187-239,1978.
24. Eisenman, R. N., Burnette, W. N., Zucco, F., Diggelmann, H., Heatern, P. Tsichlis, P. and Coffin, J. M., In Biosynthesis, Modification and Processing of Cellular and Viral Polyproteins (EDS. G. KOCH AND D. Richter) Academic Press, New York, 1980.

25. Baltimore, D., Nature (London) **226**:1209-1211,1970.
26. Temin, H. and Mizutani, S., Nature (London) **226**:1211-1213,1970.
27. Linial, M. and Mason, W. S., Virology **53**:258-273,1973.
28. Verma, I. M., Mason, W. S., Drost, S. D. and Baltimore, D., Nature (London)**251**:27-31,1974.
29. Verma, I. M., Varmus, H. E. and Hunter, E., Virology **74**:16-29,1976.
30. Lai. M.T., Verma, I. M., Tronick, S. R. and Aaronson,S. A., J. Virol. **27**:823-825,1978.
31. Grandgnett,D. P., Vora, A. C. and Schiff, R. D., Virology **89**:119-132,1978.
32. Grandgnett, D. P., Gerard, G. F. and Green, M., Proc. Natl. Acad. Sci. USA **70**:230-234,1973.
33. Gerard, G. F. and Grandgnett, D. P., In Molecular Biology of RNA Tumor Viruses (Ed. J. R. Stephenson) Academic Press, New York,1980.
34. Gibson, W. and Verma, I. M., Proc. Natl. Acad. Sci. USA **71**:4991-4994,1974.
35. Lai, M. T. and Verma, I. M., J. Virol. **25**:652-663, 1978.
36. Verma, I. M., J. Virol. **15**:843-854,1975.
37. Mölling, K., J. Virol. **18**:418-425,1976.
38. Arlinghaus, R. B., Jamjoom, G. A., Kopchick, J. and Naso, R. B., In Cell Differentiation and Neoplasia (Ed. G. F. Saunders), Raven Press, New York,1978, pp. 271-295.
39. Witte, O. N. and Baltimore, D., J. Virol. **26**:750-761, 1978.
40. Schwartz, D., Tizard, R. and Wilbert, W., Cell (in press),1982.
41. Shinnick, T. M. Larner, R. A. and Sutcliffe, J. G., Nature (London) **293**:543-548,1981.
42. Rosner, M. R., Tung, J.-S., Hopkins,N. and Robbins, P.W., Proc. Natl. Acad. Sci. USA **77**:6420-6424,1980.
43. Rothenberg,E., Donoghue, D. J. and Baltimore, D., Cell **13**:435-451,1978.
44. Fan, H. and Verma, I. M., J. Virol. **26**:468-478,1978.
45. Weiss, S. R., Varmus, H. E. and Bishop, J. M., Cell **12**:983-992,1977.
46. Hackett, P. B., Swanstrom, R., Varmus, H.E. and Bishop, J. M., J. Virol. **41**:527-534,1982.
47. Billeter, M. A., Parsons, J. T. and Coffin, J. M., Proc. Natl. Acad. Sci. USA **72**:3560-3564,1974.
48. Beemon, K. L., Faras, A. J., Haase, A. T., Duesberg, P.H. and Maisel, J. E., J. Virol. **17**:525-537,1976.
49. Kung, H.-J., Bailey, J. M., Davidson, N., Vogt, P. K., Nicolson, M. O. and McAllister, R. M.,Cold Spring Harbor Symp. Quant. Biol. **39**:827-834,1975.
50. Bender, W. and Davidson, N., Cell **7**:595-607,1975.
51. Bender, W., Chien, Y.-H., Chattopadhyay, S., Vogt, P. K., Gardner, M. B. and Davidson, N., J. Virol. **25**: 888-896,1978.
52. Waters, L. C., Biochem. Biophys. Res.Commun. **65**:1130, 1975.
53. Panet, A., Haseltine, W. A., Baltimore, D., Peters,

G., Harada, F. and Dahlberg, J., Proc. Natl. Acad. Sci. USA **72**:2535-2539,1975.
54. Rose, J. K., Haseltine, W. A. and Baltimore, D., J. Virol. **20**:324-329,1976.
55. Schwartz, D. E., Zaemcnik, P. C. and Weith, H. L., Proc. Natl. Acad. Sci. USA **74**:994-998,1977.
56. Stoll, E., Billeter, M. A., Palmenberg, A. and Weissmann, C., Cell **12**:57-72,1977.
57. Coffin, J. M., Hageman, T. C., Maxam, A. M. and Haseltine, W. A., Cell **13**:761-773,1978.
58. Van Beveren, C., vanStraaten, F., Galleshaw,J. A. and Verma, I. M., Cell **27**:97-108,1981.
59. Linial, M., Medeiros, E. and Hayward, W. S., Cell **15**: 1371-1381,1978.
60. Shank, P. R. and Linial, M., J. Virol. **36**:450-456, 1980.
61. Sutcliffe, J.G., Shinnick, T., Green, N., Liu, F.-T., Niman, H. L. and Lerner, R. A., Nature (London) **287**: 801-805,1980.
62. Sutcliffe, J. G., Shinnick, T. M., Verma, I. M. and Lerner, R. A., Proc. Natl. Acad. Sci. USA **77**:3302-3306,1980.
63. Dhar, R., McClements, W. L., Enquist, L. W. and vande Woude, G. F., Proc. Natl. Acad. Sci. USA **77**:3937-3941, 1980.
64. Van Beveren, C., Goddard,J. G., Berna, A. and Verma, I. M., Proc. Natl. Acad. Sci. USA **77**:3307-3311,1980.
65. Reddy, E. P., Smith, M. J. and Aaronson, S. A., Science **214**:445,1981.
66. Ju, G. and Skalka, A. M., Cell **22**:379-386,1980.
67. Donehower, L. A., Huang, A. L. and Hagen, G. L., J. Virol. **37**:226-238,1981.
68. Van Beveren, C., Rands, E., Chattopadhyay, S. K., Lowy, D. R. and Verma, I. M., J. Virol. **41**:541-556, 1982.
69. Dickson, C. and Peters, G., J. Virol. **37**:36-47,1981.
70. Kennedy, N., Knedlitschek, K., Gronver, B., Hynes, N. E., Heelich, P., Michalides, R. and Van Ooyen, A. J. J., Nature (London) **295**:622-624,1982.
71. Gillespie, D., Marshall, S. and Gallo, R.C., Nature New Biol. **236**:227-231,1972.
72. Weiss, R. A. and Wong, A. L., Virology **76**:826-834, 1977.
73. Bishop, J. M. and Varmus, H. E., In Molecular Biology of Tumor Viruses, Part III ((Eds. R. A. Weiss, N. Teich, H. E. Varmus and J. M.Coffin), Cold Spring Harbor Laboratory, 1982.
74. Van Beveren, C., Galleshaw, J. A., Jonas, V., Berns,A.J.M., Doolittle, R.F., Donoghue, D. J. and Verma, I. M., Nature (London) **289**:258-262,1981.
75. Blair, D. G., McClements, W. L. Oskarsson, M. K., Fischinger, R. J. and Vande Woude, G. F., Proc. Natl. Acad. Sci. USA **77**:3504-3508,1980.
76. Goff, S. P., Gilboa, E., Witte, O. N. and Baltimore, D., Cell **22**:777-785,1980.
77. Vennstrom, B., Moscavici, L., Goodman, H. M. and

Bishop, J. M., J. Virol. **39**:625-631,1981.

78. Yoshida, M., Kawai, S. and Toyoshima, K.,Nature (London) **287**:653-654,1980.
79. Wang, L.-H., Hanafusa, H., Holler, M.F.D. and Balduzzi, P. C., J. Virol. **41**:833-841,1982.
80. Ghysdael, J., Neil, J. C. and Vogt, P. K., Proc. Natl. Acad. Sci. USA **78**:5847-5851,1981.
81. Hanafusa, T., Wang, L.-H., Anderson, S. M., Karess, R. E., Hayward, W. S. and Hanafusa, H., Proc. Natl. Acad. Sci. USA **77**:3009-3113,1980.
82. Sherr, C.J., Fedele, F. A., Oskarsson, M., Maizel, J. and Vande Woude, G., J. Virol. **34**:200-212,1980.
83. Perbal, B. and Baluda, M. A., J. Virol. **41**:250-257, 1982.
84. Bister, K., Nunn, M., Boscovici, C., Perbal, B., Baluda, M. A. and Duesberg, P. H., Proc. Natl. Acad. Sci. USA **79**:3677-3681.
85. Oskarsson, M., McClements, W. L., Blair, D. G., Maizel, J. V. and Vande Woude, G. F., Science **207**:1222-1224,1980.
86. Curran, T., Peters, G., Van Beveren, C., Teich, N. and Verma, I. M., J. Virol. **44**:674-682,1982.
87. Ellis, R. W., DeFeo, D, Maryak, J. M., Young, H. A., Shih, T.Y., Chang, E.H., Lowy, D. R. and Scolnick, E. M., J. Virol. **36**:408-420.
88. Young, H. A., Gouda,M. Aml, DeFeo, D., Ellis, R. W., Nagashima, K. and Scolnick, E.M., Virology **107**:89-99, 1980.
89. Chien, Y.-H., Lai, M., Shih, T. Y., Verma, I. M., Scolnick, E. M., Roy-Burman, P. and Davidson, N., J. Virol. **31**:752-760,1979.
90. Privalsky,M. L. and Bishop, J. M., Proc. Natl. Acad. Sci. USA **79**:3958-3962,1982.
91. Bosselman, R. A., van Griensven, L.J.L.D., Vogt, M. and Verma, I. M., Virology **102**:234-239,1980.
92. Linemeyer, D. L., Ruscetti, S. K., Menke, J. G. and Scolnick, E. M., J. Virol. **35**:710-721,1980.
93. Dresler, S., Ruta, M., Murray, M. J. and Kabat, D., J. Virol. **30**:564-575,1979.
94. Ruscetti, S. K., Linemeyer, D., Field, J., Troxler, D. and Scolnick, E. M., J. Virol. **30**:787-798,1979.
95. Hartley, J. W., Wolford, N. K., Old, L. J. and Rowe, W. P., Proc. Natl. Acad. Sci. USA **74**:789-792,1977.
96. Elder, J.H., Gautsch, J. W., Jensen, F. C., Lerner, R. A., Hartley, J. W. and Rowe, W. P., Proc. Natl. Acad. Sci. USA **74**:4676-4680,1977.
97. Bosselman, R.A., van Griensven, L.J.L.D., Vogt, M. and Verma,I. M., J. Virol. **32**:968-978,1979.
98. Lung, M.L., Hering, C., Hartley, J. W., Rowe, W. P. and Hopkins, N., Cold Spring Harbor Symp.Quant. Biol. **44**:1269-1274,1980.
99. Bosselman, R. A., vanStraaten, F., Van Beveren,C., Verma, I. M. and Vogt, M., J. Virol. **44**:19-31,1982.
100. Varmus, H. E., Guntka, R. V., Fan, W., Heasley, S. and Bishop, J. M., Proc. Natl. Acad. Sci. USA **71**:3874-3878,1974.

101. Varmus, H. E. and Swanstrom, R., In The Molecular Biology of Tumor Viruses, Part III (Eds. R. A. Weiss, N. Teich, H. E. Varmus and J. M. Coffin)Cold Spring Harbor Laboratory, 1982, Chapter 5.
102. Hsu, T. W., Sabran, J. L., Mark, G. E., Guntaka, R. V. and Taylor, J. M., J. Virol. **28**:810-818,1978.
103. Sabran, J. L., Hsu, T. W., Yeater, C., Kaji, A., Mason, W. S. and Taylor, J., J. Virol. **29**:170-178, 1979.
104. Shank, P. R., Hughes, S. H., Kung, H.-J., Majors, J.E., Quintrell, N., Guntaka, R. V., Bishop, J. M. and Varmus, H.E.,Cell **15**:1383-1395,1978.
105. Hughes, S. H., Shank, P. R., Spector, D. H., Kihg, H.-J., Bishop, J. M., Varmus, H. E., Vogt, P. K. and Brietman, M. L., Cell **15**:1397-1410,1978.
106. Gilboa, E., Goff, S., Shields, A., Yoshimura, F., Mitra, S. and Baltimore, D., Cell **16**:863-874,1979.
107. Gilboa, E., Mitra, S. W. Goff, S. and Baltimore, D., Cell **18**:930100,1979.
108. Haseltine, W. A. and Baltimore, D., In ICN-UCLA Symposium on Molecular and Cellular Biolgoy, Vol. 14 (Eds. D. Baltimore, A.S. Huang and C.Fred Fox) Academic Press, New York, 1976,pp. 175-213.
109. Verma, I.M., J.Virol. **26**:615-629,1978.
110. Mitra, S. W., Goff, S., Gilboa, E. and Baltimore, D.,Proc. Natl. Acad. Sci. USA **76**:4355-4359,1979.
111. Taylor, J.M. and Hsu, T.W.,J. Virol. **33**:531-534,1980.
112. Bosselman, R. A. and Verma, I. M., J. Virol. **33**:487-493,1980.
113. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P., Science **209**:1406-1414,1980.
114. Majors, J. E. and Varmus, H. E., Nature (London) **289**: 253-258,1981.
115. Hishinuma, F., DeBona, P. J., Astrin, S. and Shalka, A. M., Cell **23**:155-164,1981.
116. Hughes, S. H., Mutschler, A., Bishop, J. M. and Varmus, H. E., Proc. Natl. Acad. Sci. USA **78**:4299-4303,1981.
117. Hayward, W. S., Neel, B. G. and Astrin, S. M., Nature (London) **290**:475-480,1981.
118. Payne, G., Bishop, J. M. and Varmus, H. E., Nature (London) **295**:207-213,1982.
119. Noori-Daloii, M. R., Swift, R. A., Kung, H.-J.. Crittenden, L. B. and Witter, R. L., Nature (London) **294**:574-576,1981.
120. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W. and Baltimore, D., Proc. Natl. Acad. Sci. USA **77**:3932-3936,1980.
121. Shapiro, J. A., Proc. Natl. Acad. Sci. USA **76**:1933-1937,1979.
122. Doehmer, J., Barinaga, M., Vale, W., Rosenfeld, M. G., Verma, I. M. and Evans, R. M., Proc. Natl. Acad. Sci. USA **79**:2268-2272,1982.
123. Verma, I. M., Jones, M., vanStraaten, F., Bosselman, R. A. and Van Beveren, C., In UCLA-CETUS Symposium on

Tumor Viruses and Differentiation, Squaw Valley, USA, 1982.

124. Calas, M. P. and Miller, J. H., Cell **20**:579-595,1980.
125. Farabaugh, P. J. and Fink, G. R.,Nature (London) **286**: 352-356,1980.
126. Dunsmuir, P., Broreir, W. J., Jr., Simon, M. A. and Rubin, G. M., Cell **21**:575-579.
127. Levis, R., Dunsmuir, P. and Rubin, G. M., Cell **21**: 581-588,1980.
128. Kulgushkin, V.V., Iiyin, Y. V. and Georgiev, G. P., Nucl. Acids Res. **9**:3451-3464.
129. Temin, H. M., Natl. Cancer Inst. Monogr. **17**:557-570, 1964.
130. Boettiger, D. and Temin, H.M., Nature (London) **228**: 622-425,1970.
131. Jacquet, M., Groner, Y., Monroy, G. and Hurwitz, J., Proc.Natl. Acad. Sci. USA **71**:3045-3049,1974.
132. Rymo, L., Parsons, J. T., Coffin, J. M. and Weissmann, C., Proc. Natl. Acad. Sci. USA **71**:2782-2786,1974.
133. Yamamoto, T., deCrombrugghe, B. and Pastan, I., Cell **22**:787-797,1980.
134. Fuhrman, S., Van Beveren, C. and Verma, I. M., Proc. Natl. Acad. Sci. USA **78**:5411-5415,1981.
135. Ostrowski, M.C., Berard, D. and Hagen, G. L., Proc. Natl. Acad. Sci. USA **78**:4485-4489,1981.
136. Harbers, K., Schnieke, A., Stuhlman, H., Johner, D. and Jaenish, R., Proc. Natl. Acad. Sci. USA **78**:7609-7613,1981.
137. Groudine, M., Eisenman, R. and Weintraub, H., Nature (London) **292**:311-317,1981.
138. Fan, H. and Baltimore, D., J. Mol. Biol. **83**:93-117, 1973.
139. Yoshinaka,Y. and Luftig, R. B., Proc. Natl. Acad. Sci. USA **74**:3446-3450,1977.

14

HOST CONTROL OF THE REPLICATION OF GROUP I RNA PHAGES

HANNA ENGELBERG-KULKA AND RACHEL SCHOULAKER-SCHWARZ

SUMMARY

In this article we summarize our studies showing the involvement of Escherichia coli proteins in the in vivo replication of the group I RNA phages (MS2, f2 and R17). We used two methods of investigation: in the first method, we followed the replication of RNA phages in the presence of rifampin, a drug which completely inhibits host protein synthesis; in the second method, we followed the effect of a specific bacterial mutation (traD) on the replication of these phages. According to both methods, the replication of only group I RNA phages was inhibited while that of the RNA phage Qβ (group III) was unaffected. This indicates that Qβ and group I RNA phages use different host proteins for their replication. In addition, we showed that the traD mutation also affects the bacterial host's membranes. We discuss the possibility that bacterial membranes are involved in group I phage RNA replication.

INTRODUCTION

The RNA containing phages are variously classified into three to five serological groups (1,2,3). Group I includes phages such as MS2, f2 and R17 which are quite similar to one another primarily in their particle properties (4) and which differ considerably from phage Qβ, a member of group III (5). Though all four phages have been studied intensively, most of the information available on RNA replication comes from studies on Qβ alone (6,7). RNA phage replication has been comprehensively reviewed by Stavis and August (6) and Weissman et al (7). In general, phage RNA replication proceeds in two steps both catalyzed by phage RNA replicase: a) synthesis of minus strands on the parental plus strand template and b) synthesis of progeny plus strands complementary to the minus strands. More detailed information on the stages of phage RNA replication and on the structure of the RNA replicase comes mainly from studies of Qβ RNA replication. The core of Qβ RNA replicase is made up of four subunits (8,9) of which only subunit II is a phage-specified protein, while the other three subunits (I,III and IV) are host specified. Subunits III and IV are the protein synthesis elongation factors Tu and Ts (10) and subunit

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

I is the 30S ribosomal protein S1 (11). In addition, another bacterial protein, host factor F (HF) is required by the purified "core" of Qβ RNA replicase to copy the parental strand _in vitro_ (12,13).

In contrast to Qβ RNA replicase, very little is known about the RNA replicases of group I RNA phages since they are unstable and difficult to purify away from the endogenous template RNA. Only the f2 poly G polymerase has been extensively purified (14). Structurally appearing similar to the core of Qβ replicase, the f2 poly G polymerase also contains three bacterial specified proteins which cosediment with the bacterial proteins found in the Qβ replicase (14); however, the purified f2 polymerase requires the presence of an additional unidentified factor for activity with either f2 or f2 complementary strands as template. This f2 factor is presumably a host protein(s) like the host factor (HF) required for Qβ RNA replication. That the Qβ host factor cannot substitute for the f2 factor (15) indicates that the _in vitro_ synthesis of Qβ and f2 RNAs require different bacterial proteins.

Here we summarize our studies showing the involvement of host proteins in the _in vivo_ replication of group I RNA phages. We relied on two approaches: in the first approach, we followed group I RNA phage replication in the presence of rifampin, a drug which completely inhibits host protein synthesis; in the second approach, we followed the effect of a specific bacterial mutation on the replication of these phages.

RESULTS

I. Effect of rifampin on the replication of RNA phages.

In _Escherichia coli_, the antibiotic rifampin inhibits the initiation of RNA synthesis (16,17) and consequently protein synthesis (18), by acting directly on the β-subunit of the enzyme RNA polymerase (19). Rifampin also interferes with the intracellular growth of RNA phages (20-26) and, in addition, inhibits the release of mature phage particles from the host (21,27). Phage RNA synthesis and release are not affected by the antibiotic when the host possesses a rifampin-resistant RNA polymerase (23,25,27), implying that host specified components participate in phage development.

We examined whether the drug affects phage RNA replication. We studied phage RNA replication by examining (^{3}H)–labeled phage RNA phenol extracted at various times from cells grown with or without rifampin. As previously

reported (28,29), when phage-infected cells are phenol deproteinized for RNA isolation, the first step of RNA replication, the synthesis of minus strands, appears as the conversion of parental RNA into a double-stranded form; the second step, the synthesis of progeny strands, appears as the displacement of the parental RNA from the duplex.

As shown in Fig. 1, in phage-infected bacterial cultures not treated with rifampin, the amount of parental RNA converted into double-stranded form increased until about 15 minutes after infection, when 16 to 32% of injected parental phage RNA was in duplex form. At this time the parental RNA became progressively displaced from the duplex form, and by 40 minutes after infection only 5% of the parental RNA remained in the double-stranded structure.

Adding rifampin (100 µg/ml) to host cultures 30 minutes before infection with (^{3}H)-labeled phage had a remarkably different effect on each of the phages studied. In the case of MS2, injected parental RNA was converted to duplex form almost as much as in the untreated control culture, and by 15 minutes after infection, the amount of parental RNA in duplex form was 80% of the maximum found in the control. Displacement of the parental RNA from the duplex form, however,was completely arrested by the drug. In fact, complete inhibition of the displacement of the parental RNA from duplex form was already achieved when rifampin was added at the time of infection (Fig. 1b) rather than 30 minutes before infection. Thus, in the case of MS2, rifampin barely affects the synthesis of RNA minus strands, but arrests the synthesis of progeny plus strands, causing the parental RNA to remain quantitatively undisplaced from the double-stranded form.

As shown in Fig. 1c, addition of rifampin 30 minutes before infection by phage f2 inhibited the conversion of injected parental RNA into duplex form by 75%. When rifampin was added at the time of infection (Fig. 1d) however, f2 RNA was converted into the double-stranded form and then displaced from the duplex as in the control cultures. Thus, contrary to the effect of rifampin on MS2 RNA replication, the drug mainly inhibited the conversion of f2 RNA into the double-stranded structure, whereas the displacement of f2 RNA from the duplex was almost unaffected. Therefore, the antibiotic seemed to inhibit the synthesis of f2 minus strands and barely affected the synthesis of f2 progeny RNA.

In the case of phage R17, the addition of rifampin 30 minutes before infection inhibited the conversion of parental RNA to the duplex form by

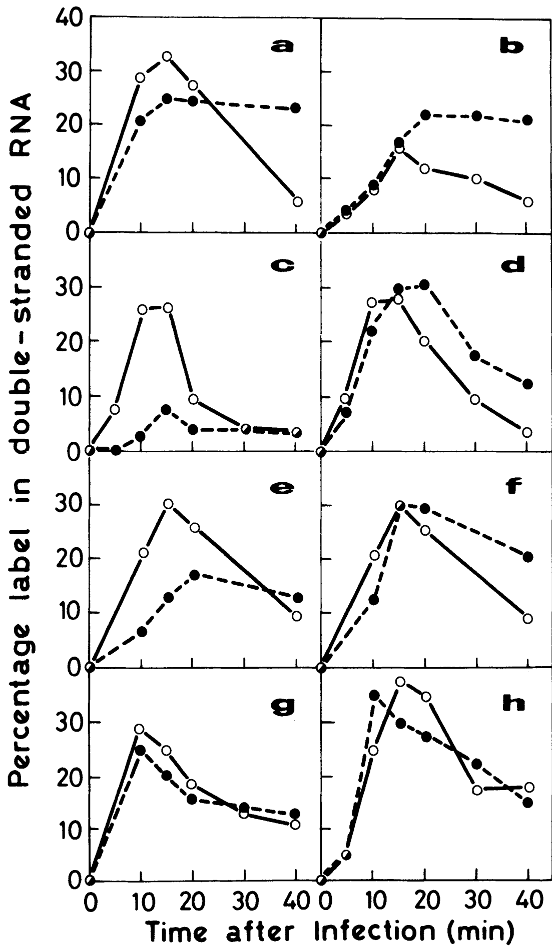


Fig. 1. Effect of rifampin on the amount of phage parental RNA in double-stranded form at different times after infection. E. coli cells were grown to a density of $2x10^8$ cells/ml in a shaking bath at 37^oC and infected at the same temperature with phage containing (^3H)-labeled RNA at a multiplicity of 5. Rifampin (100 μg/ml) was added either 30 min. before infection (a,c,e,g) or at the time of infection (b,d,f,h). At different times after infection the cells were harvested, washed four times with phosphate buffered saline and converted to spheroplasts (54). The RNA was extracted and analyzed for the presence of parental RNA in the double-stranded form by chromatography on cellulose columns according to Franklin (55). The samples of RNA applied to the column contained 4000 to 5000 counts/min. Symbols: (●) treated with rifampin; (O) control, not treated with rifampin; cells infected with MS2 (a,b), f2 (c,d), R17 (e,f), Qβ (g,h).

a maximum value of 50% (Fig. 1e). The drug also inhibited the displacement of the parental RNA from the duplex form by 50% (Fig. 1f). Increasing the concentration of rifampin from 100 μg/ml to 200 μg/ml did not cause an increased inhibitory effect (data not shown).

As seen in Fig. 1g and 1h, rifampin had no effect on the replication of group III phage Qβ RNA.

In summary, the present data clearly show that rifampin inhibits the in vivo replication of phages belonging to group I (MS2, f2, R17), but does not affect the group III phage Qβ RNA replication. Furthermore, the drug has a discriminative effect on the replication of MS2, f2, and R17 RNAs. With the method employed, we showed that rifampin interferes with the synthesis of phage f2 RNA minus strands, but has almost no effect on the synthesis of progeny plus strands. On the other hand, in the case of MS2, rifampin differentially arrests the synthesis of progeny plus strands and barely affects the synthesis of minus strands. In the case of R17, both steps of its RNA replication were affected by the drug, although each step was only partially (approximately 50%) inhibited.

II. Effect of an Escherichia coli traD (Ts) mutation on the replication of group I RNA phages.

Most of the tra cistrons of sex factor F are involved in the synthesis of F pili (30,31,32) which are implicated in the transport of nucleic acids during bacterial conjugation (33) and during male-specific phage infection (34,35,36). Thus, many of the tra mutants are defective in bacterial conjugation and resistant to male-specific phages (RNA phages and filamentous DNA phages). The traD mutants represent a special class of tra mutants. Although they make F pili, they are defective in DNA transport during bacterial conjugation (30,31,37). Furthermore, the traD mutants are resistant to group I RNA phage f2, but not to the group III RNA phage Qβ nor the filamentous DNA phage f1 (30,38).

We previously reported (39) that the traD (Ts) mutant JCFL39, temperature-sensitive in conjugational transfer (37), is also temperature-sensitive for the group I RNA phages (MS2, f2, R17) but not to phage Qβ. In addition, we have shown by temperature shift-up experiments that an intracellular step in the growth of the group I RNA phage MS2 is affected at the high temperature. Therefore, we assumed that the traD mutation, like rifampin, probably also affects the replication of group I RNA phages

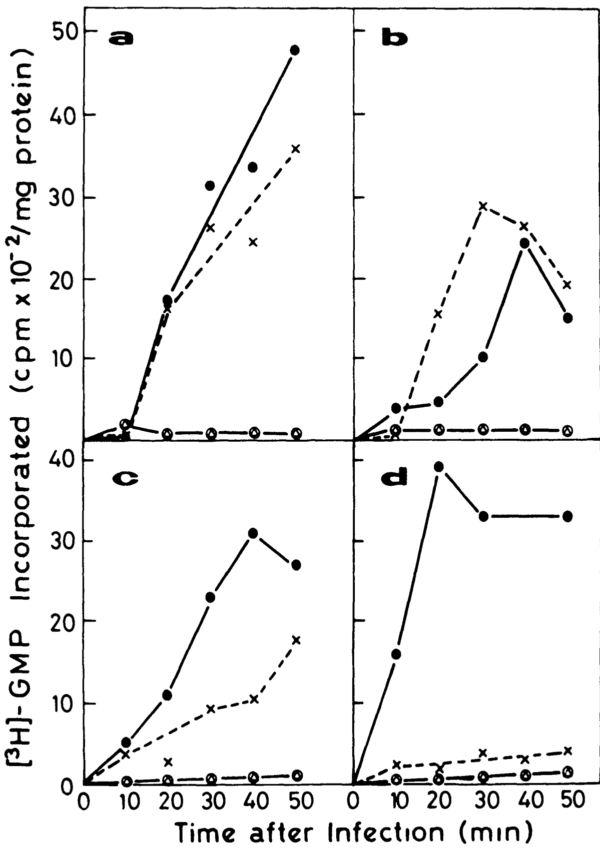


Fig. 2. Phage RNA-polymerase activity in extracts from mutant and parental cells grown and infected with RNA phages at various temperatures. Cells were grown in nutrient broth at different temperatures to a density of $2x10^8$ cells/ml and infected with the phages at a multiplicity of 10. At various intervals after infection, an aliquot containing approximately $5x10^9$ cells was centrifuged, lysed in a 0.1 M Tris-HCI, 10% (wt/vol) sucrose buffer, pH 8.5, by addition of 50 µg/ml lysozyme and a final concentration of 5 mM EDTA. After incubation for 30 min. at 4°C, final concentrations of 0.1% (wt/vol) Brij 58, 10 mM $MgCl_2$ and 25% glycerol were added. The lysate was assayed for phage RNA polymerase activity in a reaction mixture containing 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 4 mM 2-mercaptoethanol, 1 mM each of ATP, UTP, CTP and 10 µM (^{3}H)-GTP ($1.5x10^5$ CPM / nmole), 10 mM phosphoenolpyruvate, 2.5 µg pyruvate kinase, 4 µg actinomycin D, 1 µg rifampin and 0.1 ml of the Brij-lysate (0.1-0.4 mg protein). The reaction mixture was incubated for 20 min at 25°C. The assay measured the incorporation of (^{3}H)-GMP from GTP into acid insoluble materials. Symbols: phage RNA polymerase activity in mutant extracts (x), in parental extracts (•). Controls: (^{3}H)-GMP incorporated by extracts of uninfected mutant (Δ) or parental (o) cells. Qβ-infected cells grown and infected at 37°C (a) or 42°C (b). MS2-infected cells grown and infected at 33°C (c) or 42°C (d).

but not that of Qβ. We studied phage RNA replication by testing the activity of phage RNA polymerase produced in mutant and parental strains. Lysates prepared from cells grown and infected at various temperatures were examined for the incorporation of (^{3}H)-GMP into acid-insoluble material at 25°C in the presence of rifampin and actinomycin D. It can be seen that these drugs inhibit RNA polymerase activity in uninfected cells (Fig. 2). The incorporation of GMP, therefore, represents the activity of phage-specified RNA polymerases. A similar specific activity of Qβ RNA polymerase was observed in extracts of mutant and parental strains infected at low (37°C) or high (42°C) temperature (Figs. 2a and 2b). With respect to group I RNA phages (MS2, f2, R17) RNA polymerase activity was only observed when growth and infection were at 33°C (Fig. 2c). At 42°C, however, very little enzymatic activity was found in mutant extracts, while the elevated temperature stimulated activity in parental extracts (Fig. 2d). Since we found similar results with all three phages, we chose phage MS2 as a representative of group I. At the non-permissive temperature (42°C) the parental RNAs of group I RNA phages penetrate into the mutant host and are even translated (data not shown). This shows that at 42°C, the activity of phage RNA polymerase is the first step in group I RNA phage development which is affected by the mutation.

The lack of MS2 RNA polymerase activity in mutant cells grown at 42°C seemed to indicate that the first step of MS2 RNA replication was affected in the mutant. This was, in fact, the case: as shown in Fig. 3, in mutant cells grown at 42°C and infected at the same temperature with MS2 containing (^{3}H)-labeled RNA, the conversion of the parental RNA into double-stranded form is inhibited.

MS2 replicative intermediates are tightly membrane bound (40). We suspected that at 42°C, the membranes of mutant JCFL39 are changed and as a result, MS2 as well as other members of group I RNA phages are inhibited in the replication of their RNAs. This assumption was supported by the fact that at 42°C, the mutant is more sensitive than the parent to drugs such as acriflavin, rifampin, and deoxycholate whose action is affected by membrane permeability (41-45). We compared membranal proteins of mutant and parental cells grown at low (33°C) and high (42°C) temperature, by SDS gel electrophoresis. There were several differences in the electrophoretic patterns of inner and outer membrane proteins of mutant and parental cells grown at 42°C (Fig. 4). Most remarkable is the absence of

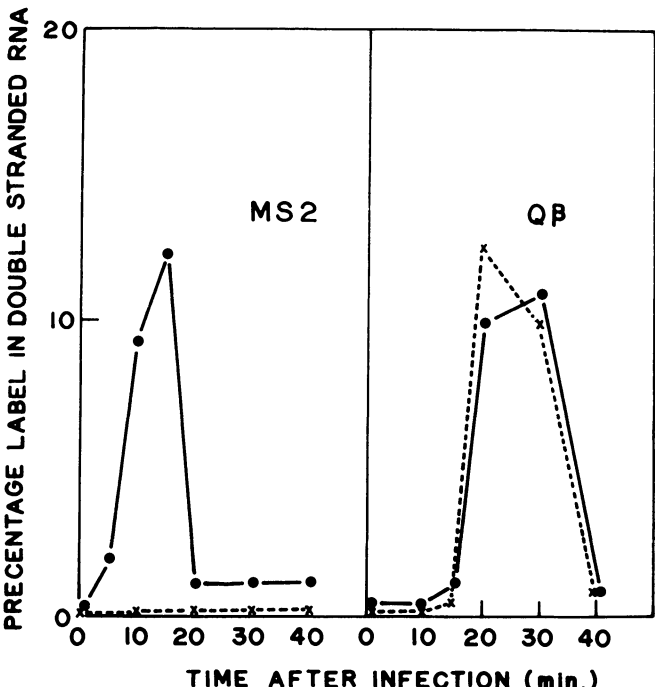


Fig. 3. Percentage of parental MS2 and Qβ RNAs in the double-stranded form at various times after infection of mutant cells grown at 33°C or 42°C. Cells were grown in tryptone broth at 33°C or 42°C to a density of $2x10^8$ cells/ml and infected at the respective growth temperatures with either MS2 or Qβ containing (^{3}H)-labeled RNA at a multiplicity of 5. At different times after infection the cells were harvested and washed with phosphate buffered saline. For other experimental conditions see Fig. 1. Symbols: growth at 33°C (•); at 42°C (x).

four inner membrane proteins (molecular weight range 11,000-18,000) from the preparation of the mutant grown at 42°C (Fig. 4a slot 1 arrows a to d). In the electrophoretic pattern of the outer membrane protein of the mutant grown at 42°C, we noticed three main alterations (Fig. 4b slot 1 versus 3) in the presence of three additional bands (e to g) which are absent from the corresponding membranal preparation of parental cells grown at 42°C.

In summary, we showed that in the traD (Ts) mutant JCFL39, group I RNA phage replication is inhibited at high temperature (42°C). Furthermore, the elevated temperature affects the membrane of the host mutant.

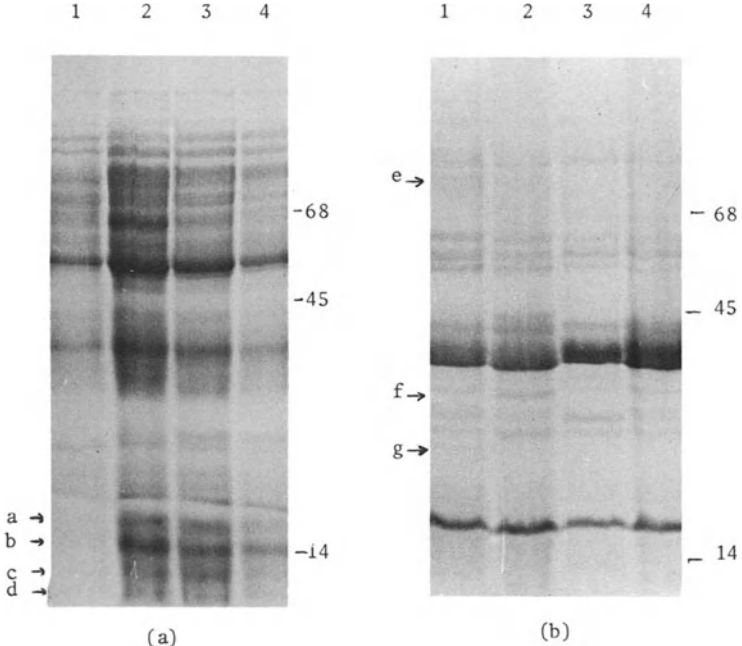


Fig. 4. SDS-polyacrylamide gels of membranal proteins from mutant JCFL39 and parental strain M176 grown at 33°C and 42°C. Cells were grown at 33°C or 42°C in tryptone broth to a density of 10^9 cells/ml. The membranes were isolated according to the procedure of Schnaitman (56) and the inner and outer membrane fractions were separated by centrifugation in sucrose gradients as described by Osborn et al (57). Inner and outer membrane fractions were recovered from the gradient by precipitation with TCA followed by washing with acetone and then dissolved in the sample buffer for electrophoresis. The samples were resolved by gel electrophoresis in the discontinuous buffer system of Laemmli (58) through a 5.4% acrylamide stacking gel and a 12.5% acrylamide resolving gel. After electrophoresis, gels were fixed and stained as described by Fegrest and Jackson (59). (a) Inner membrane proteins, (b) outer membrane proteins. Mutant JCFL39 grown at 42°C (slot 1) or 33°C (slot 2), parental strain M176 grown at 42°C (slot 3) or 33°C (slot 4). The numbers to the right of the figures are the molecular weights ($x10^{-3}$) of standard proteins with the indicated mobility. The arrows indicate changes in the proteins of the mutant versus those of the parent at 42°C. Equal volumes of membranal proteins were layered on each slot.

DISCUSSION

In the present study we attempted to clarify the role of bacterial proteins in the RNA replication of group I RNA phages *in vivo*. For this purpose we used two methods of investigation. The first approach was based on the fact that rifampin totally inhibits host protein synthesis but does not affect phage protein synthesis (23,25,27), permitting the examination of phage replication without the distraction of bacterial protein synthesis. The second approach was based on the use of an *E. coli traD* mutant which inhibits the development of group I RNA phages (39).

According to both methods of investigation, the RNA replication of only group I RNA phages was inhibited while that of Qβ was unaffected. This indicates that *in vivo*, group I RNA phages and Qβ require different host protein(s) for their replication. Although we could not point to any specific bacterial protein involved in group I RNA phage replication, several conclusions can be deduced from our studies.

1) In contrast to Qβ, the group I RNA phages, MS2, f2 and R17, seem to be dependent on a membranal organization for their RNA replication. As shown here, at 42°C, the *traD* (Ts) mutation of *E. coli* strain JCFL39 inhibits group I RNA phage replication and causes several alterations in the membrane proteins. We suggest that these changes may be responsible for the inhibition of group I RNA phage replication at high temperature. This is supported by previous studies showing that MS2 replicative intermediates are tightly bound to the membrane fraction (40). In addition, we assume that the involvement of a membranal structure in the replication of group I RNA phages might account for the lability of group I RNA replicases and the difficulties in their purification.

2) Each member of the group I RNA phages seems to require different host proteins for RNA replication. This was shown here by the use of rifampin which differentially inhibits the replication of MS2, f2 and R17. Thus far, no physiological differences among the group I phages MS2, f2, and R17 have been described. Nevertheless, the present study reveals, for the first time, differences in their replication processes which seem in extreme contrast to the very close similarity in their particle properties (4). These phages are not identical, however, as revealed by the partial data on their proteins and RNAs sequences (7,46, 47,48). The variation among the RNA sequences was deduced from partial sequence analysis and was estimated to be 2.8% for f2 versus MS2, and

R17 versus MS2, and 3.4% for f2 versus R17 (47,48). We suggest that although these variations are small, they may be considerably amplified by the spatial arrangement of the RNA, that is, the formation of hairpins and loops, and as a result, template recognition would be affected. In the case of phage Qβ, the host subunits of the Qβ replicase appear to play an essential role in the recognition of Qβ RNA as a template (7). We suggest that in the case of group I RNA phages, the host proteins play a similar role. Therefore, the differences which we suspect to exist in the RNA conformation of various group I RNA phages may be responsible for the use of different bacterial proteins for the recognition of their RNA as template. Host proteins showing an affinity for nucleic acids are good candidates for template recognition in the process of RNA phage replication. Such host proteins may be ribosomal proteins, membrane proteins involved in DNA replication (49,50) and DNA transfer during conjugation (51,52,53) or other cellular DNA or RNA binding proteins. Phages Qβ, f2, and presumably other members of group I RNA phages use ribosomal and ribosomal-bound proteins for their replication (10,11,14). We suggest that for this purpose group I RNA phages also use membranal proteins which have an affinity for nucleic acids. This assumption is based on the study of the *E. coli traD* (Ts) mutant JCFL39 which is affected (at high temperature) in its membranal proteins, inhibits group I RNA phage replication (see section II) and inhibits DNA transfer during bacterial conjugation (30,31,37).

ACKNOWLEDGMENTS

We thank F.R. Warshaw Dadon for reading the manuscript.

This research was supported by a grant from the United States-Israel Binational Science Foundation.

REFERENCES

1. Miyake, T., Haruna, I., Shiba, T., Itoh, Y.H., Yamane, K., and Watanabe, I. Proc. Natl. Acad. Sci. U.S.A. 68:2022-2024. 1971.
2. Sakurai, T., Miyake, T., Shiba, T., and Watanabe, I. J. Microbiol. 12:544-546. 1968.
3. Watanabe, I., Miyake, T., Sakurai, T., Shiba, T., and Ohno, T. Proc. Jpn. Acad. 43:204-209. 1967.
4. Scott, D.W. Virology 26:85-88. 1965.
5. Overby, L.R., Barlow, G.H., Doi, R.H., Jacob, M., and Spiegelman, S. J. Bacteriol. 91:442-448. 1966.
6. Stavis, R.L., and August, J.T. Annu. Rev. Biochem. 39:527-560. 1970.
7. Weissmann, C., Billeter, M.A., Goodman, H.M., Hindley, J., and Weber, H.

Annu. Rev. Biochem. 42:303-328. 1973.
8. Kamen, R. Nature (London) 228:527-533. 1970.
9. Kondo, M., Gallerani, R., and Weissmann,C. Nature (London) 228:525-527. 1970.
10. Blumenthal, T., Landers, T.A., and Weber, K. Proc. Natl. Acad. Sci. U.S.A. 69:1313-1317. 1972.
11. Wahba, A.J., Miller, M.J., Niveleau, A., Landers, T.A., Carmichael, G.G., Weber, K., Hanley, D.A., and Slobin, L.I. J. Biol. Chem. 249:3314-3316. 1974.
12. Carmichael, G.G., Weber, K., Niveleau, A., and Wahba, A.J. J. Biol. Chem. 250:3607-3612. 1975.
13. Shapiro, L., Franze de Fernandez, M.T., and August, J.T. Nature (London) 220:478-480. 1968.
14. Federoff, N.V., and Zinder, N.D. Proc. Natl. Acad. Sci. U.S.A. 68:1838-1843. 1971.
15. Federoff, N.V., and Zinder, N.D. Nature (London) New Biol. 241:105-108. 1973.
16. Lancini, G.C., and Sartori, G. Experientia 24:1105-1106. 1968.
17. Umezawa, H., Mizuno, S., Yanazaki, H., and Nitta, K. J. Antibiot. 21:234-235. 1968.
18. Reid, P., and Speyer, J. J. Bacteriol. 104:376-389. 1970.
19. Rabussay, D., and Zillig, W. FEBS Lett. 5:104-106. 1969.
20. Engelberg, H. J. Mol. Biol. 68:541-546. 1972.
21. Engelberg, H., and Soudry, E. J. Virol. 7:847-848. 1971.
22. Fromageot, H.P.M., and Zinder, N.D. Proc. Natl. Acad. Sci. U.S.A. 61:184-191. 1968.
23. Mario, P., Baldi, M.I., and Tocchini-Valentini, G.P. Cold Spring Harbor Symp. Quant. Biol. 33:125-127. 1968.
24. Meir, D., and Hofschneider, P.H. FEBS Lett. 25:179-183. 1972.
25. Passent, J., and Kaesberg, P. J. Virol. 8:286-292. 1971.
26. Rothwell, J.D., and Yamazaki, H. Biochemistry 11:3333-3338. 1972.
27. Engelberg, H., and Soudry, E. J. Virol. 8:257-264. 1971.
28. Erikson, R.L., Fenwick, M.L., and Franklin, R.M. J. Mol. Biol. 10:519-529. 1964.
29. Weissmann, C., Borst, P., Burdon, R.H., Billeter, M.A., and Ochoa, S. Proc. Natl. Acad. Sci. U.S.A. 51:682-690. 1964.
30. Achtman, M. Curr. Top. Microbiol. Immunol. 60:79-123. 1973.
31. Helmuth, R., and Achtman, M. Nature (London) 257:652-656. 1975.
32. Everett, R., and Willets, N. J. Mol. Biol. 136:129-150. 1980.
33. Brinton, C.C., Jr., Gemski, P., Jr., and Carnahan, J. Proc. Natl. Acad. Sci. U.S.A. 52:776-783. 1964.
34. Caro, L.G., and Schnos, M. Proc. Natl. Acad. Sci. U.S.A. 56:126-132. 1966.
35. Crawford, E.M., and Gesteland, R.F. Virology 22:165-167. 1964.
36. Valentine, R.C., and Strand, M. Science 148:511-513. 1965.
37. Achtman, M., Willetts, N., and Clark, A.J. J. Bacteriol. 110:831-842. 1972.
38. Achtman, M., Willetts, N., and Clark, A.J. J. Bacteriol. 106:529-538. 1971.
39. Schoulaker, R., and Engelberg-Kulka, H. J. Virol. 25:433-435. 1978.
40. Haywood, A.M. Proc. Natl. Acad. Sci. U.S.A. 70:2381-2385. 1973.
41. Coleman, W.G., Jr., and Leive, L. J. Bacteriol. 139:899-910. 1979.
42. Havekes, L.M., Lugtenberg, B.J.J., and Hoekstra, W.P.M. Mol. Gen. Genet. 146:43-50. 1976.
43. Miyoshi, Y., and Yamagata, H. J. Bacteriol. 125:142-148. 1976.

44.Nagel de Zwaig, R., and Luria, S.E. J. Bacteriol. 94:1112-1123. 1967.
45.Nakamura, H. J. Bacteriol. 96:987-996. 1968.
46.Min Jou, W., Haegman, G., Ysebaert, M., and Fiers, W. Nature (London) 237:82-88. 1972.
47.Nichols, J.L., and Robertson, H.D. Biochim. Biophys. Acta 228:676-681. 1971.
48.Robertson, H.D., and Jeppesen, P.G.N. J. Mol. Biol. 68:417-428. 1972.
49.Parker, D.L., and Glaser, D.A. J. Mol. Biol. 87:153-168. 1974.
50.Geider, K. Curr. Top. Microbiol. Immunol. 74:56-112. 1976.
51.Gilbert, W., and Dressler, D. Cold Spring Harbor Symp. Quant. Biol. 33:473-484. 1968.
52.Willets, N., and Skurray, R. Annu. Rev. Genetics 14:41-76. 1980.
53.Falkow, S., Tompkins, L.S., Silver, R.P., Guerry, P., and LeBlanc, D.J. Am. New York Acad. Sci. 182:153-171. 1971.
54.Engelberg, H., Brudo, I., and Israeli-Reches, M. J. Virol. 16:340-347. 1975.
55.Franklin, R.M. Proc. Natl. Acad. Sci. U.S.A. 55:1504-1511. 1966.
56.Schnaitman, C.A. J. Bacteriol. 104:882-889. 1970.
57.Osborn, M.J., Gander, J.E., Parisi, E., and Carson, J. J. Biol. Chem. 247:3962-3972. 1972.
58.Laemmli, U.K. Nature (London) 227:680-685. 1970.
59.Fegrest, J.P., and Jackson, R.L. Methods Enzymol. 28:54-63. 1972.

15

THE TOBACCO MOSAIC VIRUS GENOME: ORGANIZATION AND GENE FUNCTION

MILTON ZAITLIN AND PETER PALUKAITIS

SUMMARY

Tobacco mosaic virus (TMV) RNA codes for at least 4 polypeptides, only one of which, the coat protein, has a known function. Evidence is presented that one of these proteins (30 K protein) may be involved in potentiating the cell-to-cell movement of the virus in the plant. Virions of TMV also contain an average of one molecule of another protein (H protein) in addition to the 2130 molecules of coat protein. H protein is postulated to be an unusual branched molecule, containing the coat protein plus an additional peptide moiety of either host or viral origin. Analyses of single-stranded and double-stranded RNAs found in TMV-infected tissue by gel electrophoresis, blotting and hybridization revealed the presence of the genomic TMV and its double-stranded replicative form, and 9 subgenomic mRNAs, at least three of which also had double-stranded forms. Low molecular weight TMV-related RNAs representing different regions of the genome were also observed to arise during virus replication. Sequences derived from a number of TMV strains have confirmed strain relationships previously established, and the arrangement of the four TMV genes and have suggested the presence of other genes. Moreover, the sequence data indicate that several of the genes overlap, both in the same reading frame and in different reading frames.

INTRODUCTION

Tobacco mosaic virus is the type member of the tobamovirus plant virus group. TMV virions are rod-shaped particles (300 x 18 nm) with a sedimentation co-efficient of 200S, a single-stranded RNA genome of M_r 2.1×10^6 (6.4 kb), comprising 5% of the virus weight, and 2130 copies of the coat protein monomer (M_r 17,500) (1, 2).

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

The genomic RNA contains a 5'-7-MeGpppGp cap (3, 4), and a t-RNA-like 3'-end (5). The 3'-end "cloverleaf structure" and -CCA terminus permits the common strain of TMV to be aminoacylated in vitro by histidyl aminoacyl-ligase (6); the cowpea strain of TMV may be aminoacylated with valine (7).

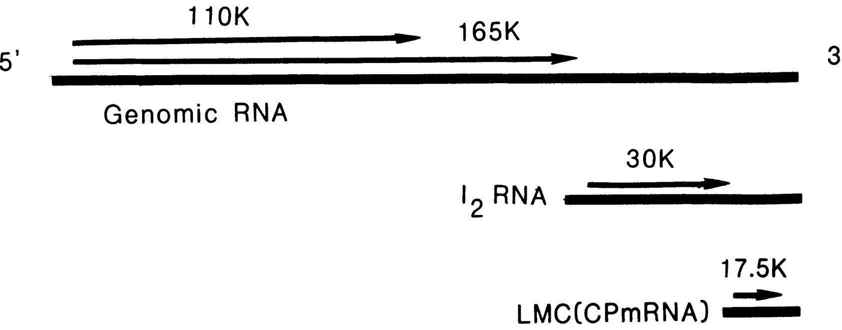


FIGURE 1. Major RNA species involved in TMV replication (heavy lines) with their known translational products (arrows).

There are at least four genes in the TMV genome (5, 7;Figure 1); however, only the two genes on the 5'-end of the genomic RNA are expressed during in vitro translation of the genomic RNA (7, 8, 9). These two genes, coding for the 110K (formerly called 130K to 145K) and the 165K proteins, overlap; they initiate together but terminate at different points (10). The other two genes, coding for the 30K and the 17.5K proteins, are expressed only on the sub-genomic RNAs that contain these genes at their 5'-ends; the I_2-mRNA and the LMC (coat protein mRNA), respectively (7, 9). The organization of these genes, their expression and their functions will be discussed below. Most of the data presented will be with reference to the common strain (U1) of TMV. Where applicable, reference will be made to results obtained with other tobamoviruses (relationships described in 11 and 12).

TMV RNA GENOME ORGANIZATION

Sub-genomic RNAs.

Until recently, only two sub-genomic RNAs of TMV had been isolated and characterized: the I_2 mRNA (M_r 6.8 x 10^5; 2.0 kb) and the LMC (M_r 2.3 x 10^5; 0.69 kb) (7, 8, 9). Both of these RNAs contain the 3'-end of the genomic RNA (Figure 1), but extend toward the 5'-end to different lengths (7). Since the I_2-mRNA contains the viral "origin of encapsidation" (OREN) region - 900 nucleotides from the 3'-end - it was found to be encapsidated by viral coat protein and could be isolated from virion preparations. The LMC did not contain the OREN and was not encapsidated (5); however, the LMC could be extracted from TMV-infected leaves (8).

Using the more sensitive techniques of transferring RNA from gels to nitrocellulose and hybridization with TMV-specific probes ("Northern Blotting"; 13, 14), a number of other sub-genomic RNAs have recently been observed, both in virions and in extracts from infected leaves (Figure 2; Reference 15). Using probes specific for different regions of the TMV genome, we were able to detect nine subgenomic RNAs (all seen in Figure 2, lane D; Table 1). With the exception of the LMC, all of these sub-genomic RNAs contain the OREN region; presumably, they all contain the 3'-terminus of TMV RNA, since they all contained the coat protein gene, although this has yet to be established.

It is not certain that any of the new sub-genomic RNAs are actually functional _in vivo_, or what viral functions they may code for. Some of them may represent extraction or electrophoretic artefacts or molecules cleaved prior to encapsidation at particularly RNase-sensitive sites; either _in vivo_, or during the isolation process. For example, Table 1 lists the sub-genomic RNAs detected in three separate analyses (9, 15, 16) and
the sizes of RNAs prepared from various partially stripped virus particles [under conditions of high pH (17) or in the presence of SDS (18), TMV virions of the U1 strain de-encapsidate from the 5'-end of the RNA towards the 3'-end, producing a number of discrete-sized isolatable intermediates known as partially stripped virions - PSVs]. This suggests that some of the so-called sub-genomic RNAs may play no relevant role in replication.

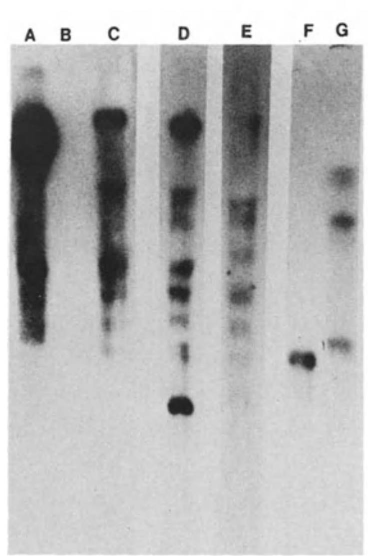


Figure 2

Figure 3

FIGURE 2. Northern Blots of various RNAs (on nitrocellulose) from a 1.5% agarose gel containing 20 mM MOPS, 5 mM Na acetate, 1 mM EDTA, pH 7.0, 2 M formaldehyde . A = TMV encapsidated RNA; B = healthy tobacco RNA; C - E = TMV-infected tobacco, LiCl insoluble RNA; F = PSV_6 RNA; G = cucumber mosaic virus (CMV) RNAs (markers). Probes were A, B. C = OREN-specific probe made with a specific-primer prepared such that it represents sequences 965-1011 nucleotides from the 3'-end; D, F = PSV_6 cDNA (3'-terminal 1050 nucleotides); E = random primed cDNA to genomic RNA; G = CMV cDNA. Leftward (of A) pointers indicate positions of (in decending order) genomic TMV RNA, I_2 RNA and LMC. Rightward (of G) pointers indicate position of CMV RNA markers: 1.35 x 10^6, 1.15 x 10^6, 0.85 x 10^6, 0.35 x 10^6.

FIGURE 3. Fluorogram of a gel (6% acrylamide:7 M urea) analysis of low molecular weight RNAs from TMV-infected and uninfected tobacco. A and B samples represent the "single-stranded" CF 11 cellulose fraction from a 2 M LiCl-soluble nucleic acid extract (24). A). ^{3}H-uridine-labeled RNAs from uninfected leaves. Labeling without actinomycin D. B) ^{3}H-uridine-labeled RNAs from TMV-infected leaves (labeled for 24 hr in the presence of 50 μg/ml actinomycin D, starting 3 days after inoculation). The LMC and 4S RNA bands are identified. C) Size markers of ^{3}H-labeled HpaII-restricted fragments of plasmid pBR322. Denatured fragments range in size from 624 nucleotides (migrating slightly below the LMC of Track B) to 69 nucleotides at the bottom of the gel. Restricted fragments were labeled using reverse transcriptase and ^{3}H-dCTP. Samples in all 3 lanes were heated for 1 min at 100° in 30% formamide before application to the gel.

Table 1. Molecular weights of genomic TMV RNA and less-than-full length TMV RNAs.

Partially stripped virion (PSV) RNAs[1]			Viron encapsidated RNAs[2]		
RNA	M_r (x 10^6)	Nucleotide length	RNA	M_r (x 10^6)	Nucleotide length
Genomic PSV-1	2.1	6400	Genomic	2.1	6400
PSV-2	1.4	4200	I_{1A}	1.55 - 1.6	4700-4850
			I_{1B}	0.9 - 1.3	2750-3950
PSV-3	0.79	2400			
PSV-4	0.65	1900	I_{2A}	0.68 - 0.69	2050-2100
PSV-5	0.50	1500	I_{2B}	0.51	1550
PSV-6	0.35	1050			

Intracellular RNAs

Goelet and Karn (15)			Palukaitis, *et al*. (16)		
RNA	M_r (x 10^6)	Nucleotide length	RNA	M_r (x 10^6)	Nucleotide length
Genomic	2.1	6400	Genomic	2.1	6400
SG-1[3]	1.15	3500	SG-1[3]	1.05 - 1.1	3200-3350
SG-2	1.0	3000	SG-2	0.95 - 1.0	2900-3050
			SG-3	0.8 - 0.85	2400-2600
SG-3	0.60	1800	SG-4	0.6 - 0.65	1800-1950
SG-4	0.50	1500	SG-5	0.48 - 0.52	1450-1550
SG-5	0.43	1300	SG-6	0.4 - 0.44	1200-1350
SG-6	0.36	1100	SG-7	0.34 - 0.35	1050
			SG-8	0.29 - 0.3	900
SG-7[4]	0.23	700	SG-9[4]	0.21 - 0.22	650

1 See text. M_r supplied by Dr. A. Asselin.
2 From Beachy and Zaitlin (9).
3 SG = Sub-genomic RNAs
4 LMC

Other TMV RNA fragments are generated during the process of infection. Zelcer et al. (19) observed the synthesis of at least three low molecular weight, apparently single-stranded RNAs, during TMV replication in tobacco in the presence of actinomycin D (an inhibitor of host RNA synthesis). We have observed a larger number of these low mol. wt. RNAs, both by labelling RNA in TMV-infected tissues (Figure 3) and by blotting the RNAs (separated in polyacrylamide gels) from gels to nitrocellulose membranes and hybridizing various TMV-specific probes to the RNAs on the nitrocellulose blots (16). Our results suggest that these low molecular weight RNAs, although single-stranded, contain sequences from both (+) TMV RNA and (-) TMV RNA strands (see following section), encompassing much of the viral genome. These fragments may represent "viral RNA trash", which is formed during the generation of genome RNA or of some of the sub-genomic RNAs described above. The formation of sub-genomic RNAs will be dealt with in the following section.

Double-stranded RNAs

The central dogma of RNA virus replication is based on studies on the analogous RNA bacteriophage Qβ system (20); i.e., a negative-stranded (-) RNA is generated, complementary to the positive-stranded genomic (+) RNA. This double-stranded (ds) RNA complex is termed the replicative form (RF). [When ds RNA is isolated by phenol extraction, it is believed that proteins separating the two partially H-bonded strands are removed, resulting in the formation of the tightly bonded ds RNA (21)]. From the RF, numerous positive-strands of TMV RNA [(+) TMV RNA] are generated. These complexes of multiple (+) strands at various stages of synthesis on a (-) strand are termed replicative intermediates (RI). Both RFs and RIs have been observed (22, 23); however, the enzymes involved in the formation of either RFs or RIs have neither been purified nor characterized.

The formation of sub-genomic RNAs also occurs by an unknown mechanism. Zelcer et al. (24) observed four ds TMV RNAs in infected tissues. The estimated molecular weights, using ds RNAs as size markers on gels were: 4.0, 2.25, 1.1 and 0.23 x 10^6; these were termed RF, ds-1, ds-2 and ds-3, respectively. They determined that all of

these RNAs contained the 5'-end of the (-) strand of TMV, but concluded, based on the calculated molecular weights, that only the RF and ds-1 could be considered the ds counterparts of the single-stranded RNAs found in infected tissues. They concluded further that there was no ds RNA related to the LMC. We (16) and others (15) now find by blotting and hybridizing denatured ds RNAs that ds-3 is of a size expected for a ds LMC. In fact, the four ds RNAs observed are double-stranded forms of the genomic RNA and three subgenomic RNAs of M_r 1.05 x 10^6, 0.65 x 10^6 (I_2) and 0.23 x 10^6 (LMC); we do not observe ds RNAs for any of the other sub-genomic RNAs (Figure 2). Goelet and Karn (15), on the other hand, claim (+) and (-) forms of the TMV genomic RNA and seven other sub-genomic RNAs.

It is still not known whether these ds RNAs are replicative forms of the corresponding (+) strand sub-genomic RNAs, or whether they fortuitously act as "one-way templates" for the viral replicase, since they all contain a common 3'-end; however, as they do not share the same 5'-end the concomitant (-) strands will not share the same 3'-end. If we assume that both (+) and (-) strand 3'-ends contain recognition sequences for the viral replicase, then by this mechanism, only one (-) strand copy and no (+) strand copies would be synthesized on sub-genomic RNAs. However, it is conceivable that the newly-generated (-) strand could contain secondary recognition sequences for the replicase (see below).

Alternatively, sub-genomic RNAs might be generated by cleavage of (+) stand genomic RNA. This cleavage event would also generate sub-genomic RNAs containing the 5'-end of the RNA genome. Using 5'-end-specific TMV RNA and cDNA probes, we have been unable to detect such fragments in blots of RNA from agarose gels; however, small RNA fragments representing different regions of the TMV genome were detected in blots from polyacrylamide gels (16). As these fragments also contain sequences of the (-) strand TMV RNA, their existence is compatible with a model in which the genomic RF (in a partly ds RNA form) is cleaved to produce sub-genomic RFs.

Another model for sub-genomic RNA synthesis is one in which the viral replicase binds to regions other than the 3'-end of the (-) strand genomic TMV RNA and catalyzes the synthesis of sub-genomic RNAs. In this system, the single-stranded 3'-end of the (-) strand is

degraded to small fragments and produces ds RNAs equivalent in length to the sub-genomic RNAs.

TMV SEQUENCES: OVERLAPPING GENES.

Prior to the recent advent of rapid sequencing techniques (25, 26), the elucidation of the TMV RNA sequence was limited to a few short stretches. The first sequences determined were those of the approximately 70 nucleotides at the 3'-end of two strains of TMV RNA (27, 28). These strains shared only a limited amount of sequence homology, but both RNAs could be folded into cloverleaf-like structures. The sequence of 236 nucleotides at the 5'-end of the common strain of TMV (29, 30), revealed a long RNase T_1-resistant fragment (Residues 1-71) that terminated at its 3'-end in an AUG. This RNase T_1 fragment (Ω; 31) was shown to be capable of binding ribosomes (32), suggesting that the 3'-terminal AUG represents the initiation codon for the first gene. A comparison of three tobamoviruses (33) revealed the presence of an Ω-fragment. Although these Ω-fragments showed extensive sequence homology with the Ω-fragment from the common strain of TMV and with each other, they were of different lengths and showed significant sequence differences; two other tobamoviruses examined in the same study did not contain any large T_1 -RNase resistant fragments, and hence, no Ω.

The sequence of the 3'-terminal 1000 nucleotides of the common strain of TMV (5) as well as the sequence of the LMC revealed the following: (a) the LMC was identical to the 3'-terminal 693 nucleotides of the genomic RNA; (b) a 3'-terminal 204 nucleotide-long untranslated region; (c) the coat protein gene (480 nucleotides) with no intervening sequences; (d) a 5'-terminal 9 nucleotide long untranslated region on the LMC; and (e) the C-terminal one-third of the 30K gene with a two-nucleotide gap between the termination codon of the 30K gene and the initiation codon of the coat protein gene; i.e., the 5'-untranslated region of the LMC overlaps with the 30K gene, but the two genes are in different reading frames.

The cloning and sequencing of 2000 nucleotides at the 3'-end of the Japanese common (OM) strain of TMV (34), which shows very few sequence differences from the U_1-strain of TMV (North American and European common strain), revealed the same gene arrangements and the

complete sequence of all of the 30K gene (now 29,791 in mol. wt.). On the other hand, the cloning and sequencing of 1060 nucleotides at the 3'-end of the cowpea strain of TMV (35), which has almost no sequence homology with the common strain of TMV (12, 35), showed one major difference in the arrangement of the 30K and the coat protein genes: they overlap by 29 nucleotides (initiation and termination sequences included, but not 5'-end leader sequences). As with the common strain, the 30K and coat protein genes of the cowpea strain are in different reading frames. In the case of the cowpea strain, the OREN region is approximately 400 nucleotides from the 3'-end (35); i.e., well within the coat protein gene. Therefore, in this strain, the LMC is found encapsidated. The LMC of the cowpea strain (711 nucleotides long) has a 7 nucleotide 5'-end untranslated region, the coat protein gene (495 nucleotides) and a 209 nucleotide 3'-end untranslated region. The 3'-end of the cowpea strain can also be folded into a tRNA-like structure.

Recently, using cloning into M13 bacteriophage and rapid DNA sequencing techniques (36), Butler _et al_. (37) have confirmed the 3'-end sequences of the common strain of TMV and have extended the sequence information well into the 110K gene (Figure 1). Their observations also confirmed earlier _in vitro_ translation data that showed the 110K and 165K proteins overlapping (10); i.e., these two proteins begin at the same initiation site and translate in the same reading frame, with the 165K protein being a read-through of the 110K protein. Since both of these proteins have been observed _in vivo_ (38, 39), it suggests they have functional roles. The sequence data of Butler _et al_. (37) also suggest that the 165K gene overlaps with the 30K gene, but in a different reading frame.

Although these are the only four proteins coded for by the TMV genome that have been characterized, the multitude of sub-genomic RNAs generated during infection suggests there may be more than just four proteins. For example, of the four ds RNAs detected [presumably ds RNAs of the genomic RNA (RF), I_2 RNA, LMC and the 1.05×10^6 RNA] three are related to well-characterized RNAs, while the fourth is not well-characterized. The 1.05×10^6 species of RNA was first detected as an actinomycin D-resistant (and presumably TMV-related) RNA synthesized in TMV-infected tobaccos (40). It was also detected as a

sub-genomic encapsidated TMV RNA (I_{1B} RNA; 9). Upon translation, a mixture of polypeptides were synthesized ranging in size from 30K up to 110K, with a dominant protein of approximately 50K (9). Cell-free translation studies of sucrose-gradient fractionated TMV-infected tobacco RNAs resulted in the synthesis of a protein of 45K, from RNAs of up to 1.1 x 10^6 (3.5 kb) in molecular weight (15).

The data in Figure 2 suggest the 1.05 x 10^6 RNA species has the 3'-end common to all the other sub-genomic RNAs, but is missing the 5'-end of the molecule. This coupled with its molecular weight, would imply that the 5'-end of the 1.05 x 10^6 RNA is located approximately 3000 nucleotides from the 3'-end of TMV RNA. The sequence of the RNA of the common strain of TMV contains an AUG at 2780 nucleotides from the 3'-end (37); 3'- to the 110K gene, but within and in the same reading frame as the 165K gene. Initiation of translation from this region could proceed to the end of the 165K gene, generating a protein of M_r 47K. Therefore, it may be possible that the sub-genomic RNA of M_r 1.05 x 10^6 codes for yet another TMV-specific protein from a region of the viral genome already contributing a viral-specific product (the 165K gene).

It is clear that further studies are required on proteins synthesized _in vivo_ during TMV-infection. Such analyses can become more well-defined when proteins sought are delimited by sequence information; i.e., size and charge.

FUNCTIONS OF TMV RNA-CODED TRANSLATIONAL PRODUCTS

As shown in Figure 1, there are at least 4 proteins which are generated with TMV RNA as a messenger in _in vitro_ translational systems; it is probable that there are several others (15). Of these proteins, the only one with a known function is the coat protein; moreover, only the coat, 110K and the 165K proteins have been detected in infected tissues; the 30K has only been seen by _in vitro_ translation of viral sub-genomic RNAs (7, 9, 39).

Nothing is known regarding possible functions for the 165K, 110K translational products. They are found in infected tissues (39) and are synthesized in the presence of actinomycin D. They are thus _bona finde_ viral translational products. In studies with protoplasts, they have been shown to be synthesized early in the replication cycle (41),

thus leading to the supposition that they are somehow involved in viral RNA replication (41). The 110K is also found in those fractions of the tobacco leaf homogenates which have TMV-related replicase activity lending support to that supposition (39).

The 30K protein

Recent research in our laboratory has focused on seeking a function for the 30K TMV-coded protein. We have correlative evidence that this protein functions by promoting cell-to-cell movement of the virus (42). As discussed below, there are also recent studies from other laboratories which indicate that the movement of plant viruses from cell-to-cell in their hosts is an active process in which the virus participates. In our studies, we utilized two tomato strains of TMV which had been isolated by Nishiguchi and his colleagues in Japan (43). One of the strains, LS-1, isolated from a culture of strain L, was shown to be temperature sensitive; this defect results in its inability to move from cell-to-cell in inoculated leaves when the plants are held at a restrictive temperature (32°). The virus can replicate well at that temperature however, because it is synthesized to normal levels in *in vitro* inoculated protoplasts held at 32°. Our studies were directed towards determining if one of the viral-coded proteins was altered in strain LS-1, when compared with strain L. The only viral-coded protein which we could obtain directly was the coat protein; the others had to be analyzed as products of *in vitro* translation. In our first experiments, we found that the two-dimensional peptide maps of the coat proteins of these two viruses were identical. This was a very important consideration in this study because it is well-known that many TMV "strains" are very different from one another when their nucleic acid homologies are analyzed (12, 44); thus, if we are to find the temperature-sensitive lesion between the RNAs of L and LS-1 they should differ by only one, or at most a few nucleotides. When we analyzed the 30K, the 165K and the 110K proteins by two-dimensional tryptic analysis, the only change we could find was a one-peptide difference between the 30K proteins. With such an analysis we were able to analyze about 95% of the TMV genome; i.e., the estimate of what is translated. The only defect found was in the 30K protein product. To be sure, without being able to test directly the function of this protein, this correlation is only just that, but the

evidence certainly points to this protein - or at least the region of the RNA which codes for it - as being associated with cell-to-cell viral (or RNA) transport.

Supporting evidence for virus-mediated transport comes from studies with TMV and possibly with cowpea mosaic virus. Atabekov and his colleagues (45, 46) have shown that viruses which are competent for movement in a given host can potentiate the movement of viruses which normally don't move in that host. Viruses unrelated to one another taxonomically can be employed; further, they observed that TMV strain L can potentiate the movement of LS-1 at a restrictive temperature.

Further support for the concept may be inferred from some recent work of van Kammen and colleagues with cowpea mosaic virus (47). Cowpea mosaic virus is a two component virus (with components termed B and M) in which the genome is divided between two distinct RNA molecules, one in each component. To get an infection which results in the synthesis of virions, infection with both components or their RNAs is necessary. Interestingly, B-component RNA can replicate independently of M-component RNA as B-RNA must code for the viral moiety of the RNA replicating enzyme, and when inoculated separately it can initiate cytopathic structures characteristic of the disease. However, in that case the infectious principle (B-RNA) does not move from the site of inoculation. It may be that encapsidation is necessary (M-RNA codes for synthesis of the viral coat proteins), but it is also possible that M-RNA codes for a factor involved in cell-to-cell movement, as viral RNAs do not have to be encapsidated to move in plants (48).

There is another phenomenon which might be attributed to the fact that plant viruses probably generate a factor which enables them to move in plants. Most plants are resistant to most viruses; in fact, resistance is the rule while susceptibility is the exception. But upon closer examination, the individual cells of "resistant" plants may be susceptible to the viruses to which the plant is resistant. It has been observed for instance that mesophyll protoplasts prepared from leaves of plants resistant to certain viruses are susceptible to those viruses, and furthermore those protoplasts are capable of producing "normal" amounts of virus (49, 50).

We have found that in the leaves of two non-hosts of TMV (cotton and cowpea), upon mechanical inoculation about one out of every 50,000 to 150,000 cells becomes infected (50). No disease develops, but these few cells support substantial virus replication. In these circumstances, the virus is confined to those initially inoculated cells and can't move; the plant must play a role here, possibly by inactivating the viral-induced movement factor. The movement of cowpea chlorotic mottle virus in some varieties of cowpea is known to be genetically controlled by the plant (51).

A SECOND PROTEIN IN TMV VIRIONS (H-PROTEIN)

In 1978 (52) our laboratory reported that virions of several strains of TMV contained an average of one copy of a protein per virion, in addition to the well-known 2150 coat protein molecules. This protein of M_r 26,500 was considered to be of host origin. No function was ascribed to it then - or since - but it was found to have a close association with the virions in that it is protected from attack by proteolytic enzymes when virions are treated *in vitro*. H-protein was thought to be distributed randomly along the length of the virion; i.e., it was not confined to any given position. Based on recent work (53) we now find that the H-protein contains the viral coat protein; in addition, there is a polypeptide moiety to H-protein which is not related to the coat protein. It could be of either viral or host origin. Evidence presented (53) suggests that this additional polypeptide is fused with the coat protein, possibly through an isopeptide linkage, generating an unusual branched molecule.

H-protein has a strong affinity for TMV coat protein, a property which has made its isolation difficult. One has to use denaturing conditions on gels, or more recently we have isolated it by preparative isoelectric focusing to separate it from the coat protein. H-protein prepared by this latter method has been utilized in reconstitution experiments to assess its affinity for virions. Proteinase K-treated TMV RNA and H protein-free TMV coat protein were reconstituted with or without added H protein and the resulting virions were tested for infectivity, reconstitution efficiency and structural integrity. The H protein was found to reassociate with the virions during reconstitution but its presence had no effect on any of the parameters tested (54).

Thus, it is unlikely that H-protein functions in the early stages of viral infection, and if it does takes part in the viral replication or disease syndrome, that role remains to be determined.

REFERENCES

1. Caspar, D. Advan. Protein Chem. 18:37-121, 1963.
2. Markham, R., Hitchborn, J., Hills, G. and Frey, S. Virology 22:342-359, 1964.
3. Keith, J. and Fraenkel-Conrat, H. F.E.B.S. Lett. 57:31-33, 1975.
4. Zimmern, D. Nucleic Acids Res. 2:1189-1201, 1975.
5. Guilley, H., Jonard, G., Kukla, B. and Richards, K. E. Nucleic Acids Res. 6:1287-1308, 1979.
6. Öberg, B. and Philipson, L. Biochem. Biophys. Res. Commun. 48:927-932, 1972.
7. Beachy, R. N., Zaitlin, M., Bruening, G. and Israel, H. W. Virology 73:498-507, 1976.
8. Hunter, T. R., Hunt, T., Knowland, J. and Zimmern, D. Nature 260:759-764, 1976.
9. Beachy, R. N. and Zaitlin, M. Virology 81:160-169, 1977.
10. Pelham, H. R. B. Nature 272:469-471, 1978.
11. Gibbs, A. J. CMI/AAB Descrip. Plant Viruses No. 184, 1977.
12. Palukaitis, P. and Symons, R. H. Virology 107:354-361, 1980.
13. Alwine, J. C., Kemp, D. J., Parker, B. A., Reiser, J., Renart, J., Stark, G. R. and Wahl, G. M. Methods Enzymol. 68:220-242, 1980.
14. Thomas, P. S. Proc. Natl. Acad. Sci. USA 77:5201-5205, 1980.
15. Goelet, P. and Karn, J. J. Mol. Biol. 154:541-550, 1982.
16. Palukaitis, P., Garcia-Arenal, F., Sulzinski, M. A. and Zaitlin, M. Manuscript in preparation.
17. Perham, R. N. and Wilson, T. M. A. Virology 84:293-302, 1978.
18. Wilson, T. M. A., Perham, R. N., Finch, J. T. and Butler, P. J. G. F.E.B.S. Lett. 64:285-289, 1976.
19. Zelcer, A., Weaber, K. F., Balázs, E. and Zaitlin, M. Unpublished data.
20. Kamen, R. I. In RNA Phages (Ed. N. D. Zinder), Cold Spring Harbor, 1975, pp. 203-234.
21. Thatch, S. S. and Thatch, R. E., J. Mol. Biol. 81:367-380, 1973.
22. Nilsson-Tillgren, T. Mol. Gen. Genet. 109:246-256, 1970.
23. Jackson, A. O., Mitchell, D. M. and Siegel, A. Virology 45:182-191, 1971.
24. Zelcer, A., Weaber, K. F., Balázs, E. and Zaitlin, M. Virology 113:417-427, 1981.
25. Maxam, A. M. and Gilbert, W. Proc. Natl. Acad. Sci. USA 74:560-564, 1977.
26. Sanger, F., Nicklen, S. and Coulson, A. R. Proc. Natl. Acad. Sci. USA 74:5463-5467, 1977.
27. Guilley, H., Jonard, G. and Hirth, L. Proc. Natl. Acad. Sci. USA 72:864-868, 1975.
28. Lamy, D., Jonard, G. Guilley, H. and Hirth, L. F.E.B.S. Lett. 60:202-204, 1975.
29. Richards, K., Guilley, H., Jonard, G. and Hirth, L. Eur. J. Biochem. 84:513-519, 1978.

30. Jonard, G., Richards, K., Mohier, E. and Gerlinger, P. Eur. J. Biochem. 84:521-531, 1978.
31. Mandeles, S. J. Biol. Chem. 243:3671-3684, 1968.
32. Konarska, M., Filipowicz, W., Domdey, H. and Gross, H. J. Eur. J. Biochem. 114:221-227, 1981.
33. Kukla, B. A., Guilley, H. A., Jonard, G., Richards, K. E. and Mundry, K. W. Eur. J. Biochem. 98:61-66, 1979.
34. Meshi, T., Ohno, T. and Okada, Y. J. Biochem. 91:1441-1444, 1982.
35. Meshi, T., Ohno, T., Iba, H. and Okada, Y. Mol. Gen. Genet. 184:20-25, 1981.
36. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. and Roe, B. A. J. Mol. Biol. 143:161-178, 1980.
37. Butler, P. J. G., personal communication.
38. Scalla, R., Boudon, E. and Rigaud, J. Virology 69:339-345, 1976.
39. Scalla, R., Romaine, P., Asselin, A., Rigaud, J. and Zaitlin, M. Virology 91:182-193, 1978.
40. Siegel, A., Hari, V., Montgomery, I. and Kolacz, K. Virology 73:363-371, 1976.
41. Paterson, R. and Knight, C. A. Virology 64:10-22, 1975.
42. Leonard, D. A. and Zaitlin, M. Virology 117:416-424, 1982.
43. Nishiguchi, M., Motoyoshi, F. and Oshima, N. J. Gen. Virology 39:53-61, 1978.
44. Van De Walle, M. J. and Siegel, A. Phytopathology 72:390-395, 1982.
45. Taliansky, M. E., Malyshenko, S. I., Pshennikova, E. S., Kaplan, I. B., Ulanova, E. F. and Atabekov, J. G. Virology 122, 318-326, 1982.
46. Taliansky, M. E., Malyshenko, S. I., Pshennikova, E. S. and Atabekov, J. G. Virology 122, 327-381, 1982.
47. Rezelman, G., Fronssen, H. J., Goldbach, R. W., Ia, T. S. and Van Kammen, A. J. Gen. Virology 60:335-342, 1982.
48. Atabekov, J. G. Comprehensive Virology 11:143-200, 1977.
49. Beier, H., Siler, D. J., Russell, M.L. and Bruening, G. Phytopathology 67:917-921, 1977.
50. Sulzinski, M. and Zaitlin, M. Virology 121:12-19, 1982.
51. Kuhn, C. W., Wyatt, S. D. and Brantley, B. B. Phytopathology 71:1310-1315, 1981.
52. Asselin, A. and Zaitlin, M. Virology 91:173-181, 1978.
53. Collmer, C. W., Vogt, V. M. and Zaitlin, M. Virology, (Accepted) 1983.
54. Collmer, C. W. and Zaitlin, M. Virology, (Accepted) 1983.

ACKNOWLEDGEMENTS

The original work described in this paper was supported in part by grants from the National Science Foundation (81-04150) and the United States Department of Agriculture, CGO 81-00638.

NOTE ADDED IN PROOF

The complete sequence of TMV RNA has recently been determined (Goelet, P. et al., Proc. Natl. Acad. Sci. USA 79, 5818-5822, 1982). As deduced from the RNA sequence, the molecular weights of the 110K and 165K proteins are 125,941 and 183,253, respectively.

16

BIOSYNTHESIS OF INFLUENZA VIRUS RNA

Brian W.J. MAHY and Thomas BARRETT

SUMMARY

Using fowl plague virus as a model highly virulent influenza A virus, we have studied the transcription and replication of the eight segments of the negative-stranded RNA genome. Initiation of transcription requires a 5' cap-containing RNA primer molecule which is recognised by PB2, the virion polypeptide product of RNA segment 1. During transcription, the 5' cap plus 10-15 nucleotides are transferred to influenza virus mRNAs, resulting in sequence heterogeneity at their 5' termini. Transcription is terminated by polyadenylation at a tract of uridine residues 17-22 nucleotides from the 5' terminus of each segment of the genome RNA template.

The site of RNA transcription appears to be the cell nucleus. Addition of toyocamycin to infected cells results in accumulation of virus-specific transcripts in the nucleus, and blocks the production of spliced mRNAs normally processed from transcripts of segments 7 and 8. Transcription is strictly regulated during normal replication, resulting in temporal control of polypeptide synthesis. This regulation can be abolished by cycloheximide, indicating that newly synthesised polypeptides (so far undefined) control transcription.

Little is known concerning the synthesis of progeny virion RNAs. Production of template cRNAs involves read-through of the polyadenylated site by an unknown mechanism. Since these cRNAs can be synthesised in the absence of new virion RNA synthesis, it is likely that only the input vRNAs are required to act as templates in their production.

Y. Becker (ed.), REPLICATION OF VIRAL AND CELLULAR GENOMES. Copyright © 1983 Martinus Nijhoff Publishing, Boston. All rights reserved.

INTRODUCTION

The influenza viruses are responsible for widespread epidemic disease in the human as well as avian, swine and equine species. In the two most recent pandemic years, 1957 and 1968, hundreds of millions of human cases occurred throughout the world with a high rate of excess mortality, and in the pandemic of 1918, more than twenty million persons died of acute influenzal pneumonia (1). The most virulent influenza virus so far isolated is the avian strain known as fowl plague virus (FPV), infection with which causes one hundred percent mortality in susceptible birds such as chickens (2). Our molecular studies of influenza virus replication have concentrated on FPV infection of chick embryo fibroblast (CEF) cells; this fully productive virus-cell system provides a model for the most virulent influenza infections in man.

Using FPV, we established that the influenza virus genome consists of eight distinct single-stranded RNA molecules (3) which are complementary to polysomal mRNA (4). Similar findings were made with other influenza virus strains (reviewed in 5). Subsequently, the sizes and coding assignments of the segments were determined (Table 1). The expression of influenza virus RNA segments involves transcription into capped, polyadenylated mRNAs (A(+) cRNA) by a virion-bound RNA transcriptase. Production of genome RNAs, on the other hand, requires newly-synthesised virus-specific proteins and full-length non-polyadenylated template molecules (A(-) cRNA). Although these reactions are extremely complex, many of the steps involved in influenza RNA biosynthesis are now understood at least in outline, and will be described here with reference to our studies using the fowl plague virus model.

EXPERIMENTAL

Initiation of mRNA synthesis studied in vitro

Initiation of transcription by fowl plague virion RNA transcriptase requires a primer molecule. This was originally suggested by the greatly enhanced rate of transcription observed when the dinucleotides GpG or GpC were added to an *in vitro* reaction

Table 1. The Influenza A (Fowl Plague) Virus Genome

Virion RNA Segment	Chain Length[a] (Nucleotides)	Encoded[b] Polypeptide	Polypeptide[d] Length (amino acids)	Function
1	2341	PB2	759	Component of RNA transcriptase: binds capped RNA molecules.
2	2341	PB1	757	Component of RNA transcriptase: elongation of capped primer RNA molecules.
3	2233	PA	716	Present in RNA transcriptase complex. Function unknown.
4	1742	Haemagglutinin	566	Adsorption and fusion to cell membranes.
5	1565	Nucleoprotein	498	Association with RNA segments to form ribonucleoprotein.
6	1413	Neuraminidase	454	May be involved in adsorption and/or release.
7	1027	Matrix	252	Structural component underlying the lipid bilayer
		M_2	97	Unknown.
		M_3[c]	9	Unknown.
8	890	NS_1	230	May be involved in replication.
		NS_2	121	

[a]From sequence analysis of the fowl plague virus genome, except segments 2, 5 and 6, which are as determined for the PR8 strain (reviewed in 32).
[b]From Almond et al. (40,41) and Inglis et al. (27,28,42).
[c]The existence of peptide M_3 has not been established, though the mRNA is present in infected cells (28).
[d]Predicted from the mRNA nucleotide sequence.

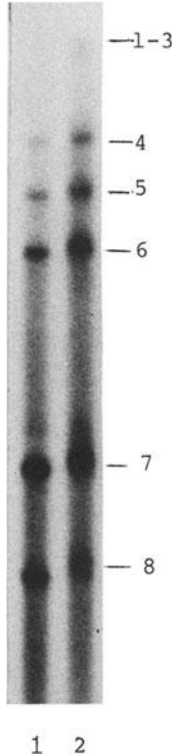


FIGURE 1. FPV mRNAs transcribed in vitro using ApG and globin mRNA as primer molecules.

The numbers on the right of the figure indicate the genome segment from which the RNA was transcribed. Each reaction mixture (150 μl) contained 100 μg purified virus, 0.3% NP40, 150 mM KCl, 50 mM Tris-HCl (pH 8.2), 5 mM $MgCl_2$, 2 mM ATP, 0.4 mM CTP, 0.2 mM GTP, 0.4 mM UTP, 5 mM DTT and 10 μCi α^{32}P-UTP. Incubation was for 1 hr at 31°C and the reaction mixtures were pronase treated before extraction of the product RNAs with phenol-SDS. The product RNAs were deadenylated with RNAse H before electrophoresis on a 30 cm polyacrylamide gel (4%) containing 7 M urea, 0.01 M Tris-boric acid (pH 7.3) and 3 mM EDTA. Electrophoresis was performed at 300V for 16 hr (43). Track 1 shows the product RNAs from a reaction with ApG (0.4 mM) as primer. Track 2 shows the product RNAs from a reaction with rabbit globin mRNA (100 μg/ml) as primer. It can be seen that the FPV RNAs synthesised in the presence of the globin mRNA primer migrate more slowly due to the addition of the cap structure and extra globin-derived nucleotides to the 5' ends.

mixture containing detergent-disrupted virions (6,7). Later it was found that ApG, which is exactly complementary to the 3' termini of the genome RNA segments (8) was the most efficient dinucleotide primer, and that its presence, together with the use of Mg^{++} as the divalent cation, resulted in the in vitro synthesis of full-length A(+) cRNA molecules by detergent-disrupted fowl plague virions (Figure 1). The product RNAs made in vitro in the presence of an ApG primer molecule lack a 5' cap structure, and the influenza virion has no detectable capping and methylating enzymes such as those present in other negative strand viruses such as vesicular stomatitis virus (VSV) (9). When a capped mRNA molecule is added to the virion transcriptase reaction, the resulting A(+) cRNA product is found to contain, at the 5' terminus, the cap structure plus 10-14 nucleotides of the RNA molecule, resulting in a lengthening of the RNA transcripts (Figure 1). A variety of capped

Table 2. Effect of various RNA molecules on in vitro RNA transcription by fowl plague virus

Added RNA (μg per 25 μl reaction)		Relative Transcriptase Activity
No primer	-	1.0
ApG	4.5 μg	8.5
Rabbit globin mRNA	2.5 μg	8.8
	1.25 μg	8.1
	0.63 μg	6.6
	0.13 μg	3.4
Reovirus mRNA	6.0 μg	6.5
Tobacco rattle virus small RNA	12.5 μg	6.2
	5.0 μg	3.9
	3.5 μg	2.7
Chick cell ribosomal RNA	10 μg	1.0

The standard reaction mixture (25 μl) contained 50 mM Tris HCl (pH 8.2), 150 mM KCl, 8 mM $MgCl_2$, 0.2 mM $MnCl_2$, 5 mM dithiothreitol, 0.5% Nonidet P40, 2 mM ATP, 0.4 mM CTP, 0.4 mM UTP, 0.2 mM GTP, and 25 μg purified virus. Data from Nichol (39).

mRNAs have been found to act as donors in this priming reaction (10), and the stimulation of fowl plague virus transcription observed with various RNA molecules is shown in Table 2. Apart from specific dinucleotides, only cap-containing RNA molecules were found to stimulate the *in vitro* reaction.

5' Sequence of mRNA molecules synthesised in vivo

That 5'-terminal capped extensions were also present on virus mRNAs synthesised in infected cells *in vivo* was suggested by size analysis (10). This phenomenon was investigated by a detailed analysis of the nucleotide sequence present at the

5' ends of fowl plague virus mRNAs extracted from infected cells. Poly(A)-containing RNA (mRNA) was isolated from chick embryo fibroblast cells infected with fowl plague virus and reverse transcribed into 'negative strand' DNA using an oligo dT primer. 'Positive strand' DNA was similarly synthesised, using the same primer with *in vitro* polyadenylated influenza virion RNA segments. The DNAs were separated by polyacrylamide gel electrophoresis and the negative and positive strand DNAs corresponding to RNA segment 7 (matrix protein gene) were isolated and hybridised to each other as described (11). The resulting DNA duplex was then cloned after tailing with poly(dA), by insertion into plasmid pBR322 in *E. coli* K12 strain HB101 using standard procedures (13). Repair of these molecules in *E. coli* should retain the complete terminal sequences including any additional host-derived primer sequences originally present at the 5' ends of the mRNA molecules. From one *E. coli* clone, which contained a full-length insert of DNA derived from segment 7 (clone pFPV.7.1) terminally-labelled restriction fragments were obtained and sequenced by the method of Maxam and Gilbert (13). The terminal virus-specific sequence of this clone is shown in Figure 2, together with the predicted sequence for the 5' end of mRNA and the known sequence at the 3' end of genome RNA.

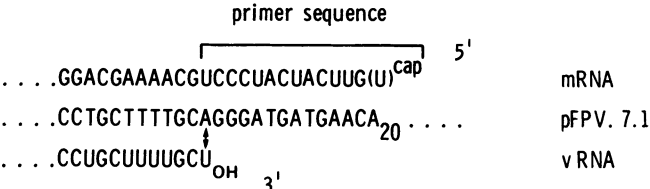


FIGURE 2. 3' Terminal sequence of clone pFPV.7.1 compared with the 3' terminus of RNA segment 7 (8). From Caton and Robertson (14).

It is clear that clone pFPV.7.1 contains an additional 13 non-virus-coded nucleotides at the end of the plasmid insert, supporting the predictions from priming of *in vitro* transcription by eukaryotic cell mRNA molecules. The cap structure is not transcribed during the initial synthesis of cDNA and so is not present in the cloned DNA. Another feature of fowl plague virus RNA transcription revealed by this sequence analysis is the absence of the complement of the 3' terminal U residue of the genome RNA in mRNA as represented in clone pFPV.7.1 (14).

Further experimental evidence for this phenomenon was obtained by direct sequence analysis of a population of infected cell segment 7 mRNAs by the dideoxy-chain termination method (15), using an 87 nucleotide restriction fragment of pFPV.7.1 DNA as a primer molecule (14). These studies clearly showed that the 5' ends of a population of matrix gene mRNAs were heterogeneous in length, over a range of 5-6 nucleotides. The last virus-specific nucleotide transcribed was the penultimate virion RNA nucleotide (C), and the next 9-15 nucleotides were clearly heterogeneous in sequence and non-virus-coded (14). Thus transcription of influenza virus RNA *in vivo* begins at the penultimate nucleotide of the genome RNA and is primed by heterogeneous RNA species (presumably capped host cellular mRNAs) which do not have any obvious common sequence.

Polyadenylation of influenza virus mRNA

Polyadenylation of each of the mRNA species synthesised by VSV occurs at a specific virion RNA sequence AUAC followed by a tract of six or seven uridine residues (16-18). Terminal sequence analysis of fowl plague virus RNA revealed a similar tract of 5-7 uridine residues approximately 17-22 nucleotides from the 5' end of each segment, although no common nucleotide sequence which might act as a signal for polyadenylation was found prior to the U tract (8). Assuming that the U tract was nevertheless involved in polyadenylation, an approach similar to that employed by Schubert *et al*. (18) was used to test the assumption experimentally. The 5' terminal nucleotide sequences of the eight segments of fowl plague virus (8) indicate that

in segments 4, 5, 7 and 8 the U tracts are situated at or close to the 5' end of large T_1 oligonucleotides (27, 16, 13 and 19 residues long respectively).

After T_1 RNase digestion of fowl plague virion RNA, these large oligonucleotides specifically hybridised with the 3'-terminal region of virus-specified mRNA, including the first 5-7 adenine residues of the 3' poly(A) tail (19).

Selection of the predicted oligonucleotides confirms that the 5' terminal U tracts in at least four of the fowl plague virion RNA segments are sites of polyadenylation of mRNA during transcription, and it seems reasonable to assume that this is true for all eight segments. Thus, with the exception that no common sequence which might signal polyadenylation can be found 3' proximal to the U tract, mRNA synthesis by influenza virus appears to terminate in the same way as by other negative strand viruses. A possible mechanism for generating long poly(A) tails from a short stretch of uridine residues would be a 'stuttering' by the RNA transcriptase at this sequence, resulting in repetitive copying of the uridine tract (19).

Polyadenylation occurs normally both in vitro, and in the presence of cycloheximide in vivo, and so does not require any host cell factor or virus-specified protein synthesis. Since cycloheximide blocks synthesis of A(-) cRNA (20) it is likely that new virus protein synthesis is required to prevent polyadenylation and allow normal transcription to the 5' ends of the genome RNA segments. Possible candidates for modification of the virus polymerase or the genome template so as to allow read-through are the non-structural proteins (NS_1, NS_2 and M_2), encoded by genome RNA segments 7 and 8, which are not present in the virion. Phenotypic analysis of two temperature-sensitive mutants with lesions in segment 8 has not, however, revealed any defect in A(-) cRNA production at the restrictive temperature although vRNA synthesis was blocked (21).

Kinetics of mRNA synthesis in vivo

Fowl plague virus mRNAs were measured in infected cells following selection of the polyadenylated fraction using

oligo-dT columns (4). The concentration of virus-specific sequences in this fraction was measured at various times post-infection by hybridisation to ^{32}P-labelled fowl plague virus genome RNA (20,22). The RNA could be fractionated by polyacrylamide gel electrophoresis to enable individual mRNA species to be detected. The procedure used was to determine the amount of polyadenylated cRNA needed to saturate a constant small amount of the labelled vRNA probe. A ranging hybridisation was first carried out to determine the amount of cRNA needed to give about 30% hybridisation with the selected amount of vRNA. This amount of cRNA was then hybridised to saturation with increasing amounts of vRNAs. Double reciprocal plots of the data were used to estimate the amount of cRNA present: extrapolation of the straight line reciprocal plots to an infinite number of input counts was used to obtain the maximum number of labelled vRNA counts bound to the cRNA at saturation, and expressed as counts per minute vRNA bound per μg cRNA. From these results, and taking into account the specific activity of the virion RNA probe and the recovery of total nucleic acid from the infected cells, it is possible to calculate genome copy numbers present in the samples. Analysis of the total genome copies in the nuclear and cytoplasmic fractions of fowl plague virus-infected cells showed that at early times there was a high proportion of transcripts in the nuclear fraction, but from 2 hours post-infection there was a dramatic rise in total polyadenylated transcripts in the cytoplasm (Table 3).

Site of viral RNA synthesis

Influenza virus, unlike other negative strand viruses, requires a functional host cell nucleus for replication (23). It is now clear that the functions which the host nucleus provides relate to the capping and priming of virus mRNA transcripts as outlined in the previous section. Proteins which are known to be involved in mRNA transcription migrate to the nucleus shortly after synthesis (24). It is, therefore, probable that the nucleus is the site of virus mRNA transcription. Further evidence for this has come from our recent studies

Table 3. Estimated amounts of influenza cRNA present in the polyadenylated RNA of nucleus and cytoplasm

Time post-infection (hours)	μg total nucleic acid recovered		^{32}P-vRNA counts/min hybridised per μg polyadenylated RNA		Ratio of nuclear to cytoplasmic	Number of cRNA copies per		Percentage of total cRNA copies present in nucleus
	Cytoplasm	Nucleus	Cytoplasm	Nucleus	Specific activity	Cytoplasm	Nucleus	
0.5	690	325	548	4,511	8.2	0.8	1.4	63
1.0	780	230	2,231	9,125	4.1	3.6	3.0	46
1.5	900	345	20,759	25,244	1.2	37	29	44
2.0	600	220	158,068	105,479	0.67	297	113	28
2.5	760	325	185,534	86,582	0.47	373	78	17
3.0	600	155	931,262	17,572	0.019	2,009	27	1.3

RNA was extracted from nuclei and cytoplasm of primary chick embryo fibroblasts (CEF) monolayers infected with fowl plague virus at a multiplicity of 50 pfu/cell. Cells were harvested into ice-cold saline. The cell pellets were resuspended in RSB (0.1M NaCl, 0.0015M $MgCl_2$, 0.01M Tris HCl, pH 7.4), kept in ice for 15 min and homogenised in a Dounce homogeniser. Crude nuclei were pelleted at 1000 x g for 1 min. The supernatant (cytoplasm) was kept and the nuclei resuspended in RSB. Deoxycholate was added to 0.2% w/v and NP40 to 1% v/v. The nuclei were vortexed for 1 min and pelleted at 1000 x g for 1 min. The supernatant was pooled with the previous cytoplasmic supernatant and centrifuged at 7000 x g for 10 min to remove mitochondria. The nuclei were washed three times with 0.32M sucrose, 1 mM $MgCl_2$ and the final nuclear pellet resuspended in RSB. Equal volumes of pronase (1 mg/ml) in 0.05M NaCl, 0.01M EDTA, 0.5% SDS, 0.1M Tris HCl, pH 7.5) were added to both the nuclear and cytoplasmic fractions which were then incubated at 37° for 1 hr. The nucleic acid was then extracted with equal volumes of chloroform-isoamyl alcohol (25:1) for 5 min at 37°. The aqueous phase was reextracted with chloroform-isoamylalcohol, and the final aqueous phase precipitated with ethanol at -20°. The nucleic acid fractions were resuspended in 10 x RSB and incubated with DNase I (50 μg/ml final concentration) for 2 hr at 28°. The nucleic acid was then purified by pronase treatment and chloroform-isoamyl alcohol extraction as before. RNA samples were dissolved in 1 mM EDTA, 10 mM Tris HCl pH 7.4, and 90% with DMSO, heated to 45° for 20 min, and ethanol-precipitated. Polyadenylated RNA was selected in oligo d(T) columns. RNA samples were hybridised with ^{32}P-labelled vRNA (1-10 x 10^6 cpm/μg) as described (20,22).

with toyocamycin. This drug is an analogue of adenosine which is phosphorylated in the cell and incorporated into RNA in place of adenosine.

Toyocamycin inhibits replication of fowl plague virus by 99% at a concentration of 10 μg/ml in MDCK cells but even at concentrations four times as great virus RNA transcription is not abolished. Under these conditions 90% of the virus-specific RNA transcripts produced remain in the nucleus (25). The fowl plague virus RNA transcripts produced in toyocamycin-treated cells are polyadenylated and function normally as templates for protein synthesis in cell-free translation (Figure 3) although short pulses with ^{35}S methionine in vivo fail to show virus-specific proteins above the host cell background. These results provide further evidence for the nucleus as the site of influenza virus mRNA synthesis.

Treatment of adenovirus-infected HeLa cells with toyocamycin also resulted in an accumulation of virus-specific HnRNA in the nucleus (26).

Synthesis of spliced mRNA molecules

The primary mRNA transcripts of fowl plague virus consist of eight RNA species which are almost full-length transcripts of their corresponding genes. In the presence of cycloheximide these are the only transcripts that can be detected. On secondary transcription three subgenomic or spliced mRNA species appear, one from gene 8 and two from gene 7 (27,28). In the presence of toyocamycin these spliced mRNAs fail to appear and no proteins corresponding to these mRNAs can be detected by cell-free translation (Figure 3). Unlike cycloheximide, toyocamycin does not inhibit splicing by preventing the synthesis of essential proteins. If cycloheximide is added at 2½ hr post-infection, when the proteins required for splicing have accumulated, the spliced RNAs can be produced from newly synthesised RNA. However, if toyocamycin is added at this time, synthesis of the spliced mRNAs cannot be detected (Figure 4). This is analogous to the situation described in adenovirus-infected cells treated with toyocamycin, where HnRNA accumulates in the nucleus. One effect of toyocamycin is to reduce the

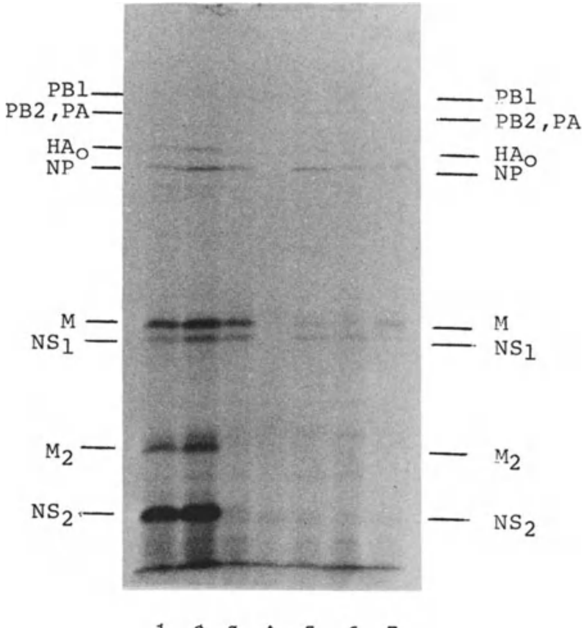


FIGURE 3. Virus-specific proteins produced in a cell-free translation system using polyadenylated RNA extracted from FPV-infected MDCK cells treated with toyocamycin and cycloheximide.

Polyadenylated RNA was extracted from the cytoplasm of FPV infected MDCK cells as described in the legend to Table 3. RNA from 5 10 cm confluent dishes of MDCK cells was pooled and dissolved in 50 μl water. Each reaction mixture (10 μl) contained 5 μl preincubated wheat germ extract, 10 mM HEPES buffer (pH 7.6), 2 mM DTT, 50 μg/ml creatine phosphokinase, 10 mM creatine phosphate (disodium salt), 1 mM ATP, 0.8 mM spermidine hydrochloride, 100 μM GTP, 100 mM KCl, 50 μM each of unlabelled amino acids, except methionine, 10 μCi [^{35}S]methionine, 1.5 mM $MgCl_2$, and 2 μl of the test RNA (42). Reactions were carried out for 1 hr at 30°C. An equal volume of SDS gel loading buffer was added and the protein mixture analysed by the Laemmli discontinuous gel system (44) using a 17.5% gel. The gels were fluorographed, dried and exposed to X-ray film. The lettering on the side of the figure indicates the positions of the virus-specific proteins. Track 1, untreated cells; 2, cells treated with DMSO (0.5%); 3, cells treated with toyocamycin (40 μg/ml) dissolved in DMSO (final concentration 0.5%); 4, no RNA added; 5, cells treated with cycloheximide (100 μg/ml); 6, cells treated with cycloheximide (100 μg/ml) and DMSO (0.5%); 7, cells treated with cycloheximide (100 μg/ml) and toyocamycin (40 μg/ml) dissolved in DMSO (final concentration 0.5%).

methylation of RNA. In the case of fowl plague virus-infected MDCK cells, we found that methylation of all RNA species was reduced by greater than 70%. It has been suggested that for adenovirus, internal methylation is involved in the formation of mature mRNA by splicing (29) and such a mechanism is possibly also involved in the synthesis of spliced mRNAs in fowl plague virus infections.

Control of mRNA synthesis

The synthesis of fowl plague virus-specific polypeptides is regulated in the infected cell. Early after infection of chick embryo fibroblast cells, the nucleocapsid and non-structural polypeptides are synthesised in increased amounts relative to other virus-specific proteins, whereas later in infection the synthesis of the matrix polypeptide predominates. Our early results showed that translation of cytoplasmic poly(A)-containing RNA from infected cells, using a wheat germ cell-free system, resulted in a similar pattern of virus-specific polypeptides *in vitro* to that observed at the corresponding times by pulse-labelling *in vivo* (30). Thus the amounts of virus-specific proteins formed in fowl plague virus-infected cells are a direct reflection of the levels of mRNA in the cells.

Direct measurement of the amounts of individual mRNAs was carried out by hybridisation of A(+) cRNA, extracted from cells at various times post-infection, to individual ^{125}I-labelled virus RNA segments (20). The amount of each A(+) cRNA produced in infected cells did not simply reflect the size of the gene from which it was transcribed. At 1 hr post-infection, the transcripts were present in approximately inverse proportion to the size of the corresponding RNA segment from which they were derived (Table 4). However by 2 hr, there was a relatively greater concentration of RNA transcribed from segments 5 and 8, and at 4 hr the RNA transcribed from segment 7 predominated.

The transcriptional controls observed in these experiments were abolished during infections in the presence of cycloheximide (Table 4) indicating that newly synthesised, probably virus-specific, proteins are essential for the observed

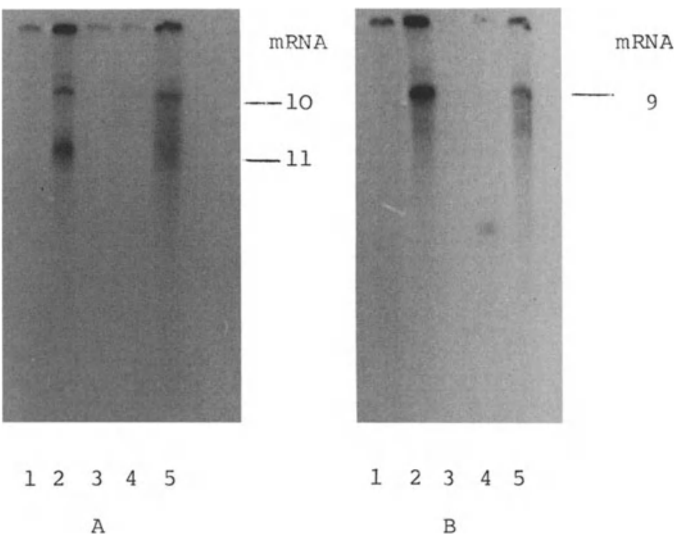


FIGURE 4. Effect of toyocamycin and cycloheximide on the synthesis of spliced FPV mRNA species.

MDCK cells were infected with FPV and phosphate-free maintenance medium added. Cycloheximide and toyocamycin were added either immediately after infection or at 2½ hr after infection as indicated. At 3½ hr post-infection 2 mCi ^{32}P orthophosphate was added to the cells. Cells were harvested at 6½ hr post infection and total RNA extracted from the cytoplasm as described in Table 3. ^{32}P-labelled RNAs complementary to genome segments 7 and 8 were selected using specific cloned DNAs bound to nitrocellulose discs (27,28). RNA which bound specifically was deadenylated using RNase H and the spliced RNAs analysed by electrophoresis on a 7% acrylamide gel containing 7M urea, 0.01M Tris-boric acid (pH 8.3) and 3 mM EDTA. Electrophoresis was performed at 300V for 3 hr. Gels were fixed in 10% methanol, 10% acetic acid (4 x 250 ml) for 1 hr, dried onto Whatman 3 mm filter paper and exposed to X-ray film. Panel A, RNAs selected using a clone containing a DNA copy of gene 7. Panel B, RNAs selected using a clone containing a DNA copy of gene 8 DNA. The positions of the mRNAs for protein NS_2 (mRNA 9), protein M2 (mRNA 10) and the mRNA for a putative peptide M3 (mRNA 11), are indicated. Track 1, RNA from cells treated with cycloheximide (100 μg/ml) from time 0; 2, RNA from cells treated with cycloheximide from 2½ hr; 3, RNA from cells treated with toyocamycin (40 μg/ml) from time 0; 4, RNA from cells treated with toyocamycin (40 μg/ml) from 2½ hr; 5, RNA from control cells.

Table 4. Amount of virus-specific polyadenylated cRNA in chick embryo fibroblasts infected with fowl plague virus in the presence and absence of cycloheximide

Virion RNA segments	Control 1h	2h	4h	Cycloheximide-treated 1h	2h	4h
1-3	43	163	2918	22	47	102
4	86	467	10852	17	70	116
5	138	2183	13216	32	99	174
6	172	1023	13816	29	159	222
7	176	1259	16887	28	159	244
8	186	3261	14470	42	159	228

The concentration of individual virus-specific A(+) cRNAs was measured by hybridisation to separated ^{125}I-labelled vRNA segments. Segments 1-3 were not well separated and were combined in these experiments. Hybridisation was carried out as described in the text and elsewhere (20). The results are expressed as pg virus-specific RNA/μg of total cytoplasm RNA (20). Annealing was carried out in 2 x SSC at 68°C for 18 hr. The extent of hybridisation was determined by treatment of the hybrids with a mixture of RNase A and T_1 (50 μg/ml and 50 U/ml, respectively). Where indicated, cycloheximide was present at a concentration of 100 μg/ml from 1 hr prior to, and throughout the period of, infection.

regulation (20,31).

Synthesis of virion RNA

The RNA templates for synthesis of new negative-stranded virion RNA molecules are complete, non-polyadenylated transcripts of the genome RNA segments. Although these A(-) cRNAs are not found in association with cell polysomes, and are presumed not to act as mRNAs in the infected cell, they will direct the synthesis of virus-specific polypeptides *in vitro* (31).

To measure the kinetics of synthesis of A(-) cRNAs, we analysed the non-polyadenylated fraction of RNA extracted from the cytoplasm and nucleus of infected cells at various times using ^{125}I-labelled virion RNA as a probe. At the same time, we measured the amount of vRNA present in the samples by hybridisation to a ^{3}H-cDNA copy of virion RNA, prepared by reverse transcription *in vitro* with oligo(dG) as the primer molecule (20).

Table 5. Estimation of the number of genome equivalents of non-polyadenylated cRNA and vRNA in the nucleus and cytoplasm of fowl plague virus-infected chick embryo fibroblasts

	1h	1.5h	2h	3h	4h	6h
cRNA						
cytoplasm	100	337	1125	1425	1437	1687
nucleus	50	112	156	312	475	737
vRNA						
cytoplasm	112	237	262	662	1050	1400
nucleus	50	62	75	262	512	1150

Cells were infected, and RNA was extracted, as described in Table 3. The concentration of virus-specific template cRNA (non-polyadenylated cRNA) and vRNA in the RNA fraction which did not bind to oligo(dT) cellulose was determined by annealing the RNA to both a ^{3}H-labelled cDNA copy of virion RNA, prepared by reverse transcription of vRNA _in vitro_ with oligo(dG) as the primer molecule, and to ^{125}I-labelled vRNA (20).

From 1.5 to 2 hr there was a dramatic rise in A(-) cRNA accumulation, after which there was little further increase in the cytoplasm (Table 5). A(-) cRNA accumulated steadily in the nuclear fraction throughout infection, but at a much lower concentration. The kinetics of new vRNA synthesis were different. After remaining at input level during the first hour after infection, there was a slow increase up to two hours then a faster accumulation from 2.5 hours, when the bulk of template A(-) cRNA had been formed (Table 5).

DISCUSSION

The segmented nature of the influenza virus genome has greatly facilitated its analysis at the nucleotide sequence level. The conservation of 3' and 5' terminal sequences for 12 or 13 nucleotides respectively (8) has enabled specific primers to be used for reverse transcription into DNA which can then be sequenced by a variety of methods. To date, the sequences of more than 40 influenza RNA segments have been published, including the complete sequence of the A/PR/8/34 strain (reviewed in 32).

The high degree of conservation between the terminal RNA sequences of a wide range of influenza virus strains suggests that these sequences are crucial to the biosynthetic mechanism.

Although they presumably contain binding sites for transcription and replication, their exact significance is unknown. However, it seems almost certain that each segment is transcribed and replicated as a separate unit.

The influenza virion transcriptase is capable of faithful transcription of all RNA segments both in vitro and in vivo. The transcripts are polyadenylated at their 3' termini, the signal for polyadenylation being the tract of U residues present 17-22 nucleotides from the 5' terminus of the virion RNA template (19). The polyadenylated transcripts therefore lack the sequence complementary to approximately 16 nucleotides at the 5' terminus of virion RNA, and so cannot act as templates for the synthesis of new genome RNA segments. Since all the in vitro transcripts are polyadenylated, it is likely that a virus-specific protein is necessary to prevent polyadenylation and so to allow read-through to the 5' terminus of the virion RNA template. There are several candidates for this protein, as at least three non-structural proteins (NS_1, NS_2 and M_2) are synthesised in influenza virus-infected cells. However, analysis of two fowl plague virus mutants with a temperature-sensitive lesion in RNA segment 8 did not reveal a defect in template cRNA synthesis, although vRNA synthesis was inhibited (21). Only one temperature-sensitive mutant with a defect in RNA segment 7, encoding the M2 protein, has been reported, but this displayed no RNA synthetic defect at the non-permissive temperature (33).

The RNA transcriptase of fowl plague virus, as with other influenza viruses (10), requires a primer molecule for optimum activity in vitro. The most efficient primer molecules are capped mRNAs, and these appear to be the primer molecules for RNA transcription in infected cells. One of the components of the virion RNA transcriptase, the PB2 polypeptide, binds specifically to cap-containing RNA molecules (34), and this appears to be the first transcriptional event during normal virus replication. Viruses with a temperature-sensitive mutation in RNA segment 1, which encodes PB2, all show defects in RNA transcription, and many such mutants have been found to be

defective in the recognition of capped primer RNA molecules (25).

Although capped RNA molecules stimulate the rate of transcription by influenza virions *in vitro*, we have obtained evidence that priming and stimulation of transcription are not necessarily linked. Priming involves base-pairing to the 3' terminus of the template vRNA, resulting in stimulation of virus-specific RNA synthesis in a manner similar to that of dinucleotides such as ApG. However, some cap analogues such as m^7GpppN^6mAm apparently stimulate the virus transcriptase by direct interaction with the cap-binding site, possibly involving a conformational change in the PB2 polypeptide (35). Sequence analysis of mRNAs synthesised in fowl plague virus-infected cells showed considerable heterogeneity in the primer regions (14) and it would appear that virtually any cell mRNA species can donate its 5' terminal 10-15 nucleotides in the cap-transfer reaction. However, since inhibition of mRNA synthesis by drugs such as α-amanitin or actinomycin D results in such a rapid cessation of influenza virus RNA transcription (5) it seems that only newly synthesised capped mRNA molecules can participate as cap donors. We have presented evidence here and elsewhere (23) that the nucleus is the site of influenza mRNA synthesis, and the cap transfer reaction may therefore require nascent host cell mRNA molecules, those which have left the nucleus being no longer available to participate in influenza RNA transcription.

Another function provided by the host cell nucleus is that of splicing, since transcripts of RNA segments 7 and 8 are each spliced to form additional mRNAs in the infected cell (reviewed in 32). We found that the production of spliced mRNAs was inhibited by toyocamycin, an adenosine analogue which also prevents the formation of spliced RNAs in adenovirus-infected cells (26). Toyocamycin is effective in mammalian cells such as MDCK cells, but in chick embryo fibroblast cells it has less inhibitory effect on influenza virus replication, perhaps because it is not efficiently incorporated into such cells. RNA methylation is reduced in MDCK cells treated with toyocamycin, and it is possible that this reduction in methylation

is responsible for the failure in RNA splicing. It is not clear whether virus-specific proteins play a role in the splicing of transcripts from segments 7 and 8. Although splicing is blocked by addition of cycloheximide early in infection, the spliced RNA encoding M3 has been detected from cloned DNA copies of segment 7 introduced into Cos-1 cells using an SV40 vector (36). The splicing signals can therefore be recognised by host cell enzymes.

Transcription of the fowl plague virus genome is regulated in infected cells, and this results in strict temporal control of polypeptide synthesis. By translating cytoplasmic RNA from fowl plague virus-infected cells in a wheat germ cell-free system, it is possible to analyse the relative amounts of functional mRNAs at various times post-infection (31). These experiments show that there are three phases in RNA biosynthesis. The first, primary transcription phase results in close to equimolar amounts of each RNA transcript, but as soon as virus-specific proteins have been synthesised (probably within 30 minutes post-infection), early secondary transcription begins. During this second phase the mRNAs for segments 5 and 8 become amplified relative to the other segments. The final phase, late secondary transcription, is characterised by amplification of the mRNAs for segments 4, 6 and 7, together with a reduction in segment 8 mRNA. A recent analysis by Smith and Hay (37) has shown a similar pattern in the regulation of vRNA synthesis to that described above for mRNA, and it is possible that amplification and selective transcription occurs at the level of vRNA synthesis. Template cRNA is synthesised early in infection, probably from input virion RNA molecules, since normal amounts of cRNA are synthesised in cells infected with temperature-sensitive mutants which fail to make vRNA at the non-permissive temperature (21,38).

Which virus-specific proteins are responsible for the regulation of vRNA and mRNA synthesis is presently an unsolved question. Even less well understood is the process of assembly, whereby the progeny influenza virions acquire their correct complement of RNA segments. The identical 3' and 5' termini

present on each RNA segment would seem to preclude a role for these regions in regulated assembly. However the RNAs appear to be assembled in the form of ribonucleoproteins, and it is also possible that correct packaging involves a protein-protein interaction. Despite the wealth of information on the structure and biosynthesis of the influenza virus, these and many other unsolved questions remain open to future research.

ACKNOWLEDGMENTS

We wish to thank our many past and present colleagues, especially Andrew Caton, Stephen Inglis, Stuart Nichol, Charles Penn, Jamie Robertson and Adrian Wolstenholme, to whose work we have referred. We are also grateful to our typist, Mrs. Mary Wright, who prepared the camera-ready copy. Our work in Cambridge is supported by the Medical Research Council of Great Britain.

REFERENCES

1. Grist, N.R. Brit. med. J. ii: 1632-1633, 1979.
2. Barry, R.D., Cruickshank, J.G. and Wells, R.H.J. Vet. Rec. 76: 1316-1322, 1964.
3. McGeoch, D., Fellner, P. and Newton, C. Proc. Natl. Acad. Sci. USA 73: 3045-3049, 1976.
4. Glass, S.E., McGeoch, D. and Barry, R.D. J. Virol. 16: 1435-1443, 1975.
5. Barry, R.D. and Mahy, B.W.J. Brit. med. Bull. 35: 39-46, 1979.
6. McGeoch, D. and Kitron, N. J. Virol. 15: 686-695, 1975.
7. Carroll, A.R., McGeoch, D.J. and Mahy, B.W.J. Colloq. Inst. Nat. Sante Rech. Med. 47: 95-102, 1975.
8. Robertson, J.S. Nucleic Acids Res. 6: 3745-3759, 1979.
9. Banerjee, A.K., Abraham, G. and Colonno, R.J. J. gen. Virol. 34: 1-8, 1977.
10. Krug, R.M. Curr. Topics Microbiol. Immunol. 93: 125-149, 1981.
11. Caton, A.J. and Robertson, J.S. Nucleic Acids Res. 7: 1445-1456, 1979.
12. Sleigh, M.J., Both, G.W. and Brownlee, G.G. Nucleic Acids Res. 6: 1309-1321, 1979.
13. Maxam, A. and Gilbert, W. Proc. Natl. Acad. Sci. USA 74: 560-564, 1977.
14. Caton, A.J. and Robertson, J.S. Nucleic Acids Res. 8: 2591-2603, 1980.
15. Sanger, F., Nicklen, S. and Coulson, A.R. Proc. Natl. Acad. Sci USA 74: 5463-5467, 1977.
16. McGeoch, D.J. Cell 17: 673-681, 1979.
17. Rose, J.K. Cell 19: 415-421, 1980.

18. Schubert, M., Keene, J.D., Herman, R.C. and Lazzarini, R.A. J. Virol. 34: 550-559, 1980.
19. Robertson, J.S., Schubert, M. and Lazzarini, R.A. J. Virol. 38: 157-163, 1981.
20. Barrett, T., Wolstenholme, A.J. and Mahy, B.W.J. Virology 98: 211-225, 1979.
21. Wolstenholme, A.J., Barrett, T., Nichol, S.T. and Mahy, B.W.J. J. Virol. 35: 1-7, 1980.
22. Mahy, B.W.J., Carroll, A.R., Brownson, J.M.T. and McGeoch, D.J. Virology 83: 150-162, 1977.
23. Mahy, B.W.J., Barrett, T., Briedis, D.J., Brownson, J.M. and Wolstenholme, A.J. Phil. Trans. Roy. Soc. B. 288: 349-357, 1980.
24. Briedis, D.J., Conti, G., Munn, E.A. and Mahy, B.W.J. Virology 111: 154-164, 1981.
25. Mahy, B.W.J., Penn, C.R., Nichol, S.T., Briedis, D.J. and Barrett, T. In: Genetic Variation among Influenza Viruses (Ed. D.P. Nayak), Academic Press, New York, pp.113-125.
26. Swart, C. and Hodge, L.D. Virology 84: 374-389, 1978.
27. Inglis, S.C., Gething, M.-J. and Brown, C.M. Nucleic Acids Res. 8: 3575-3589, 1980.
28. Inglis, S.C. and Brown, C.M. Nucleic Acids Res. 9: 2727-2740, 1981.
29. Chen-Kiang, S., Nevins, J.R. and Darnell, J.E. J. mol. Biol. 135: 733-752, 1979.
30. Inglis, S.C., Carroll, A.R., Lamb, R.A. and Mahy, B.W.J. Virology 74: 489-503, 1976.
31. Inglis, S.C. and Mahy, B.W.J. Virology 95: 154-164, 1979.
32. McCauley, J.W. and Mahy, B.W.J. Biochem. J. 211: in press, 1983.
33. Ghendon, Y.Z., Markushin, S.G., Klimov, A.I., Lotte, V.D. and Ginzburg, V.P. J. gen. Virol. 64: 291-304, 1983.
34. Penn, C.R., Blaas, D., Kuechler, E. and Mahy, B.W.J. J. gen. Virol. 62: 177-180, 1982.
35. Penn, C.R. Influenza virus messenger RNA synthesis. Ph.D. thesis, University of Cambridge, 1983.
36. Lamb, R.A. and Lai, C.-J. Virology 123; 237-256, 1982.
37. Smith, G.L. and Hay, A.J. Virology 118: 96-108, 1982.
38. Mahy, B.W.J. In: Genetics of Influenza Viruses (ed. Palese, P), Springer Verlag, in press, 1983.
39. Nichol, S.T. In vitro transcription of fowl plague virus RNA. Ph.D. thesis, University of Cambridge, 1980.
40. Almond, J.W. and Barry, R.D. Virology 92: 407-415, 1979.
41. Almond, J.W., McGeoch, D. and Barry, R.D. Virology 92: 416-427, 1979.
42. Inglis, S.C., McGeoch, D.J. and Mahy, B.W.J. Virology 78: 522-536, 1977.
43. Carter, M.J. and Mahy, B.W.J. Arch. Virol. 71: 13-25, 1982.
44. Laemmli, U.K. Nature 227: 680-685, 1970.

Index

Abelson murine leukemia virus (AbLV), 287
Acriflavin, and group I RNA phage replication, 321
Actinomycin D
  group I RNA phage replication and, 321
  influenza virus RNA biosynthesis and, 362
  parvovirus DNA replication and, 263
  tobacco mosaic virus (TMV) replication and, 334, 337, 338
Adenine, in *Drosophila* mitochondrial DNA, 132–136
Adeno-associated viruses (AAV), 257, 258, 260–262, 267–268, 271
Adenovirus DNA replication, 237–253
  chain elongation in, 240, 242, 247, 249–251, 252
  description of adenoviruses in, 237–239
  DNA-binding protein (DBP) in, 242–246, 252
  initiation complex in, 239, 241
  in vitro systems for, 239–242
  isolated nuclei in, 239–240
  nuclear extracts in, 240–242
  priming mechanism in, 252–253
  replication complexes in, 240
  replication protein in, 246–249
  spliced mRNA molecules in, 357, 362
African swine fever virus, 209
A gene, and T antigen, 29
α-amantin
  influenza virus RNA biosynthesis and, 362
  parvovirus DNA replication and, 263
Amber mutant, in *E. coli* replication, 97
Amidase, in bacteriophage T7 replication, 185
Aphidicolin
  adenovirus DNA replication with, 249, 250–251, 253
  SV40 DNA replication with, 62
Arabinosyl cytosine, in *E. coli* replication, 107
ATP
  DNA synthesis and, 62
  as promoter in ØX174 replication, 7–8, 18
Avian erythroblastosis virus, 289
Avian myeloblastosis virus (AMV), 287
Avian retrovirus, 293
Avian sarcoma virus (ASV), 277, 278, 280, 283, 301, 302
Avipoxviruses, 206
Bacteriophage 4 DNA replication, 164
Bacteriophage f2, 315, 316
  *E. coli tra*D (Ts) mutation and, 319, 324, 325
  rifampin and, 317, 319
Bacteriophage G4 replication, 163
  priming site in, 11, 12
Bacteriophage Group I RNA replication, *see* Group I RNA phage replication
Bacteriophage λ DNA replication, 42, 43, 164
  induction of replication in, 86, 88
  polyoma (Py) virus insertion with, 90
Bacteriophage M13, 198, 337
  replication of, 11, 12, 163
Bacteriophage MS2, 315
  *E. coli tra*D (Ts) mutation and, 321, 324
  rifampin and, 317, 319
Bacteriophage P2, 102, 104
Bacteriophage Qβ, 315, 316, 334
  rifampin and, 319
  template recognition in, 325
Bacteriophage R17 replication
  *E. coli tra*D (Ts) mutation and, 319, 324
  rifampin and, 317–319
Bacteriophage T4
  E. coli replication with, 100
  helicase activity with, 184–185
Bacteriophage T7 DNA replication, 163–200
  chromosome of bacteriophage T7 in, 164–165
  cloned origin of, 194–196
  dispensable proteins in, 182–185
  DNA binding protein in, 184
  DNA ligase in, 183
  DNA polymerase in, 172, 175
  essential proteins in, 170–182
  E. coli mutants and, 168
  E. coli replication with, 100, 101
  eye forms in, 165, 189, 193
  gene 1.2 protein in, 184–185
  gene 2 protein in, 182
  gene 3 endonuclease in, 180–181
  gene 4 protein in, 175–180
  gene 6 exonuclease in, 181–182
  genetic analysis of, 166–168
  helicase activity in, 176–178
  initiation of, 189–199
  in vitro system of, 168–170
  in vivo replication and, 164–168
  lagging strand synthesis in, 187–189
  leading strand synthesis in, 186–187

Bacteriophage T7 DNA replication (continued)
mutants in, 166–168
primary origins in, 189–194
primase activity in, 178–180
replication bubbles in, 189, 194, 196, 197, 199
replication fork in, 185–189
RNA polymerase in, 171–172, 191–194
rolling circle mechanism in, 187
terminal redundancy in, 165
Y forms in, 165, 189, 195
Benzo(a)pyrene, in polyoma virus synthesis, 84
BKV virus replication, 30, 31
Bovine parvovirus (BPV), 257, 259
Bromodeoxyuridine (BUdR)
mammalian cell DNA replication and, 118, 120–123
parvovirus DNA replication and, 261
polyoma (Py) virus DNA replication induction with, 81, 84
poxvirus DNA replication and, 218

Canary pox virus, 206
Canine parvovirus (CPV), 258
Carcinogens, and DNA replication, 46
Catenanes, 149
groups in, 40–41
mitochondria DNA replication with, 41
monomers and decay of, 43–44
papovavirus replication termination and, 40–41
SV40 DNA replication and, 43
Chloramphenicol, and *E. coli* replication, 105
Circular dimers, 43
Circular molecules, in *Drosophila* mitochondrial DNA (mtDNA) replication, 137, 138, 139, 142–143
Coat protein, in tobacco mosaic virus (TMV) replication, 329, 330, 331, 337, 339, 340, 341
Concatemers
bacteriophage T7 DNA replication with, 165, 166, 167, 169, 170
parvovirus DNA replication with, 270
poxvirus DNA replication with, 222, 225
Cowpea mosaic virus, 337, 340, 341
Cowpox, 206, 210
Cycloheximide
catenated dimers with, 43
influenza virus RNA biosynthesis with, 345, 352, 355, 357–359, 363
parvovirus DNA replication with, 263
Cytosine methylation, 128–129

Deletion mutants
bacteriophage T7 replication with, 174
SV40 DNA replication termination with, 41
T antigen and, 32–34, 35
Deoxycholate, and group I RNA phage replication, 321
Dideoxithymidine triphosphate (ddTTP), in SV40 DNA replication, 62
Dimers, 149–151
papovavirus replication termination with, 41
parvovirus DNA replication with, 270–271
replication and, 43–44
retrovirus DNA replication with, 280, 282
SV40 DNA replication with, 43
7,12-dimethylbenz(a)anthracene, in polyoma virus synthesis, 84
DNA-binding protein (DBP)
adenovirus DNA replication with, 238, 240, 242–246, 252
bacteriophage T7 replication with, 184, 193
parvovirus DNA replication with, 266, 267, 268
DnaB protein, as replication promoter for Primase, 7, 11, 13
DNA endonuclease, in retrovirus DNA replication, 278–279, 300
DNA gyrase, in *E. coli* replication, 111
DNA ligase
bacteriophage T7 replication with, 182, 183
E. coli replication with, 95, 97, 102, 182, 183
parvovirus DNA replication with, 269
ØX174 DNA replication with, 18, 19, 20
DNA polymerase
adenovirus DNA replication with, 239
bacteriophage T7 replication with, 167, 172–175, 176, 178, 179, 186, 194, 198
parvovirus DNA replication with, 262, 266–267, 268, 269
poxvirus DNA replication and, 218
retrovirus DNA replication with, 278
DNA polymerase α
adenovirus DNA replication with, 249, 250, 253
E. coli and Okazaki fragments and, 185
Okazaki fragments and, 38–39
parvovirus DNA replication with, 266, 267
polyoma DNA replication with, 36, 42
SV40 DNA replication with, 36, 42, 55, 62, 66–67
viral chromosomes with, 64–67

DNA polymerase β
dideoxithymidine triphosphate (ddTTP) and, 62
mammalian mitochondrial DNA (mtDNA) replication with, 157, 160
parvovirus DNA replication with, 266
viral chromosomes and, 64–66
DNA polymerase γ
adenovirus DNA replication with, 249–251, 253
dideoxithymidine triphosphate (ddTTP) and, 62
parvovirus DNA replication with, 266, 267
viral chromosomes with, 64–66
DNA polymerase I, in *E. coli* replication, 95, 97, 100, 102
DNA polymerase III
E. coli replication with, 97
ØX174 replication with, 15, 18, 19, 21, 24
DNA-TP, in adenovirus DNA replication, 238, 242, 251, 253
Drosophila mitochondrial DNA (mtDNA) replication, 131–147, 300
adenine + thymine (A+T) region in, 131, 132–136
circular molecules in, 137, 138, 139, 142–143
materials and methods in study of, 132
mitochondrial rRNA (mt-rRNA) genes in, 143
origin and direction of, 132, 137–141, 146
primers for, 101
replication intermediaries in, 136–137, 138, 139, 141
replication modes in, 141–143
replicative forms in, 136–137
sequence divergence in, 144–145, 146–147
Drosophila melanogaster mitochondrial DNA (mtDNA) replication
adenine + thymine (A+T) region in, 132, 134
origin and direction of, 138, 139–141, 146
replicative forms in, 136
sequence divergence in, 144
Drosophila yakuba mitochondrial DNA (mtDNA), 145–146
D2 protein, in papovavirus replication, 32, 33
Duplex replicative forms
bacteriophage DNA replication with, 163–164, 181
ØX174 DNA replication with, 2, 4, 18–20
dUTPase (*dut*), in *E. coli* replication, 95

Ectromelia DNA, 208
Escherichia coli chromosome
gene 3 and endonuclease and gene 6 exonuclease and degradation of, 181–182
ØX174 replication and, 2, 3
rifampin and, 31
Escherichia coli chromosome replication, 93–112, 300
discontinuous chain elongation in, 94–104
discontinuous replication in, 95–99
dual origin of nascent DNA fragments in, 94–95
excision repair sequences in, 102–104
initiation-defective mutants and, 105
initiation of, at origin, 105–112
mapping procedures in *oriC* region in, 108
model for semidiscontinuous replication in, 23–24
origin in, 93, 94
primer RNA in, 95–99, 101, 107, 109, 111
priming sites in, 107–111
replication forks in, 90
replication promoter with, 6
rifampin and, 316
*Escherichia coli tra*D (Ts) mutation, and group I RNA phage replication, 319–323, 324

FBJ murine osteosarcoma virus (FBJ-MSV), 288
Feline panleukopenia virus (FPV), 257, 258
Feline sarcoma viruses, 287
Fluorodeoxyuridine (FdU)
polyoma virus synthesis and, 84
primers at replication origins and, 123–124
replicon size estimates with, 118
Fowl plague virus (FPV), 345, 346, 360, 361
5′ sequence of mRNA molecules in, 350, 351
kinetics of mRNA synthesis in, 352–353
polyadenylation of, 351
regulation of, 363
site of, 355
Fowlpox virus, 206

Gene A protein, 160
Gene 1.2 protein, in bacteriophage T7 replication, 182, 184–185
Gene 2 protein, in bacteriophage T7 replication, 166–167, 182
Gene 3 endonuclease, in bacteriophage T7 replication, 180–181, 185

Gene 4 protein, in bacteriophage T7 replication, 167, 170, 175–180, 185
  helicase activity in, 174, 175, 176–178
  lagging strand synthesis with, 188
  leading strand synthesis with, 186–187
  primase activity with, 178–180
  RNA polymerase with, 192
Gene 5 protein, in bacteriophage T7 replication, 172, 173, 174
Gene 6 exonuclease, in bacteriophage T7 replication, 167, 169, 170, 181–182, 188
Glioblastoma cells, SV40 infection of, 41
Group I RNA phage replication, 315–325
  *Escherichia coli tra*D (Ts) mutation and, 319–323, 324
  host control of, 315–325
  primary origin with, 196, 197, 198
  rifampin and, 316–319
Gyrase, with duplex ØX DNA, 20

Helicase
  gene 4 protein in bacteriophage T7 replication with, 174, 175, 176–178
  leading strand synthesis with, 186–187
  *opt*A protein and bacteriophage T4 with, 184–185
  replication forks and, 99, 110
H-protein, in tobacco mosaic virus (TMV) replication, 341–342
Hydroxyurea (HU), in poxvirus DNA replication, 213–214, 228

Influenza virus RNA biosynthesis, 345–364
  control of, 357–359
  description of virus in, 346
  5′ sequence of mRNA synthesis in, 349–351
  in vitro studies of, 346–349
  kinetics of mRNA synthesis in, 352–353
  polyadenylation of, 351–352
  site of, 353–355
  spliced mRNA molecules in, 355–357, 362–363
  virion RNA synthesis in, 359–360
Inhibitors, in viral DNA replication, 213–214
Initiation-defective mutants, 105
Initiation sites, 115
  adenovirus DNA replication with, 239, 241, 247, 249
  bacteriophage T7 DNA replication, 179, 180
  in egg or early embryo, 116
  retrovirus DNA replication, 294–295
  tobacco mosaic virus (TMV) replication, 337
  vaccinia DNA replication, 217
  see also Origin of replication
Intermediates, *see* Replication intermediates
Iododeoxyuridine, in polyoma virus synthesis, 84

Kilham rat virus (KRV), 257, 258, 259, 260, 263, 264
λ repressor, 86, 88

Mammalian mitochondrial DNA (mtDNA) replication, 115–129, 149–160
  bidirectional origins of replicon in, 119–120, 128
  displacement-loop (D-loop) in, 149, 151, 154, 156–159
  light-strand (L-strand) synthesis in, 149, 151, 159–160
  materials and methods in study of, 151
  moderately repeated sequences in origins of, 116–117
  primers in, 120–126, 157–158
  replication mode in, 151
  size of replicon in, 117–119
  stop points of D-loop strands in, 157
  10 kb replication intermediate in, 127–128
  unicircular dimer mitochondrial DNA (mtDNA) in, 151–155, 156
Methylation of cytosine, 128–129
20-methylcholanthrene, in polyoma virus synthesis, 84
Mink cell focus-forming (MCF) specific sequences, 289
Mink encephalitis virus (MEV), 258
Minichromosomes, viral, as primers, 54
Minute virus of mice (MVM), 257, 259, 262
Missense mutation, in bacteriophage T7 replication, 174
Mitochondrial DNA replication
  adenovirus DNA replication similar to, 253
  catenanes in, 41
  daughter strand analysis in, 40
Mitomycin C (MCC), in polyoma (Py) virus replication, 72–75, 79, 80, 84, 85
Molluscum contagiosum virus, 209
Moloney murine leukemia virus (Mo-MLV), 275, 277, 279, 280, 283, 292, 294, 295, 297
Moloney murine sarcoma virus (Mo-MSV), 275, 287, 297
Monkeypox, 208, 210
Monomers
  catenane formation with, 41, 43–44

extrachromosomal viral DNA with, 72
papovavirus integration with, 82
parvovirus DNA replication with, 271
polyoma (Py) DNA replication insertion sites and, 77
replication pathways with, 42, 45
Mouse mammary tumor virus (M-MTV), 276, 283, 293
Murine leukemia viruses, 278–279
Mutagens, and DNA replication, 46
Mutants
bacteriophage T7 replication with, 166–168
see also Deletion mutants

Nitrogen mustard, in polyoma virus synthesis, 84
N-nitrosomethylurea, in polyoma virus synthesis, 82
Nucleosomes, in SV40 chromosomes, 58

Okazaki fragments
bacteriophage T7 DNA replication and, 166, 170, 180, 181, 189
DNA polymerase α and, 38–39
DNA synthesis incubation time and, 62
E. coli DNA polymerase I and, 185
E. coli replication with, 24
initiation sites in, 37
initiator RNA primase in, 39
nascent DNA fragments in *E. coli* replication and, 95
papovavirus replication with formation of, 36–39
primers at replication origins and, 123–125
replication fork and synthesis of, 38–39
RNA primers in initiation of, 37–39, 56
SV40 DNA replication with, 58, 60
10 kb replication intermediate in, 128
terminal bush in, 126
Oligomers of viral DNA, 72, 82, 149
Onion skin model of replication, 45, 46, 87
OptA protein, in bacteriophage T7 replication, 184
Orf virus, 206
Origin of replication
bacteriophage T7 replication, 189–194
Drosophila mitrochondrial DNA replication, 132, 137–141, 143, 146
Drosophila yakuba mitochondrial DNA, 145
E. coli replication, 93, 94, 105–112
light-strand (L-strand) synthesis in mammalian mitochondrial DNA (mtDNA) replication, 149, 151, 159–160
moderately repeated sequences in, 116–117
papovavirus replication, 29–32
parvovirus DNA replication, 259
polyoma (Py) virus DNA replication, 35, 89
primers at, 120–126
SV40 DNA replication, 55
T antigen binding at, 32, 34, 35
termination signal and replicating forks and, 41
unicircular dimer mitochondrial DNA (mtDNA) replication, 151
Xenopus laevis replication, 143
see also Initiation sites
Orthopoxvirus, 207

Papovavirus replication, 29–48
bidirectionality of, 35–36
circular minichromosomes in, 54
initiation of, 29–36
integration sites in, 82, 83
Okazaki fragment formation in, 36–39
replication time in, 41–42
T antigen and deletion mutants in, 32–34, 35
termination intermediate in, 40
termination of, 39–44
Parapoxvirus, 207, 210
Parvovirus DNA replication, 257–271
autonomy of parvovirus DNA and, 258–260
defective replication in, 260–262, 267–268
DNA polymerases in, 266–267
model for, 268–271
origin of replication in, 259
rolling hairpin model for, 270
tandem repeat in, 260
Phages, *see* Bacteriophage *headings*
ØX174 DNA replication, 1–24
dnaC protein and, 105
DNA polymerase α in, 36
duplex replicative form in, 2, 4, 18–20
initiation of complementary strand replication in, 4–5
mobility of priming system in, 13–14
model for semidiscontinuous replication of *E. coli* chromosome from, 23–24
multiple primers on single-stranded DNA template in, 4–5
prepriming intermediate complex in, 3
priming at unique site in, 11–13, 99
primosome at replication fork in, 20–23
protein n′ recognition of specific sequence in, 10–11

ØX174 DNA replication (continued)
  replication promoter in, 6–8
  replication proteins preceding priming event in, 2–3
  RNA primer location in, 19–20, 101
  stability of prepriming intermediate in, 8–9
Plasmid chimeras, in *E. coli* mapping, 106, 110
Plasmid replication, 41, 43
Poly ADP ribose polymerase, 47
Polymerase, *see* DNA polymerase; RNA polymerase
Polynucleotide kinase, with primer RNA in *E. coli* replication, 99
Polynucleotide ligase, in poxvirus DNA replication, 218
Polyoma virus (Py) DNA replication, 69–90
  amplification of DNA cosedimentation with chromosomal DNA in, 75–76
  cellular deletion next to polyoma insertions in, 78–79
  cloning of cell DNA sequences flanking polyoma insertions in, 76–77
  covalent linkage between viral sequences and DNA in, 71–72
  DNA polymerase α in, 36
  induction of polyoma virus synthesis in, 84–89
  initiation of, 29–36
  integration site mapping in, 76, 81–84, 88–89
  LPT cells and polyoma insertions in, 77–78
  mitomycin C (MCC) induction of, 72–75, 85
  nonpermissive cell replication studies with, 45–46
  organization of sequences in, 69–90
  origin region in, 35
  origin sequences in, 30, 31, 35, 89
  polyoma (Py) DNA in LPT cells in, 70–72
  replicon organization and control model for, 120
  rolling circle intermediates in, 82
  Southern blotting technique used with, 76, 77, 79
  tandem repeat of insertions in, 81, 82
  T antigen and, 86
  termination of, 41, 79–81, 89–90
Poxvirus DNA replication, 205–231
  cell fractionation in, 214–216
  core in, 220
  cross-linked termini in, 209
  description of viruses in, 205–207
  directions for future research in, 227–231
  elementary bodies (ER) in, 212
  experimental systems with virus in, 211–212
  host-cell nucleus in, 226–227
  infection of cells in, 212–213
  inhibitors of replication and, 213–214
  nicks in, 220, 223
  parental genomes and, 219–226
  rolling circle mechanism in, 225
  steps in viral DNA synthesis in, 217–219
  structure of vaccinia genome in, 208–211
  tandem repetitive sequences in, 208, 210
  terminal repetitions in, 208, 210–211, 227, 230–231
Prepriming intermediate complex
  association of, in primosome of DNA replication, 14–18
  ØX174 DNA replication with, 3, 4, 5, 8–9
  priming at unique site and, 13
  protein n′ associated with, 11
  stability of, 8–9
Primase
  assembly of prepriming proteins with, 14, 15
  gene 4 protein in bacteriophage T7 replication with, 178–180
  phage 4 complementary DNA replication with, 12
  ØX174 replication with, 3, 4, 5, 21, 23, 99
  replication promoter for, 7, 11, 13, 17
Primers
  adenovirus DNA replication with, 239
  bacteriophage T7 replication with, 178
  DNA polymerase α in replication with, 36, 66
  influenza virus RNA biosynthesis with, 345, 346, 348, 359, 361, 362
  mammalian mitochondrial DNA (mtDNA) replication with, 157–158
  parvovirus DNA replication with, 266, 268
  ØX174 DNA replication with, 4–5, 99
  replication origins and, 120–126
  10 kb replication intermediates with, 128
  unique site of, 11–13
  viral minichromosomes as, 54
  see also RNA primer
Primosome
  association of prepriming proteins in, 14–18
  ØX174 DNA replication with, 2, 3, 13, 23, 24, 99
  replication fork of duplex ØX DNA and, 20–23
  termination forks and, 110
Promoters
  bacteriophage T7 DNA replication with, 166, 191–193, 195, 197, 199

E. coli replication with, 111
ØX174 DNA replication with, 6–8, 14
primase and, 17
Proteinase K
parvovirus DNA replication with, 265–266
tobacco mosaic virus (TMV) replication with, 341
Protein n′, in ØX174 DNA replication, 10–11, 13, 15–17, 19, 21, 99
Protein R, in retroviruses, 283
Proteins
adenovirus DNA replication with, 246–249
bacteriophage T7 replication with, 170–182
DNA-binding, *see* DNA-binding protein (DBP)
ØX174 replication priming event with, 2–3, 4, 5
retrovirus replication with, 277
stability of, 8, 9
SV40 DNA replication with, 62, 64
Pseudotyping, in retrovirus DNA replication, 284

Rabbitpox virus, 206, 210, 211, 223
Radiation, and DNA replication, 46
Replicase, in tobacco mosaic virus (TMV) replication, 335, 339
Replication complexes, in adenovirus DNA replication, 240, 242
Replication forks, 89–90
adenovirus DNA replication with, 239, 246
bacteriophage T7 DNA replication with, 164, 165, 167, 185–189, 195
DNA polymerase α in replication with, 36
Drosophila mitochondrial DNA (mtDNA) replication with, 136
E. coli replication with, 94, 95, 102, 104, 107, 109–110
Okazaki fragment synthesis and, 38–39
phage λ, 42, 43
ØX174 DNA replication with, 14, 18, 19, 20–23, 24, 99
polyoma virus (Py) DNA replication with, 87
size of, 118
SV40 DNA replication with, 55
termination signal with, 41
Replication forms (RF)
bacteriophage replications with, 163
Drosophila mitochondrial DNA (mtDNA), 136–137
parvovirus DNA, 262
replication intermediates and, 20–21
tobacco mosaic virus (TMV), 329, 334, 335
Replication intermediates
adenovirus DNA replication with, 239, 240–241, 248, 253
bacteriophage T7 DNA replication with, 164, 165–166, 185, 186
direction of priming system and, 14
Drosophila mitochondria DNA (mtDNA) replication with, 136–137, 138, 139, 141
duplex replicative form multiplication and formation of, 18
E. coli replication with, 93
group I RNA phage replication with, 324
mammalian mitochondrial DNA (mtDNA) replication with, 151, 160
papovavirus replication termination with, 40
parvovirus DNA replication with, 264, 266, 268, 269
ØX174 DNA replication with, 2
poxvirus DNA replication with, 215, 217, 225, 230
primosomes and RF replication and, 20–21
RNA primers and, 20
rolling circle, in viral DNA replication, 82
SV40 minichromosomes as, 43, 55, 57
10 kb, 127–128
termination time in papovavirus replication with, 42
tobacco mosaic virus (TMV) replication with, 331, 334
Replication proteins, *see* Proteins
Replicons, 89
bidirectional origins of, 36, 119–120, 128
initiation sites for replication and, 115
model for organization and control in, 120
size of, 117–119
Replisome, in ØX174 DNA replication, 23, 24
Retrovirus DNA replication, 275–307
analogy to transposons in, 300–302
events in life cycle of retrovirus and, 303–306
future research trends in, 306–307
general structure of viruses in, 275–284
integration in, 296–300
landmarks of viruses in, 280–284
long terminal repeat (LTR) in, 294–296
strategies of viral genomes in, 284–290
RFII duplexes, in ØX174 DNA replication, 18, 19
Ribonuclease H, in retrovirus DNA replication, 278

Rifampins
bacteria replication and, 214
E. coli replication and, 105
group I RNA phage replication with, 316–319, 321
RNA polymerase
bacteriophage T7 replication with, 166, 171–172, 190, 191–194, 195–197
E. coli and, 167, 168, 191
group I RNA phage replication and, 316
ØX174 DNA replication with, 3, 5
poxvirus DNA replication with, 220
RNA polymerase II
parvovirus DNA replication with, 263
retrovirus DNA replication with, 294–295, 304
RNA primase, in Okazaki fragment initiation, 39
RNA primers
advenovirus DNA replication with, 239
bacteriophage T7 replication with, 178–179, 181, 189, 198
closed duplex RFI DNA with, 19
E. coli replication with, 94, 95–101, 107, 109, 111
location of, 19–20
Okazaki fragments and, 37–39, 56
parvovirus DNA replication with, 269, 270
structure of, 99–101
RNA replicase, in group I RNA phage replication, 315
RNA transcriptase, in influenza virus RNA biosynthesis, 346, 352
RNase H, and *E. coli* replication, 100, 101
Rolling circle mechanism
bacteriophage T7 replication with, 187
poxvirus DNA replication with, 225
viral DNA replication with, 82
Rous sarcoma virus (RSV), 279, 282, 283

Salt concentrations, in SV40 DNA replication, 59
Sea urchin replication, 101, 143
Southern blotting technique, in polyoma (Py) DNA replication studies, 76, 77, 79
Spleen focus-forming virus (SFFV), 289
SSB, and replication forks, 110
Stomatitis papulosa virus, 206
SV40 chromosomes
nucleosomes in, 58
preparation and characterization of, 56–58
SV40 DNA replication, 53–67
aphidicolin as inhibitor in, 249
catenated dimers in, 43
development of system for, 54–55
DNA polymerase α in, 36, 42, 55, 66–67
initiation of, 29–36, 44
integration into cell DNA in, 82
in vitro systems for, 53–67
nonpermissive cell replication studies with, 45–46
origin sequences in, 30–31, 55
preparation and characterization of SV40 chromosomes in, 56–58
purification of replicating chromosomes in, 60–64
replicon organization and control model for, 120
rolling circle intermediates in, 82
salt concentrations in, 59
T antigen binding in, 32, 44, 47, 64, 86
termination of, 41, 55
Xenopus laevis protein and, 55, 61, 63

T antigen
deletion mutants and, 32–34, 35
initiation of viral replication with, 44, 54
monoclonal antibodies directed against, 34
onion skin model of replication with, 45, 46
origin of papovavirus replication with, 32, 34, 35
replication and binding of, 29, 33–34
SV40 replication and, 47
viral DNA replication with, 85–86
Terminal protein
adenovirus DNA replication with, 237, 239
parvovirus DNA replication with, 268
Termination codons, in *Drosophila* mitochondrial DNA (mtDNA) replication, 146
Termination intermediate, in papovavirus replication, 40–41
Termination sites
chromatin replication and, 89–90
polyoma (Py) virus DNA replication with, 79–81
tobacco mosaic virus (TMV) replication with, 337
vaccinia DNA replication with, 217
Thioredoxin, and bacteriophage T7 replication, 168, 170, 172, 173, 174
Thymine, in *Drosophila* mitochondrial DNA, 132–136
Tobacco mosaic virus (TMV) replication, 329–342
coat protein in, 329, 330, 331, 337, 339, 340, 341

description of virus in, 329–330
double-stranded RNAs in, 334–336
H-protein in, 341–342
overlapping genes in sequences in, 336–338
replicative form (RF) in, 329, 334, 335
replicative intermediates (RI) in, 331, 334
RNA-coded translational products in, 338–341
RNA genome organization in, 331–336
sub-genomic RNAs in, 331–334
30K protein in, 339–341
Topoisomerase
catenane formation with, 40, 41, 44
termination forks in replication and, 99, 110
Toyocamycin, with influenza virus RNA biosynthesis, 345, 355, 362

Vaccinia virus, 207, 213
Vaccinia virus DNA replication
cell fractionation in, 214–216
host-cell nucleus in, 226–227
in vitro synthesis in, 228
parental genomes and, 219–226
rolling circle mechanism in, 225
steps in viral DNA synthesis in, 217–219
structure of genome in, 208–209, 210
terminal repeats in, 227, 230–231
Variola virus DNA, 209, 211, 230
Vesicular stomatitis virus (VSV), 284, 348, 351
Vicia faba, and replication origins, 120–122
Viral chromosomes, and DNA polymerase, 64–67
Viral DNA replication
inhibitors in, 213–214
parental genomes and, 219–226
T antigen in, 85, 86
Viral proteins, in retroviruses, 277

Xenopus laevis mitochondrial DNA (mtDNA), 131, 133
closed circular DNA replication with, 115–116
moderately repetitive sequences in origins of, 116–117
origin of, 143
SV40 DNA replication and protein from, 55, 61, 63